ARCHEO.FOSS XIV 2020

Open software, hardware, processes, data and formats in archaeological research

Proceedings of the 14th International Conference
15-17 October 2020

edited by
Julian Bogdani, Riccardo Montalbano, and Paolo Rosati

ArcheoFOSS 14 | 2020

edited by

Julian Bogdani,
Riccardo Montalbano,
and Paolo Rosati

Archaeopress Archaeology

Archaeopress Publishing Ltd
Summertown Pavilion
18-24 Middle Way
Summertown
Oxford OX2 7LG
www.archaeopress.com

ISBN 978-1-80327-124-8
ISBN 978-1-80327-125-5 (e-Pdf)

Image cover: Theatrum Pompei and surrounding areas, from SITAR - Sistema Informativo Territoriale Archeologico di Roma, Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma (Ministero della Cultura)

All contributions of this volume have undergone a double peer-review process.

WIKIMEDIA ITALIA

MINISTERO DELLA CULTURA
Direzione Generale Educazione, ricerca e istituti culturali

The publication of this volume was made possible by the generous supports of *Wikimedia Italia* and *Direzione Generale Educazione, ricerca e istituti culturali* of the Italian Ministry of Culture

This book is available in print and as a free download from www.archaeopress.com

Contents

Foreword

This volume represents the editorial outcome of the 14th edition of ArcheoFOSS international conference, which took place on 15–17 October 2020 on the World Wide Web. The event has been held annually since 2006 and is dedicated to the theoretical framework and actual application of free and open-source software solutions and the promotion and encouragement of the Open Data paradigm for archaeology and, more generally, for Cultural Heritage.

Compared to the past editions, the 2020 conference introduced some substantial changes. First of all, the pandemic crisis due to the COVID-19 outbreak, forced us for the first time to hold a Web-based conference, a solution that introduced some important advantages. For example there was the facility to overcome geographical distances and therefore greatly broadening participation, both in terms of speakers (presentations, workshops, demos) and audience. Additionally we had the possibility to share thoughts on the specific topics of the conference with foreign colleagues from different backgrounds (universities, research centres), widening consequently the network of collaboration. We hope that the effort to open the ArcheoFOSS conference beyond Italian national borders – and beyond a small circle of individuals who in the last years have tenaciously and with great difficulty tried to keep alive the spirit of the conference – will not remain isolated, but will be further pursued in the next editions.

Another innovative aspect was the introduction of a panel dedicated to open data, open formats and open standards. While these topics have not been absent in the previous editions, the main focus has always been on the development and application of FLOS software and hardware solutions for Cultural Heritage. By specifically calling for papers dealing with the free sharing of data, we tried to go beyond software and technological development. Open and reusable data publishing platforms, available in open formats, and distributed with open licenses with no bias on the tools with which the data were created. The aim was to encourage and enhance the creation and publication of open archaeological archives, easily re-usable by the community.

This volume well represents the approach taken at the conference and the extensive participation it received. Eighteen high-level and peer-reviewed papers, well distributed in two thematic sections – application cases and development, and open data – contributed by more than forty Italian and foreign scholars, researchers and freelance archaeologists working in the field of Cultural Heritage. For an event organized at no cost, without funding or support of any kind, these are significant numbers, which reward us for the great organizational and editorial effort. The most important budget line was invested in releasing this book as open-access, using a CC BY license. We strongly believe that the conference proceedings must strictly follow the spirit of the event, and that the free distribution and sharing of the volume is a *conditio sine qua* non for its publication. This also marks a break with the past, when open-access was not always a prerequisite. It is worth noting, furthermore, that these proceedings are being published only one year after the conference. This is a decisive turnaround, which testifies to the strong will to revitalize the ArcheoFOSS community. Technology is evolving very fast, and it is not uncommon to read on fresh publications about outdated software and workflows or scripts that have already disappeared, greatly reducing or nullifying the utility of the publication, if not (perhaps) for the academic careers of its authors.

Not strictly related to this book, but important to the ArcheoFOSS community, was the decision to accelerate the publication of the 2019 edition, which was neglected due to financial issues and the outbreak of the COVID-19 pandemic. Furthermore, it was decided to alternate 'lighter' versions of the conference, mostly focused on workshops and hands-on sessions, demos, etc. and more 'traditional' ones, based on paper presentations. This will hopefully facilitate the prompt publication of the proceedings and regain a closer relationship with younger and frequently more active researchers.

Looking to the future, we all hope for the end of the current pandemic emergency, but it is clear how much this crisis sped up many cultural processes already ongoing, by changing our lives, our way of researching, teaching, experiencing and communicating archaeology.

In the coming years, the financial resources earmarked to fund digital projects in the field of Cultural Heritage will be substantial (consider, for example, the Italian National Recovery and Resilience Plan). The challenge for our community is therefore to stand ready to proactively suggest solutions to govern and guide this change, rather than passively undergo it.

As editors, we would like to thank WikiMedia Italia and the Ministero della Cultura – Direzione Generale Educazione, ricerca e istituti culturali for financially supporting the publication of these Proceedings; the University of Pisa, which granted us the use of the infrastructure and support for the streaming of the three-day conference; the colleagues of the Organising Committee, who shared with us the organization of the conference; the scholars and researchers who supported us as reviewers in the evaluation process for the conference and for the publication. Last but not least, we thank the members of the outgoing and current Scientific Committee, whose experience and competence guarantee the scientific quality of ArcheoFOSS initiatives.

The Editors

Julian Bogdani
Riccardo Montalbano
Paolo Rosati

ArcheoFOSS 2020 Committees

Scientific committee (alphabetical order)

Luca Bezzi
luca.bezzi@arc-team.com
Independent researcher and member of ArcTeam, Cles (TN)
Founder of the ArcheoFOSS conference

Paola Buzi
paola.buzi@uniroma1.it
Full Professor of Egyptology and Coptic Civilization at Sapienza University of Rome
Honorary Professor of Egyptology at Hamburg University
Director of the Research Centre 'DigiLab' - Sapienza University of Rome

Elena Calandra
ic-archeo@beniculturali.it
MiBACT, Director of Service II - Excavations and Protection of the Archaeological Heritage, Directorate General for Archaeology, Fine Arts and Landscape
Director ad interim of the Central Institute for Archaeology

Marco Callieri
marco.callieri@isti.cnr.it
First Researcher at the Visual Computing Lab, CNR-ISTI, Pisa

Enzo Cocca
enzo.ccc@gmail.com
GIS consultant for the Honor Frost Foundation (Southampton) for which he is in charge of data management of maritime and land archaeological surveys in Lebanon.

Andrea D'Andrea
dandrea@unior.it
Director of the Interdepartmental Centre for Archaeology Services of the University of Naples L'Orientale

Benjamin Ducke
benjamin.ducke@dainst.de
Director of the Department of Information Technology – German Archaeological Institute (DAI), Berlin Headquarters

Responsible for the development and implementation of Open-Source software for archaeology and the research infrastructure of the Institute (iDAI.world).
Gabriele Gattiglia
gabriele.gattiglia@unipi.it
Researcher of Archaeological Research Methodologies at the University of Pisa

Antonella Guidazzoli
a.guidazzoli@cineca.it
CINECA and University of Bologna
Head of VISIT (Visual Information Technology) Lab CINECA

Anna Maria Marras
annamaria.marras@unito.it
University of Turin

Paola Moscati
paola.moscati@cnr.it
CNR, Research Manager at ISPC

Antonella Negri
antonella.negri@beniculturali.it
Central Institute for Catalogue and Documentation (ICCD) – Ministry for Cultural Heritage and Activities and Tourism (MiBACT)
Architect officer, responsible for the 'General Catalogue Quality Service' and contact person for the General Catalogue Information System (SIGECweb)

Silvia Orlandi
silvia.orlandi@uniroma1.it
Associate Professor of Epigraphy at Sapienza. University of Rome
Director of the Epigraphic Database of Rome (EDR)

Sofia Pescarin
sofia.pescarin@cnr.it
CNR and University of Bologna; former head of 'Virtual Heritage: integrated digital technologies for the knowledge, enhancement and communication of cultural heritage through virtual reality systems'.
Chief Editor since 2017 of 'Elsevier Digital Application in Archaeology and Cultural Heritage' (DAACH)

Alessandro Pintucci
alessandro.pintucci@gmail.com
Archaeologist

Serena Sensini
serena.sensini@gmail.com
Roma Tre University

Mirella Serlorenzi
mirella.serlorenzi@beniculturali.it
MiBACT officer
Since 2007 he has been directing and coordinating the SITAR (Sistema Informativo Territoriale Archeologico di Roma) project of the Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma, for the creation of the first archaeological cadastre of the City.

Valeria Vitale
valeria.vitale@sas.ac.uk
University of London
Pelagios Commons Education Director

Organizing committee (alphabetical order)

Julian Bogdani
Sapienza University of Rome
julian.bogdani@uniroma1.it

Emanuel Demetrescu
VHLab, CNR-ISPC Rome
emanuel.demetrescu@gmail.com

Sara Gonizzi Barsanti
Politecnico di Milano
sara.gonizzi@gmail.com

Saverio G. Malatesta
DigiLab, Sapienza University of Rome
saveriogiulio.malatesta@uniroma1.it

Luca Mandolesi
adArte Srl. Rimini
luca@adartesnc.com

Riccardo Montalbano
University of Pisa
ricca.montalbano@gmail.com

Augusto Palombini
CNR, Rome
augusto.palombini@gmail.com

Paolo Rosati
Sapienza University of Rome
paolo.rosati@uniroma1.it

Strumenti digitali open-source per la documentazione della cultura visuale paleo-mesolitica: dati preliminari da un flusso di lavoro sulle decorazioni incise su supporto calcareo dalla Grotta di Santa Maria di Agnano (Ostuni, BR)

Michele Pellegrino, Donato Coppola

Michele Pellegrino, PhD student, Università degli Studi di Salerno; mipellegrino@unisa.it
Donato Coppola, Università degli Studi di Bari Aldo Moro; donato.coppola@libero.it

Abstract
The aim of this work is to contribute to the definition of a multidisciplinary methodological approach and a functional workflow that can meet the requirements of archaeological documentation and research on prehistoric art. The RTI technique (Reflectance Transformation Imaging) proved effective in improving the record of specific details and in processing an interactive false3D-image of a sample of engraved pebbles, found during the archaeological excavation at Grotta di Santa Maria di Agnano (Ostuni, BR – Italy).

Keywords: REFLECTANCE TRANSFORMATION IMAGING; PHOTOGRAPHIC METHOD; INTERACTIVE RE-LIGHTING; PALEO-MESOLITHIC ART; ENGRAVED PEBBLES.

Introduzione

La complessità genericamente riscontrata nella visualizzazione e analisi dei segni grafici, pertinenti al repertorio delle manifestazioni artistiche di età preistorica, rappresenta un ostacolo ricorrente alla comprensione dei caratteri del sistema grafemico-simbolico condiviso dai gruppi di cacciatori-raccoglitori di cultura paleo-mesolitica (Leroi-Guran 1971; Graziosi 1973; Martini 2016), espresso con una varietà di tecniche grafiche (pittura, incisione, graffitura) su una molteplice diversità di supporti (pareti rocciose, ciottoli, ossa, elementi vegetali): la difficoltà nella riproduzione dell'articolata sintassi di segni ha portato a rilevanti ripercussioni sull'analisi delle rappresentazioni e, talvolta, con l'intento di una chiara leggibilità, ha deviato la 'trascrizione' oggettiva dei segni verso una privilegiata osservazione soggettiva delle tematiche; se tuttavia per le espressioni naturalistico-figurative l'individuazione macroscopica di modelli iconografici non ha apportato modifiche sostanziali nell'interpretazione generale dei contenuti, uno squilibrio dell'accuratezza metodologica è dichiaratamente manifesto per il panorama non figurativo e geometrico-lineare dove, oltre al lacunoso rapporto con il supporto, si tradisce una comprensione semiotica organica e complessiva dei segni (Minellono 1992: 131–134; Vialou 1998).

La necessità diagnostica di fornire dati sensibili della cultura visuale paleo-mesolitica quanto più oggettivi, autentici e multiscalari ha reso fondamentale sviluppare tecniche di rilievo diretto e strumentale accurati e contestuali (Chabredier 1966; D'Errico e Vanhaeren 1999): se da un lato il calco cartaceo a bassorilievo (estampage o lottino plastica) e per sfregamento

Figura 1: a) Henri Breuil: rilievo diretto di un graffito parietale; b) riproduzione a pastello di bovidi presso la Grotta di Altamira (Groenen 2018: figg. 10, 22).

(frottage) si sono dimostrati validi metodi, seppur invasivi, dai pionieri del know-how della seconda metà del XIX secolo, le più efficaci restituzioni ricalcate su carta opaca/trasparente di H. Breuil dell'inizio del secolo scorso (Figura 1) sono apparse idonee ed adeguate sino alla larga diffusione dei supporti fotografici (Groenen 1999 e bibliografia; Dufayet 2008).

Le tecniche 'analogiche' di rilievo grafico e fotografico consolidate della ricerca archeologia sono oramai integrate da pratiche metodologiche e strumenti di acquisizione digitale dei dati mediante procedimenti di lavoro ed elaborazioni informatiche (Ruiz Lopez 2019), che sono volte ad individuare nuove modalità di osservazione ed interazione con i prodotti digitali spostando il baricentro dell'analisi da uno studio essenzialmente iconografico verso un'analisi globale del fenomeno (Arrighetti e Martini 2017; Fritz e Tosello 2007; Fuentes, Lepelé e Pinçon 2019).

Con il sostegno delle attuali metodologie e applicazioni informatiche, l'analisi della 'grammatica dei segni' di età preistorica è virata verso l'indagine interattiva degli oggetti, ovvero verso approcci sempre meno invasivi che tengano conto della visione contestuale dei valori plastici che assumono i singoli segni in relazione al proprio supporto e alle tecniche/strumenti di realizzazione (Gil-Docampo, Peña-Villasenín e Ortiz-Sanz 2019; 2020; Greaves et al. 2020; Porter et al. 2016).

M.P.

I supporti calcarei incisi dalla Grotta di Santa Maria di Agnano

L'occasione per una sperimentazione metodologica è stata offerta dal rinvenimento di un nutrito corpus di incisioni con motivi geometrico-lineari su supporto calcareo mobile, recuperate tra il 2015-2017 (Coppola et al. 2016; 2017) nel corso delle indagini stratigrafiche presso la Grotta di Santa Maria di Agnano (Ostuni, BR, Coppola 2012) e attribuite ad un contesto rituale del primo Olocene (Figura 2); con la documentazione fotografica ad alta risoluzione, seppur integrata da una visione monoculare al microscopio digitale, si sono riconosciuti i limiti della metodologia utilizzata: i supporti calcarei si presentano con volumi e superfici irregolari, con tessiture non omogenee e caratterizzate da forti alterazioni (erosione, incrostazioni carbonatiche, ecc.); le decorazioni sono definite da segni incisi talvolta molto

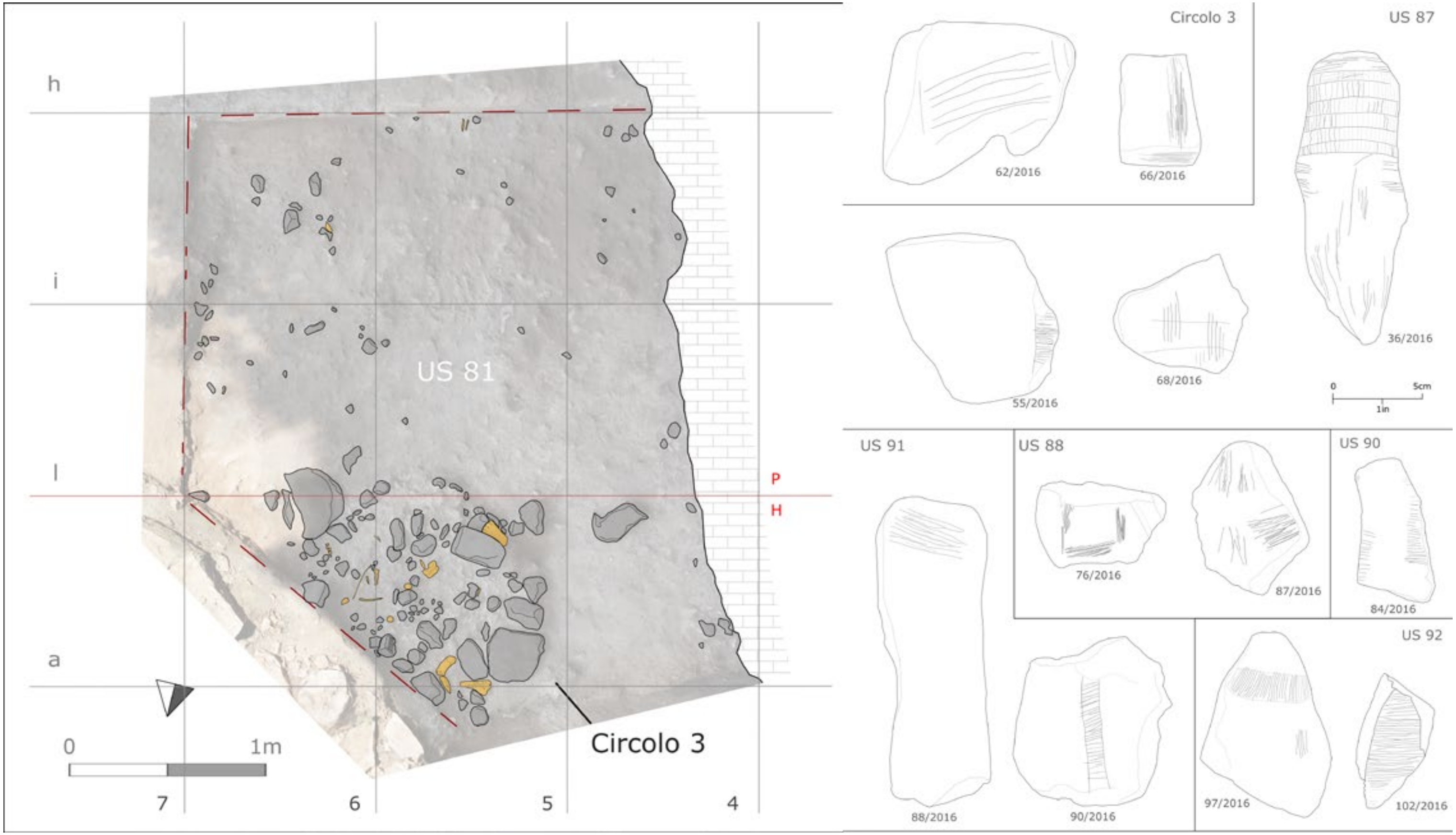


Figura 2: Grotta di Santa Maria di Agnano (Ostuni, BR): a) saggio di scavo archeologico (anno 2016); b) supporti calcarei con sintassi decorative geometrico-lineari rinvenute nel corso della campagna di scavo 2016 (Coppola et al. 2017: Figura 8).

deboli (simmetrici, asimmetrici, superficiali, profondi), con pattern grafici non costanti e caratterizzati dalla sovrapposizione di tratti multi-direzionali, che ne rendono problematica la visualizzazione sistematica della sintassi grafica con una unica fonte di luce.

D.C.

La Reflectance Transformation Imaging (RTI)

La tecnica di acquisizione e di elaborazione di immagini, denominata Reflectance Transformation imagining (RTI), si è dimostrata la più valida tra le tecniche a supporto della documentazione scientifica: si tratta di una pratica non invasiva che, mediante l'elaborazione di una serie di immagini fotografiche con un soggetto statico riproposto in condizioni di illuminazione variabile, produce una mappatura polinomiale delle superfici riprese, sintetizzandole dalle informazioni sulle fonti di luce registrate (Earl, Kik e Malzbender 2010; Earl et al. 2011; 2012; Fiorini 2018; Florindi et al. 2020).

La struttura primaria dell'algoritmo Polinomial Texture Map è stata sviluppata dai laboratori della Hewlett Packard nel 2001 (Malzbender, Gelb e Wolters 2001), poi ripresa e ottimizzata pochi anni dopo con una Hemisferical Harmonics map (HSH) dall'Università della California a Santa Cruz in collaborazione con la fondazione Cultural Heritage Imaging©: entrambi gli algoritmi sono implementati nell'applicativo di elaborazione open-source RTIbuilder[1] che codifica il valore di ogni pixel del frame fotografico e calcola le funzioni di riflettanza dai dati acquisiti, rendendo possibile una fruizione analitica del soggetto in modo interattivo nel software RTIviewer.[2]

Nel solco di flussi di lavoro open-source e strumentazioni low-cost facilmente reperibili (Porter et al. 2016), per le fasi di sperimentazione, è stato adottato un set per fotografie composto da uno stativo da laboratorio per sostenere e fissare una fotocamera (Canon EOS 1100D, E-FS 18-55) su un semplice lightdome autoprodotto: la complessità delle superfici e dei segni incisi degli 8 esemplari selezionati come campioni ha condizionato la scelta di utilizzare l'ambiente scarsamente illuminato di una cupola semirigida, dotata di una maglia di fessure su cui posizionare e muovere una sorgente luminosa LED, agevolando così il processo digitale di estrazione del calcolo delle normali di incidenza della luce di ciascuna fotografia. Per ogni progetto di ripresa sono stati realizzati 36 scatti (focus manuale, EFL ~25, ISO 100; 12 MP, formato JPEG), corrispondenti alla trama del sistema geodetico di punti luminosi equidistanti del lightdome (con un'ampiezza compresa tra 15° e 65°), mediante un criterio di autoscatto remoto e con una media di esecuzione di circa 8 minuti.

L'algoritmo HSH fitted presente nel software RTIbuilder, mediante una pipeline di lavoro definita ad alte luci (Highlight based), compila le stringhe relative alla struttura della superficie riflettente in base al valore della riflettanza, ottenuta analizzando ogni pixel dell'intera immagine e avente come guida l'espressione matematica variabile della fonte di luce di ogni singola immagine (Figura 3a): l'output è un file di formato libero (estensione .rti) e gestibile con il software RTIviewer che, come un applicativo terminale visualizzatore, oltre alla valutazione del flusso di lavoro eseguito, permette di adottare diverse modalità di potenziamento di

[1] http://culturalheritageimaging.org/What_We_Offer/Downloads/Process/index.html (accesso 01/08/2021).
[2] http://culturalheritageimaging.org/What_We_Offer/Downloads/View/index.html (accesso 01/08/2021).

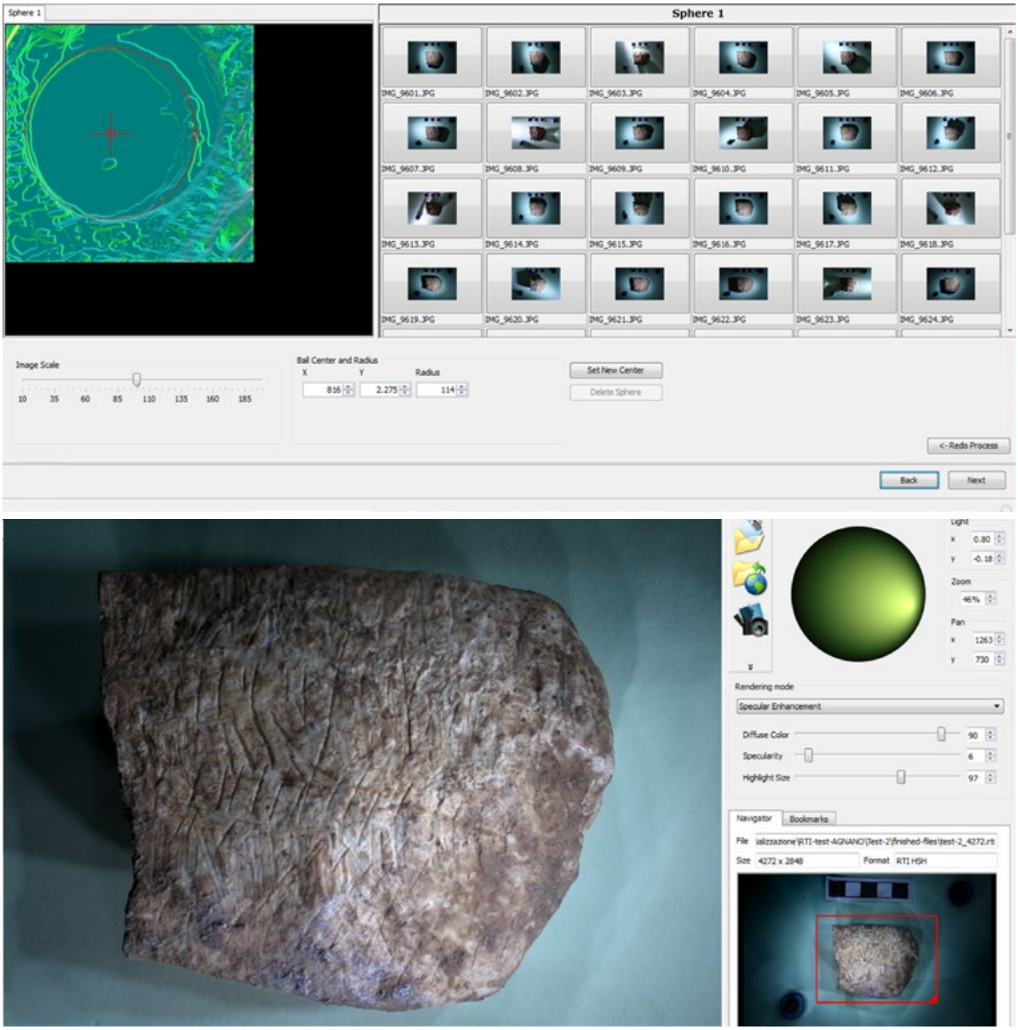


Figura 3: Grotta di Santa Maria di Agnano (Ostuni, BR): a) RTIbuilder, fase di detecting sphere; b) RTIviewer: visualizzazione in modalità Specular Enhancement.

percezione dei dettagli (Specular Enhancement e Normal visualization) e personalizzare i rendering mode delle superfici al fine di ottimizzarne la riproduzione (mediante parametri di Diffuse color, Specularity e Highlight size)(Figura 3b); un limite alla comprensione analitica dei segni potrebbe essere rappresentato dall'assenza di un tool vettoriale capace di registrare contestualmente gli elementi riconosciuti direttamente nel visualizzatore.

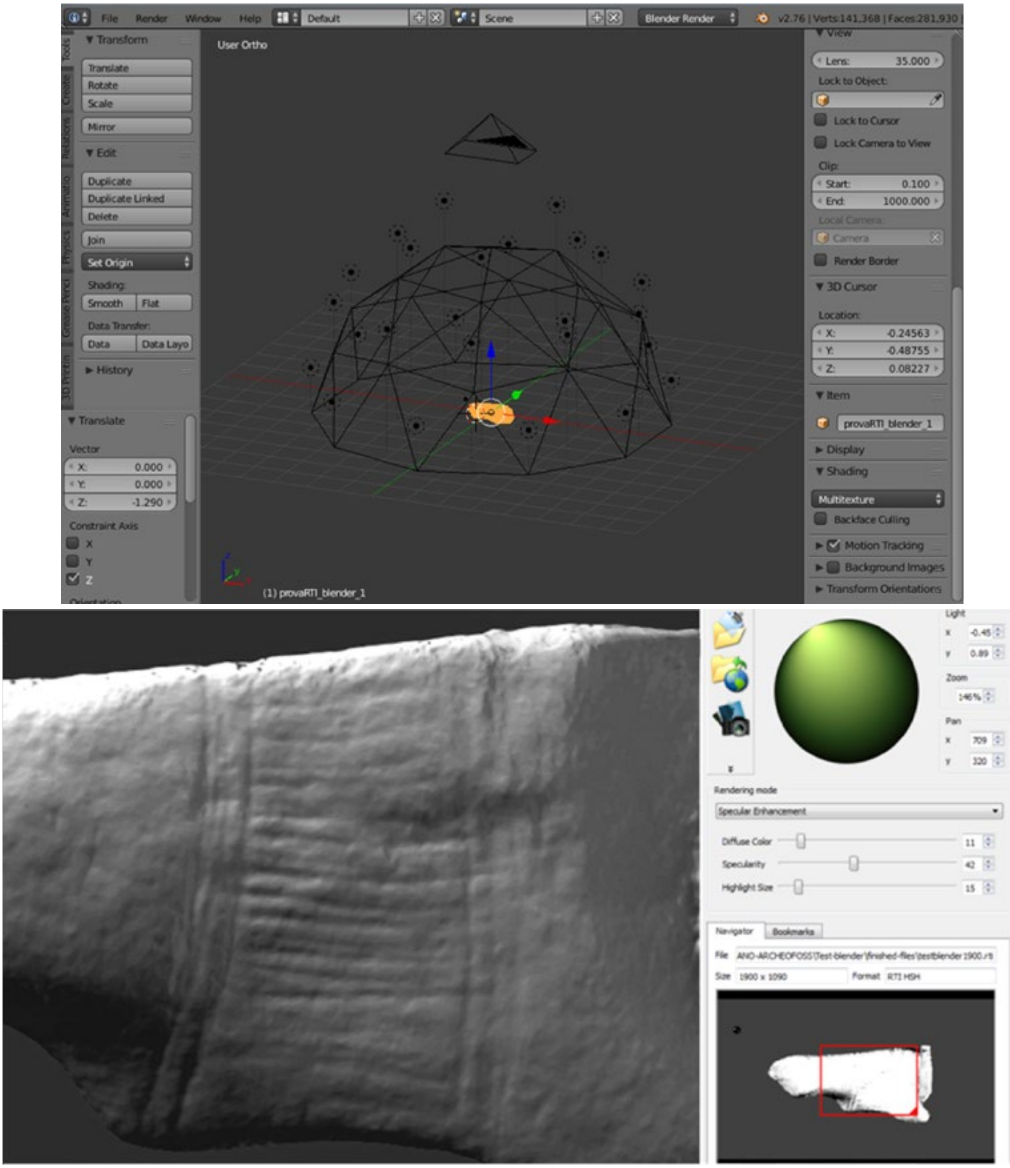


Figura 4: Grotta di Santa Maria di Agnano (Ostuni, BR): a) Blender: lightdome virtuale con camera ortografica e mesh dell'oggetto; b) RTIviewer: visualizzazione in Specular Enhancement del procedimento v-RTI.

La virtual-RTI

Ulteriore sperimentazione è stata realizzata effettuando rilievi fotogrammetrici close-range di 3 soggetti campione mediante pipeline open-source con il software OrtogonBlender (in media si sono elaborate 35 foto, dense-cloud da circa 2M point e mesh con circa 300.000 faces)(Figura 4a): come da pratiche ormai consolidate nella digitalizzazione del patrimonio culturale, si è

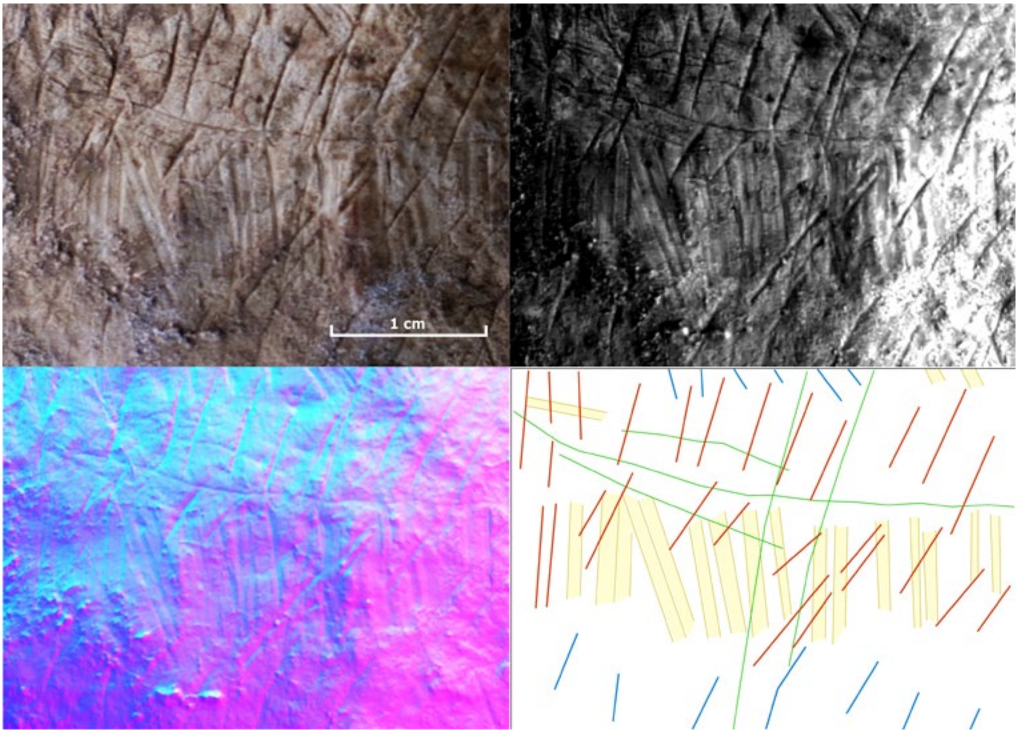


Figura 5: Grotta di Santa Maria di Agnano (Ostuni, BR): a-b) visualizzazione del dettaglio di un supporto calcareo con incisioni lineari non-figurative (SMA-test_2) in modalità Specular Enhancement [a) d.c. 90, sp. 05, h. s. 95; b) d.c. 0, sp. 30, h. s. 15]; c) modalità Normal visualization; d) rappresentazione grafica della sequenza dei segni individuati [giallo-blu (I ordine), verde (II ordine), rosso (III ordine)].

utilizzato una macchina fotografica digitale (Canon EOS 1100D, E-FS 18-55) fissata su treppiedi e un set fotografico composto da due spot luminosi, una base rotante e uno sfondo di colore uniforme.

In questo caso, la tecnica RTI è stata replicata con le medesime operazioni descritte ma in ambiente virtuale (v-RTI) nell'applicativo di grafica 3D Blender, realizzando un lightdome simulato all'interno del software e riprendendo i rendering del soggetto al variare della sorgente luminosa (tipo Sun) con una camera ortografica (3840x2160 px, formato JPEG, Figura 4b, Earl et al. 2012 e bibliografia; Pires, Rubio e Elorza Arana 2015: 418–419). Un aspetto rilevante è costituito dalla possibilità di posizionare e ruotare la mesh del campione, orientandolo secondo modalità funzionali e specifiche per ogni reperto analizzato.

L'acquisizione ed elaborazione fotogrammetrica, variabile in base alla definizione dei punti triangolati e delle superfici richieste, vanno ad incidere sensibilmente sul parametro di rapidità di esecuzione della tecnica RTI (con valori compresi tra i 20–45 minuti), che tuttavia risulta favorita e agevolata dallo svolgimento digitale del processo, soprattutto nella fase di detecting sphere.

Risultati e conclusioni

L'analisi approfondita degli elementi sottoposti al test multi-lighting della RTI ha portato verso la definizione di un approccio metodologico multidisciplinare e un workflow funzionale che risponde alle esigenze della documentazione grafica e della ricerca archeologica sulla cultura visuale preistorica: l'indagine ha virato l'attenzione formale verso tecniche di visualizzazione artificiale RTI, interessando elementi di produzione figurativa paleo-mesolitica mobiliare.

L'interazione immediata con la rappresentazione in false3D delle forme ha garantito e sostenuto un deciso miglioramento nella restituzione di maggiori dettagli rispetto alla osservazione empirica delle superfici complesse, con un grado di definizione 'aumentato' rispetto la fotografia con una sorgente luminosa unica (Figura 5a–c).

Il rendering non foto-realistico prodotto dall'ambiente di visualizzazione virtuale di illuminazione non è difatti riproducibile nel 'mondo reale': il fine scientifico della tecnica è, pertanto, definire una misura di perfezionamento che gestisca un elevato grado di nitidezza dell'immagine e allo stesso tempo ne preservi la luminosità diffusa, garantendo un miglioramento generale della percezione della forma e dei dettagli più particolari, incrementando lo scenario di dati da esaminare.

L'approccio macrofotografico 'aumentato' della RTI, tuttavia, ha manifestato alcuni limiti nell'applicazione della tecnica: se da un lato la fase di acquisizione dei frame fotografici è risultata piuttosto speditiva (anche in assenza di una automazione del processo), dall'altro è stata ostacolata più volte dalla morfologia irregolare delle superfici che, nello svolgimento degli spot luminosi, ha generato dei coni d'ombra molto invasivi; questi, che talvolta hanno reso illeggibile porzioni considerevoli delle superfici riprodotte, hanno indotto a ripensare il processo con diverse angolazioni e ottimizzare i risultati con soluzioni di compromesso.

Le medesime considerazioni fin qui espresse rimangono valide anche per la riproduzione della RTI in ambiente virtuale che, seppur potenzialmente limitata e calibrata sull'accuratezza dello strumento e del metodo utilizzato per la scansione delle superfici (fotocamere più perforanti, obiettivi macro, laser scanner ecc..), permetterebbe l'applicazione del test multi-lighting su possibili molteplici orientamenti dell'oggetto, divenendo una pipeline in fase di sperimentazione altrettanto efficace e alternativa all'utilizzo degli algoritmi di radiance scaling, presente anche nel software libero Meshlab (Pires, Rubio e Elorza Arana 2015; Gil-Docampo, Peña-Villasenín e Ortiz-Sanz 2020).

L'intero processo v-RTI è caratterizzato, in realtà, dalla riproducibilità della tecnica: l'opportunità offerta dalla riproduzione 3D delle superfici dei campioni, oltre a snellire le fasi di acquisizione, rende possibile una loro manipolazione in remoto; questo assunto, seppur carattere assodato per gli oggetti digitali, impreziosisce la tecnica con elementi di replicabilità e multiscalarità: l'analisi, difatti, può essere replicata di volta in volta adeguando la visualizzazione alle necessità di lettura delle superfici, dalla loro totalità alle porzioni di esse più complesse, annullando il limite delle morfologie non omogenee e i coni d'ombra indirettamente prodotti, garantendo un migliore approccio allo studio oggettivo dei segni.

La sperimentazione avviata, allo stato attuale, non permette di instaurare confronti solidi tra le due tecniche adoperate poiché le superfici riprodotte non risultano fisicamente congrue, ma fissa dei punti su cui impostare futuri livelli di sviluppo della RTI: a fronte di strumentazioni più performanti (tra cui obiettivi fotografici macro, lightdome e luci automatizzate, PC con migliori prestazioni per l'elaborazione dei dati), la tecnica può essere individuata come un indiscutibile metodo fotografico analitico altamente competitivo di visualizzazione artificiale, ideale per soggetti con profili tendenzialmente regolari e per l'evidenziazione di solchi multiformi e microtracce; la v-RTI, invece, nonostante un grado non elevato di definizione della mesh con cui è stata elaborata, ha sorpreso per i rilevanti standard quantitativi/qualitativi degli esiti prodotti e per le notevoli potenzialità di sviluppo suggerite, considerandola non solo come alternativa ma virtuosamente più prospettica e degna di ulteriori approfondimenti.

Tra i risultati più interessanti vi è che la lettura di ulteriori porzioni non comprensibili e una visione contestuale dei segni hanno permesso la revisione dei rilievi eseguiti in precedenza, avviando così un'ulteriore fase di indagine semiotica che registri non solo gli aspetti grafici ma che integri anche aspetti riguardanti le dinamiche paleografiche di esecuzione dei segni (D'Errico 1992; Fritz 1999). Evidenziate anche dalla rimozione del colore e dalla modifica interattiva della radenza del fascio luminoso personalizzata dall'utente, le microtracce individuate rivelano una complessa varietà dimensionale e morfologica dei solchi (v-shape, u-shape, flat, bifid, multiple, codebar ecc..), esito dell'utilizzo diversificato di uno o più strumenti; una successione relativa tra i segni ha permesso, inoltre, una lettura stratigrafica delle azioni avvenute, ponendo ulteriori spunti di riflessione sulle modalità e tempi di esecuzione delle sintassi decorative geometrico-lineari (Figura 5d).

In conclusione, l'obiettivo di auspicabili programmi di ricerca è quello di proseguire e ampliare il campione di indagine sui supporti mobili (intra-sito ed in altri corpora ancora parzialmente editi (Martini 2017: 299–301), tentando di recuperare informazioni qualitativamente migliori sulla grammatica dei segni dei gruppi di cacciatori e raccoglitori paleo-mesolitici del sud-est italiano. Inoltre, lo sviluppo e la disponibilità di strumenti open-source capaci di gestire ed elaborare prodotti e volumetrie digitali ha permesso una documentazione low-cost competitiva ed efficace di digitalizzazione per studi futuri, ne garantisce una conservazione digitale del record (alterazioni e deterioramento superficiale, ecc..) ma, soprattutto, liberamente condivisibile con il pubblico di ricerca accademica, ovvero facilmente riproducibile e accessibile nei percorsi di fruizione museale.

M.P.

Ringraziamenti

Si ringraziano il Comune di Ostuni (BR) e l'Istituzione Museo delle Civiltà Preclassiche della Murgia Meridionale di Ostuni, per il supporto alle attività di scavo e per aver reso fruibili gli ambienti del laboratorio.

Bibliografia

Arrighetti, A. e F. Martini. 2017. Registrazione tridimensionale delle incisioni mobiliari di Grotta del Cavallo (Lecce). In F. Radina (ed.) Preistoria e Protostoria della Puglia, Studi di Preistoria e Protostoria 4: 651–655. Firenze: Istituto Italiano di Preistoria e Protostoria.

Chabredier, L. 1966. Étude méthodologique des relevés d'art pariétal préhistorique. Bulletin de la Société préhistorique française 63: 501–512.

Coppola, D. (ed.) 2012. Il Riparo di Agnano nel Paleolitico Superiore. La sepoltura di Ostuni e i suoi simboli. Roma: Università di Roma Tor Vergata.

Coppola, D., N. de Pinto, M. Pellegrino, H. Baills, M. Dewailly, J. Russo e J. Bartolomeo 2016. Santa Maria di Agnano (Ostuni, Puglia). Chronique des activités archéologiques de l'École française de Rome. Italie du Sud (Février 22): 1–17. https://doi.org/10.4000/cefr.1482.

Coppola, D., N. de Pinto, M. Pellegrino, H. Baills, V. Stasolla, M. Dewailly, J. Russo e Jessica Bartolomeo 2017. Santa Maria di Agnano (Ostuni, Puglia). Chronique des activités archéologiques de l'École française de Rome. Italie du Sud (Mars 01): 1–31. https://doi.org/10.4000/cefr.1669.

D'Errico, F. 1992. Dal tratto inciso al gesto, dal gesto al significato. Atti della XXVIII Riunione Scientifica dell'Istituito Italiano di Preistoria e Protostoria: 151–157.

D'Errico, F. e M. Vanhaeren 1999. Les méthodes d'analyse de l'art mobilier paléolithique. Quelques exemples issus de la region Cantabrique. Anthropologie et Préhistoire 110: 31–45.

Dufayet, C. 2008. Les méthodes de relevé dans l'art préhistorique. Hominides.com: l'évolutions de l'homme, November 1, 2008. https://www.hominides.com/html/art/methode-releve-art-prehistorique-avant-1966.php (accessed 14/10/2021).

Earl, G., M. Kirk e T. Malzbender 2010. Archaeological applications of polynomial texture mapping: analysis, conservation and representation. Journal of Archaeological Science 37 (8): 2040–2050.

Earl, G., P.J. Basford, A. Bischoff, A. Bowman, C. Crowther, J. Dahl, M. Hodgson, K. Martinez, L. Isaksen, H. Pagi, K.E. Piquette and E. Kotoula 2011. Reflectance transformation imaging systems for ancient documentary artefacts. In J.P. Bowen, S. Dunn and K. Ng (eds) Proceedings of EVA London 2011: Electronic Visualisation and the Arts, July 6–8, 2010, 147-154. London: BCS.

Earl, G., G. Beale, K. Martinez e H. Pagi 2012. Polynomial texture mapping and related imaging technologies for the recording, analysis and presentation of archaeological materials. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives. 38: 218–223.

Fiorini, A. 2018. Il metodo fotografico RTI (Reflecrance Transformation Imaging) per la documentazione delle superfici archeologiche. L'applicazione ai materiali di età protostorica. Archeologia e Calcolatori 29: 241–258.

Florindi, S., A. Revedin, B. Aranguren e V. Palleschi 2020. Application of Reflectance Transformation Imaging to Experimental Archaeology Studies. Heritage 3 (4): 1279–1286. https://doi.org/10.3390/heritage3040070.

Fritz, C. e G. Tosello 2007. The hidden meaning of forms: methods of recording Paleolithic Parietal Art. Journal of Archaeological Method and Theory 14 (1): 48–80.

Fritz, C. 1999. Towards the Reconstruction of Magdalenian Artistic Techniques: the Contribution of Microscopic Analysis of Mobiliary Art. Cambridge Archaeological Journal 9 (2): 189–208.

Fuentes, O., J. Lepelé e G. Pinçon 2019. Transferts méthodologiques 3D appliqués à l'étude de l'art paléolithique: une nouvelle dimension pour les relevés d'art préhistorique. In Situ. Revue des patrimonies 39: 1-24. https://doi.org/10.4000/insitu.21510.

Gil-Docampo, M., S. Peña-Villasenín e J. Ortiz-Sanz 2019. Professional SfM and TLS vs a simple SfM photogrammetry for 3D modelling of rock art and radiance scaling shading in engraving detection. Journal of Cultural Heritage 37: 238–246. https://doi.org/10.1016/j.culher.2018.10.009.

Gil-Docampo, M., S. Peña-Villasenín e J. Ortiz-Sanz 2020. An accessible, agile and low-cost workflow for 3D virtual analysis and automatic vector tracing of engravings: Atlantic rock art analysis. Archaeological Prospection 27: 153–168. https://doi.org/10.1002/arp.1760.

Graziosi, P. 1973. L'arte preistorica in Italia. Firenze: Sansoni Editore.

Greaves, A.M., S. Duffy, J.R. Peterson, Ş.R. Tekoğlu e A. Hirt 2020. Carved in Stone: Field Trials of Virtual Reflectance Transformation Imaging (V-RTI) in Classical Telmessos (Fethiye). Journal of Field Archaeology 45 (7): 542–555. https://doi.org/10.1080/00934690.2020.1804135.

Groenen, M. 1999. Du relevé de la représentation à la lecture du dispositif pariétal. Anthropologie et Préhistoire 110: 3–23.

Groenen, M. 2018. Reproduire l'art des grottes ornées paléolithiques: du relevé au fac-similé. Koregos. Revue et encyclopédie multimédia des arts, April 6, 2018. http://www.koregos.org/fr/marc-groenen-reproduire-art-grottes-ornees-paleolithiques-du-releve-fac-simile/13362.

Leroi-Gourhan, A. 1971. Préhistoire de l'art occidental. Paris: Mazenod.

Malzbender, T., D. Gelb e H. Wolters 2001. Polynomial Texture Maps. Proceedings of the 28th Annual Conference of Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH), August 12–17, 2001. Los Angeles-USA: 519–528. https://doi.org/10.1145/383259.383320.

Martini, F. 2016. L'arte paleolitica e mesolitica in Italia. Millenni. Studi di Archeologia Preistorica 12. Firenze: Museo e Istituto Fiorentino di Preistoria 'Paolo Graziosi'.

Martini, F. 2017. La produzione figurativa paleo-mesolitica in Puglia nel contesto culturale italiano. In F. Radina (ed.) Preistoria e Protostoria della Puglia, Studi di Preistoria e Protostoria 4: 101–111. Firenze: Istituto Italiano di Preistoria e Protostoria.

Minellono, F. 1992. Problemi di lettura delle incisioni preistoriche. Atti della XXVIII Riunione Scientifica dell'Istituito Italiano di Preistoria e Protostoria: 131–141.

Pires, H., J.M. Rubio e A. Elorza Arana 2015. Techniques for Revealing 3d Hidden Archeological Features: Morphological Residual Models as Virtual-Polynomial Texture Maps, ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W4: 415–421. https://doi.org/10.5194/isprsarchives-XL-5-W4-415-2015.

Porter, S.T., N. Huber, C. Hoyer e H. Floss 2016. Portable and low-cost solutions to the imaging of Paleolithic art objects: A comparison af photogrammetry and reflectance transformation imaging. Journal of Archaeological Science: Reports 10: 859–863. https://doi.org/10.1016/j.jasrep.2016.07.013.

Ruiz Lopez, J.F. 2019. Tecnologías actuales al servicio de la documentación, estudio, conservación y divulgación del arte rupestre. In R. Viñas i Vallverdú (ed.) I Jornades Internacionals d'Art Rupestre de l'Arc Mediterrani de la Península Ibèrica. XXè Aniversari de la Declaració de Patrimoni Mundial: 341–373. Montblanc: Centre d'Interpretació de l'Art Rupestre de les Muntanyes de Prades (CIAR).

Vialou, D. 1998. Problématique de l'interpretation de l'art paléolitique. Rivista di Scienze Preistoriche XLIX: 267–281.

Valutazione integrata delle dinamiche di rischio di erosione del suolo in presenza di depositi archeologici. Il metodo proposto dal progetto RESEARCH (REmote SEnsing techniques for ARCHaeology)

Stefano De Angeli, Fabiana Battistin, Federico Valerio Moresi, Philip Fayad, Matteo Serpetti

Stefano De Angeli, Università degli Studi della Tuscia, p.le dell'Università s.n.c., 01100 Viterbo, Italy; deangeli@unitus.it

Fabiana Battistin, Università degli Studi della Tuscia, p.le dell'Università s.n.c., 01100 Viterbo, Italy; f.battistin@unitus.it

Federico Valerio Moresi, Università degli Studi della Tuscia, p.le dell'Università s.n.c., 01100 Viterbo, Italy; f.v.moresi@unitus.it

Philip Fayad, ALMA SISTEMI Srl, via della tenuta del Cavaliere 1, Building B, 00012 Guidonia, Roma, Italy; philip.k.fayad@gmail.com

Matteo Serpetti, ALMA SISTEMI Srl, via della tenuta del Cavaliere 1, Building B, 00012 Guidonia, Roma, Italy; mse@alma-sistemi.com

Abstract

Soil erosion poses a serious threat to the preservation of archaeological deposits. The RESEARCH project defined an articulated and innovative workflow that integrates archaeological and geological/environmental data to produce a detailed mapping of soil erosion risk by using open-source software and datasets. This article illustrates the risk assessment methodology through the results of some simulations carried out on the site of Falerii Novi (IT) by comparing several open-source models dedicated to soil erosion and testing a new plug-in specifically designed to process high-resolution GPR data.

Keywords: ARCHAEOLOGICAL DEPOSIT; SOIL EROSION; SIMWE; USPED; ARCHAEOLOGICAL VULNERABILITY; GIS TECHNOLOGY; GPR DATA; RISK ASSESSMENT.

Introduzione

Il patrimonio archeologico europeo è oggi sempre più a rischio, messo in pericolo da processi ambientali e pressioni antropiche, che agiscono sia a breve che a lungo termine. I pericoli naturali, legati anche ai cambiamenti climatici, combinati ed amplificati dalle pressioni antropiche, costituiscono infatti delle concrete minacce per la conservazione dei beni archeologici esposti e sepolti, amplificando il naturale deterioramento dei materiali e riducendo la capacità del suolo di preservare le evidenze sepolte (Rao e Colette 2007). Quest'ultima categoria di patrimonio archeologico, la cui conservazione viene spesso trascurata, comprende aree caratterizzate dalla presenza di depositi archeologici composti di strutture, stratigrafie, manufatti ed ecofatti. Un ruolo principale nella sua conservazione è svolto proprio dal suolo, laddove ogni cambiamento del suo equilibrio fisico-chimico può comportare danni alle evidenze archeologiche in esso sepolte. Di fronte a questo scenario, è di estrema importanza poter disporre di buone pratiche e metodologie di previsione del rischio e di specifici strumenti informatici che in maniera sistematica ed efficace, siano in grado di monitorare simili processi di degrado, così da consentire alle istituzioni preposte alla tutela una manutenzione preventiva del patrimonio archeologico.

A questa specifica esigenza intende rispondere il progetto RESEARCH (*Remote Sensing techniques for Archaeology*, H2020-MSCA-RISE,[1] grant agreement n. 823987), riguardante la progettazione e lo sviluppo di una piattaforma *multi-task*, che, combinando tecnologie avanzate di telerilevamento con applicazioni GIS, si propone da un lato la mappatura del patrimonio archeologico e dall'altro il monitoraggio a lungo termine e su diverse scale di minacce legate al suolo, quali il movimento, l'erosione e il cambiamento d'uso, al fine di generare specifiche mappe di rischio (Battistin e De Angeli 2020; De Angeli *et al.* 2020). Il progetto RESEARCH coordina le competenze e le attività di ricerca di un consorzio composto da diverse istituzioni accademiche e PMI: *l'Università degli Studi della Tuscia* (IT), quale coordinatore del progetto, il *Technologiko Panepistimio Kyprou* (CY), *Alma Sistemi Srl* (IT), la *Foundation for Research and Technology Hellas* (GR), *Space Systems Solutions Ltd* (CY), *Geosystems Hellas* (GR), e l'*Uniwersytet im. Adama Mickiewicza w Poznaniu* (PL). La metodologia e gli strumenti di valutazione del rischio proposti dal progetto vengono testati su diversi casi di studio in Italia, Grecia, Cipro e Polonia.

La metodologia di valutazione del rischio

Valutare il rischio per il patrimonio archeologico significa comprendere e stimare le possibilità di un determinato fenomeno, qualsiasi sia la sua origine, di mettere a repentaglio la conservazione dei beni. Per tale ragione, al fine di considerare tutti i fattori coinvolti, la metodologia di valutazione del rischio adottata dal progetto è fondata sul principio che il rischio è la risultante di tre fattori (Cardona *et al.* 2012), ossia:

- la fonte di rischio, la minaccia (*hazard*), relativa al verificarsi di un determinato fenomeno in un certo momento e luogo e con una variabile intensità;
- la vulnerabilità (*vulnerability*), relativa al bene a rischio, la quale varia a seconda delle caratteristiche del bene stesso che possono renderlo vulnerabile rispetto alla minaccia considerata;
- l'esposizione (*exposure*), che indica quali siano gli oggetti effettivamente esposti a rischio, ossia cosa potrebbe essere danneggiato o andare perduto.

Nel contesto del progetto RESEARCH, il rischio è calcolato su base cartografica e la procedura verrà gestita all'interno della piattaforma GIS appositamente realizzata, che consentirà di mettere in relazione i tre fattori (Figura 1). Relativamente alla minaccia e ai beni a rischio, i dati raccolti e prodotti attraverso catene di processamento esterne alla piattaforma verranno trasformati rispettivamente in mappe di minaccia e di vulnerabilità, mediante l'applicazione di specifici gradienti (*ranking*) definiti dalla teoria in base al tipo di minaccia considerata. Le

[1] https://www.re-se-arch.eu (accesso 23/07/2021).

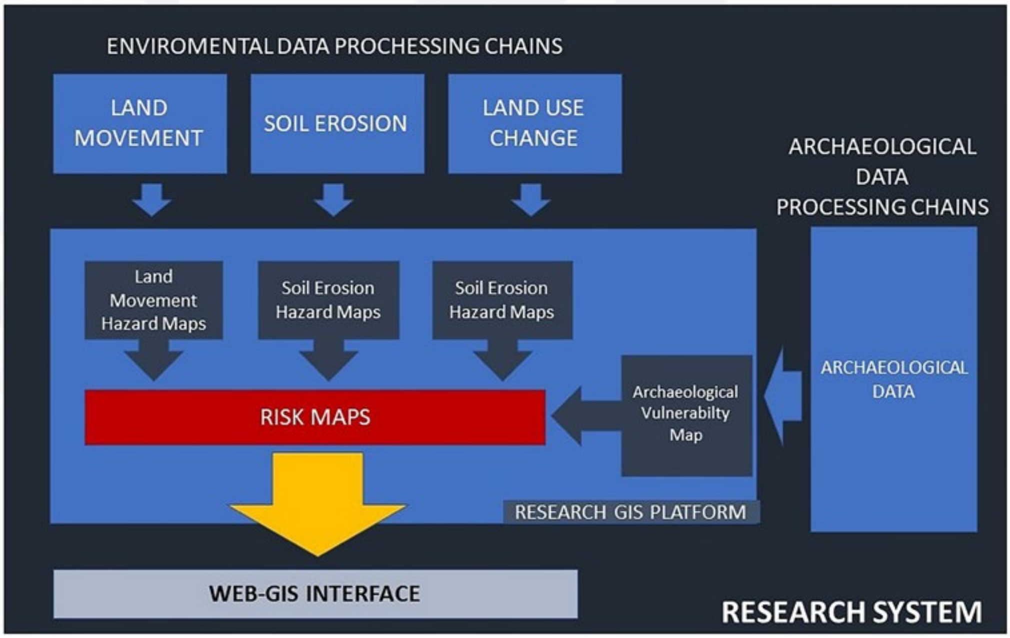


Figura 1: Struttura generale del sistema RESEARCH: catene di processamento e piattaforma Web-GIS (@RESEARCH Project).

minacce saranno classificate in relazione alla loro intensità, mentre il livello di vulnerabilità è stato definito considerando diverse tipologie di evidenze archeologiche e valutando quali siano più o meno vulnerabili ed esposte a una determinata minaccia.

Nel caso dell'erosione del suolo, l'esposizione dei depositi archeologici viene graduata seguendo il principio che più un oggetto è conservato in profondità, ossia distante dalla superficie, più è protetto dall'erosione superficiale. Pertanto, ovunque sia nota la profondità del deposito archeologico, si può applicare un coefficiente in grado di rappresentare le differenze nell'esposizione. Data un'esposizione massima pari a 1, il coefficiente sarà compreso tra 0,01 (corrispondente alle strutture più profonde rilevate utilizzando la geofisica) e 0,99 (assegnato alle strutture più superficiali).

La tabella 1 illustra i valori di vulnerabilità ed esposizione definiti da RESEARCH per la valutazione del rischio derivante da fenomeni di erosione e assegnati alle diverse tipologie di evidenze archeologiche identificate.

Una volta che la piattaforma avrà prodotto le mappe di minaccia e vulnerabilità, il passaggio successivo sarà la messa in relazione di questi due tipi di mappe per la produzione di mappe di rischio finali.

Tipologie evidenze archeologiche	Metodi di identificazione	Gradiente	Valore	Specifiche sulla esposizione alla minaccia
Strutture conservate fuori terra	Fotografie aeree e satellitari, immagini LiDAR, ricognizione di superficie e scavo	Nullo	0	Nel caso di strutture conservate fuori terra, la minaccia è nulla poiché non è presente suolo a coprire le strutture.
Earthwork	Fotografie aeree e satellitari, immagini LiDAR, ricognizione di superficie	Basso	1	Gli *earthwork* non sono normalmente interessati da attività agricole che facilitano il processo di erosione del suolo.
Aree di materiali di superficie	Ricognizione di superficie	Medio	2	Le aree di frammenti corrispondono a depositi archeologici già compromessi. Ciò implica una riduzione del livello di vulnerabilità e di esposizione, rispetto ai depositi archeologici ancora intatti.
Depositi archeologici (strutture e stratigrafia archeologica)	Prospezioni GPR per identificazione strutture sepolte; fotografie aeree e satellitari per identificazione di *crop mark*	Medio-Alto	2,05-2,95	La vulnerabilità diminuisce con il diminuire dell'esposizione alla minaccia, correlata alla distanza dalla superficie del deposito. Nel caso del *soil mark* si tratta di un'evidenza archeologica molto prossima alla superficie e dunque particolarmente esposta alla minaccia.

Tabella 1: Valori di vulnerabilità ed esposizione alla minaccia delle varie tipologie di evidenze archeologiche.

Il caso di studio: *Falerii Novi*

Ai fini di testare metodologia e strumenti di valutazione del rischio proposti dal progetto, con particolare riferimento al problema dell'erosione del suolo nei siti archeologici (Meylemans, Poesen, e In't Ven 2014), si è scelto come caso di studio il sito archeologico italiano di *Falerii Novi*. Si tratta di una città romana, fondata *ex novo* nel 241 a.C., su un pianoro caratterizzato da un lieve declivio con direzione nord-sud, prospiciente la forra creata dal Rio Purgatorio. La città conserva quasi per intero il perimetro delle antiche mura urbane, costruite in grandi blocchi squadrati di tufo, mentre le sue strutture giacciono quasi completamente interrate, fatta eccezione per un'area scavata nella seconda metà del secolo scorso. Scavi ottocenteschi hanno messo in luce la presenza di un teatro (ora di nuovo interrato) all'interno dell'area urbana, mentre nel settore suburbano posto immediatamente a nord della città si conservano ancora i resti di un acquedotto e dell'anfiteatro. Il sito ha conosciuto una frequentazione, forse ininterrotta, fino all'età medievale, come testimonia la costruzione della chiesa romanica di Santa Maria in Falleri e del convento annesso. L'intera area urbana è stata rilevata con prospezioni geofisiche magnetometriche e con strumentazione GPR (*Ground Penetrating Radar*), consentendo una ricostruzione dell'intero assetto urbano e l'individuazione di alcuni importanti edifici e strutture della città (Keay *et al.* 2000; Millett *et al.* 2019). Oltre ad altre minacce connesse a movimenti del suolo e alla crescita incontrollata della vegetazione, il sito, di proprietà privata, è soggetto a un uso agricolo del suolo che, nonostante i limiti imposti dalla Soprintendenza (lavorazione non superiore ai 30 centimetri di profondità), facilita in diversi punti, unitamente alla naturale pendenza del pianoro, fenomeni di erosione del suolo.

Modelli di stima della erosione del suolo e creazione della mappa di minaccia

La stabilità dei pendii è un fenomeno complesso, che non solo include frane e alluvioni (Sidle *et al.* 2016; Moresi *et al.* 2020) ma anche processi meno evidenti come l'erosione del suolo (Harmon *et al.* 2019). Negli ultimi anni sono stati sviluppati in GIS alcuni modelli numerici capaci di simulare l'evoluzione nel tempo di cambiamenti topografici (elevazione del suolo) dovuti ad attività erosive connesse a fenomeni meteorici (Malik 2008). *r.sim.terrain* è uno di questi modelli che opera per regimi di flusso sia stazionari che dinamici su una gamma di diverse scale spaziali (Harmon *et al.* 2019). Si tratta di un *plug-in open-source* implementato in GRASS GIS[2], che riconosce i cambiamenti topografici causati dal flusso idrico superficiale utilizzando i modelli empirici RUSLE3D (*3-Dimensional Revised Universal Soil Loss Equation*, Mitasova *et al.* 2013), USPED (*Unit Stream Power-based Erosion Deposition*, Mitasova *et al.* 2004) e SIMWE (*Simulated Water Erosion*), maggiormente basato sulle caratteristiche fisiche del suolo (Mitasova *et al.* 2013).

Per il funzionamento dei tre modelli, *r.sim.terrain* necessita di un DTM (*Digital Terrain Model*) come base topografica per la simulazione del deflusso dell'acqua in funzione dei dati di intensità e di durata della pioggia. Inoltre, al fine di definire i flussi di trasporto del terreno, esso richiede alcuni dati caratteristici del suolo e relativi al suo uso, dal momento che l'erosione sarà più accentuata laddove il suolo è privo di una copertura vegetale o sottoposto ad attività agricole (Harmon *et al.* 2019). *r.sim.terrain* consente di utilizzare per i diversi parametri presi in esame un unico valore per l'intera area di studio, oppure di utilizzare file raster per dati distribuiti spazialmente.

Tra i vari risultati finali dell'elaborazione dei tre i modelli quelli che maggiormente interessano la nostra valutazione sono:

- *Altitudine* [m]: variazione dell'altitudine dei singoli pixel del DTM in relazione al tempo di durata dell'evento piovoso;
- *Erosione-deposizione* [kg/m^2·s]: quantità di suolo eroso e depositato nel tempo e nello spazio;
- *Differenza netta* [m]: differenza tra i valori di altitudine dei pixel del DTM prima dell'inizio e dopo la fine dell'evento.

Come area in cui testare la funzionalità di *r.sim.terrain* e produrre una prima mappa di minaccia si è scelto la parte centrale dell'area meridionale del sito di *Falerii Novi*. Al fine di dimostrare le differenze tra i modelli USPED e SIMWE, ritenuti più efficaci della RUSLE3D, sono state avviate una sequenza di simulazioni basate su di uno stesso evento piovoso, individuato nel valore più inteso di precipitazione rilevato dalla stazione meteorologica di Civita Castellana (Viterbo), nel periodo 2014–2015 (dati ARPA).

Abbiamo prima testato il modello USPED per simulare le variazioni topografiche del suolo (altitudine) sulla base di un regime di erosione dinamico, con una capacità di trasporto relativa ad un evento piovoso della durata di 60 min. e intensità costante di 50 mm/h^{-1}. Abbiamo quindi usato il modello SIMWE per simulare le variazioni topografiche del suolo (altitudine) sulla

[2] https://grass.osgeo.org/grass78/manuals/addons/r.sim.terrain.html (accesso 23/07/2021).

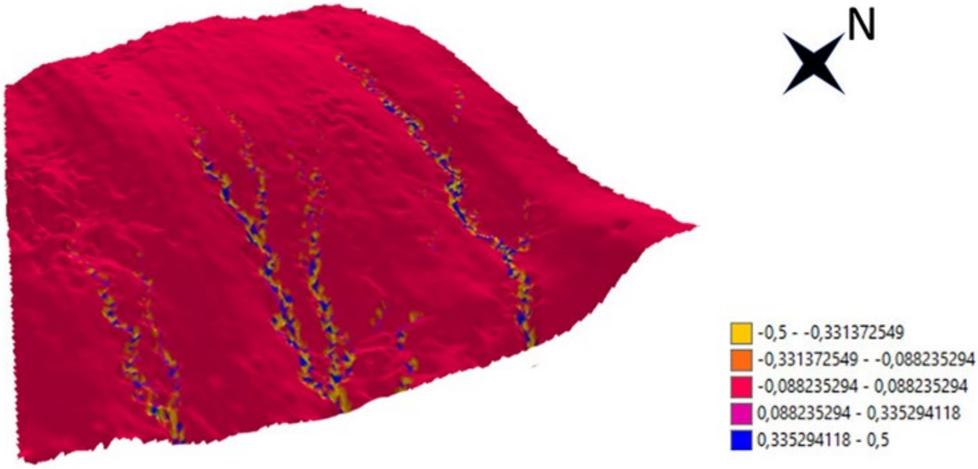


Figura 2: Falerii Novi, area campione. Modello USPD: DTM finale con variazioni dei valori di altitudine (@RESEARCH Project).

base di un regime di erosione statico, tramite un sistema di erosione/deposizione variabile, relativamente ad una pioggia della durata di 60 min. con un'intensità di pioggia costante di 50 mm/h^{-1}. In entrambi i casi è stato previsto un intervallo per ogni elaborazione dei file di output di 10 minuti.

A conclusione di entrambe le simulazioni, sono stati confrontati i risultati delle elaborazioni, analizzando in particolare le differenze nette tra il DTM iniziale e quello finale e della variazione volumetrica. Da tale confronto è emerso che la simulazione del modello USPED evidenzia un'erosione superficiale del terreno all'interno di alcuni canali di erosione (*gullies*) già esistenti, allargandoli piuttosto che approfondendoli, mentre la deposizione avviene in canali più piccoli posti più a valle, mostrando uno schema distribuito di erosione e deposizione. La simulazione SIMWE ha evidenziato invece un processo di erosione consistente nell'approfondimento e allargamento graduale dei canali e nella generazione di creste deposizionali ai lati di questi. Il confronto delle due metodologie mostra come il modello USPED evidenzi una maggiore attività erosiva rispetto al modello SIMWE (Figura 2), rendendo il primo più adatto ad un approccio cautelativo rispetto alla protezione dei depositi archeologici sepolti.

Una volta ottenute le mappe di erosione, queste sono state lavorate in GIS per essere trasformate in vere e proprie mappe di rischio, in cui i valori assoluti ottenuti vengono classificati seguendo un gradiente di minaccia: un livello basso è assegnato alle aree che presentano un livello di erosione inferiore alla media locale, medio laddove coincidano, alto quando il livello di erosione è più alto del valore medio locale (Figura 3).

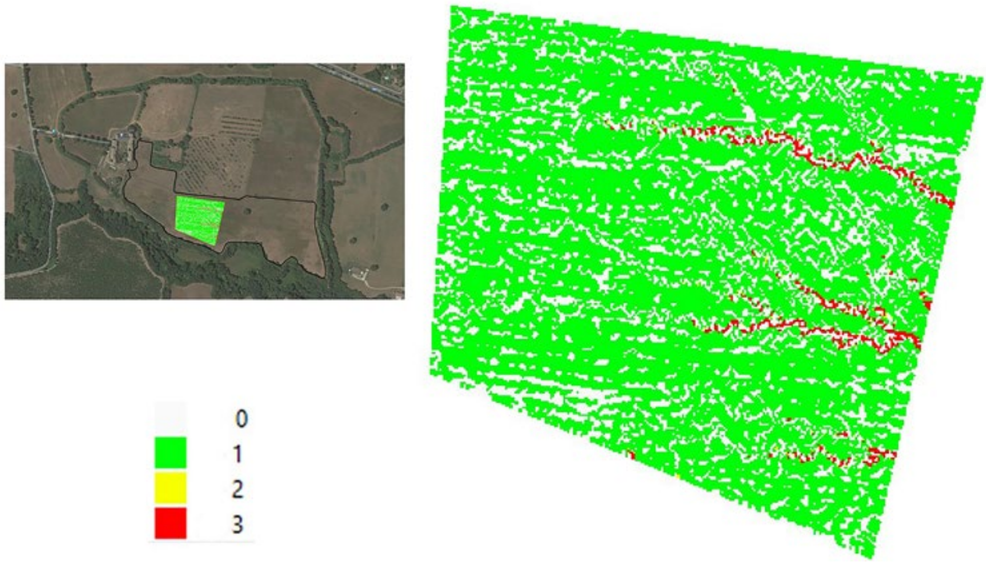


Figura 3: *Falerii Novi*, area campione. Mappa di minaccia dell'erosione del suolo (@RESEARCH Project).

Post-processamento dei dati di GPR per la relizzazione di mappe di vulnerabilità

Ai fini della creazione di una mappa di vulnerabilità archeologica dell'area presa in esame sono stati utilizzati i dati GPR ottenuti dalle recenti prospezioni geofische svolte dalle università di Cambridge e Ghent, disponibili in *open-access* per finalità di studio (Millett *et al.* 2019). Si tratta in particolare di una sequenza di *timeslides* georeferite ad alta risoluzione (5 cm per pixel), che consentono la visualizzazione delle strutture sepolte individuate, ad un intervallo regolare di profondità di circa 5 centimetri, e che, partendo dalla superficie, raggiungono i 3 metri di profondità. Poiché le simulazioni di cui al paragrafo 4 sono state effettuate nella parte centrale dell'area meridionale del sito, sono stati selezionati i risultati delle prospezioni geofisiche di questa parte della città (Millett *et al.* 2019, area 4), in cui sono presenti unicamente strutture sepolte.

In relazione alle necessità del progetto, i dati sono stati utilizzati non tanto per la rilevazione di strutture archeologiche, ma piuttosto per la definizione della profondità alla quale è posto il deposito archeologico potenzialmente ancora non compromesso dalle attività agricole e a rischio dei fenomeni di erosione del suolo.

Al fine di generare una mappa capace di restituire un simile risultato è stato sviluppato uno strumento software specifico (*plug-in*) in ambiente QGIS, che consiste in una catena di operazioni raccolte in un unico algoritmo, scandito da tre differenti processi. Il primo seleziona i pixel pertinenti alle evidenze archeologiche delle singole *timeslice*, unendole in un unico *raster* che consente di ottenere una restituzione 2D delle evidenze archeologiche individuate con valori di profondità dei singoli pixel distinti secondo una scala di grigi, che facilita una lettura tridimensionale della restituzione (Figura 4). Il secondo processo, sulla base dei valori di profondità e del relativo gradiente di vulnerabilità, compreso tra 2,05 e 2,95 (vedi sopra tabella 1), attribuisce ad ogni singolo pixel della restituzione 2D uno specifico valore,

Figura 4: *Falerii Novi*, area meridionale. Restituzione 2D delle evidenze archeologiche individuate con valori di profondità dei singoli pixel (@RESEARCH Project).

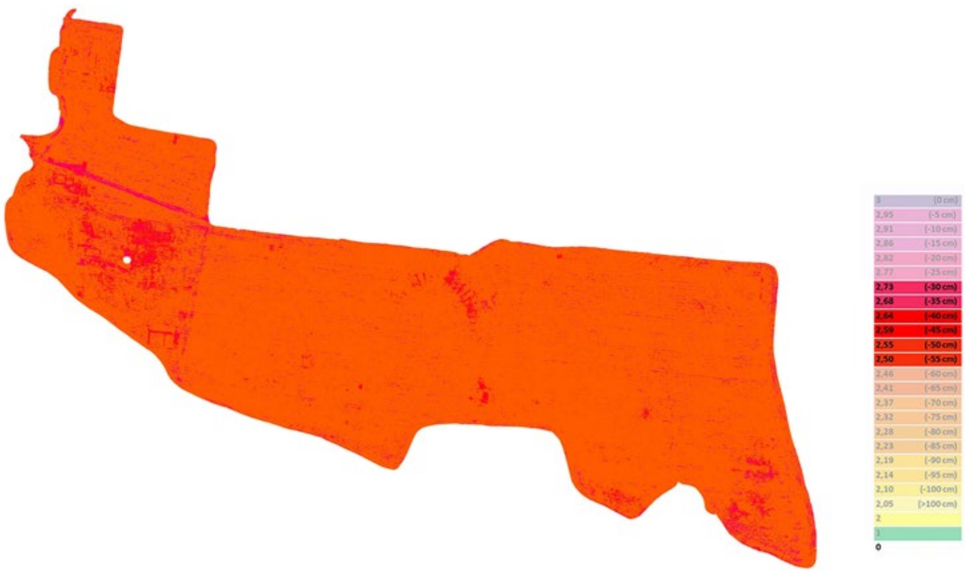


Figura 5: *Falerii Novi*, area meridionale. Mappa di vulnerabilità archeologica correlata alla minaccia di erosione del suolo (@RESEARCH Project).

Soil Erosion Risk rating		Hazard rating			
		0	1	2	3
Vulnerability rating	3	0	3	6	9
	2,95	0	2,95	5,9	8,85
	2,91	0	2,91	5,82	8,73
	2,86	0	2.86	5,72	8,58
	2,82	0	2,82	5,64	8,46
	2,77	0	2,77	5,54	8,31
	2,73	0	2,73	5,46	8,19
	2,68	0	2,68	5,36	8,04
	2,64	0	2,64	5,28	7,92
	2,59	0	2,59	5,18	7,77
	2,55	0	2,55	5,1	7,65
	2,5	0	2,5	5	7,5
	2,46	0	2,46	4,92	7,38
	2,41	0	2,41	4,82	7,23
	2,37	0	2,37	4,74	7,11
	2,32	0	2,32	4,64	6,96
	2,28	0	2,28	4,56	6,84
	2,23	0	2,23	4,46	6,69
	2,19	0	2,19	4,38	6,57
	2,14	0	2,14	4,28	6,42
	2,10	0	2,10	4,2	6,3
	2,05	0	2,05	4,1	6,15
	2	0	2	4	6
	1	0	1	2	3

Figura 6: *Falerii Novi*, area campione. Mappa di rischio finale (@RESEARCH Project).

distinguibile anche per il colore, generando così una mappa di vulnerabilità archeologica del deposito sepolto. Infine, per offrire una restituzione realistica e continua dell'interfaccia superiore del deposito archeologico, alle aree 'vuote', in cui le prospezioni GPR non hanno restituito evidenze archeologiche chiare, è stato automaticamente assegnato un valore unico pari a -55 cm., corrispondente alla profondità media a cui è possibile ipotizzare la presenza di una stratigrafia archeologica sulla base delle evidenze sepolte individuate dai dati geofisici (Figura 5).

La mappa di rischio finale

Una volta ottenute la mappa di minaccia dell'erosione del suolo dell'area presa in esame e quella della vulnerabilità archeologica dell'intera area meridionale del sito, queste sono state sovrapposte, limitatamente all'area centrale in cui è stata condotta la simulazione di erosione del suolo, e tramite una semplice operazione di *raster calculator* è stato possibile generare automaticamente, come risultato finale, la relativa mappa di rischio (Figura 6).

Sviluppi futuri del progetto

Dopo queste prime attività volte ad applicare la metodologia di valutazione del rischio e a sperimentare la funzionalità della piattaforma GIS nei suoi diversi automatismi, gli sviluppi futuri del progetto, limitatamente alla valutazione del rischio dell'erosione del suolo presso il sito di *Falerii Novi*, saranno i seguenti:

- produrre un nuovo DTM, con risoluzione di almeno 2–3 cm, dell'area urbana del sito di *Falerii Novi* utilizzando la tecnologia LiDAR;
- completare la raccolta di dati metereologici (in corso, con l'ausilio di una stazione metereologica in loco);
- produrre una nuova mappa di erosione del suolo dell'intera area urbana sulla base del nuovo DTM e rispetto ad una serie multitemporale di eventi piovosi;
- produrre una mappa di rischio dovuto all'erosione del suolo dell'intera area urbana.

Finanziamento

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823987.

Bibliografia

Battistin, F. e S. De Angeli 2020. Da CLIMA a RESEARCH. Monitoraggio e valutazione del rischio nei siti archeologici mediante l'applicazione di tecnologie di *remote sensing* e GIS. In A. Russo e I. Della Giovampaola (eds) *Monitoraggio e manutenzione delle aree archeologiche. Cambiamenti climatici, dissesto idrogeologico, degrado chimico-ambientale. Atti del Convegno Internazionale di Studi (Roma, Curia Iulia, 20-21 Marzo 2019)*: 275–279. Bibliotheca Archaeologica 65, Roma: L'Erma di Bretschneider. ISBN: 9788891319470.

Cardona, O.-D., M.K. van Aalst, J. Birkmann, M. Fordham, G. McGregor, R. Perez, R.S. Pulwarty, E.L.F. Schipper e B.T. Sinh 2012. Determinants of risk: exposure and vulnerability. In C.B. Field *et al.* (eds) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*: 65–108. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139177245.005.

De Angeli, S., F. Battistin, M. Serpetti, A. Di Iorio e F.V. Moresi 2020. The RESEARCH project. Soil-related hazards and archaeological heritage in the challenge of climate change. *IOP Conf. Series: Materials Science and Engineering*, no. 949, 012058. https://dx.doi.org/10.1088/1757-899x/949/1/ 012058.

Harmon, B.A., H. Mitasova, A. Petrasova e V. Petras 2019. r.sim.terrain 1.0: a landscape evolution model with dynamic hydrology. *Geosci. Model Dev.* 12(7): 2837–2854. https://doi.org/10.5194/gmd-12-2837-2019.

Keay, S, M. Millett, S. Poppy, J. Robinson, J. Taylor e N. Terrenato 2020. *Falerii Novi*: a New Survey of the Walled Area. *Papers of the British School at Rome* 68: 1–93. https://dx.doi.org/10.1017/S0068246200003871.

Malik, I. 2008. Dating of small gully formation and establishing erosion rates in old gullies under forest by means of anatomical changes in exposed tree roots (Southern Poland). *Geomorphology* 93: 421–436. https://doi.org/10.1016/j.geomorph.2007.03.007.

Meylemans, E., J. Poesen e I. In't Ven (eds) 2014. *The Archaeology of Erosion, the Erosion of Archaeology, Proceeding of the Brussels Conference, April 28–30, 2008. Relicta Monografieën* 9. Brussels: Flanders Heritage Agency.

Millett, M., L. Verdonck, N. Leone e A. Launaro 2019. *Beneath the surface of Roman Republican cities*. Archaeology Data Service. https://doi.org/10.5284/1052663.

Mitasova, H., C. Thaxton, J. Hofierka, R. McLaughlin, A. Moore e L. Mitas 2004. Path sampling method for modeling overland water flow, sediment transport, and short term terrain evolution in Open Source GIS. *Dev. Water Sci.* 55: 1479–1490. https://doi.org/10.1016/S0167-5648(04)80159-X.

Mitasova, H., M. Barton, I. Ullah, J. Hofierka e R.S. Harmon 2013. 3.9 GIS-Based Soil Erosion Modeling. In J.F. Shroder (ed.) *Treatise on Geomorphology*: 228–258. San Diego: Elsevier. https://doi.org/10.1016/B978-0-12-374739-6.00052-X.

Moresi, F.V., M. Maesano, A. Collalti, R.C. Sidle, G. Matteucci e G. Scarascia Mugnozza 2020. Mapping Landslide Prediction through a GIS-Based Model: A Case Study in a Catchment in Southern Italy. *Geosciences* 10 (8): 309. https://doi.org/10.3390/geosciences10080309.

Rao, K. e A. Colette (eds) 2007. *Climate Change and World Heritage*. World Heritage Reports 22. Paris: UNESCO World Heritage Center. https://whc.unesco.org/en/documents/8874.

Sidle, R.C. e T.A. Bogaard 2016. Dynamic earth system and ecological controls of rainfall-initiated landslides. *Earth-science reviews* 159: 275–291. https://doi.org/10.1016/j.earscirev.2016.05.013.

Rome – NE Palatine slopes: open-source methodologies and tools for the analysis of ancient architectures

Emanuele Brienza, Giovanni Caratelli, Lorenzo Fornaciari, Cecilia Giorgi

Emanuele Brienza, University of Enna 'Kore'; emanuele.brienza@unikore.it

Giovanni Caratelli, CNR-ISPC; giovanni.caratelli@cnr.it

Lorenzo Fornaciari, University of Salerno; lorenzo.fornaciari88@gmail.com

Cecilia Giorgi, CNR-ISPC; cecilia.giorgi@cnr.it

Abstract

The excavations of the NE slopes of the Palatine Hill in Rome has been the subject of a long archaeological research. Under the direction of C. Panella, the remains of many architectures and buildings have been unearthed, part of an urban tissue made by a continuous stratigraphic overlapping. Within this framework, our research has been focused on 3D survey, building archaeology, WebGIS and data sharing, experiencing the migration of our data set, analogical and digital, into free and open-source products. This article takes into consideration not only technical aspects but also the possibility of renewing methods for collecting, managing and sharing archaeological information.

Keywords: 3D SURVEY; BUILDING ARCHAEOLOGY; WEBGIS; DATA SHARING.

Introduction

The area of the NE slopes of the Palatine, next to the Colosseum valley, has been investigated by archaeological research carried out by 'Sapienza' University of Rome for more than 30 years. This research gave light to countless material remains, ancient buildings and monuments, testifying an environmental and topographical *continuum* of different urban systems from the Iron Age to contemporary era. All kinds of new and updated methods, techniques, instruments, and devices have been involved, following, on a parallel theoretic line, the most valid and innovative research branches in archaeology. In this brief contribution we want to illustrate some new developments of our research, relating 3D survey, data-sharing and dissemination, using free and open-source tools as a philosophical and deontological matter.

E.B.

3D surveys of the 'Baths of Elagabalus'

CNR-ISPC (formerly CNR-ITABC) is involved in the survey, documentation and study activities of the archaeological structures found on the northeast slope of the Palatine and, especially, those around the so-called 'Baths of Elagabalus'. This involvement is neither fortuitous nor recent, since the authors professionally trained in this sector of the hill, which was – and still is – the best opportunity to address the archaeological heritage of the city of Rome (Caratelli 2013; Giorgi 2013).

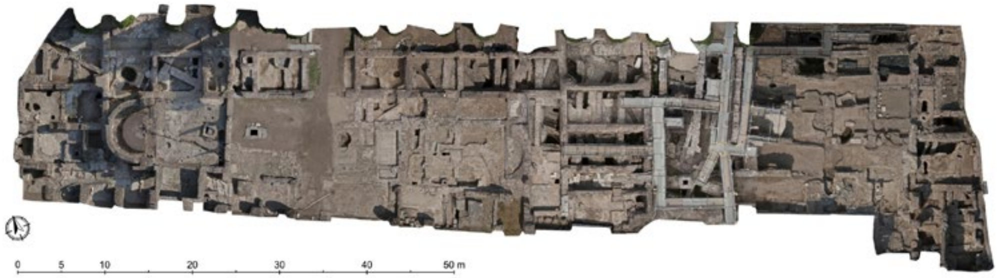


Figure 1: Rome, NE Palatine slopes. Orthophotomosaic produced by CNR at the end of the 2012 excavation campaign.

As is well known, the 'Baths of Elagabalus' – systematically investigated by 'Sapienza' University of Rome from 2007 to 2013 – are the largest portion of the excavation on the northeast slope of the Palatine (conventionally named *Area IV*). This excavation, configured as a logical extension of the fruitful archaeological investigations started in 1986 near the *Meta Sudans* (Panella 1996), is certainly one of the most complex and significant urban archaeology operations of recent years. In addition to the quality of the findings, due to the extraordinary uniqueness of the topographical context, and to the exemplary conduction of the excavation (which trained dozens of professionals), we also must acknowledge the director's openness towards colleagues from other universities and research institutes, which allowed CNR-ISPC's early involvement.

This involvement manifested itself especially in the frenetic documentation activity and analysis of the hundreds of archaeological structures that were found, by using and experimenting with modern survey technologies (laser scanner and digital 3D photogrammetry). To have an idea of the quantity of unearthed wall structures, it is sufficient to look at the orthophoto mosaic produced at the end of the 2012 excavation campaign (Figure 1). It clearly shows the complex layering of walls, resulting from the constant alternation of construction and destructive actions through time. However, it is also an excellent example of the intense and extended activity of photogrammetric survey, systematically performed by CNR-ISPC's Survey Lab at the end of each excavation campaign, from 2010 to 2012. The surveys performed with a 3D laser scanner in the context of two Master and specialization theses should also be considered as part of this activity (Panella, Gabrielli, and Giorgi 2011; Giorgi 2013).

In this context, the recent memorandum of agreement stipulated between CNR-ISPC, 'Sapienza' University of Rome and University of Enna 'Kore' is offering stimuli to launch a new season of studies and research, which should lead to the complete and systematic publication of the archaeological data in a few years. Indeed, a complete edition has not yet been produced. Therefore, in view of this fundamental and essential objective, the direction of the excavation has started a series of initiatives to recover, verify and reorganize the entire set of archaeological documentation that has been produced (it consists of a set of stratigraphical unit records, photographs, and drawings, both in digital and analogue formats, which fortunately have already been partly included in the intra-site GIS of the excavation). This documentation will form the database necessary to address the reading and analytic reconstruction of the various historical phases that, since protohistory until the Middle

Ages, have followed each other in this area, creating an extremely complex and significant stratigraphic and archaeological palimpsest.

In the new knowledge framework that should emerge from this substantial reconsideration of the gathered data, a punctual survey of the discovered archaeological structures surely could not be omitted. This should be understood mainly as a powerful knowledge tool, but also as a spatial-geometric basis to address, organize and support the vast amount of information that was produced. However, this objective, which could otherwise seem 'ordinary', in the context of the 'Baths of Elagabalus' involves several difficulties, generating a veritable challenge. Indeed, from 2013, when the archaeological surveys ended, to the present day, the area has undergone a still-ongoing, extended restoration operation (Asciutti 2015) to enable the public reopening of this sector of the Palatine slope, seriously compromised by the long exposure of the walls to climate factors. Therefore, these new interventions, were added to the heavy and often disrupting manipulations that affected this area throughout history until very recently (consider, for example, the circumstances of the discovery – following the indiscriminate digging of the large area between the Arch of Titus and the Arch of Constantine – and last century's restorations). The new actions have necessarily altered or – in the worst cases – covered the unearthed wall structures.

In these circumstances, and because the excavation is no longer directly accessible in many sectors, the photogrammetric surveys by CNR-ISPC are particularly valuable, the photographic archive being certainly an important and fundamental source of information for structures that are no longer visible or accessible.

In fact, the digital photos taken in past years have recently been recovered, in the context of the above-mentioned agreement. This data set, already used to produce general views of the excavation and in the analysis of areas investigated by the graduation theses, are being reprocessed using current Image-Based Modelling and Rendering techniques to represent the area as it was when the archaeological surveys ended, specifically in 2012.

The objective is to virtually unearth all archaeological structures at the time of discovery, to better understand stratigraphic relations, construction, and period phases, but also to provide a new way to make them usable, interpretable, and available to posterity, especially now that they are no longer visible or are undergoing integrations due to restoration.

The reprocessing of images became necessary since the photogrammetric technique used in the past was based on the principle of stereoscopic vision, using three digital images to generate point clouds with RGB values. The point cloud was then used to generate orthophotos of planimetries and façades. However, the 3D model itself – although useful for measurements and elevations – is not as readable for represented wall facings and planes, being only a point cloud and missing a representation with surfaces. The new 3D model, instead – rendered using 'structure from motion' algorithms – is a realistic high-definition representation, since it contains meshes and photo textures. It allows easy and correct reading of the represented structures and can be easily managed and used, since it can be exported to any format without the need for sophisticated computer equipment (Figure 2).

Figure 2: Rome, NE Palatine slopes. Orthographic view of the new 3D model representing the 'Baths of Elagabalus' and Vigna Barberini's substructions.

Figure 3: Rome, NE Palatine slopes. (a) Portion of the 3D model recently reprocessed with the CNR photographic archive, using multi-image photogrammetry; (b) the same portion of 3D model integrated by processing dataset acquired in the last topographic and photogrammetric survey.

CNR's photographic archive enabled indeed the rendering of this model, but there are some gaps. Therefore, in the last months a new topographic and photogrammetric survey was performed to integrate the model, especially to document some wall façades that were not studied by the graduation and Master theses (Figure 3).

Moreover, a parallel on-going activity is the 3D representation of Vigna Barberini's substructions outlining the examined excavation area to the south. During the already mentioned photogrammetric surveys performed in 2010 and 2012, many frontal shots of the enormous construction were obtained, using an elevated basket along what is now *Via Sacra* (Figure 2).

The research we are preparing to tackle is not simple or obvious at all, but it certainly represents a methodological challenge worthy to be addressed, even if only for the strong knowledge component that always stems from a dedicated and pondered survey operation.

G.C., C.G.

GIS for Building Archaeology data sharing

All the documentation produced by research (file-cards, photos, maps and drawings of many types) required the development of an IT system dedicated to managing, organizing and assembling data, in order to propose new and useful elements of research.

Our archive is managed by an intra-site GIS, used for data-retrieving, spatial analysis and for the elaboration of archaeological themes and/or reconstructive models. Here all evidence and contexts can be distinguished by their unique number and grouped by relative sequences, typologies, and absolute chronology.

Over the years this system has been implemented in software, using different products. Today, our spatial database manages together digital and analogical documentation, following the specific purpose, mainly deontological, to maintain the integrity of the research archive and its history (Brienza 2006; Panella and Brienza 2009; Panella, Fano and Brienza 2015; Brienza 2016: 32–47).

In particular, paying great attention to the issues of open-data and to the dissemination of open-source software and tools, together with the support of a wide community of developers, we have been pushed to adopt open technologies for archaeological practice.

Inside this framework, having in mind the crucial open-data issue (Arizza *et al.* 2018: 9–116), we have experienced the migration of the entire data set and its interrogation system to the open-source geodatabase PostgreSQL/PostGIS, using QGIS as an interface and analysis platform.

The new geodatabase is divided into different sets (schemas) based on the type of data (Tables, shapefiles, rasters) and reproposes the system of relations defined in the conceptual structure.

First, all available data were exported in table format, reorganised and critically summarised. After that, each table was imported into QGIS, linked to the respective vector geometries using the SQL *join* function and then re-exported as new shapefile. Finally, using the plugin DB Manager, each shapefile was easily migrated into the new geodatabase in PostgreSQL/PostGIS (Figure 4).

A special mention should be made for raster data that represent a fundamental component for archaeological analysis and require a special migration procedure. Due to the size of this type of data, it is often preferred not to transfer them into geo-databases and to consume aerial/satellite images and/or cartographic base-map as WMS services.

This process has involved not only technical aspects related to the nature of some formats but also indispensable reflections on the possibility of renewing methodologies and techniques for data collection, management, and analysis.

Focusing on the study of ancient buildings, we proceeded with the reorganization of integrated survey-filing criteria, to collect information related to construction methods such as, for example, structural expedients for static stability, specific materials selection in relation to

Figure 4: Rome, NE Palatine slopes. The new database on PostgreSQL/PostGIS performed on QGIS.

needs or quantification of the work in terms of time and number of the workers, following the guidelines of the archaeology of architecture and building archaeology (DeLaine 1997 for quantification, the proceedings of *Arqueología de la Construcción* published since 2008, the Italian journal *Archeologia dell'Architettura* and the Spanish journal *Arqueología de la Arquitectura;* see also Brogiolo and Cagnana 2012).

We have planned a new database for ancient structural contexts (USM) and designed file-card formats directly by the new QGIS version of our archaeological intra-site GIS, having the chance to easily perform data entry, editing, and retrieving operations in a single environment.

We have normalized the data-entry creating taxonomies and controlled vocabularies: in addiction, a particular section is dedicated to samples analysis (normally their size is 1 m^2), and detailed morphometric data collection from autoptic analysis of wall facades: here each 'constituent' (i.e. brick, block, etc.) is organized by type, size, material, manufacture, finishing. In this way it is possible to enter new information through an updated and pre-packaged entry form.

This meticulous data collection is integrated and strictly linked to orthophotos of wall façades, obtained with photogrammetric techniques with very high accuracy and definition, and then subjected to a vectorization process, validated by on-site examination.

The quantifications can be obtained from the samples following traditional procedures already widely consolidated and widespread, but also using calculation tools available in the GIS platform. Through a series of expressions, it is in fact possible to automatically calculate the dimensions of the components and quickly their degree of homogeneity and variability but also the variable of the constituent/conglomerate ratio. The computed data are then recorded in a specific sample-card giving the interpreted synthesis from which it is possible to elaborate synoptic tables and graphs (Medri *et al.* 2016; Medri 2017: 41–67). For these last ones, we use the QGIS plugin DataPlotly which allows one to build customized charts without using other applications.

Ancient wall stratigraphic data are linked to the 3D surveys at the level of detail, for the evaluation of walls samples, and at a broader level, for the diachronic and typological evaluation of the large architectural complexes of the area. An external connection with the open-source software CloudCompare[1] is a first attempt to show the graphic quality of photogrammetric acquisition and to partially overcome the 2.5-dimensional perspective of GIS.

Particular attention was paid to the question of the 'philological transparency' of information. In the new version of our intra-site GIS a new exploring process of the collected documentation has been planned for users.

Furthermore, we are developing a WebGIS for the online publication of archaeological data already structured in a GIS platform, implementing access control, customizable user interfaces, advanced tools for data querying and analysis and, finally, the functionality to export data. It will be necessary to obtain explicit permission from the *Parco Archeologico del*

[1] https://www.danielgm.net/cc/ (accessed 01/08/2021).

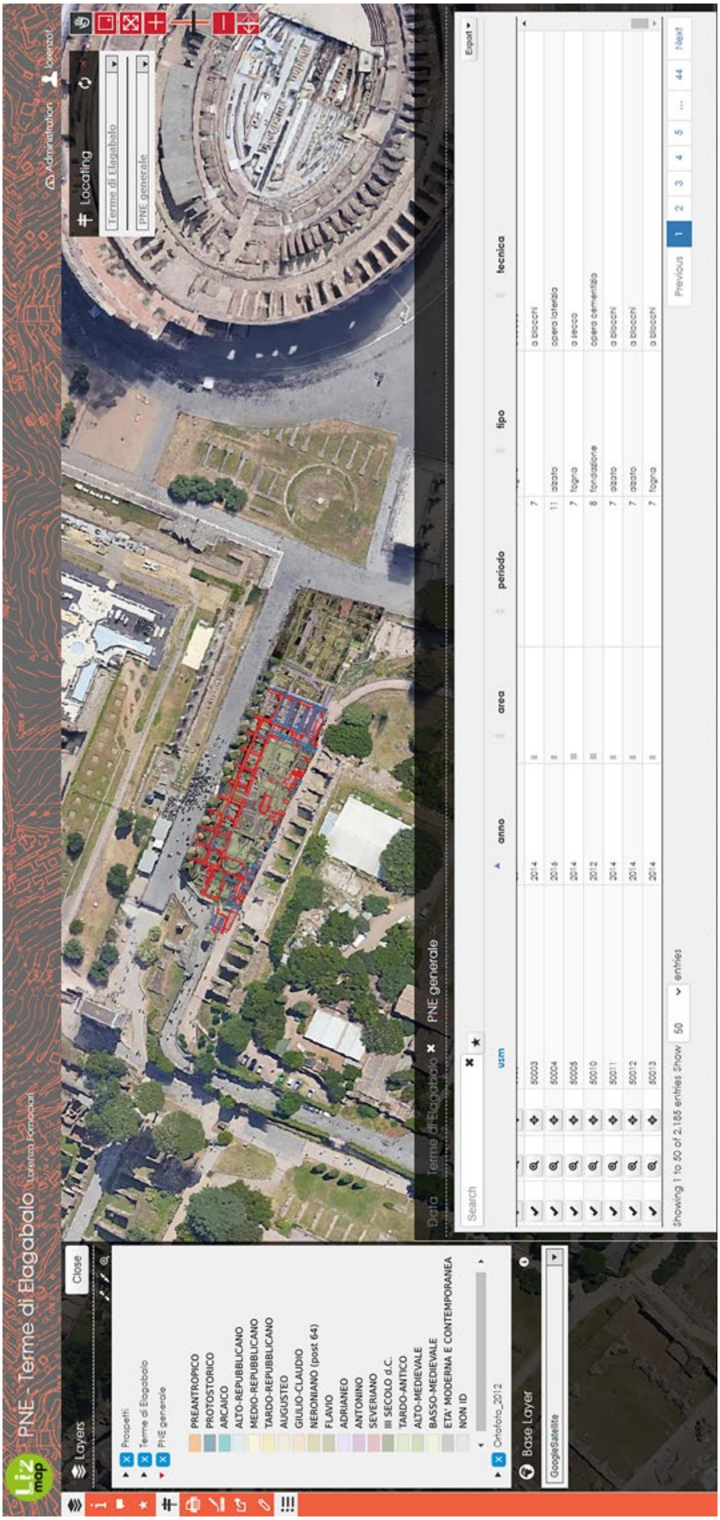


Figure 5: Rome, NE Palatine slopes. A first WebGIS development carried out thanks to the gishosting service of the Gter company (https://www.gishosting.gter.it/home/).

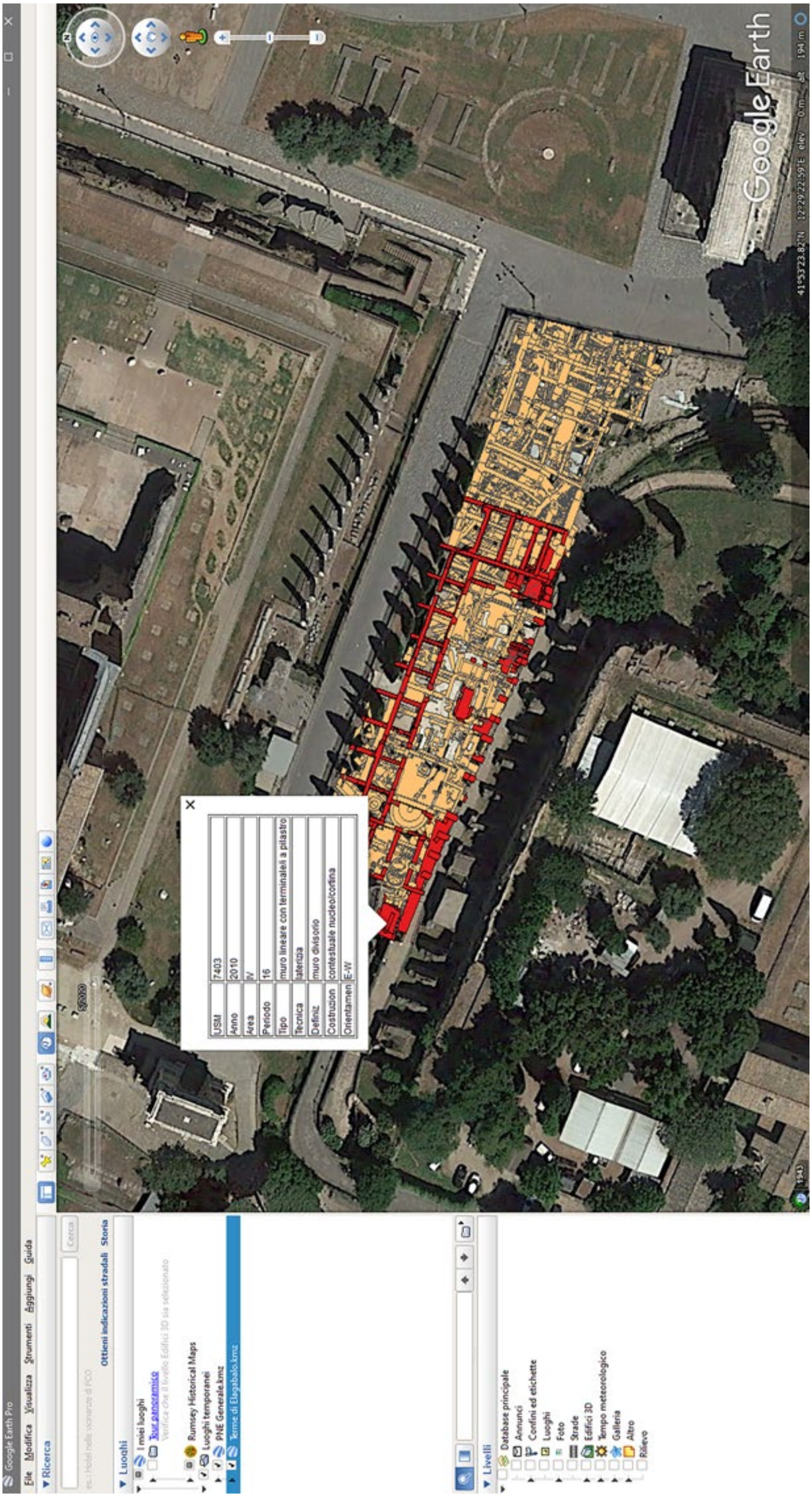


Figure 6: Rome, NE Palatine slopes. An example of exporting data in KML format and their integration in Google Earth platform.

Colosseo for the contents to be published. It is our intention, as well, to adapt our system with other already in use in Rome, such as the SITAR project.

An interesting perspective seems to be offered by the *Forma Romae* project carried out by the Capitoline Superintendency and financed with EU funds from the National Operational Programme for Metropolitan Cities.[2] The project consists in the creation of an information system for the historical, archaeological, and architectural heritage of Rome that integrates the existing geodatabases on the history, archaeology, architecture of the city. Compared to other experiences, the *Forma Romae* proposes a new way of managing the problem of fragmentary and specific information in the field of cultural heritage: the solution adopted does not involve the creation of a new database that absorbs all the others, but rather the creation of a system for accessing existing databases, leaving them unchanged. Through the combined use of data lake and AI solutions, it will be possible to access for each object, all the information contained in many distributed databases. By following this paradigm, our data set will be easily integrated in the network without losing its structure and specificity.

The new WebGIS platform that we are implementing makes use of free open-source software. In particular, QGIS Server and Lizmap Web-Client, were chosen, a combination that offers the possibility of quickly and easily publishing data as they have been defined and configured within the desktop QGIS platform

Once the data to be published have been selected (Tables, rasters, vectors), it is necessary to activate the WMS and WFS services from the project properties for each data type. Subsequently, to configure the on-line platform, we used the Lizmap plugin,[3] whose user-friendly interface allows us to set a series of variables, including interactive functions (measuring, drawing, printing, etc.), data view, query tools and front-end export functions (Figure 5). As for this last function, exporting, different file formats will be made available, including KML, SHP, GeoJSON, etc., according to OGC standards[4] (Figure 6).

We are also working on the integration of the online data set with 3D models: in this task we are testing the use of the 3DHOP software (Potenziani *et al.* 2015: 129–141).

In conclusion, we are trying to create a platform to display and share the complex set of information that fed the interpretative process. Our challenge is not only making the archaeological experience accessible in terms of final results, but also to highlight the process of knowledge and the interpretative reasons for each hypothesis.

L.F., E.B.

Acknowledgements

We very much want to thank the Parco Archeologico del Colosseo, especially Dr.ssa Giulia Giovanetti, for supporting us in this part of the work.

[2] http://www.formaromae.it/ (accessed 01/08/2021).
[3] https://github.com/3liz/lizmap-plugin (accessed 01/08/2021).
[4] https://www.ogc.org/docs/is (accessed 01/08/2021).

Bibliography

Arizza, M., V. Boi, A. Caravale, A. Palombini and A. Piergrossi (eds) 2018, I dati archeologici. Accessibilità, proprietà, disseminazione (Roma, CNR, 23 maggio 2017). *Archeologia e Calcolatori*, 29: 9–116.

Asciutti, M. 2015. Restauro, riqualificazione e valorizzazione di aree di scavo archeologico: Valle del Colosseo/Palatino nord-orientale. *The Journal of Fasti Online: Archaeological Conservation Series* 2: 1–25

Brienza, E. 2006. Un GIS Intra Site per la valle del Colosseo: concetti, metodi, applicazione. *Scienze dell'Antichità,* 13: 123–139.

Brienza, E. 2016. *Valle del Colosseo e pendici nord-orientali del Palatino. La via tra valle e foro: dal dato stratigrafico alla narrazione virtuale (64 d.C.-138 d.C.)*. Roma: Quasar.

Brogiolo, G.P. and A. Cagnana 2012. *Archeologia dell'architettura. Metodi e interpretazioni*. Firenze: All'Insegna del Giglio.

Caratelli, G. 2013. L'aula cruciforme e lo *stibadium* delle Terme di Elagabalo. In C. Panella and L. Saguì (eds) *Valle del Colosseo e pendici nord-orientali del Palatino. Materiali e contesti 1*: 87–120. Roma: Scienze e Lettere.

DeLaine, J. 1997. *The baths of Caracalla: a study in the design, construction, and economics of large-scale building projects in imperial Rome*. Portsmouth: R.I.: Journal of Roman Archaeology.

Giorgi, C. 2013. Terme di Elagabalo. Il balneum tardoantico: studio archeologico e rilievo 3D. In C. Panella and L. Saguì (eds) *Valle del Colosseo e pendici nord-orientali del Palatino. Materiali e contesti 1*: 55–86. Roma: Scienze e Lettere.

Medri, M., V. Di Cola, S. Mongodi and G. Pasquali 2016. Quantitative analysis of brick-faced masonry: example from some large imperial buildings in Rome. *Arqueología de la Arquitectura,* 13: 1–8.

Medri, M. 2017. Studio dei paramenti laterizi delle Mura Aureliane. 1. Osservazioni generali. In D. Esposito *et al.* (eds) *Le Mura Aureliane nella storia di Roma. 1. Da Aureliano a Onorio (Roma, 25 marzo 2015)*: 41–67. Roma: Edizioni Roma TrE-Press.

Panella, C. (ed.) 1996. *Meta Sudans I. Un'area sacra in Palatio e la Valle del Colosseo prima e dopo Nerone*. Roma: Istituto poligrafico e Zecca dello Stato.

Panella, C. and E. Brienza 2009. Il geodabase della valle del Colosseo e del Palatino nord-orientale ed il trattamento digitale del dato archeologico. *Bollettino della Società Italiana di Fotogrammetria e Topografia* 3: 9–18.

Panella, C., M. Fano and E. Brienza 2015. 30 years of urban archaeology: measuring, interpreting and reconstructing. In *Proceedings of the 1st International Conference on Metrology for Archaeology (Benevento, 21–23 ottobre 2015)*: 49–54.

Panella, C., R. Gabrielli, and C. Giorgi 2011. Le Terme di Elagabalo sul Palatino: sperimentazione di un metodo fotogrammetrico 3D applicato allo scavo archeologico. *Archeologia e Calcolatori* 22: 243–260.

Potenziani, M., M. Callieri, M. Dellepiane, M. Corsini, F. Ponchio and R. Scopigno 2015. 3DHOP: 3D Heritage Online Presenter. *Computer & Graphics* 52: 129–141.

Un *workflow* open-source per l'elaborazione delle immagini termiche da drone

Gabriele Ciccone

Gabriele Ciccone, Dipartimento di Storia, Patrimonio Culturale, Formazione e Società.
Università di Roma Tor Vergata. cccgrl01@uniroma2.it

Abstract
In the first phase of the PhD project 'Flying off-site. New methodologies for the analysis of historical landscape' the main goal was to build a workflow to elaborate 3D models and orthophotos by thermographic images taken by drone. In line with the scientific literature outside the archaeological world, a first workflow was developed using proprietary software. The next step consisted in converting this workflow to use exclusively open-source software, aiming both at evaluating an effective possibility of an exclusively open workflow, and also at verifying the pros and cons of open-source solutions compared to proprietary software.

Keywords: THERMAL IMAGING; UAV; WORKFLOW.

Introduzione

Questo articolo ha origine nell'ambito del progetto di dottorato dell'autore 'Flying off-site: nuove metodologie di indagine per l'analisi dei paesaggi storici' (XXXV ciclo, Dottorato in Storia e Scienze filosofiche-sociali, Università Tor Vergata), la cui la parte iniziale è stata dedicata all'individuazione di un *workflow* efficiente per l'elaborazione di rilievi 3D e ortofoto ricavati dai sensori termici.

Nel dettaglio, i sensori termici registrano la Radiazione Termica Infrarossa (TIR), ossia una parte dello spettro elettromagnetico compresa nel *range* dell'infrarosso. Mentre la lunghezza d'onda del visibile (cioè percepibile ad occhio nudo) è compresa tra 0,4 a 0,78 µm, quella dell'infrarosso è maggiore e va da 0,78 a 1000 µm. A sua volta questo *range* è diviso in *near infrared* (NIR, infrarosso vicino, tra 0,78 a 2,5 µm), *mid infrared* (MIR, infrarosso medio, tra 2.5 a 10 µm) e *far infrared* (FIR, lontano infrarosso, tra 10 e 1000 µm). La TIR, cioè la banda dello spettro elettromagnetico rilevata dai sensori termici, si colloca a cavallo delle ultime due categorie, posizionandosi tra 8 e 15 µm (Cool 2015: 5). Si è infatti evidenziato come questo sia il range nel quale può essere rilevata la temperatura di un terreno, sotto forma di radiazione termica, senza che ci sia alcuna interferenza da parte della temperatura dell'aria (Scollar *et al.* 1990: 610).

'*Soil temperature depends on solar radiation and exchange with the atmosphere*' (Périsses, Tabbagh 1981: 169): la temperatura che misuriamo su di un terreno deriva sia dalla radiazione termica infrarossa emessa dal sole (quella che è riuscita ad attraversare indenne, senza essere riflessa indietro, i gas che formano l'atmosfera) che dalla radiazione termica infrarossa che viene assorbita e riflessa dai materiali presenti sopra o vicino il terreno, e dal terreno stesso. Infatti, le diverse componenti fisiche e chimiche di un terreno non omogeneo comportano un differente assorbimento e una conseguente emissione della radiazione solare che si evidenzia in una differente radiazione termica misurata in superficie (Cool 2015: 6). Ed è proprio sulla base di questo fenomeno che si fondano le analisi delle immagini termiche con finalità archeologiche: i materiali che si trovano sopra e sotto la superficie di un terreno assorbono, emettono o riflettono i raggi infrarossi in modi differenti, in base alla propria composizione, densità e contenuto di umidità. Secondo tale principio una gran quantità di elementi archeologici può essere teoricamente individuata nelle immagini termiche (Casana *et al.* 2017: 311).

Gli archeologi sono a conoscenza, già dagli anni '70 dello scorso secolo, delle potenzialità delle immagini termiche a infrarossi per l'individuazione di strutture non visibili in superficie, concentrazione di materiale, o elementi del paesaggio come strade, confini dei campi coltivati, paleoalvei, ecc. L'ostacolo principale che ha determinato uno scarso utilizzo di questa tecnica fino a poco tempo fa è individuabile nel costo elevato delle missioni dovuto alle prime tecniche di ripresa termica: tra gli anni '70 e '80 infatti si faceva uso di un ampio radiometro a scansione raffreddato con azoto liquido che registrava le immagini termiche su lunghi rotoli di pellicola.

Una nuova fase si è avuta nei primi anni Duemila, quando iniziarono ad essere utilizzate manualmente camere termiche convenzionali utilizzate da un elicottero o montate su di una piattaforma per analizzare le differenti concentrazioni di manufatti sparsi sul terreno (Buck *et al.* 2003).

In seguito diversi studiosi iniziarono ad utilizzare le stesse camere termiche, montandole questa volta sotto ad un aquilone (Wells 2011) o a un pallone aerostatico ad elio (Giardino and Haley 2006).

Nel recentissimo periodo le innovazioni tecnologiche hanno permesso l'elaborazione di camere termiche commerciali di dimensioni e peso minori ma migliorate nella risoluzione. Questo ne ha permesso l'installazione al *payload* degli UAV (*Unmanned aerial vehicle*), diminuendo da un lato il costo delle missioni svolte per avere immagini termiche ad alta risoluzione e permettendo dall'altro la capacità di programmare e svolgere i voli in breve tempo, elemento fondamentale quando c'è bisogno di effettuare voli in un preciso arco temporale. Ad oggi, i migliori esempi di rilievi termici da drone in campo archeologico vengono dai lavori svolti da J. Casana e la sua equipe in New Mexico, Illinois (Casana *et al.* 2014, Casana *et al.* 2017) e New Hampshire (Hill *et al.* 2020). In particolare la ricerca svolta nel sito di Blue J in New Mexico ha permesso di evidenziare l'importanza di avere più voli durante la stessa giornata in modo da individuare la fascia oraria migliore, in base alle variazione del ciclo termico giornaliero. In questo caso i ricercatori hanno effettuato quattro voli con camera termica tra il 20 e il 21 giugno del 2013 (ore 9:58, 5:18, 6:18, 7:18) e un volo con camera standard in RGB (ore 5:45) e i risultati hanno mostrato che solo nel volo effettuato prima dell'alba, alle 5:18, risultavano evidenti il maggior numero di anomalie di carattere archeologico, probabilmente perché le pietre delle creste murarie e i relativi crolli trattengono più a lungo il calore rispetto al suolo desertico circostante e questo risulta più visibile nelle ore più fredde della mattina rispetto che l'inizio della notte, quando il terreno trattiene ancora il calore dell'irraggiamento solare giornaliero (Casana *et al.* 2014). Anche il lavoro svolto dalla stessa equipe nel sito di Enfield Shaker Village nel New Hampshire, ha evidenziato l'importanza di effettuare una ricognizione termica multi temporale al fine di avere un'analisi termografica completa dell'area di indagine. In questo caso i voli sono stati svolti in differenti orari dello stesso giorno, ma anche

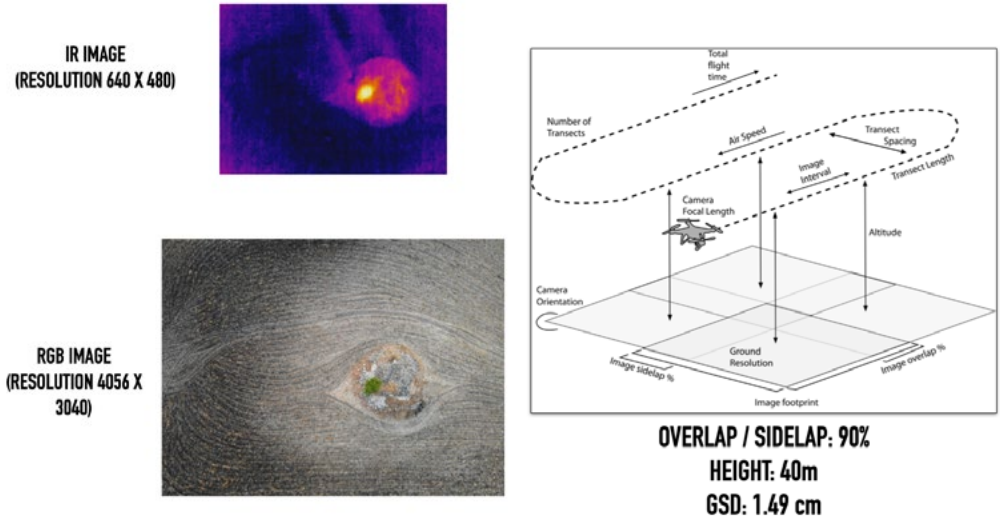


Figura 1: Esempi di immagine RGB e IR e schema dei voli effettuati.

in differenti mesi dell'anno (ottobre 2016, maggio e giugno 2017, ottobre 2017), e i risultati ottenuti hanno mostrato, per quest'area, la massima visibilità delle anomalie archeologiche nell'arco temporale che va da 'subito dopo il tramonto' ad 'alcune ore dopo', mentre il contrasto, seppur ancora visibile, andava diminuendo nei voli effettuati 'prima dell'alba', per poi scomparire quasi del tutto 'dopo il sorgere del sole' (Hill *et al.* 2020: 11–12).

Metodologia

In linea con gli esempi appena citati si è deciso, nella prima fase della ricerca, di testare l'UAV con termocamera programmando riprese termiche della stessa area con voli effettuati in diversi momenti della giornata, al fine di valutare le variazioni termiche dovute ai differenti orari di ripresa. A causa della bassa risoluzione delle immagini termiche ricavate dalla camera termica FLIR LEPTON 3.5, sono stati effettuati voli aventi strisciate sovrapposte con una sovrapposizione di 90% tra gli scatti, lateralmente e longitudinalmente (Figura 1).

La seconda fase, quella di elaborazione delle immagini, è stata svolta con differenti software sia proprietari che open-source, al fine di valutare le potenzialità in termini di precisione del dato e di velocità dell'elaborazione: è apparso subito evidente come l'orario di ripresa determinasse la buona riuscita o meno della fase di allineamento delle immagini, in tutti i software fotogrammetrici proprietari e open-source testati.

Le elaborazioni hanno evidenziato una situazione differente tra i set di immagini scattate in condizioni di buona luminosità, rispetto ai set ricavati dalle riprese effettuate dopo il tramonto e prima dell'alba.

Se nel primo caso, infatti, tutti i software hanno correttamente individuato i punti in comune tra le immagini in fase di allineamento, riuscendo ad elaborare sia le ricostruzioni 3D che le

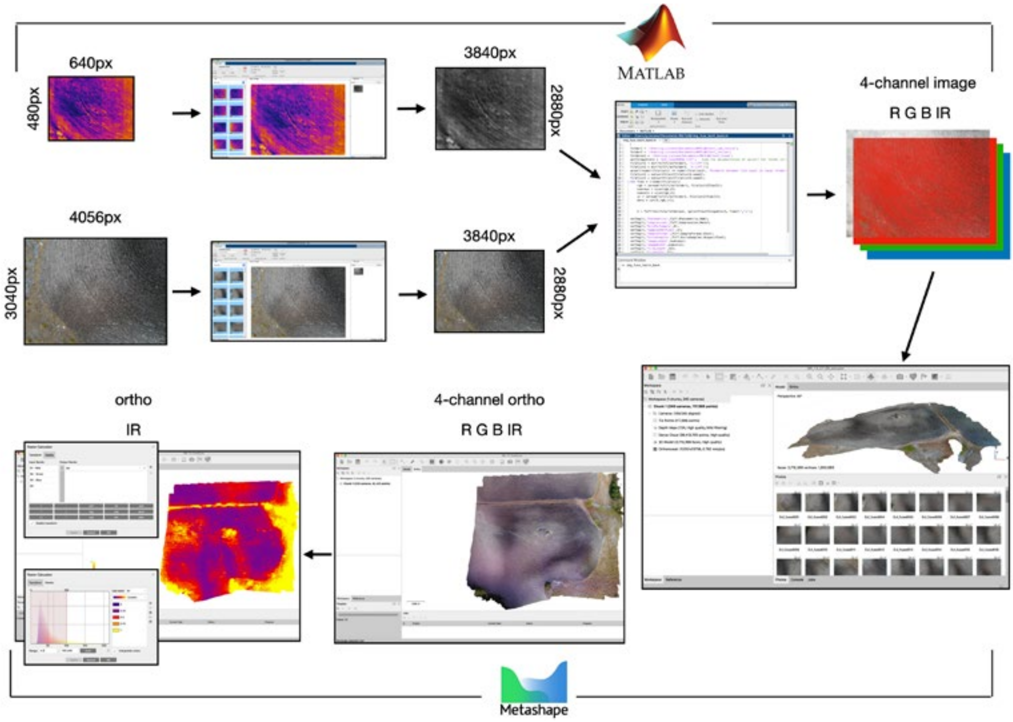


Figura 2: Schema del workflow con software proprietari per l'elaborazione di immagini termiche.

successive ortofoto, nel secondo, la scarsa luminosità delle immagini non ne ha permesso un corretto allineamento, impedendo l'avanzamento delle elaborazioni con tutti i software proprietari e open-source testati.

In linea con la letteratura scientifica in materia di fotografia termica al di fuori del mondo archeologico (Yang and Lee 2019), è stato elaborato un *workflow* che permette di unire le immagini RGB e IR (termiche), in formato JPG, in singole immagini in quattro bande (R, G, B, IR) in formato TIFF, al fine di riuscire ad elaborare le immagini TIFF con i software di fotogrammetria per la ricostruzione del rilievo 3D e la successiva elaborazione delle ortofoto, da cui infine estrarre la banda relativa all'immagine termica (Figura 2).

Inizialmente, il workflow è stato elaborato utilizzando software proprietari: si è utilizzato Matlab (software con licenza chiusa ma distribuito gratuitamente per la ricerca universitaria)[1] per le fasi di conversione in scala di grigi e ridimensionamento delle immagini IR (da 640x480 a 3840x2880), per il taglio delle immagini RGB (da 4056x3040 a 3840x2880), e infine per l'unione nelle singole immagini TIFF a 4 bande. In seguito le immagini sono state processate in Agisoft Metashape,[2] dalla ricostruzione 3D fino all'estrazione della singola ortofoto termica.

L'utilizzo del software Matlab, nella fase di trasformazione delle immagini originali, ha facilitato di molto l'elaborazione dei codici dedicati, grazie all'utilizzo dell'*Image Batch*

[1] EULA: https://it.mathworks.com/pricing-licensing.html?intendeduse=comm (accesso 21/07/2021).
[2] EULA: https://www.agisoft.com/pdf/metashape-pro_eula.pdf (accesso 21/07/2021).

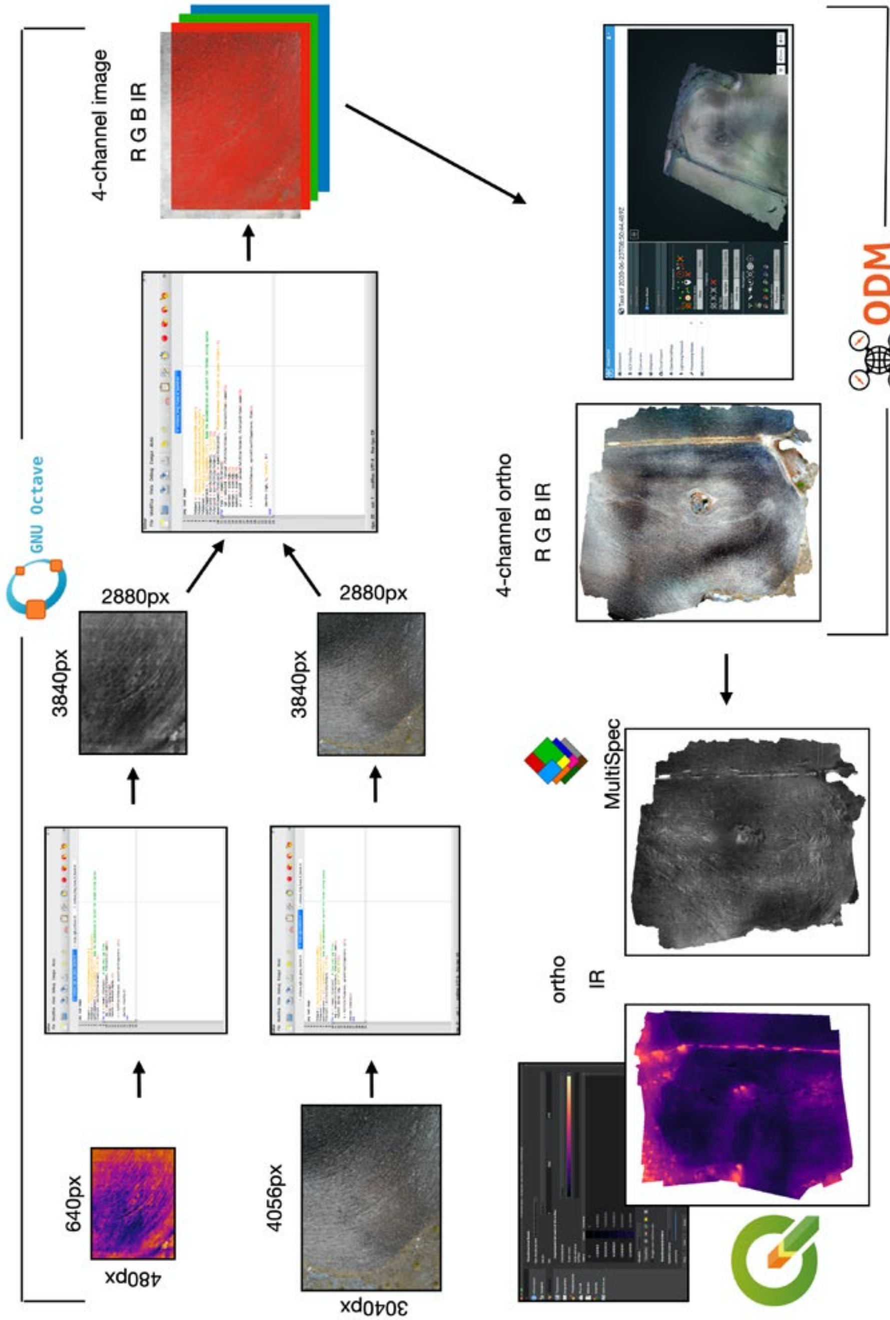


Figura 3: Schema del workflow con software *free* e *open-source* per l'elaborazione di immagini termiche.

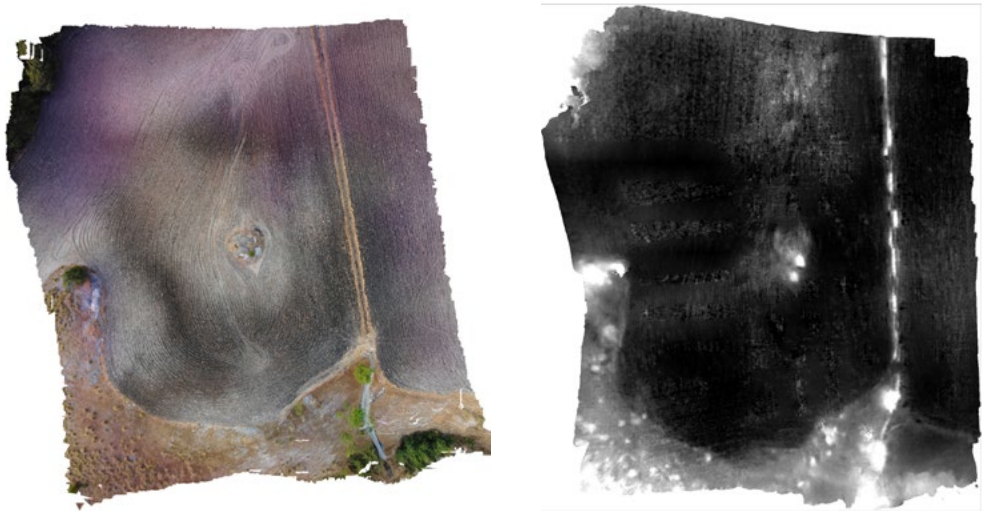


Figura 4: (a) Ortofoto in 4 bande (R, G, B, IR); (b) Ortofoto nella singola banda IR.

Processor, un'applicazione di Matlab che permette in maniera automatica di elaborare lo stesso codice per tutte le immagini presenti in una cartella. Il successivo passaggio, quello relativo alla fusione delle immagini IR e RGB (in formato JPEG) in immagini TIFF a quattro canali, non potendo passare tramite *Image Batch Processor*, ha determinato la compilazione di un codice leggermente più articolato. Una volta ottenute le immagini TIFF a quattro canali, queste sono state caricate nel software di fotogrammetria Metashape il quale è riuscito facilmente nell'operazione di allineamento potendo determinare i punti in comune dai livelli RGB delle singole immagini. Questo ha permesso, quindi, le successive fasi di elaborazione della nuvola di punti, del modello 3D e delle ortofoto. Il passo finale è stato poi quello di estrapolare dall'ortofoto solo l'informazione relativa al quarto canale, quello delle immagini IR. È stato possibile effettuare questo step direttamente all'interno del software Metashape, il quale permette la distinzione dei diversi canali di un'immagine TIFF.

A obiettivo raggiunto, nella successiva fase della ricerca è stato elaborato un medesimo processo di lavoro utilizzando soltanto software free o open-source, sia per valutare un'effettiva possibilità di *workflow* esclusivamente aperto, che per testarne pregi e difetti rispetto al procedimento svolto con software proprietari (Figura 3).

In questa fase sono stati testati differenti software sostitutivi di Matlab (GNU Octave,[3] Scilab,[4] SageMath,[5] R[6]) arrivando infine a preferire GNU Octave in quanto il più simile nella sintassi a Matlab, cosa che ha permesso di non dover rielaborare né stravolgere troppo le funzioni già utilizzate.

[3] https://www.gnu.org/software/octave/index (accesso 21/07/2021).
[4] https://www.scilab.org/ (accesso 21/07/2021).
[5] https://www.sagemath.org/ (accesso 21/07/2021).
[6] https://www.r-project.org/ (accesso 21/07/2021).

Una prima differenza, rispetto al *workflow* fatto con soli software proprietari, riguarda l'assenza, all'interno di GNU Octave, di un applicativo con le stesse funzioni automatiche dell'*Image Batch Processor* di Matlab. Questa differenza ha determinato però soltanto un'integrazione dei codici utilizzati in Octave, con un'aggiunta di poche righe relative all'individuazione delle rispettive cartelle contenenti le immagini IR e quelle RGB e la successiva indicazione delle cartelle in cui andare a salvare le immagini trasformate. Un'ulteriore differenza, relativa sempre all'utilizzo di GNU Octave come software sostitutivo di Matlab, è dovuta all'assenza in Octave delle funzioni relative alla libreria TIFF. Questo ha determinato, non un'aggiunta di codice, come nel caso precedente, ma una totale sostituzione di alcune righe, cambiando le funzioni della LibTIFF con altre funzioni presenti in Octave.[7] Una volta completato questo passaggio e ottenute le immagini TIFF a quattro livelli si è utilizzato il software OpenDroneMap (tramite l'interfaccia WebODM[8]) per l'allineamento delle immagini e le successive elaborazione della nuvola di punti, del modello 3D e delle ortofoto (Figura 4a).

Tuttavia, per l'estrazione del solo canale IR dalle ortofoto (Figura 4b), non essendo presente tale funzione in ODM, sono state testate differenti possibilità, tra le quali QGIS[9] e il freeware Multispec[10] sono risultate le migliori.

Conclusione

Innanzitutto va evidenziato che entrambi i *workflow* testati in questa prima fase della ricerca risultano funzionali al raggiungimento dell'obiettivo finale: sia il *workflow* fatto di software proprietari che quello che utilizza solo software free o open-source, permettono l'elaborazione di ortofoto termiche da immagini scattate in momenti della giornata con scarsa o nulla luminosità. L'utilizzo di uno dei due *workflow* permette però di confrontare tutte le immagini scattate prima dell'alba o dopo il tramonto con quelle scattate in presenza di sole (Figura 5), permettendo quindi l'individuazione dell'orario migliore in cui effettuare i successivi rilevamenti termici: quello in cui le anomalie termiche nel terreno risultano cromaticamente più evidenti sulle immagini.

Per quanto riguarda le differenze tra Matlab e GNU Octave, evidenziate sopra, dai test si evince che non risultano essere determinanti. L'assenza di un applicativo nativo in Octave come l'*Image Batch Processor* di Matlab è facilmente superato aggiungendo alcune righe di codice e il tempo di utilizzo risulta essere piuttosto bilanciato: il tempo che si impiega in Octave nel modificare le righe di codice relative all'esatta indicazione delle cartelle da cui caricare le immagini e in cui salvare a fine processo, risulta non essere maggiore di quello impiegato nel ricercare le stesse cartelle tramite *Image Batch Processor*. Mentre, per quanto riguarda l'assenza della LibTIFF tra le funzioni di Octave rispetto a Matlab, questo determina una mancanza finale di alcuni metadati aggiuntivi nelle immagini TIFF a quattro canali, che comunque non determinano differenze né nel canale relativo alle informazioni termiche né nella successiva elaborazione tramite software fotogrammetrico.

[7] https://github.com/gabcicc/OSSWorkflowForAllignIRImages (accesseo 21/07/2021).
[8] Licenza GNU Affero General Public License v3.0, https://github.com/OpenDroneMap/WebODM (accesso 13/08/2021).
[9] Licenza GNU GPL, https://www.qgis.org/it/site/about/index.html (accesso 13/08/2021).
[10] Licenza Apache License, Version 2.0, https://github.com/larrybiehl/MultiSpec (accesso 13/08/2021).

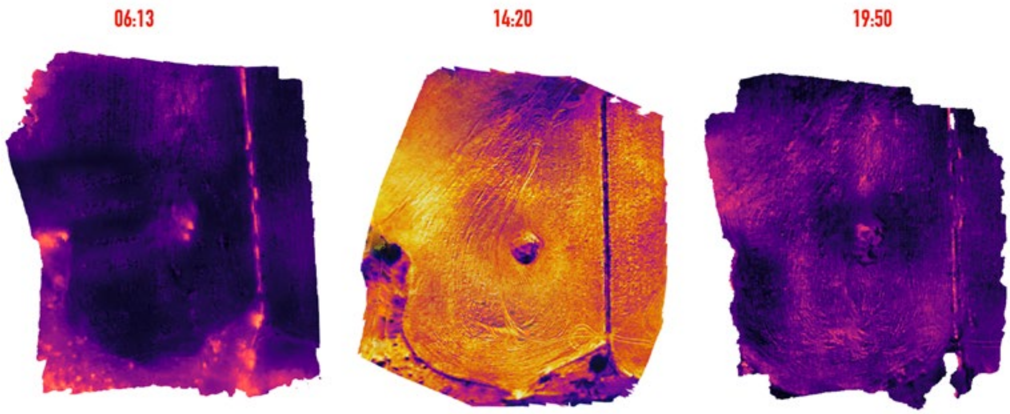


Figura 5: Confronto di ortofoto in banda IR in differenti orari della stessa giornata.

La principale e, ad oggi, maggiormente determinante differenza riguarda invece il tempo di elaborazione del software fotogrammetrico. I pro a favore della scelta di utilizzare WebODM, rispetto ad altri software di fotogrammetria free o open-source (come Meshroom,[11] PPT,[12] OpenMVP,[13] MicMac,[14] ecc.), sono stati principalmente il suo essere multi piattaforma (quindi utilizzabile su sistemi Linux, Mac e Windows) e la sua semplicità di utilizzo, rendendolo un software facilmente accessibile a chiunque.

Il contro, ad oggi unico ma determinante nel tempo di impiego, è quello di non sfruttare le GPU (Graphics Processing Unit), ma di utilizzare ancora unicamente le CPU per l'elaborazione fotogrammetriche (Burdziakowski 2017). Questo ha fatto si che, nei test effettuati, il software abbia impiegato circa tre volte il tempo utilizzato dai maggiori software fotogrammetrici proprietari (Metashape, Pix4D,[15] ecc.). La differenza di tempo può essere fino a otto volte maggiore in macchine meno performanti (Burdziakowski 2017: 105).

Nell'economia di un lavoro con molti voli effettuati, ognuno con un alto numero di immagini, è indubbio che un rallentamento di circa tre volte il tempo di elaborazione finale non sia da sottovalutare; va però subito evidenziato come questo problema probabilmente sarà superato abbastanza in fretta: lo sviluppo di librerie per l'utilizzo delle GPU è nella *roadmap* degli sviluppatori di ODM da alcuni anni, e la popolarità in aumento del software, e di conseguenza del bacino di utenti, sviluppatori e finanziatori, fanno ben sperare in un rapido miglioramento dei tempi di elaborazione.

Infine, l'ultima differenza riscontrata tra i due *workflow* riguarda l'ultimo passaggio: l'estrazione del solo canale IR dalle ortofoto. Nel *workflow* con software proprietari questo passaggio viene svolto sempre all'interno del software Metashape, mentre nel *workflow* FLOS viene richiesto

[11] https://alicevision.org/#meshroom (accesso 21/07/2021).
[12] https://github.com/archeos/archeos-ppt-gui (accesso 13/08/2021).
[13] https://openmv.io/ (accesso 21/07/2021).
[14] micmac.ensg.eu (accesso 21/07/2021).
[15] https://www.pix4d.com/ (accesso 21/07/2021).

l'utilizzo di un altro software. In questo caso sono stati utilizzati fruttuosamente sia QGIS che Multispec[16] e i test evidenziano che più che un contro rispetto al primo *workflow*, questo passaggio in un software aggiuntivo deve essere visto come un'opportunità, in quanto sia QGIS che Multispec permettono diverse possibilità di elaborazione delle ortofoto IR che Metashape non presenta.

Bibliografia

Buck, P.E., D.E. Sabol and A.R. Gillespie 2003. Sub-pixel artifact detection using remote sensing. *J. Archaeol. Sci.* 30 (8): 973-989.

Burdziakowski, P. 2017. Evaluation of Open Drone Map Toolkit for geodetic grade aerial drone mapping – Case study. *17 International Multidisciplinary Scientific GeoConference SGEM 2017: Informatics, Geoinformatics and Remote Sensing*, no. 23 (2017): 101–109. https://doi.org/10.5593/sgem2017/23/s10.013.

Casana, J., J. Kantner, A. Wiewel e J. Cothren 2014. Archaeological Aerial Thermography: A Case Study at the Chaco-Era Blue J Community, New Mexico. *Journal of Archaeological Science* 45 (maggio): 207–219. https://doi.org/10.1016/j.jas.2014.02.015.

Casana, J., A. Wiewel, A. Cool, A.C. Hill, K.D. Fisher e E.J. Laugier 2017. Archaeological Aerial Thermography in Theory and Practice. *Advances in Archaeological Practice* 5 (4): 310–27. https://doi.org/10.1017/aap.2017.23.

Cool, A. 2015. Aerial Thermography in Archaeological Prospection: Applications & Processing. *Master's thesis, Department of Anthropology*. Fayetteville: University of Arkansas.

Giardino, M.J. e B. Haley 2006. Airborne remote sensing and geospatial analysis. In J. Johnson (ed.) *Remote Sensing in Archaeology: an Explicitly North American Perspective*: 47–77. Tuscaloosa: University of Alabama Press.

Hill, A.C., E.J. Laugier e J. Casana 2020. Archaeological Remote Sensing Using Multi-Temporal, Drone-Acquired Thermal and Near Infrared (NIR) Imagery: A Case Study at the Enfield Shaker Village, New Hampshire. *Remote Sensing* 12 (4): 690. https://doi.org/10.3390/rs12040690.

Perisset, M.C. e A. Tabbagh 1981. Interpretation of thermal prospection on bare soils. *Archaeometry* 23 (2): 169–187. https://doi.org/10.1111/j.1475-4754.1981.tb00304.x.

Scollar, I., A. Tabbagh, A. Hessse e I Herzog 1990. Archaeological Prospecting and Remote Sensing. *Topics in Remote Sensing* 2. Cambridge: Cambridge University Press.

Wells, J. 2011. Kite aerial thermography. *Int. Soc. Archaeol. Prospect. News* 29: 9–10.

Yang, Y. e X. Lee 2019. Four-Band Thermal Mosaicking: A New Method to Process Infrared Thermal Imagery of Urban Landscapes from UAV Flights. *Remote Sensing* 11 (11): 1365. https://doi.org/10.3390/rs11111365.

[16] https://engineering.purdue.edu/~biehl/MultiSpec/ (accesso 13/08/2021).

Analysis of urban mobility in 18th-century Rome: a research approach through GIS platform

Renata Ago, Domizia D'Erasmo

Renata Ago, Sapienza University of Rome; renata.ago@uniroma1.it
Domizia D'Erasmo, Sapienza University of Rome; domizia.derasmo@uniroma1.it
Correspondence: domizia.derasmo@uniroma1.it

Abstract
The Geographic Information System (GIS) has been commonly used to analyse and elaborate the big data of urban mobility in modern cities. At the same time a range of studies has been utilizing it to examine various kinds of movements in old times, especially those on a large territorial scale. This article aims to highlight the use of GIS in the study of urban mobility of Rome during the 18th century, drawing on textual evidence and archives. The results of this work will offer new hypothesises around the most trafficked zones and the main routes of movement traveled in the Italian capital during this period.

Keywords: URBAN MOBILITY; MOVEMENT DATA; MOVEMENT ANALYTICS; ROAD NETWORK; HISTORICAL PATHS; 18TH CENTURY; MODERN HISTORY; GEOGRAPHICAL INFORMATION SYSTEMS.

Introduction

In the last few years, many different studies have addressed the use of Geographic Information Systems (GIS) to carry out analyses on the mass movement data. Undoubtedly, geo-informatics is the most recent source for the management of mobility and transport planning. However, the data set available to people who study these modern dynamics is different from that used by those who study the same phenomena in older times. It is essential to recall that the use of GIS and analyses of Least Coast Paths (LCP) is not lacking in the studies of the mobility of the past (van Lanen *et al.* 2015; Fonte *et al.* 2017). Especially, the LCP approach that utilizes the digital models of the territories and combines them with other factors (rivers, enemy borders, etc.), is often applied. Nevertheless, the review of the available bibliography, demonstrates that this type of methodology has been used widely for larger-scale research (regional or national). On the other hand, those who studied smaller contexts, such as cities, did not base their approach on information that precisely describing paths (Branting 2007), as in the present case.

The main concern of this research is to study the use of urban space through the reconstruction of people's movement and transportation of Rome during the 18th century. Therefore, it is necessary to list the sources from which data on urban mobility have been extracted. The crucial texts are booklets which were published to describe ceremonial events. These booklets provide information about the itinerary of imperial and religious processions in Rome. In addition, there are sources written by the criminal and civil justice (about 600 documents). The study of this sort of evidence gives endless information on the toponomy of streets and people's movements as the incidents have been registered in detail. For example, during

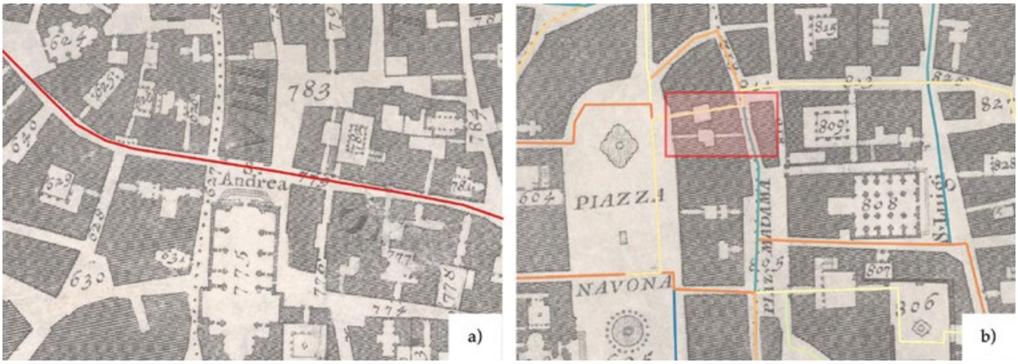


Figure 1: a) A section of the *strada della Valle* as depicted in Nolli's map; b) Example of a path that involves passing through a courtyard of a building in *piazza Navona* (base map: *Nuova Topografia di Roma*).

criminal investigations, the testimony of people demonstrates where a certain incident happened, where people came from, where they lived, etc. The case presented in this article focuses precisely on these data source. Not only they tell stories about the old times and a different urban layout of the city, but also provide precise descriptions about a certain path and the movements from start to end of it.

Therefore, considering these data, one of the goals of this study was to highlight the available information as much as possible using a GIS-based methodology. For this reason, each path has been recreated following step by step the description given by the sources. The outcome of this mapping made it possible to register all the ceremonial movements and daily transports and to evaluate its impact on the city and people's lives.

Last but not least, it should be emphasized that this research is entirely based on the use of an open-source solution through the use of QGIS. This choice is also in line with the release of an open-data test dataset to offer the possibility to replicate the methodology presented in this paper[1] and assumes a simple approach to the study of urban mobility that can then be reproduced or implemented by other researchers.

Materials and Methods

The cartography

First of all, an effective way to reproduce the urban grid of 18th century Rome was sought. Regarding this, in the field of studies investigating the past, one of the most effective responses to the challenge of reproducing the old appearance of places has been to exploit historical cartography. This constitutes a powerful means, especially when combined with the use of GIS (Bogdani 2021; D'Erasmo 2019; Rumsey and Williams 2001).

For this reason, the *Nuova Topografia di Roma* made by the surveyor G.B. Nolli in the middle of the 1700s was used in this project (Bevilacqua 2013). This map (downloaded from Rumsey

[1] https://codeberg.org/lad/roma-nel-settecento (accessed 29/07/2021).

2021) was georeferenced through the plugin of QGIS 'Freehand Raster Georeferencer' (Guilhem 2016) and used as a basic map for drawing the itineraries on it. The map has an almost perfect correspondence with the modern urban grid, since the historical center of Rome has undergone few changes in the last 250 years. Moreover, this map provides precise information about the alleys, secondary entrances in palaces, etc. that no longer exist. In order to recognize all these places, the index available within 'The Nolli Map' project of the University of Oregon GIS Laboratory was also consulted (Tice and Steiner 2021). All these topographical elements construct a scenery of daily movements of Roman citizens, whereby, the paths have been mapped. This has allowed to work in absolute respect of the descriptions of the city given by the sources.

The other type of base cartography used is satellite imagery (Bing Satellite in this case). However, their use was limited to few cases where Nolli's map could not provide complete coverage for the entire itinerary (Figure 4).

The data structure

The mobility data differed by route type and by source. For the processional paths there is a precise sequence of stages. However, even in cases like these where a path is well described, there is the possibility of making assumptions about some roads that may or may not have been taken between stages (when they are not mentioned in the sources). An example could be the procession of the S.mo *Crocifisso* in 1700. The sequence of the stages indicated on the path by the booklets is: *S. Marcello al Corso*, *strada del Gesù*, *piazza dei Cesarini*, *strada della Valle*, *piazza Pasquino*, *ponte S. Angelo*, *Borgo Nuovo* and *S. Pietro*. Furthermore, with the analysis of the tales of the chronicler of that time, is possible to highlight further details: a huge machine was brought to the procession that needed more than sixty men to support it. The chronicler also mentions that when the brothers passed through the *strada della Valle* (Figure 1a), there were great difficulties in transporting the device, and to make it pass had to turn it to one side (Ago 2021). This procession was not the only one to use with such devices. Consequently, during the vectorization of these events, only the wide streets were selected when discussing their use.

Regarding the paths of individuals, a great part of their paths was documented by the texts. These reported a starting point, intermediate points, and an end point. Where there was a lack of information regarding the roads taken to get from point A to point B, alleys and side entrances in the courtyards of the buildings were favored (Figure 1b). The nature of this choice is dictated by the movements of all these men and women. In fact, they were released from having to follow an 'obligatory' path unlike the cases of processions. In addition, they most likely preferred to take shorter routes to get to the point of arrival.

The preference for alleys was not the only arbitrary choice made during the research. In fact, it happened that sometimes the sources spoke of places not mentioned in Nolli's index, and obviously no longer extant. In these cases, always trying to remain faithful to the sources, an arbitrary decision was taken about the position of these places. One example is that of Giuseppe Alessandri, a coachman who denounced an assault in 1749. From his complaint we know that the man left from *piazza Tomassi* for *piazza S. Carlo*. On the way he stopped at the *osteria del Turchetto* in *via Frattina* and was robbed there. Nolli's map does not indicate the location of the aforementioned inn; we chose to place it between *via Frattina* and *piazza di*

Figure 2: Result of the vectorization of all analysed paths by GIS platform.

Spagna. Similarly, for the washerwomen's paths the sources often spoke of the point of arrival as fountains located near *Borgo S. Spirito*. In that case it was decided to use the western bank of the Tiber river in front of the remains of the *ponte Trionfale* as the arrival point. These are only a few examples of arbitrary choices, but they serve to make clear how decisive our intervention has sometimes been.

Based on the foregoing, the paths were divided into those of ceremonial and of private citizens. In terms of numbers, the available data mapped around 30 paths used during the ceremonial events and 478 paths undertook by the individual citizens (Figure 2). Moreover, there are 1,221 stages make up these itineraries (Figure 2). All these data are gathered in two vector layers, one for each type of the paths. Each layer contains two types of vectors: lines, reproducing the paths, and points, reproducing the stages. Regarding the vector of lines, the vectorization procedure was to keep always active the functions 'snap' and 'tracing' to hook the vertices of the various lines and guarantee the overlap between them. Through this method, we have been able to reproduce the 'heaviness' of the paths virtually, which means when a certain trace of a street is undertaken by more than one person. Following this, it is possible to select a trace of the street and pin a number as a record to show the crossing of the individuals by that point (Figure 3).

Last but not least, each of these vectors has been associated with the following data: name and surname of those travelling the path or the name of the procession, year, job, type

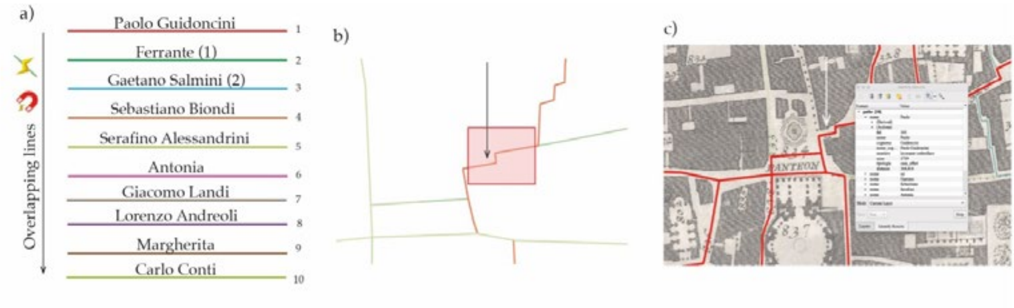


Figure 3: a) List of ten paths of private citizens passing through a street adjacent to *piazza della Rotonda*; b) vectorization result; c) list of returned records (base map: *Nuova Topografia di Roma*).

(ceremonial; home-work; home-business), and a calculation field indicating the distance travelled in meters (Table 1).

fid	name_surname	job	year	typology	distance
1	Agnese	unidentified	1739	home-work	2,336.432
2	Agostino Vigna	*scarpinello*	1739	home-work	2,892.471
3	Agostino Vigna	*scarpinello*	1739	home-business	2,892.471
4	Alessio Preti	*sbirro*	1739	home-work	835.591
5	Anastasio Ricci	*muratore*	1748	home-business	162.539
6	Anonymus (1)[1]	*beccamorto*	1739	home-work	1,077.47
7	Anonymus (2)	*maniscalco*	1739	home-work	290.135
8	Antonia	*fruttarola*	1739	home-business	2,806.896
9	Carlo Tognacci	*ortolano*	1739	home-work	1,147.85
10	Ester	*bottonara*	1739	home-work	881.372

[1] Sometimes more than one path of the same typology is associated with a person. The numbers are an expedient to count the different paths.

Table 1: Extract of the first ten records of the table of attributes of private citizens' paths.

Data analysis

Once the vectorization of the paths was completed, the data is analysed. The information was queried in a way as to divide the paths based on type and year. This way, it is possible to have a diachronic vision of the quantity of the itineraries mapped by year and type. Consequently, heatmaps were created to highlight the busiest zones of the city.

In order to use the Kernel analysis, the line vectors were transformed into points using a SagaGIS algorithm *lines to points*, through QGIS interface, which allows to insert a parameter indicating every how many meters a point must be inserted. To avoid leaving even the shortest path, a distance of 10 meters between two points has been set. As the lines had been overlaid in the vectorization process, the result was a layer that reproduced a concentration of points equal to the number of overlapping lines. In this way we reproduced the frequency of use of a certain road (Figure 4). This process was repeated for each type of path and each examined

Figure 4: a) Vectorized paths around *piazza della Rotonda*; b) transformation of lines into points set 10 m apart; c) Kernel analysis (Base map: *Nuova Topografia di Roma*).

year. Finally, the Kernel analysis was applied which led to the creation of seven heatmaps which will be discussed in the following paragraphs.

Results

The ceremonial paths

After three centuries of the building projects by both popes and noble families and elite members of the society, Rome attained a significant urban space in the 18th century. These building projects mainly developed the busy quarters and districts of the city which were all in or around *Trastevere*. The increase in the population of Rome during the third decade of the 16th century onwards affected the whole street grid of the city and extended the inhabited areas towards the north close to the *porta del Popolo*, and towards the east, close to *via del Corso* until *Trastevere*. Furthermore, *via del Corso* assumed its current shape in the mid 17th century. The noticeable architecture of the palace of *Quirinale*, the residence of the popes, had a massive impact on the development of other palaces of nobles and pontiffs. They began to rebuild and renovate their palaces on *via del Corso* and adjacent areas (Guidoni and Marino 1979; Bevilacqua and Madonna 2003). The outcome of this slow transformation is visible in the maps of Nolli that indicates the centrality of the artery, crossing through the entire town from the *porta del Popolo*, close to *Campidoglio*, and the regularity of the street grid that defines the most recent quarters of the city (Bevilacqua 1998).

The results highlighted through the GIS platform between what this grid seems to suggest about the set of ceremonial processions along the streets are striking (Figure 5). As mentioned before, the booklets which had registered the itinerary of imperial processions, make the mapping of these movements possible and helped to compare them with the existing mobility. In addition, these booklets recall the problematic issues in choosing the right itinerary that could have met the needs of people and eased their movements on the streets during such events. Thus, they refer to the feeble nature of these statements, which tried to promote propaganda only.

The ceremonial paths of the pontiffs which were held right after the election of every new pope remained untouched according to traditions. The bishop of Rome had to receive the possession from his diocese, located in *S. Giovanni Laterano*. Therefore, he had to follow the *via*

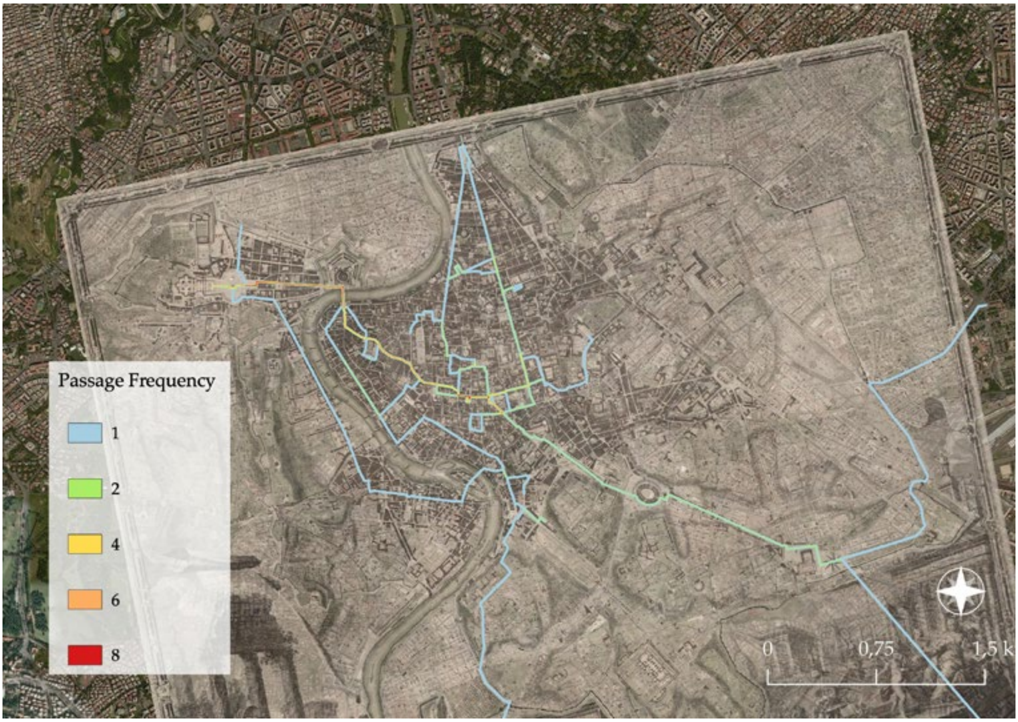


Figure 5: Heatmap of ceremonial paths (14) of the 18th century (base maps: *Nuova Topografia di Roma* and Bing Satellite).

Papale path that from the apostolic palace in the Vatican, crossed the Tiber with the bridge of *S. Angelo* to reach *Campidoglio* and then *S. Giovanni* (Visceglia 2002). The first stretch of the street that crossed through the most densely built-up neighborhoods, was never enlarged or rebuilt as it required a series of expropriations that not even a Renaissance or Baroque pope could have afforded. On the other hand, the abandonment of that street was quite impossible, given the symbolic value of buildings that were built there. Despite all these attempts, during the processions which were held for the ambassadors of other emperors to meet with the Roman court or allies who entered the city during the holy years, they barely passed through all those stunning streets with the magnificent palaces which were built to attract them. Although foreigners who did not know the city might be tempted to take a route that allowed them not to move blindly but to see their destination from afar, the presence and guidance of locals allowed them to adopt even more tortuous routes. In this way they had the advantage of passing through significant places, such as their national church or the residence of the ambassador of their sovereign.

The examination of the heatmap of the ceremonial paths also showed that the same thing happened for the processions of Catholic monarchs (or their representatives) visiting Rome. Ambassadors who entered the city through the *porta del Popolo* rarely passed through *via del Corso* for reaching Vatican on the east or *Quirinale* on the west. Sometimes they avoided certain roads because of hostile relations with their enemies (avoid passing by their buildings). The architects of that time were convinced that the straight and wide streets were more practical

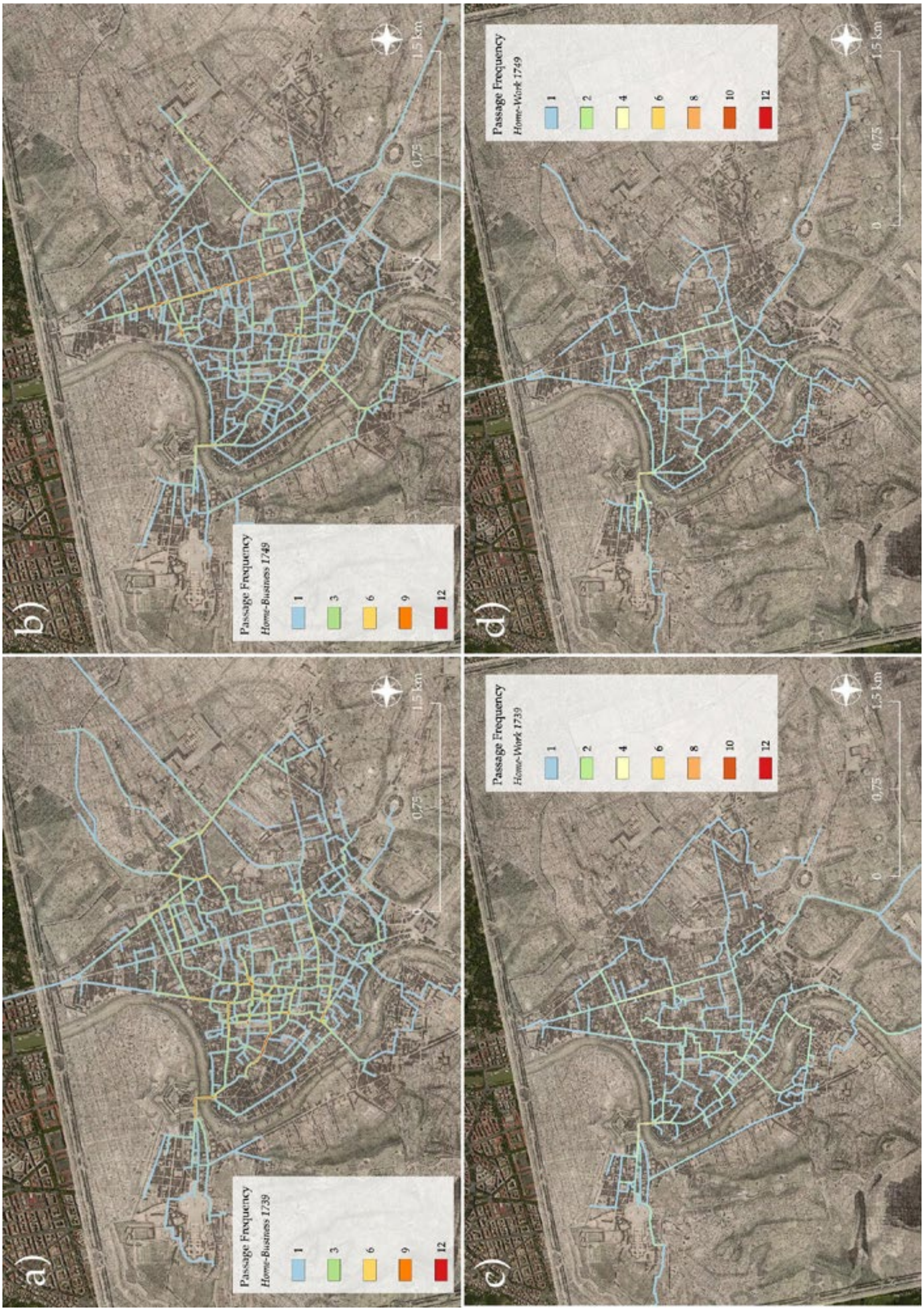


Figure 6: a) Heatmap of home-business paths in 1739; b) heatmap of home-business paths in 1749; c) heatmap of home-business paths in 1739; d) heatmap of home-business paths in 1749 (base maps: *Nuova Topografia di Roma* and Bing Satellite).

as they could allow people and carriages to pass through comfortably. However, the symbolic value of the buildings mentioned in the itinerary rewarded everything else and the processions of Rome continued to be held on that path.

The paths of private citizens

The analysis of the urban mobility of private citizens revealed that most of them did not go that far from their houses, and their work location is usually not distant from their homes. However, the landlords' shops were in the same place as their houses were. For the workers this was not the case: the numbers of apprentices living together with their masters were quite rare. The explanation is to be found in the high mobility that characterized both the home and the workplace. In fact, most of young workers could not afford a proper house, so they mostly shared with others; in some cases, they shared the whole room. Under these circumstances, finding accommodation was not that difficult and they could even find a place, close to their work. Following this, staying in one place was quite common and consistent with the economy of pre-industrial cities, where the demand for goods and services was limited and heavily relied on a small number of people (Brewer 1982; Ago 1998). Furthermore, sales were mainly on credit, thus, wages were paid partially, waiting for the final sale. There was no restriction or law regarding daily trips and transport outside of one's residency. However, it was not always appropriate to do so for fear of losing the contacts that had been established over time (Canepari 2017; Ago 2021).

In addition, the analysis of these paths also brought to light several unexpected data, such as activities outside the work sphere (e.g. going to a party, a tavern, or managing a street show). In these cases, it happened more often that people exercised their right to reside in the city, to fully enjoy what it offered without being confined in a circumscribed space.

Conclusion

Besides mapping the movements of individuals and the distances of paths and directions, processing the data with GIS allows evaluating the repletion of certain movements the frequent use of certain streets. Moreover, it helps to find out which roads were busier, which were the main directions of these movements (figs. 5, 6).

It is noteworthy that all the heatmaps, linked with any sort of movement, demonstrate the revealing direction of north-south compared to east-west which was taken during the ceremonial events (Figure 5). In other words, the city of daily affairs did not coincide with that of formal occasions which were highlighted by a few significant points of attraction such as Vatican, *Campidoglio*, and *Quirinale*. Moreover, all the economic activities and transactions were concentrated in central districts, from *piazza Navona* to *via del Corso* in the east-west direction and from *Campo Marzio* to *ponte Sisto* in the north-south direction. Also, this path had built a sort of border between the central area and the peripheral areas, while its wide, straight roads which crossed the entire city, were more frequent than *via del Corso*. Between 1739 and 1749 (Figure 6), the trafficked area of the city began to assume the brand-new zones as it was drawing the passengers' attention (Ago 2021).

In theory, paths that allow visual and long-distance control of the road are the friendliest. Moreover, they are those that regardless of other characteristics of the roads or the existence of particular poles of attraction, facilitate travel and even induce them (Hillier *et al*. 1993). To conclude, the results of this research suggest that this thesis is only partially correct: the local people who were familiar with the streets of Rome continued to use the same challenging roads for years and the new built-up part of *via del Corso* became gradually more popular over the years.

Bibliography

Ago, R. 1998. *Economia barocca. Marcato e istituzioni nella Roma del Seicento*. Roma: Donzelli.

Ago, R. 2021. *Il diritto alla città. Roma nel Settecento*. Roma: Viella.

Bevilacqua, M. 1998. *Roma nel secolo dei lumi. Architettura, erudizione, scienza nella pianta di G.B. Nolli 'Celebre Geometra'*. Napoli: Electa.

Bevilacqua, M. 2013. Nolli, Giovanni Battista. In *Dizionario Biografico degli Italiani*. Vol. 78. https://www.treccani.it/enciclopedia/giovanni-battista-nolli_(Dizionario-Biografico)/.

Bevilacqua, M. and M.L. Madonna 2003. Sistemi di residenze nobiliari a Roma e a Firenze: architettura e città in età Barocca. In M.L. Madonna and M. Bevilacqua (eds) *Il sistema delle residenze nobiliari. Stato Pontificio e Granducato di Toscana*: 9–35. Roma: De Luca.

Bogdani, J. 2021. *Atlas. Paths: Tracking Papyrus and Parchment Paths*. https://atlas.paths-erc.eu/map/ (accessed 14/10/2021).

Branting, S. 2007. Using an urban street network and a PGIS-T approach to analyze ancient movement. In J.T. Clark and E.M. Hagemeister (eds) *Digital Discovery. Exploring new frontiers in human heritage, CAA 2006*: 99–108. Budapest: Archaeolingua.

Brewer, J. 1982. Commercialization and Politics. In N. McKendrick, J. Brewer and J.H. Plumb *The birth of a consumer society: the commercialization of eighteenth-century England*: 197–262. London: Harper Collins.

Canepari, E. 2017. *La construction du pouvoir local. Élites municipales, liens sociaux et transactions économiques dans l'espace urbain. Rome, 1550–1650*. Rome: École Française de Rome.

D'Erasmo, D. 2019. Georeferencing Napoleonic cartography to reconstruct ancient Egypt landscapes: methods in comparison and the case of the island of iw-rd in the 16th nomos of Upper Egypt. *Groma* 4: 1–18.

Fonte, J., C. Parcero-Oubiña and J.M. Costa-García 2017. A GIS-Based Analysis of the Rationale Behind Roman Roads. The Case of the So-called Via XVII (NW Iberian Penisula). *Mediterranean Archaeology and Archaeometry* 17 (3): 163–189.

Guidoni, E. and A. Marino 1979. *Storia dell'urbanistica. Il Seicento*. Roma-Bari: Laterza.

Guilhem, V. 2021. Freehand Raster GeoreferencerPlugin. *QGIS Python Plugins Repository*. https://plugins.qgis.org/plugins/FreehandRasterGeoreferencer/.

Hillier, B., A. Penn, J. Hanson, T. Grajewski and J. Xu 1993. Natural Movement: or, Configuration and Attraction in Urban Pedestrian Movement. *Environment and Planning, B: Planning and Design* 20 (1): 29–66.

Lanen, R. van, M.C. Kosian, B.J. Groenewoudt, T. Spek and E. Jansma 2015. Best Travel Options: modelling roman and early-medieval routes in the Netherlands using a multi-proxy approach. *Journal of Archaeological Science*: Reports 3: 144–159.

Rumsey, D. 2021. Nuova Topografia di Roma. *David Rumsey Map Collection*. https://www.davidrumsey.com (accessed 14/10/2021).

Rumsey, D. and M. Williams. 2001. Historical Maps in GIS. In A.K. Knowles (ed.) *Past Time, Past Place: GIS for History*: 1–18. Redlands, CA: ESRI Press.
Tice, J. and E. Steiner 2021. *Nolli Map Index*, http://nolli.uoregon.edu/gazetteer.asp.
Visceglia, M.A. 2002. *La città rituale. Roma e le sue cerimonie in età moderna*. Roma: Viella.

Towards FreeCAD experimentation and validation as a FOS HBIM platform for building archaeology purposes

Filippo Diara, Fulvio Rinaudo

Filippo Diara, Politecnico di Torino – DAD department; filippo.diara@polito.it
Correspondence: filippo.diara@polito.it
Fulvio Rinaudo, Politecnico di Torino – DAD department; fulvio.rinaudo@polito.it

Abstract
The adoption of BIM (Building Information Modelling) for archaeology is limited by different issues inherited by the software, which is usually designed for AEC industry and rarely fits Heritage assets. An unconventional solution could be provided by FOS (Free and Open-Source) software, thanks to source code accessibility and modifications possibilities, by adapting thus the software to specific needs and not the opposite. HBIM (Historic Building Information Modelling) methodology could become a good solution for the documentation and analysis of building archaeology, and FreeCAD software – apart from being a well-known modeller – has the potential to become the proper tool to use, because of its adaptability (overcoming limitations of standard BIM software). This paper is focused on experimentations and adaptations of FreeCAD to building archaeology assets, as a HBIM custom platform for the documentation and analyses, trying to avoid methodological compromises related to commercial BIM software, as well as trying to preserve the original goals of specific research.

Keywords: BUILDING ARCHAEOLOGY; DOCUMENTATION; FREECAD; FOSS; HBIM.

Introduction and objectives

Informative systems and open-access data have experienced a great development in the archaeological panorama especially in order to solve research and methodology-related issues. Specifically, data fragmentation and the difficulty of connecting paper-based to digital information has affected publication and knowledge-sharing steps of excavated sites or surveyed historical buildings for many years.

Historical Building Information Modelling (HBIM) was conceived eleven years ago (Murphy *et al.* 2009), and it has had an enormous importance becoming the subject of researches regarding archaeological fields. This term is referred to the adoption of BIM methodology (informative system created for architecture design) for Cultural Heritage assets, in particular to historical buildings documentation and analyses. In fact, HBIM should be intended as the knowledge management of historical buildings and it takes place in a specific historic moment; by contrast, traditional BIM methodology is intended for new constructions and it takes place at the beginning of the life-cycle of the building (Fassi *et al.* 2013).

This project moved from two starting points, one linked to geomatics and the other to building archaeology: in recent years BIM has rapidly developed thanks to commercial software, and our question is: can we today design a BIM/HBIM system by using exclusively FOSS tools?

In parallel, informative systems have demonstrated their importance for the archaeological field allowing one to keep linked different kinds of data. Is it, therefore, possible to design a proper HBIM solution to specifically address the needs of building archaeology?

Based on these assumptions, this project focuses on experimentations and adaptations of FreeCAD FOS software[1] as an HBIM platform for the documentation and analysis of building archaeology. This unconventional solution was used especially to try to adapt tools, software and workflows to specific needs, by avoiding methodological compromises imposed by commercial BIM software and preserving researches goals (Logothetis *et al.* 2016; Diara *et al.* 2018).

FreeCAD was trialled and validated on the refectory of the medieval Staffarda Abbey, a Cistercian monastery located near Saluzzo in north-west Italy. Knowledge processes on the religious context have generated a large amount of data which needed to be managed in a smart way for achieving a good standard of documentation. For this reason, HBIM was selected as an appropriate methodology and dynamic informative platform.

HBIM and archaeology: short review

HBIM methodology for the archaeological domain is rapidly spreading thanks to its huge potential. This approach involves the representation and reproduction of archaeological sites and monuments, focusing also on semantic knowledge resulting from morphological and conceptual peculiarities, by including the dimension of time (Historic England 2017; Trizio *et al.* 2019; Diara 2020). Although powerful BIM software products such as Autodesk Revit[2] or Graphisoft Archicad[3] were designed for AEC industry, in recent years there have been interesting and well conducted applications of the software to the archaeological domain, in order to document and investigate archaeological and building stratigraphy (Scianna *et al.* 2015; Bosco *et al.* 2019; Trizio *et al.* 2019). Consequently, the time dimension is fundamental in an HBIM project. A very interesting and recent project is focused on the possibility of implementing a Harris' Matrix inside a BIM model by using ontologies (Borin *et al.* 2020). This integration becomes essential for analysing the time dimension of the case study through semantic relations with parametric objects, achieving a relative chronological vision of constructive phases and periods (Trizio *et al.* 2019; Borin *et al.* 2020; Diara *et al.* 2020). Moreover, stratigraphic graphs assume great importance when trying to compare and merge, through a new methodology, 3D data with traditional bidimensional data (Diara 2020).

However, the already mentioned commercial BIM software allows a parametric modelling based on simple and predefined architectural families. This is relevant for the digital reproduction of stratigraphic units related to archaeology or building archaeology, especially since there are no predefined parametric families corresponding to archaeological elements. For this reason, the HBIM design of archaeological sites involves the creation of *ad hoc* families by using parametric modelling tools. During this time-consuming step, important interpretations occur. In fact, shapes related to stratigraphic units or historical architectural elements need to be understood and simplified (and therefore conceptualised) in order to

[1] https://www.freecadweb.org/ (accessed 30/07/2021). FreeCAD is released with a GNU GPL v2 license.
[2] https://www.autodesk.it/products/revit/overview (accessed 30/07/2021).
[3] https://graphisoft.com/solutions/archicad (accessed 30/07/2021).

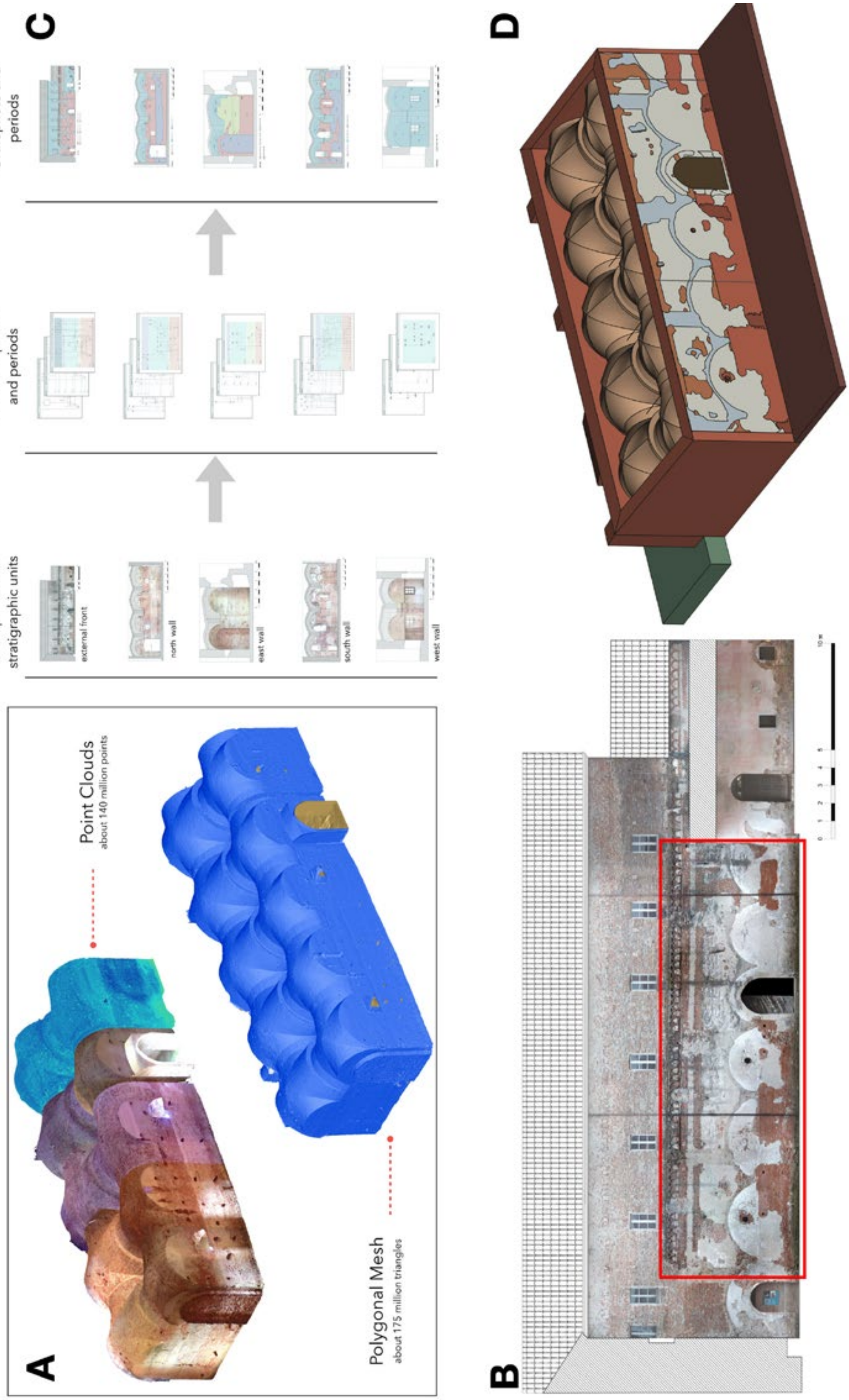


Figure 1: Knowledge processes: from metric survey (A) to stratigraphic survey and analysis (B and C), until the parametric model construction (D).

establish physical and semantic relations. However, this operation is not so easy to manage due to the overall complexity of the geometric data and the required level of accuracy (Diara *et al.* 2019; Bosco *et al.* 2019). This difficulty is largely due to modelling tools integrated into the software, and then due to software design. For this reason, unconventional solutions were the focus of different projects aiming at creating and managing *ad hoc* platforms, trying to solve these methodological issues (Logothetis *et al.* 2016; Diara *et al.* 2018; 2020). A branch of these studies is focused on the exploitation of FOS instruments in order to adapt and customize them to archaeological needs, and this paper focuses on this approach.

Starting point: data acquisition

The trial we are about to present was conducted on the refectory of the Staffarda monastery. A previous larger project (Diara 2020), took care of the data acquisition, obtained from metric and stratigraphic surveys of this religious context (Figure 1). The metric survey was carried out using total station for the topographic framework and LiDAR acquisitions of structures, surveyed by using the terrestrial laser scanner (TLS) Faro Focus 3D S120.

On the other hand, the stratigraphic survey has provided a fundamental step in order to achieve historical information from masonries. The refectory of the medieval Staffarda Abbey was chosen as a case-study for the implementation of stratigraphical recording inside the HBIM platform, encoded as semantic historical information: in fact, this building is characterized by the presence of a complex masonry stratigraphy generated between the 12th – 13th centuries and the beginning of the 20th century (Beltramo 2016; Diara 2020).

The stratigraphic survey was performed on the internal walls and on the external façade, facing the cloister of the abbey. It was conducted both on-site and in the laboratory, on 2D CAD projects to achieve a clear stratigraphic analysis on detailed orthophotos. The Harris Matrix was designed to obtain a relative chronology by isolating phases and historical periods.

Briefly, the stratigraphic analysis was crucial for the comprehension of construction and destruction phases and chronological periods. The building of the refectory could be divided into seven chronological periods, the principal being: from the early construction site of the refectory (1160) to the main construction block referred to 1180–1220; from morphological and structural changes after the Battle of Staffarda in 1690, to the maintenance interventions of the 19th – 20th centuries (Beltramo 2016; Diara 2020; Diara *et al.* 2020).

HBIM platform design

FreeCAD is a parametric modeller, integrated by different customizable tools and supported by a large community of users and developers. Moreover FreeCAD, given the premises, could also be used as complete BIM solution, entering the FOSS ecosystem of informative platforms, next to QGIS[4] and MicMac,[5] just to mention the most well-known projects.

[4] https://www.qgis.org/ (accessed 30/07/2021).
[5] https://micmac.ensg.eu/ (accessed 30/07/2021).

The customization of FreeCAD as a HBIM platform has been the project's focal point, and it was carried out by the implementation of *ad hoc* libraries, workflows, and a proposal of classification.

Initially, a smart integration of fundamental libraries was designed to be included inside the software: IfcOpenShell[6] open library was preferred since it allows managing IFC (Industry Foundation Classes) open format, a necessary requirement for BIM and data exchange. Additional packages and libraries represent a crucial base for the HBIM platform in terms of basic functionalities and specific workflows.

Apart from these additions, FreeCAD was implemented with particular *workbenches* for defining the HBIM platform as well as CAD functionalities: *BIM workbench*[7] played a fundamental role in order to establish all features of Staffarda Abbey and its refectory (including floor levels) and other general attributes related to the main construction site; *Dynamic Data workbench*[8] was included especially for managing parametric objects with dynamic properties; *Reporting workbench*[9] played a key role for the data validation since it allows designing and managing custom SQL (Structured Query Language) queries and the related output. All these additional workbenches were easily downloaded and implemented through the internal *addons manager*.

Despite these implementations, FreeCAD was also modified for IFC classification. These modifications involved the default IFC2X3 schema file located in the root directory of FreeCAD: in addition to default AEC classification the IFC2X3 schema was implemented with a proposal/demonstrative custom classification useful for building archaeology purposes through designed typological sub-divisions of existing default classification. Then, an adequate classification with the actual architectural situation of the refectory were included, such as *ifcVault* (as sub-type of ifcRoof) – *ifcPillar* (as sub-type of ifcColumn) – *ifcUSM* and *ifcUSR* (as sub-type of ifcWall) concerning macro stratigraphy evidence (mostly medieval) related to masonry (USM) and render/plaster (USR), fundamental both for archaeology and restoration purposes.

Model and data management

Despite the possibility of building a high quality 'as-built' model of the refectory, the modelling choice was affected by the intention of a voluntary simplification of the architectural elements. In fact, the HBIM model was conceived not as a *digital twin* of the real architectural context but as a conceptual and simplified representation of reality, as BIM models should be 3D conceptual database-models.

The simplification of architectural elements of the refectory was designed in order to highlight historical information represented by stratigraphic units of masonries, permitting easy handling of the management of essential chronological elements inside FreeCAD (Figure 2). For this reason, the modelling phase of HBIM projects must be principally related to semantic

[6] http://www.ifcopenshell.org/ (accessed 30/07/2021), LGPL-3.0 License.
[7] https://wiki.freecadweb.org/BIM_Workbench (accessed 30/07/2021).
[8] https://wiki.freecadweb.org/DynamicData_Workbench (accessed 30/07/2021).
[9] https://github.com/furti/FreeCAD-Reporting (accessed 30/07/2021).

Figure 2: FreeCAD platform and parametric model of the refectory with stratigraphic units.

information and how investigate them (Di Stefano *et al.* 2019; Matrone *et al.* 2019; Bruno *et al.* 2019).

The obtained metric data (point clouds) was the skeleton around which the parametric model was created after the extraction of planar sections. Inside FreeCAD, parametric volumes were modelled by extruding surfaces obtained from planar sections shapes. Stratigraphic units were modelled starting from profiles drawn on orthophotos mapped on the model. By using these profiles, flat surfaces were created and extruded as parametric volumes, having a minimum thickness value (e.g. 2 mm – 5 mm).

After volume creation, other metric attributes and information were managed by using the *Dynamic Data workbench* that allows the creation of Python containers for new custom and dynamic metric properties: when a custom property is changed, the related object will automatically and proportionally update. Once architectural elements and stratigraphic units have turned into parametric objects, the *IFC Role/Type* attribute is set for each element. In this regard, the already mentioned proposal classification is followed.

Behind the parametric model, the real actor of the platform is the historical data related to architectural peculiarities of the refectory. The process of knowledge and the stratigraphic analysis is fundamental and propaedeutic for the archaeological data integration into the platform. The attributes menu of IFC objects was correctly implemented with information, such as: *General description* concerning architectural elements and stratigraphic units detected on masonries; *Detailed photos* for each element were included, also concerning iconographic resources; *Material* of each object.

The default material database of FreeCAD was implemented by adding materials related to the refectory; *IFC roles* and architectural classification depending on single elements, including stratigraphic units by using the *ad-hoc* classification prepared during the custom modifications phase; *ID code* by using *Standard Code* text field; Architectural elements and stratigraphic units of the refectory were enriched by alphanumerical identifiers referred to stratigraphic analysis enumeration; *Stratigraphic diagrams* (Harris Matrices) – included to obtain a relative chronology analysing physical relations among stratigraphic units – were included by using the *additions* field (Figure 3); *Chronological periods* were implemented through the *tag* field (alphanumerical value) to include the chronological interpretation carried out with the Harris Matrix. Furthermore, chronological periods were also highlighted directly on parametric objects by the use of different colours. Finally, the integrated DBMS of FreeCAD was used to manage bibliographic resources and other external references concerning the case study. All these collected data relating to the refectory turned out to be crucial for revisions and updates, as well as for overcoming data dislocation.

Platform validation and results

The implementation of historical data inside the platform has proven essential for two reasons. First to create a dynamic informative environment focusing on the refectory and preventing data fragmentation, and, second, for the purpose of managing punctual analysis through custom queries for data validation, revisions and updates.

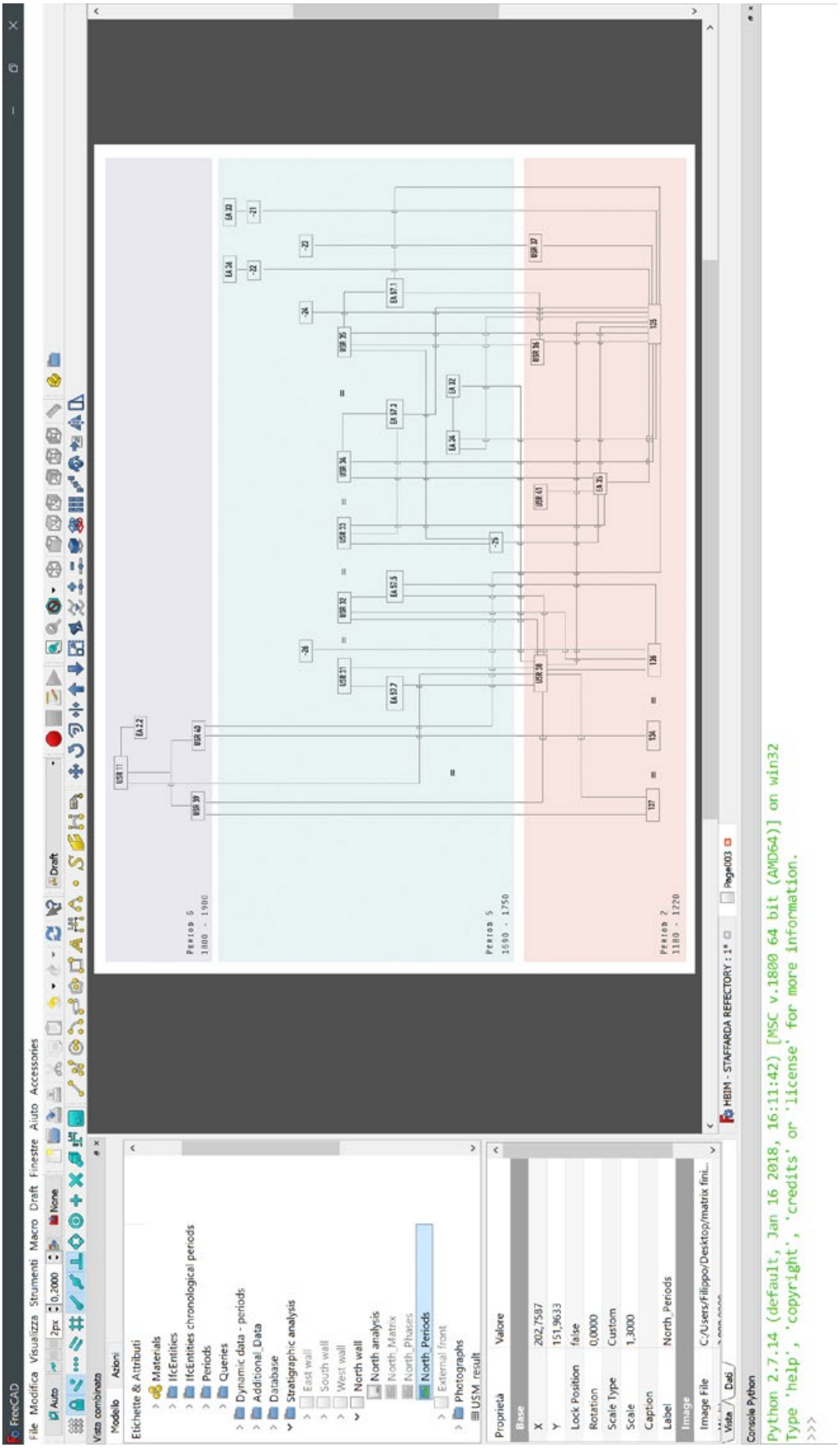


Figure 3: Stratigraphic diagrams implemented as semantic data (Harris Matrix of north wall of the refectory of medieval Staffarda Abbey).

Reporting workbench and *sqlparser* allowed the design and management of custom SQL queries to investigate the object properties of both architectural elements and stratigraphic units.

SELECT SQL queries were performed in two ways. The first is based on the *sqlparser* module, which allows queries to be executed directly on the FreeCAD Python console: the query has to be written on the console after evoking *sqlparser* module and the console returns the result using green-coloured text.

The second method relies on customizable SQL statements of *Reporting workbench*. Following this procedure, *SELECT* queries are set on the configuration window of *Reporting workbench* and saved with custom labels. In this way, the result report is obtained from double clicking on query label and the platform return a datasheet containing the query result (Figure 4). Datasheet reports can be easily exported as CSV files. This method is certainly far more dynamic and user friendly than the previous one, also facilitating data exchange processes among collaborators. Following this procedure, data storage and future analyses on the refectory are correctly ensured, especially for monitoring operations.

SQL queries are performed for investigating IFC types, refectory materials, particular layer identifiers, chronological periods via *tag* field, etc. These queries were designed to investigate stratigraphic units and their properties linked to the parametric model. In this regard, the *SELECT* syntax proved essential for facilitating the semantic selection process. This kind of query and syntax is referred to a standard query language and these queries could also be managed in other informative environments, for example GIS platforms.

Regarding data interoperability, the internal storage of FreeCAD may connect semantic data with external DMBSs and vice-versa. In fact, internal DBMS and resulted CSV from SQL queries can be exported as datasheets manageable, for example, with stand-alone database software (Libre Office Calc, Google Documents, MySQL, etc.). At the same time, external data originally managed with third-party DBMS software can be imported as CSV files into FreeCAD and further enrich the available semantic data set.

FreeCAD proved to be not only a parametric modeller and a BIM/HBIM software but also a dynamic informative platform for metric and historical data collection and validation. Moreover, this custom platform could be exploited by other users and professionals for further studies on Staffarda Abbey as well as exploited for others contexts. FreeCAD modifications and additions were used for external libraries, packages and workbenches, being easy to download and implement. Furthermore, a custom and independent bundle is under development to create a FreeCAD branch based on modifications performed for this project. In this regard, the FreeCAD community will be essential for monitoring the constant development and growth of this software, even for improving this workflow.

Conclusion

This paper was focused on the use of FreeCAD as a valid and efficient BIM/HBIM platform, in addition to being a parametric modeller. This software was customized for adapting particular libraries to building archaeology needs for monitoring and restoration, as well as further analyses. In short, FreeCAD has a huge potential as HBIM platform for building

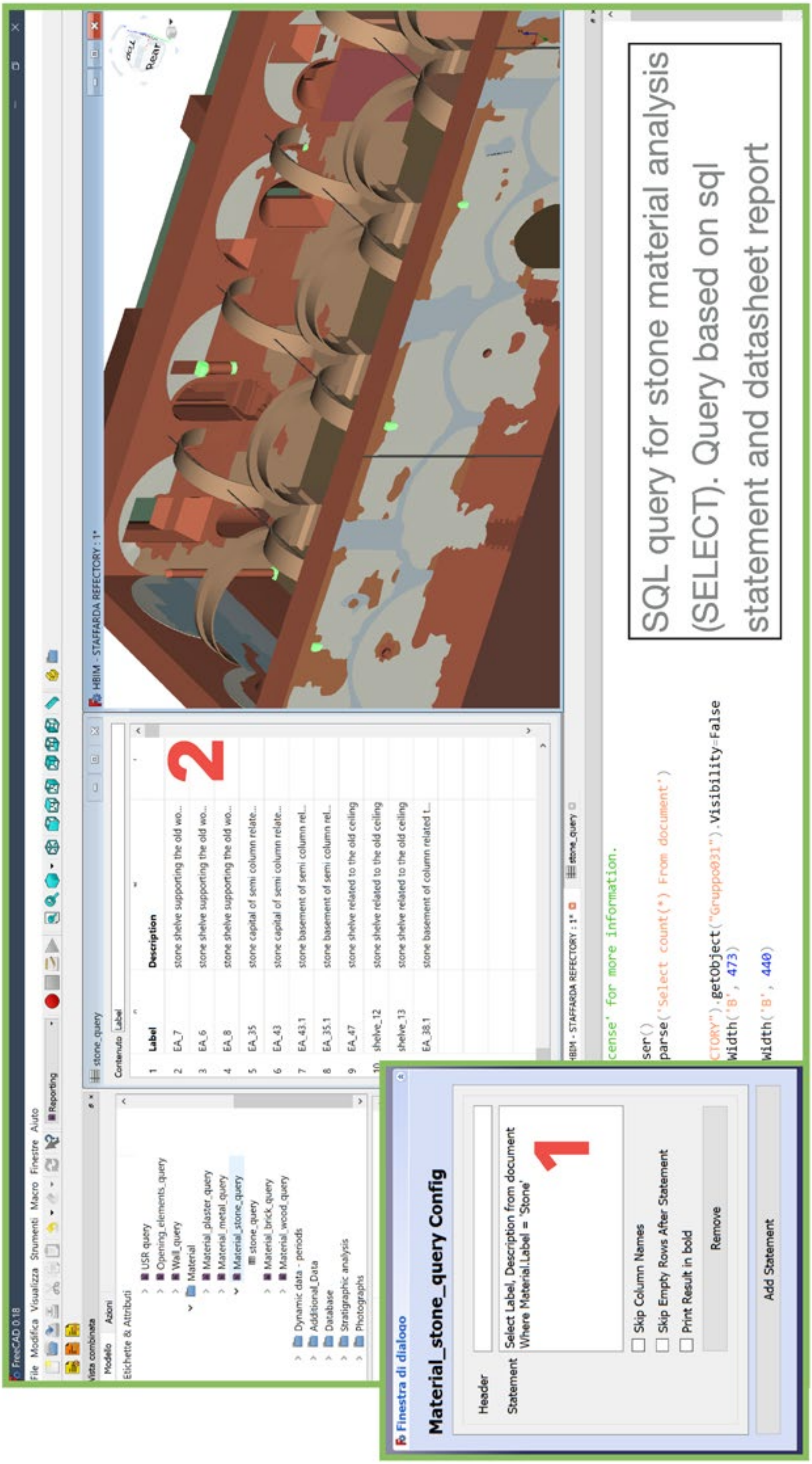


Figure 4. SQL query by using Reporting workbench and statement configuration: selection of stone elements and their description (result on CSV).

archaeology purposes. The integration of stratigraphic analysis has become possible by the new IFC classification proposal, the inclusion of stratigraphic units as parametric objects and the related analysis. As a result, custom SQL queries have proved to be fundamental for investigating and validating archaeological data.

This project has illustrated advantages and innovative points: first of all, the resulting platform was revealed to be a secure environment for collecting and managing different type of data, avoiding 2D and 3D data fragmentation and favouring historical interpretation, as well as maintenance operations. Through source code accessibility, the workflow tried to overcome traditional BIM limitations concerning IFC classification, modelling and implementing tools. But beyond these positive feedbacks, some critical issues also arose. This experimental scan-to-BIM workflow involves different file formats and software. There are risks, too, for data exchange and compatibility, and project stability in general. Moreover, it was a time-consuming workflow: from metric acquisition to parametric modelling, from stratigraphic surveys to stratigraphic analysis of the entire refectory, from FreeCAD implementation to data collecting and querying. Finally, the absence of a standard semantic classification for stratigraphic units was one of the most critical issues faced by the project. This important issue led this research towards a custom (and not perfect) solution to overcome this limitation. Notwithstanding these difficulties, the designed workflow succeeded in solving the goals set in advance, as an operative and custom BIM platform was created for building archaeology purposes via FOS software.

Proceeding in the direction of current informative systems (GIS and BIM) it is possible to ensure data protection and exchange of complex architectural and archaeological contexts, overcoming limitations to do with the juxtaposition of digital analyses and paper-based studies.

The integration of methodologies and tools should become the only procedure to overcome technical and compatibility issues, obtaining a methodological synthesis suitable for Heritage assets documentation and analysis.

However, the documentation and analysis of archaeological and architectural contexts should not be based on predefined methodological and software solutions that are not sufficient warranty of a correct data analysis. Experimenting new ways should always be an available solution to carry out *ad hoc* analyses, adapting techniques to the research, and not the other way around.

The constant increase of technological innovation leads researchers to use specific technologies and geomatics methodologies just because they are new and benchmarking, even if not every solution could be reliable for all disciplines. The HBIM methodology and FOSS adoption have proved to be the best solution for this project – FreeCAD allowing the collection of the refectory data and therefore dynamic revisions and updates – but it may not be the perfect solution for other case-studies. Replicability and feedback from the community should always be a major objective of this kind of applied research.

Bibliography

Beltramo, S. 2016. Il romanico cistercense: analisi del costruito della chiesa e del monastero di Staffarda. In S. Lomartire (ed.) *Romanico piemontese - Europa romanica. Architetture, circolazione di uomini e idee, paesaggi*: 14–25. Confronti 10. Livorno: Centro Studi Città e Territorio.

Borin, P., R.A. Bernardello and A. Grigoletto 2020. Connecting Historical Information with BIM Ontologies. HBIM Methods for the Visualization of Harris Matrix for the Torrione in Carpi. In L. Agustín-Hernández *et al.* (eds) *Graphical Heritage*: 757–770. Springer Series in Design and Innovation 5. Cham: Springer. https://doi.org/10.1007/978-3-030-47979-4_65.

Bosco, A., A. D'Andrea, M. Nuzzolo and P. Zanfagna 2019. A BIM approach for the analysis of an archaeological monument. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2/W9: 165–172, https://doi.org/10.5194/isprs-archives-XLII-2-W9-165-2019.

Bruno, N. and R. Roncella 2019. HBIM for Conservation: A New Proposal for Information Modeling. *Remote Sensing* 11 (15): 1751. https://doi.org/10.3390/rs11151751.

Di Stefano, F., E.S. Malinverni, R. Pierdicca, G. Fangi and E. Shurte 2019. HBIM implementation for an ottoman mosque. Case of study: Sultan Mehmet Fatih II mosque in Kosovo. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15: 429–436. https://doi.org/10.5194/isprs-archives-XLII-2-W15-429-2019.

Diara, F. and F. Rinaudo 2018. Open source HBIM for Cultural Heritage: a project proposal. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2: 303–309. https://doi.org/10.5194/isprs-archives-XLII-2-303-2018.

Diara, F. 2020. *Experimental workflow for the creation of a non-conventional open source HBIM platform integrating metric data and stratigraphic analysis: the case study of the refectory of Santa Maria di Staffarda Abbey*. PhD diss., Politecnico di Torino.

Diara, F. and F. Rinaudo 2019. From reality to parametric models of Cultural Heritage assets for HBIM. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2: 413-419. https://doi.org/10.5194/isprs-archives-XLII-2-W15-413-2019.

Diara, F. and F. Rinaudo. 2020. Building Archaeology documentation and analysis through open source HBIM solutions via NURBS modelling. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2: 1381-1388. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1381-2020.

Fassi, F. and S. Parri. 2013. Web 3D BIM for the Department of Cultural Heritage. Tests and experimentation on the Spire of the Duomo of Milan. In R.A. Genovese (ed.) *To know, conserve, enhance*: 271–290. Naples: Arte tipografica editrice.

Historic England 2017. *BIM for Heritage: Developing a Historic Building Information Model*. Swindon: Historic England.

Logothetis, S. and E. Stylianidis 2016. BIM Open-Source Software (OSS) for the documentation of Cultural Heritage. *Virtual Archaeology Review* 7 (15): 28–35. https://doi.org/10.4995/var.2016.5864.

Matrone, F., E. Colucci, V. De Ruvo, A. Lingua and A. Spanò 2019. HBIM in a semantic 3D GIS database. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2/W11: 857–865. https://doi.org/10.5194/isprs-archives-XLII-2-W11-857-2019.

Murphy, M., E. McGovern and S. Pavia 2009. Historic building information modelling (HBIM). *Structural Survey* 27 (4): 311–327. http://dx.doi.org/10.1016/j.isprsjprs.2012.11.006.

Scianna, A, M. Serlorenzi, S. Gristina, M. Filippi and S. Paliaga 2015. Sperimentazione di tecniche BIM sull'archeologia romana: il caso delle strutture rinvenute all'interno della cripta della chiesa dei SS. Sergio e Bacco in Roma. *Archeologia e calcolatori* Supplemento 7: 199–212.

Trizio, I, F. Savini, A. Giannangeli, R. Boccabella and G. Petrucci 2019. The archaeological analysis of masonry for the restoration project in HBIM. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W9: 715–722.

FLOS for Museums: open solutions to train communities and manage heritage sites

Paolo Rosati

Paolo Rosati , Sapienza University of Rome, SARAS Department 1; paolo.rosati@uniroma1.it
Correspondence: paolo.rosati@uniroma1.it

Abstract

The Faro Convention's adoption by the Italian Parliament changed the future of Italian public museums; the Convention regards the active participation of the community as a new resource, intrinsically linked to the conservation and enhancement of heritage. This contribution hopes to give some guidelines for the future 'techno-creative spaces' at the Italian cultural sites, presenting some cases studies from the last three years (2018–2021).

Keywords: FLOS; MUSEUM; MUCIV, MNETRU, PARTICIPATIVE-PROJECTS; BUSINESS MODEL.

Introduction

The Fab Lab experience started in 2000 and the first scientific paper that focused on this subject was published in 2002 (Mikhak *et al.* 2002). The first US article where the words 'Fab Lab' appear in connection with 'museum', can be found in a work by Sherry His (2008), who writes about the Science Museum exhibition 'Nanozone' at the Lawrence Science Hall, University of California, Berkeley. Today, the archetypal idea behind the Fab Lab experience has not radically changed since its origin, i.e. Fab Lab spaces and the 'Maker' approach, following the principles of Free Libre and Open-Source Software (FLOSS), are widely used in museum science and technology. In several parts of the world (Mexico, Russia, France, Japan, Canada), before the pandemic, there was a remarkable increase in the number and variety of digital experiences in heritage museums (Artemova and Espinoza Torres 2020).

But what about Italy, the country that, with China, has the greatest number of UNESCO heritage sites?[1] Until October 2020 there were no stable techno-creative spaces (*infra* par. 2.2) in public archaeological museums, but something new is happening (*infra* par. 3) and technological approaches geared at reaching new generations in their free time could become crucial. In the USA stable scientifical practices after school have become standard since 1986 (Hsi 2008: 896–897). The results are certainly remarkable in terms of quality and 'lifelong learning' attitudes (*infra* 2.4).

[1] https://whc.unesco.org/en/list (accessed 01/08/2021).

During pandemic times, with the deep crisis in Italian school systems,[2] the mission of the government should have been to guarantee a 'lifelong learning' network for the future of our 'twenty-first-century citizens' (Xanthoudaki 2015). This is not a vague comparison between two different educational systems forged by centuries of customs (USA and Italy), this is a term to measure the gap between them and propose a pivotal and equidistant meeting point: the opening of innovative informal educational institutions in Italy as techno-creative spaces at archaeological sites and museums, that can maximise the potential of the Italian cultural system. As His writes (2008: 898): 'Informal learning institutions – museums, community-based after-school programs, science centers – have an opportunity to work with schools to bridge the experiences of children to provide a more coherent learning experience.'

Expressions such as *museologia-comunitaria* (community museology), collaborative culture, participative culture/projects (Artemova and Espinoza Torres 2020) do not seem in current use in the Italian literature about archaeological museums. In Italy sometimes museum activities programs include laboratories of applied archaeology and history (rarely digital humanities) for children. This often occurs in Italy during school time; in the US the key point is that they occur during outside school, Hsi again (2008: 891): 'There is a growing recognition that most young people spend more of their hours in learning environments outside of school and that this informal learning time is equally important in their overall preparation for future work and lifelong learning.'

This contribution focuses on the FLOSS and Maker approach at archaeological museums and sites. The following paragraphs will show that the Italian cultural system has many latent highlights often restricted by a traditional museography; however, in the cultural epistemology the word 'tradition' involves an intrinsic contradiction: it can be a strong constraint, but it can also become a useful educational paradigm.

Methodologies

Laws and governance

For the Italian archaeological museums bridging the gap between 'science' and 'art' museums is to analyse where community museology belongs and how the Free Libre and Open-Source movement can boost that process of innovation. If in the 21st century 'museums want to remain relevant, they need to be able to grasp the changes and reflect on them before these become outdated' (Xanthoudaki 2015). The tardy adoption of the Faro Convention by the Italian Parliament was clearly a negative sign, as will be explained in the following paragraphs.

The main collections and museums of Italy are public, managed by the Ministry of Culture (MIC), the Regions, and the Municipalities. It is clear that without any clear public recognition of the role of communities it was clearly impossible for Italy to begin general projects of techno-creative spaces within its museums. The Faro Convention enshrines an effective

[2] https://www.savethechildren.it/press/la-povert%C3%A0-educativa-ai-tempi-del-coronavirus-bambini-e-adolescenti-intrappolati-tra-crisi (accessed 04/08/2021) and De Marchi 2020.

community participation in museums and cultural sites as an intrinsic and fundamental resource for the preservation and enhancement of heritage (Council of Europe 2005 2b). The Convention became law in Italy (L. 113/2020), seven years after other earlier subscriptions to it.[3]

Techno-creative spaces

Techno-creative spaces can be defined as ‘maker’ spaces inside museums, managed by the museum staff, or led by organizations and societies, aimed at facilitating the access of a wider museum community to activities and programs that can foster the development of digital co-creativity and learning through artefact making.

Therefore, since the assimilation of the Faro Convention, the fundamental role of the cultural and technological communities has been recognised by the Italian Parliament. The directors can now create techno-creative spaces or experiences using the knowledge of the relevant communities (Capron *et al.* 2020).

It is now possible to give a definition for the new contract between museums and citizens, it being crucial in changing the way in which the ‘museum public’ is perceived: the ‘public’ must become a ‘community’. In this context, the challenge is to borrow the word ‘community’ from the technological environment and keep it separate from the word ‘user’, as understood in economic terms.

Starting from the highest level, the definition of *Community museology* can be given as:

> ‘Community museology is a discipline of the social sciences whose fundamental purpose is to develop a process of community organization around the planning and operation of educational and cultural spaces dedicated to research, protection, conservation, valuation and dissemination of natural and cultural heritage, of a specific community or region, whose mission is to promote and implement teaching-learning processes that contribute to comprehensive development to improve the quality of life of the population.’ (Méndez Lugo 2011, translation from Spanish by the author).

An important benchmark that helps to determine if a museum participative-collaborative project is going in the right direction is to control the correspondence between the project targets and museum mission (Artemova and Espinoza Torres 2020: 149).

Epistemology and didactic methods

The didactic methodologies commonly adopted in experiences with children, teenagers and adults are usually made up of three points, according to Xanthoudaki (2015: 256): a) spiral learning; b) open learning; and c) inquiry.

[3] https://www.gazzettaufficiale.it/eli/id/2020/10/23/20G00152/sg (accessed 07/08/2021). For the path to Italy’s adoption of the Faro Convention, some laws and documents are: DPCM 171/2014 art. 20 (not, accessed 07/08/2021); Nizzo 2018; Nizzo and Franceschini 2018; Capitano 2019; DPCM 76/2019 art. 35/d (https://www.gazzettaufficiale.it/atto/serie_generale/caricaArticolo?art.versione=1&art.idGruppo=6&art.flagTipoArticolo=0&art.codiceRedazionale=19G00081&art.idArticolo=35&art.idSottoArticolo=1&art.idSottoArticolo1=10&art.dataPubblicazioneGazzetta=2019-08-07&art.progressivo=0#art, accessed 07/08/2021).

We should, perhaps, include a fourth, 'd', point[4] – the use of new technologies as a metalinguistic language for teaching humanities. 'Learning by doing' is a very powerful paradigm in activating the human curiosity (Marras *et al.* 2016: 102). The class, for example, could analyse digitized antiques, reproduce them, explore new cultural ideas; then, hopefully, it will be possible to link the museums with the design, fashion, food industries close to the city where the institution is located – 'New relationships between the individual and society are established, artefacts are developed as open-source, and process is valued more than the final product'. (Xanthoudaki 2015: 254).

Connecting the techno-creative space with public municipalities and private societies is crucial. Public and private funds will increase the impact of local communities on their futures in an active way, weighing the role of museums as high-level research and development (R&D) institutions for territories and cities.

Lifelong learning

Artemova and Espinoza Torres (2020: 157) describe some important practices of participative projects in museums, but there is another central task for public museum projects which can be called participative: they have to encourage and disseminate the crucial role of 'lifelong learning' for the members of the museum communities. That idea can be highly improved by the application of information technologies. For instance, it is clearly beneficial for a contemporary museum to make daily use of the video-making techniques (i.e. the MNETRU – Etruschannel[5]) or the activation of platforms such as Moodle, to share and develop classroom experiences (i.e. OpenMakers for the historical centre of Cori, LT).[6] The aim must be to make a difference in 'lifelong learning' via an informal public space such as the public museums.

For the development of techno-creative spaces it is also essential to improve an informal environment. Through an informal approach, heritage organizations can aggregate various kinds of users, transforming a substantial part of their target from 'public' to 'community', who are more engaged, as Hsi (2008: 891) points out: 'Transformed by the presence of IT, informal learning institutions [...] are creating freely available educational resources accessible over computer networks and the Web to create extended learning opportunities outside of the formal schooling. Concurrently, informal learners are assimilating new IT technologies and transforming them into new practices and applications to support their curiosity and interests.'

N.	Name	Description
1	Evolution	To evolve the museum space and its exhibits digitally, and mediate a new kind of knowledge (STEAM).
2	Empowering	To empower the scientific segments and the editorial management of the heritage institution, writing about new discoveries and filing patents.
3	Interconnection	To build an interconnection between the museum and the neighbors of the city, creating stable and operative communities nearby the institution.

[4] The 'd' point comes from classrooms in the Museum 'Rodolfo Lanciani' in Guidonia (infra 3.5); it is valid for any Digital Humanities application.
[5] https://www.youtube.com/channel/UCor2tlNzAfuAcWwmvcJkz8A/playlists?app=desktop (accessed 07/08/2021).
[6] https://www.open-makers.eu/ GNU v.3.0 (accessed 07/082021).

4	Economic growth	To teach self-employment techniques helping family economies from a start-up level.
5	Return school	To reach out to young people prone to early school leaving, projecting open spaces with a FLOS habitat for them, which can stimulate their curiosity and spirit of believing in themselves.
6	Lifestyle rank	To increase in the museum communities the need of a plain cultural existence and growth in lifestyle ranking.
7	Deep study	To explore deeply with the communities the collections.
8	Museum Economy	To enrich the museum economy with new editorial products, open-access, online catalogues, linked open-data for projecting new web services for the online communities.
9	Research	To let to the citizens, investigate the daily fundamental role of the researchers and rise the appreciation on the great developments of science.
10	Challenges	To educate communities in solidarity, equality, environmental importance, and green habits (as reuse, recycling and self-made skills).

Table 1: Top 10 practices for the 21st-century museums, based on the study of the author during the case studies (*infra* 3).

From the last two points (9, 10), in brief, the role of the learner can be defined as follows, again using the words of Xanthoudaki (2015: 256): ‘The learner can become the cornerstone, the co-creator of knowledge, but also a self-confident citizen with noble goals: political and civic re-engagement opting for a more equal distribution of wealth, opportunity and power’.

These are the tasks for the near future; the challenges of the pandemic times could hopefully improve the development of smart-green-cities in a humanistic way.

Case Studies

There are two important Italian examples of public techno-creative spaces (Mandarano 2019) in a public museum: the ‘Liquid Museum’ of Cagliari, Sardinia (Marras *et al.* 2016) and the MArTA-Lab (Taranto) – the first Italian Fab Lab space in a public Museum, which started in January 2021.[7]

What follows is a brief presentation of the results of a three-year research and application of FLOS methodologies as applied to museums and archaeological sites. The research project started in 2017 and some of its goals were:

- To organize FLOS training and engage communities to funding and managing heritage places.
- To develop participative heritage projects, involving citizens, increasing the life-long learning habits into museums and heritage places.
- To reach new generations with the application of ‘makers’ solutions.
- To help the Faro Convention become an Italian law (infra 3.2).

[7] https://www.martalab.com/ (accessed 01/08/2021).

The project was applied via six activities based on FLOSS by the NGO Una Quantum inc. using its own funds.[8]

N.	Museum/ Archaeological site	Project name	Place/s	Year/s	Main activities
1	Museo delle Civiltà	FLOSS for archaeologists	Rome EUR	2018/2019 2019/2020	Teaching FLOSS
2	Museo Nazionale Etrusco	CIRCUITI	Rome Villa Giulia	2018/2019 2019/2020 2020/2021	Teaching FLOSS Virtual tour 360°
3	Appia Antica	Tactile panels	Mausoleo di Priscilla	2018–2021	Prototype of tactile panels
4	Bomarzo/Lugnano in Teverina	Field-school of digital excavation	Trocchi/ Museo di Lugnano in Teverina (Vt-Te)	2019	Digital archaeological Excavation, FLOSS teaching for archaeologists
5	Museo Civico Archeologico Rodolfo Lanciani	Museum management	Guidonia (Rm)	2018/2019 2019/2020	'FLOSS kids' camp, teaching FLOSS for archaeologists Virtual tour 360°
6	Museo delle Culture Villa Garibaldi	Scientific Direction	Riofreddo (Rm)	2020/2023	New FLOSS website and teaching FLOSS web tools

Table 2: The six activities by Una Quantum in cultural heritage management using FLOSS tools.

Rome – Museo delle Civiltà (MUCIV)

In the spring of 2018, an agreement with the Museum Service was signed to organize 'New technologies training for museology' in MUCIV spaces and engage museum communities in the study of monuments and collections. At the end of the courses a public exam, recognized by Italian Ministry of Education and Cultural Heritage, certified the final level.

Classroom Name	Number of Classes	Software/code	Average age	Participants
Coding	3	JavaScript	32	9
Photogrammetry	6	Regard 3D Mesh Lab Cloud Compare	25	31
3D modelling	5	Blender	23	29
Geographical Information Systems (GIS)	6	QGIS Pyarchinit	26	50
Virtual Tour 360°	3	Pannellum	23	15
Tot. classrooms	Tot. courses	Tot. FLOSS Software	Average age	Tot. Participants
5	23	7	25.8	134

Table 3: FLOSS classrooms in two-year activities at MUCIV, Rome.

[8] https://www.unaquantum.com (accessed 01/08/2021).

Rome – Museo Nazionale Etrusco (MNETRU)

Thanks to Dr. Mariflora Caruso, during the spring of 2018, responding to the call for action known as 'Faro Call' (Nizzo 2018), the Una Quantum inc. NGO group guaranteed this project more than 300 hours of FLOSS workshops during a two-year experiment. All the teaching activities were free for three types of educational classes: children, high-school students, and freelance.

Classroom Name	Number of Classes	Software/code	Mean age	Participants
Coding	2	Java Python	15	16
Photogrammetry	2	Regard 3D Mesh Lab Cloud Compare	24	48
3D modelling	2	Blender	25	52
GIS and Webgis	3	QGIS Pyarchinit	23	76
3D Printing	2	Cura	2	38
Free CAD and Laser Cutter	1	QAD	25	23
Virtual Tour 360°	2	Pannellum	27	32
Tot. classrooms	Tot. courses	Tot. FLOSS Software	Average age	Tot. Participants
7	14	11	20.5	285

Table 4: Free access classrooms at MNETRU of Rome during the 'Circuiti' program

The whole experience of the 'Faro Call' has been monitored by the *Direzione Generale Archeologia, Belle Arti e Paesaggio*, for the collection and study of 'quality of experience' related to the application of the Faro Convention in terms of Italian public museums. In this way a great contribution to the cause was made, which was also recognized by the museum managements involved.[9]

Rome – Parco dell'Appia Antica

Hands-on involvement using plexiglass for children with visual impairments was planned and realized in the 'Tomb of Priscilla', Rome, using FLOS technologies, thanks to the NGO 'Italia Nostra'.

The workflow included the following software/methodologies: QGIS, QGIS2threeJS, QAD, Laser Cutter, assembling. The total number of children (average age 7–8) attending the activities was 67.

[9] Protocol MiBAC|MN-ETRU|19/11/2018|0002022-P.

Bomarzo – Lugnano in Teverina Field school of Digital Excavation

Between May and July 2019, the international Field School of Digital Excavation took place in Trocchi – Bomarzo (Viterbo). The activities were divided into stratigraphic excavations in the morning and FLOSS training in the afternoon.

Classroom Name	Number of Classes	Software/code	Average age	Participants
GIS	1	QGIS	27	4
Photogrammetry	1	Regard 3D Mesh Lab Cloud Compare	26	6
3D modelling	1	Blender	24	8
Archaeological remains lab	1	QAD	26	4
Pyarchinit	1	Pyarchinit	24	4
Tot. classrooms	Tot. courses	Tot. FLOSS Software	Tot. Average age	Tot. Participants
5	5	7	25.4	26

Table 5: The digital excavation field-school.

Guidonia – Museo Civico Archeologico 'Rodolfo Lanciani'

The best chance to practice the model tested in the previous activities was the management experience of the Museum 'Rodolfo Lanciani' in Guidonia, managed by Dr Mariflora Caruso: labs, experiments, FLOSS training, maker activities, AI, coding, gaming for children and teenagers were included in a three-month FLOSS summer camp at the museum.

Months in 2019	Weeks of camp	Mean age	Participants
June	4	11	40
July	4	10	30
September	2	8	18
Tot. Months	Tot. Weeks	Tot. Average age	Tot. Participants
3	10	9.6	88

Table 6: Summer camp at the Museo Civico Archeologico Rodolfo Lanciani in Guidonia (Rome).

These are the numbers of participants from one of the most economically depressed areas of eastern Rome. At the 'EU Maker Faire 2019', the project was awarded the prize 'Maker of Merit – Blue Ribbon 2019'. For the youngest age group, technology through innovation is the key to spark their curiosity, discovery, and learning habits.

The main other activities between October 2018 and February 2020 were FLOS training for cultural heritage, the core business of the NGO, capable of opening and managing museums in a sustainable way.

Riofreddo (Rome) – scientific direction

N.	Name	License	Website	Developed by
1	Glamkit	BSD 3-Clause 'New' or 'Revised' License (2010)	https://www.glamkit.com/	Community
2	Omeka	GNU GPL v.3.0	https://omeka.org/	Roy Rosenzweig Center for History and New Media, and George Mason University)
3	Arches	GNU v.3.0	https://www.archesproject.org/	Getty Foundation
4	Collective Access (CA)	GNU v3.0	https://www.collectiveaccess.org/	Community

Table 7: Building the Museo delle Culture 'Villa Garibaldi (MUDECU), GNU' site. Some free licensed CMS for museum sites.

N.	Name of software/ Code/Coding libraries	License	Used for training in heritage sites	Used for products at heritage sites	Number of institutions involved
1	QGIS	GNU General Public License (GPL)	Yes	Yes	6
2	Blender	GNU GPL	Yes	Yes	5
3	Pyarchinit	GNU GPL	Yes	Yes	4
3	Gimp	GNU GPL	Yes	Yes	3
4	Pannellum	MIT license	Yes	Yes	3
5	Regard 3D	MIT license / third parties can be in 3-Clause BSD 3-Clause License	Yes	Yes	6
6	Mesh Lab	GNU GPL v3	Yes	Yes	6
7	Cloud Compare	GNU GPL v2	Yes	No	2
8	Python	Python Software Foundation License (PSFL)/ GNU GPL compatible	Yes	No	1
9	Java	Most major components of Java are available under GNU GPL	Yes	No	2
10	QAD	GNU GPL v3	Yes	Yes	2
11	Cura	GNU GPL v3	Yes	Yes	2
12	Leaflet	BSD 2-Clause 'Simplified'	Yes	Yes	2
13	OSM	Open Data Commons Open Database License (ODbL)	Yes	Yes	3

Table 8: Overview of FLOS software used, debugged, developed for training and labs in public museums.

Conclusion

The NGO taught at five museums and heritage-site communities. A community of 600 people, with an average age of *c.* 20, took part in these participative projects. These results were reached in less than 18 months without a starting budget. This is a new way to approach FLOSS philosophy, very different in terms of the use and development of software and open datasets.

N.	Description	Institution (Municipalities)
1	For highly certified training	MUCIV (Rome)
2	To support activities with a high cultural purpose	MNETRU (Rome)
3	To support activities with a high social purpose	Parco Archeologico dell'Appia Antica (Rome)
4	To manage and fundraise for two archaeological excavations	Lugnano in Teverina (Tr) and Bomarzo (Vt)
5	To manage communities and fundraise the museum	Guidonia Montecelio (Rome)

Table 9: Utility of FLOSS technologies in heritage-institution management.

Finally, we may present the business model and the future of techno-creative spaces in six core points.

N	Name	Description
1	Training/labs	Training and teaching is the core business. Training is possible to improve and enrich programs with new FLOS solutions.
2	Project hub	Heritage places in this way can apply new public calls for projects and boost research.
3	Technology	Projecting with new FLOS technologies, the institution can improve services (websites, virtual tours, social media, Fab-labs, new exhibitions) and communities (students, makers, and tech/lovers); it increases the variety of solutions and experiments.
4	Marketing	Spread technology by social events, stimulate the media and social communities to publish news about the institution each week.
5	Assessment	At the end of the cycle there is the need to study the results to improve the model for new, bigger and better, cultural sites/opportunities.
6	Share	The museology community in ArcheoFOSS and other FLOSS communities must engage in the mission for creating together a greater network of tecno-creative spaces.

Table 10: The business model for developing techno-creative spaces.

Bibliography

Artemova, I. and M. Espinoza Torres 2020. Proyectos de cultura participativa en los museos y el aprendizaje para toda la vida. In J.C. Sanabria Zepeda (ed.) *Cultura participativa en museos tecno-creativos. Reflexiones a partir del proyecto LITCA*: 147–66. Guadalajara: Arquitonica.

Capitano, A. 2019. Beni culturali, il Parlamento non ratifica la Convenzione di Faro. Ed è incomprensibile. In *Blog of Antonio Capitano* (blog). 15 June 2019. https://www.ilfattoquotidiano.it/2019/06/15/beni-culturali-il-parlamento-non-ratifica-la-convenzione-di-faro-ed-e-incomprensibile/5255548/ (accessed 07/08/2021).

Capron, E., D. Sagot-Duvouroux and R. Suire 2020. Anatomy of a Techno-Creative Community – The Role of Places and Events in the Emergence of Videomapping in Nantes. In N. Economides (ed.) *Economics of Networks eJournal*: 1–25 New York: Leonard N. Stern School of Business.
Council of Europe 2005. Council of Europe Framework Convention on the Value of Cultural Heritage for Society. In *Council of Europe Council of Europe Treaty Series* 199. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680083746 (accessed 07/08/2021).
De Marchi, V. 2020. Con gli occhi delle bambine atlante dell'infanzia a rischio 2020. In *Save The Children Italian Report* 11. https://atlante.savethechildren.it/index.html (accessed 07/08/2021).
Hsi, S. 2008. Museums and Cultural Heritage Centers and the Use of IT in Digital Repatriation. In J. Voog and G. Knezek (eds) *International Handbook of Information Technology in Primary and Secondary Education, Part One*, 20: 891–899. Enschede, Denton: Springer.
Mandarano, N. 2019. *Musei e media digitali*. Roma: Carocci editore.
Marras, A.M., M. Gerolama Messina, D. Mureddu and E. Romoli 2016. A Case Study of an Inclusive Museum: The National Archaeological Museum of Cagliari Becomes 'Liquid'. In K.J. Borowiecki, N. Forbes and A. Fresa (eds) *Cultural Heritage in a Changing World*: 99–111. Cham: Springer.
Méndez Lugo, R.A. 2011. Concepción, método y vinculación de la museología comunitaria. In *Cadernos De Sociomuseologia, Questões Interdisciplinares na Museologia*, edited by Universidade Lusófona de Humanidades e Tecnologias, 41: 45–58. Lisbon: Edições Universitárias Lusófonas.
Mikhak, B., C. Lyon, T. Gorton, N. Gershenfeld, C. McEnnis and J. Taylor 2002. Fab Lab: An Alternative Model of ICT for Development. In *Bangalore ThinkCycle 2nd international conference on open collaborative design for sustainable innovation 'development by design' (dyd02)*, edited by The MIT Center for Bits and Atoms. Bangalore: ThinkCycle. https://cba.mit.edu/events/03.05.fablab/fablab-dyd02.pdf (accessed 07/08/2021).
Nizzo, V. 2018. Avviso pubblico per la presentazione di manifestazioni di interesse finalizzato alla concessione di uso temporaneo di spazi del museo a titolo gratuito per la realizzazione di eventi di alto profilo culturale presso le sedi del Museo Nazionale Etrusco di Villa Giulia - Villa di Papa Giulio III e Villa Poniatowsky - rivolto ad associazioni no profit, onlus e organismi similari. In beniculturali.it. 28 February 2018. https://storico.beniculturali.it/mibac/multimedia/MiBAC/documents/1519827069992_MNETRU_AVVISO_MANIFESTAZIONE_INTERESSE_EVENTI_NO_PROFIT.pdf (accessed 07/08/2021).
Nizzo, V. and D. Franceschini 2018. Approvazione dello Statuto del Museo Nazionale Etrusco di Villa Giulia. In beniculturali.it. 5 April 2018. https://www.beniculturali.it/mibac/multimedia/MiBAC/documents/feed/pdf/D.M.%205%20APRILE%202018%20%20REP.%20189-imported-77855.pdf (accessed 07/08/2021).
Xanthoudaki, M. 2015. Museums, Innovative Pedagogies and the Twenty-First Century Learner: A Question of Methodology. In A. Filippoupoliti and S. Sylaiou (eds) *Museum Education Today: Creative Synergies and Pedagogic Innovations in Multicultural Contexts*: 253–271 Leicester: Museum and Society. https://doi.org/10.29311/mas.v13i2.329.

The *virtual countryman.* A GRASS-GIS tool for ancient cultivation recognition

Augusto Palombini

Augusto Palombini, CNR – Istituto per le Scienze del Patrimonio Culturale;
augusto.palombini@gmail.com

Abstract
A GRASS-GIS module is presented, aimed to identify areas with the highest potentiality in terms of agricultural use. It has been conceived for VR reconstruction of past landscapes but is also suitable for predictive archaeology. It may take into account geological soil features, distance from settlements, roads, water-streams and cost map, giving to each parameter a specific weight in the process. The script is available under the GNU GPL at: www.palombini.net/sw/country/.

Keywords: GIS, OPEN-SOURCE; SPATIAL ARCHAEOLOGY; LANDSCAPE ARCHAEOLOGY.

Introduction

As 3D modelling for the reconstruction of architectural remains and monuments is a well-known practice in the field of Virtual Reality applied to Cultural Heritage, it is not the same for the virtual reconstruction of natural environments and ecosystems of the past. This is due to a series of factors: monuments and buildings clearly belong to the domain of archaeology and heritage sciences, while environmental reconstruction may be variously addressed by geology, biology, climate studies, what makes it is intrinsically interdisciplinary. Moreover, the digital reconstruction of a building in a specific time is (at least theoretically) referrable to a 'real' model, that is to say that the monument had a specific shape, structure and aspect, although it may be only partially known to us; on the contrary, the reconstruction of a wood or a grassland is difficult to define, as such entities change continuously and are never homogeneous or identical to themselves. Despite of a 'reproduction' modelling, a procedural approach is more suitable, taking into account the rules underlying the vegetation and cultivation presence, in order to obtain reconstructions as close as possible to the ancient landscape, and such a goal may be obtained both in a virtual reality perspective (aimed at drawing realistic views of ancient landscapes) and for predictive archaeology purposes (aimed at inferring the position and coverage of cultivations and farm-related infrastructures).

On the one hand, such an approach may appear more complex, because of the many possible factors influencing landscape formation, but, on the other we may refer directly to the relevant study branch of landscape archaeology, which was developed for defining ancient environmental situations in relation to human behaviour.

Landscape archaeology has a long tradition in archaeological studies, and this paper is not aimed at providing a complete overview: it may be traced back to Gordon Childe's ecological approach to prehistoric economy (Childe 1958) and to the cultural ecology as a branch of

Figure 1: Virtual reconstruction of Iron Age and Roman landscapes in the Upper Tiber Valley (Arnoldus-Huydzendveld *et al.* 2012; Pietroni *et al.* 2013).

anthropology (Harris 1968: 902–922). Its development, indeed, is linked on the one hand to the spread of spatial analysis (Hodder and Orton 1976; Clarke 1977), on the other, to the need of prehistoric archaeology, lacking written documents, to infer reliable information on the economy of ancient populations (Higgs and Vita Finzi 1970). Such an approach has then bloomed into a real study domain thanks to the push of New Archaeology (Flannery 1976; Renfrew 1973) and of the spread of the computer-based spatial analysis (Ammerman and Cavalli Sforza 1984; Gaffney and Stancic 1991). In Italy, landscape archaeology had a significant development (Cambi and Terrenato 2000; Cambi 2003), and a series of projects have been carried out in central Italy according to such an approach (Patitucci Uggeri 2007; Arnoldus and Pozzuto 2009), as well as in the specific area of the present research (Cambi 2004; Di Gennaro and Guidi 2009; 2010).

Some ten years ago, in the context of the Tiber Valley Project (Arnoldus-Huydzendveld *et al.* 2012; 2013; Pietroni *et al.* 2013), a chance to deepen such a topic was given to my research unit: the project implied the VR reconstruction of the ancient landscape of the middle Tiber valley as a context for the Iron Age settlements and monumental Roman sites (Figure 1).

The pipeline was planned on the basis of an interdisciplinary approach and thanks to the effective collaboration between a team of archaeologists and Antonia Arnoldus-Huyzendveld, a geologist expert in the evaluation and reconstruction of ancient environments.

The starting point was the need to obtain a reliable digital elevation model (DEM) of the ancient situation, on the basis of the oldest data available: the aerial photographs taken during the '50s were used to obtain a photogrammetric model of the terrain. This DEM was then 'carved' on the basis of the known elements of ancient situation, in order to reach a final model as close as possible to the ancient situation (Arnoldus-Huyzendveld *et al.* 2013) (Figure 4).

The theoretical basis

Secondly, a theoretical model was needed to create the GIS-based map of the ancient ecosystems, the future basis of the VR environment reconstruction. The distribution of the natural vegetal coverage is relatively a simple task, as it is performed thorough the analysis of soil features and the definition of an eco-landscape map. More difficult was the identification of the potential cultivation areas. The result of the study was the formalization of a polynomial approach taking into account different factors affecting such an attitude, as parameters to be weighted:

$$M = (aX+bY+cZ+dK+W) \text{ [if } J<40\%]$$

Where:

M: areas of agricultural potential
X: map of distance from settlements (round buffers at 2, 4, 6 and 10 km)
Y: map of distance from roads (buffer at 1, 2 km)
Z: map of distance from rivers and water-streams (buffers at 300 and 1000 m)
K: simplified eco-landscape map (five categories of ecological potential)
J: slope map (only values below 40% are considered)
W: cost map (cost analysis): quantifies the effort needed to reach any point from the settlements

The basic concept underlying this formula was the idea that the complexity of society may influence the relevance of some parameters (proximity to sites, water, roads) in choosing the land suitable for farming. Relevance of proximity to sites is certainly a valid criterion for simple societies, where a territorial network is not well established: for instance, the 'site catchment analysis' model relies on this assumption, and aims to evaluate the feeding resources of a human group taking into account the surrounding area of settlements (Higgs and Vita Finzi 1970; Tiffany and Abbott 1982; Gaffney and Stancic 1991). Indeed, it may be argued that such a factor may lose its relevance in the frame of more complex societies, where – for instance – proximity to roads (instead of sites) may indicate more organized social nets. Water is crucial for cultivation, but the relevance of proximity to it may be weaker for more complex societies, capable of building complex irrigation systems. The relevance of cost (i.e. the effort to reach a place from bigger sites) is a complex issue that must be evaluated for each case: it may indicate the presence of complex societies (there is more affordable land and chances to choose the best) or simple ones (it is easier to overcome obstacles thanks to

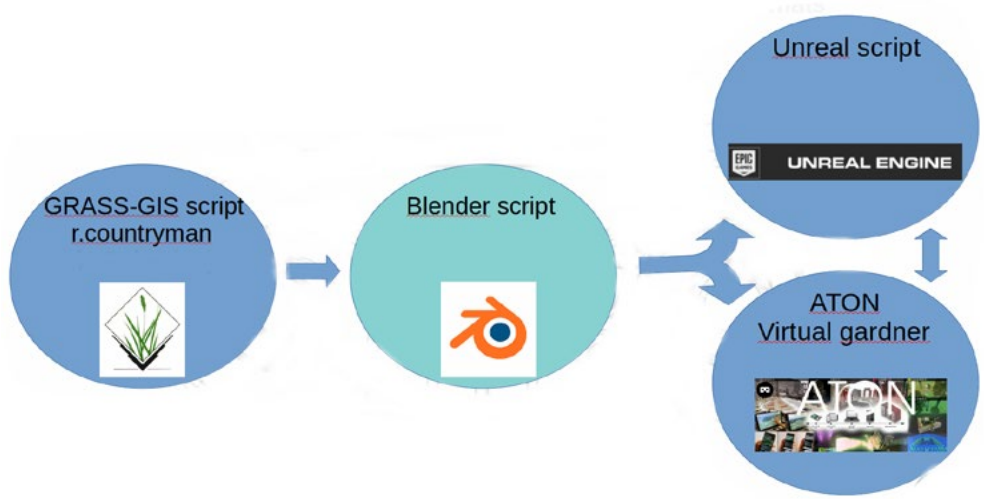


Figure 2: Flowchart of the landscape reconstruction pipeline, as conceived by the CNR Virtual Heritage Lab after the Tiber Valley Project.

a huge working crew, slaves, etc.). Thus, the above-mentioned formula was conceived to try to consider all such aspects, giving to each of them a specific weight in relation to the social context to be studied.

In the specific case study, the algorithm was applied both to Iron Age and Roman era landscapes, changing the weight of distance to settlements, roads and rivers or water-streams. The maps obtained were used as basis for Virtual Reality landscape reconstruction (Figure 1).

After the Tiber valley experience, the reflection on the topic was still discussed and deepened. Firstly, it appears clear that a tool conceived for landscape simulation could be considered as well as a resource for predictive archaeology. Then, the aim was set to move towards the perspective of a complete open-source pipeline for landscape reconstruction, covering from GIS map creation, to rendering, and to local or online browsing (Figure 2). The project went on over discussions and reconsiderations, and the proposed pipeline may change in some aspects. For instance, the need of a Blender module is still under discussion because the great capabilities of Unreal Engine 5 may partly or completely replace a Blender module.

In the meanwhile, the GIS part was developed, aimed at building a complete module in GRASS-GIS, as well as the instancing (gardening) tools for landscape reconstructions and online browsing were developed using the ATON framework (Fanini 2020; Fanini *et al.* 2019; 2021; Fanini 2020).

The new tasks and the r.countryman GRASS-GIS module

The previously built script was developed to become a complete module for GRASS-GIS (Neteler and Mitasova 2002), called *r.countryman*, and released under the GNU GPL licence, in accordance with the main GRASS-GIS software.

r.countryman identifies the most suitable areas for cultivation, i.e. the areas that most probably were cultivated in ancient times. A DEM, the water-stream network, the settlements and the road network are needed, each one having a user-defined weight. To perform the analysis, the script uses many other GRASS modules, the most relevant being: *r.sun*, *r.mapcalc*, *r.buffer*.

Besides the creation of the module interface, some changes in the general approach were made. These changes are the result of the reconsideration of the input parameters with experts in the field of ethnology and archaeology:

> **Modification of buffer zones.** According to literature on site catchment analysis, some buffer values have been modified: river and water-stream buffers have been changed from 300 and 1000 m to 1, 2 and 4 km; the road buffers changed form 1 and 2 km to 1, 2 and 3 km.
> **Addition of the mean annual solar radiance.** The previous parameter of sun exposition (aspect), was replaced by the more accurate yearly radiance factor, calculated through the GRASS-GIS module *r.sun* (Hofierka 1997; Hofierka and Suri 2004). This requires a longer calculation time, but it offers certainly more reliable results. The radiance calculations is computed for two mid-season days of the year.
> **High slope values are not excluded.** As in the previous version slope values higher than 40% were excluded from calculation, in the new version the slope threshold was set at 50%. This choice has been due to many reasons. On the one hand, in particular conditions, it is possible to arrange cultivations even in steep areas; on the other, the slope value is strongly connected to the map resolution, that in some cases may hide specific areas because of the interpolation.

The result may be summarized in the new version:

$$M = (aX+bY+cZ+dC+K+R+J)*[\text{if } J<50\%]$$

Where:
M: areas of agricultural potential
X: map of closeness to settlements (round buffers at 2, 4, 6 and 10 km)
Y: map of closeness to roads (buffer at 1, 2, 3 km)
Z: map of closeness to rivers and water-streams (buffers at 1, 2 and 4 km)
K: simplified eco-landscape map (categories of ecological potential)
J: slope map
R: mean annual solar radiance
C: cost map (cost analysis): quantifies the effort needed to reach any point from the settlements

The *r.countryman* module's graphic user interface consists of three main tabs (Figure 3):

> *Basic*: the 'basic' tab (Figure 3a) requires the indication of the input DEM and of the soil attitude geological map (at least one of the two is mandatory), as well as of the output map name (mandatory). These are the minimal required information for the module to be able to perform an output.

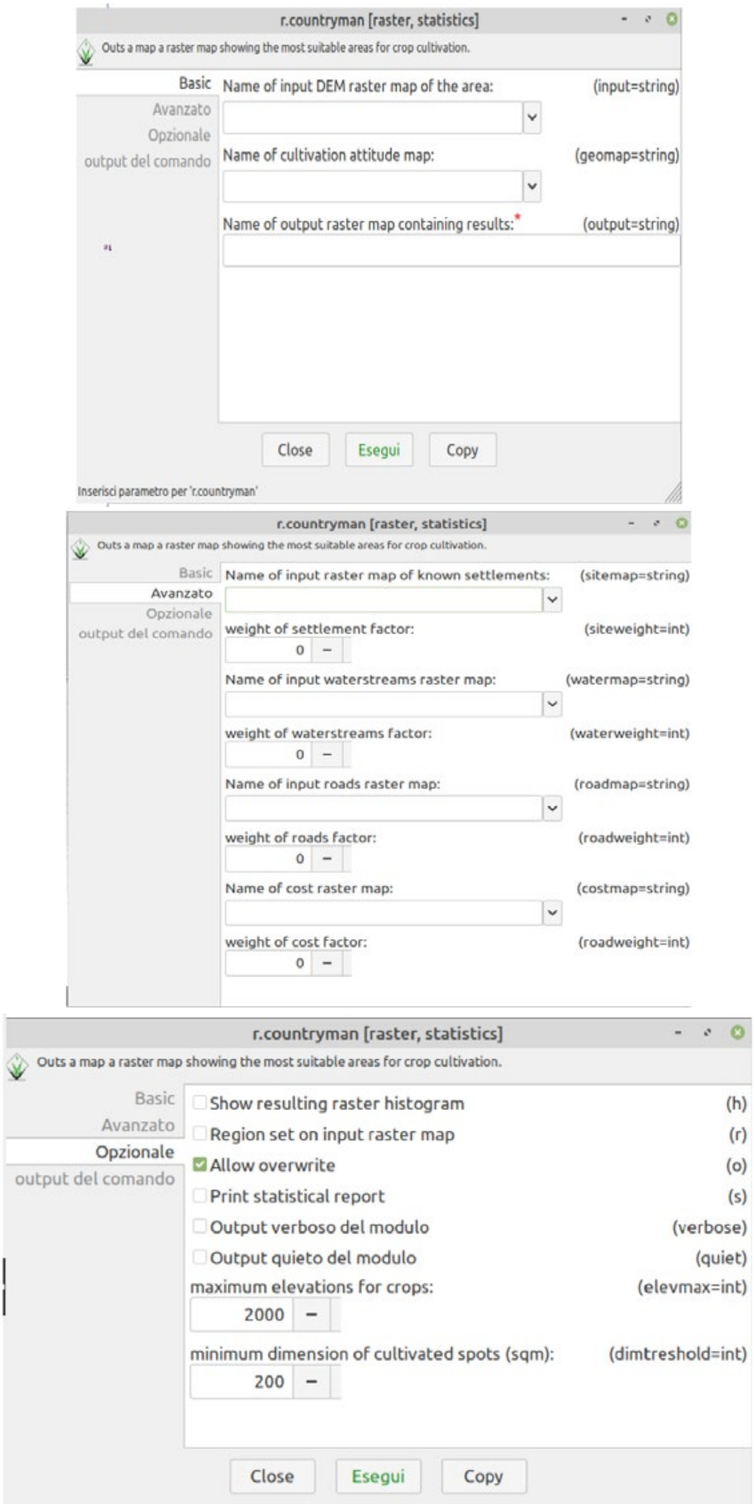


Figure 3: Main tabs of the GRASS-GIS graphic user interface of the *r.countryman* module (Basic, Advanced, Optional).

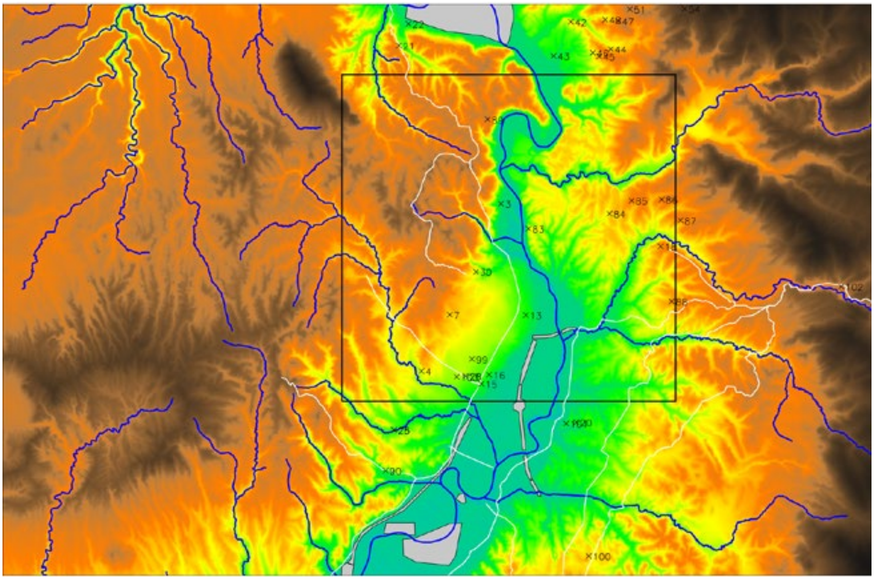


Figure 4: Map of the data sets considered in the analysis. The water-streams net (blue, with the modified Tiber path), The road net (white), the late imperial Roman sites (black crosses). The central square represents the extension of the DEM obtained by photogrammery from IGM Aerial photos, as the grey areas are the ones modified carving modern interventions (see text for explanation). Legend for the sites (see text for explanation): 3: Meana; 4: Villa Del Sasso; 7: Collina Belvedere; 13: Villa Baciletti; 15: Lucus Ferioniae; 16: Villa dei Volusii; 18: Cures Sabini; 20: Eretum; 21: Villa Coste Sor Nicola; 22: S.Andrea In Flumine; 25: Villa Rustica Monte Casetta; 28: Villa Prato La Corte; 30: Vicus Flavianus; 42: Villa Marcigliana; 43: Villa Cerro; 44: Villa S. Angelo; 45: Villa Formello; 46: Villa S. Antonio; 47: Villa S. Nicola; 48: Villa Colle Marrone; 50: Villa S. Luigi; 51: Villa S. Francesco; 54: Villa S. Valentino; 83–88:Villas at Cures 1–6; 89: Seperna; 90: Fontanile Vaccihereccia; 91: Fidene; 94: Villa Castel Giubileo; 97: Villa Tor S. Giovanni; 98: Nomentum; 99: Villa Standa; 100: Grotta Marozza; 101: Eretum; 102: Villa Nerola; 103: Villa Prato La Corte Bigatti.

Advanced: the 'advanced' tab (Figure 3b) requires data for a more accurate elaboration: settlements, water-streams, roads and cost raster maps may be added with the relative weight, to be considered in relation to social complexity.
Optional: the 'optional' tab (Figure 3c) gives access to other functions such as: output statistics on the cultivation potential areas; the possibility to include thresholds for maximum elevation above the sea level for cultivated areas (which may differ for crops and contexts); the option to set a minimum dimension in square meters of spots, in order to avoid or allow the identification of smaller areas.

The countryman at work

A first test of the module has been performed on the same data set of the Tiber Valley Project, but limiting the analysis to the Roman settlements of middle and late Imperial era (Figure 4). The aim was to check the effectiveness by changing different parameters of the agricultural landscape. The maps in Figure 5 show the results of running the script with the different

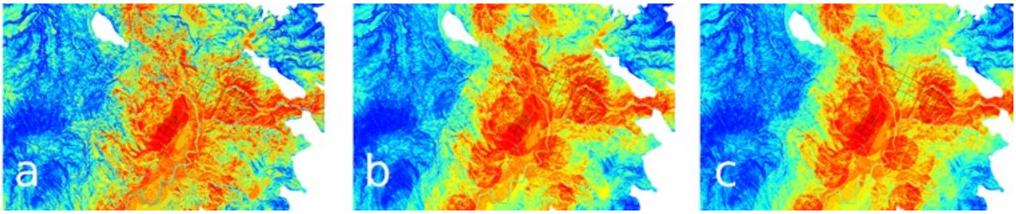


Figure 5: Maps resulting from running *r.countryman* module, giving different weights to the parameters, specified in the Table 1, in comparison to the traces of Roman centuriation identified in the archaeological landscape (see part 4 and table. 1 for details).

parameters as listed in table. 1, in comparison to the traces of Roman centuriation identified in the archaeological landscape.

What follows is a description of the dataset used for the analysis is provided, only partly corresponding to one by the Tiber Valley Project:

- The DEM was obtained through a laborious workflow, in order to reach a result as close as possible to the ancient situation (Pietroni *et al.* 2013). The central part of the study area was reconstructed thanks to the photogrammetric elaboration of a set of 12 aerial images by IGM (1954), digitized at 20 micron (*c.* 2600 dpi; camera Fairchild KF246 with 152 mm lens; height of flight: *c.* 5600 m; ground sampling distance: *c.* 40 cm; *c.* 30 GCP for georeferencing). For the surrounding area, modern satellite elevation data were used (res. 10 m), and the DEM was 'carved' in correspondence of areas clearly altered by anthropic actions, which were 'smoothed' so as to interpolate the natural geomorphological conditions (Figure 4).
- Site position refers to the sites attested for the middle and late Roman Empire (since the late first century AD onwards). Settlements in a broader sense, included villas, were considered as sites, after a consideration of the archaeological remains sufficient to infer a stable human presence (Figure 4) (Castagnoli *et al.* 1985; Carbonara and Messineo 1994; Cambi 2004; Sternini 2004; De Franceschini 2005).
- Roman roads were digitized from the *Carta dell'Agro* (Carta dell'Agro 1990).
- Rivers and water-streams net (Figure 4) correspond to the modern ones, with the exception of the Tiber course, which has been changed according to the adjusted DEM and paleoenvironmental considerations (Arnoldus *et al.* 2013; Pietroni *et al.* 2013).

Figure 5a shows an ideal situation in which the same weight is given to all parameters. Figure 5b refers to a situation with a higher weight (3) for the proximity from sites and a lower one for the proximity to roads. This is the typical situation of a primitive society, lacking an articulated organization. On the contrary, Figure 5c may fit to a situation of a more advanced social organization, where proximity to sites is still relevant (3) but proximity to water becomes less relevant (1) and roads increase their importance (2).

These maps have been compared to the centuriation network (Castagnoli *et al.* 1985), and a calculation has been performed on the total surface of the *centuriae* falling into each map.

Figure 5a is clearly the most distant from the centuriation landscape; Figure 5b and Figure 5c are certainly more fitting. Even if they may look quite similar, a more detailed analysis shows that maps with higher values for the road system (c), a higher space is covered by *centuriae* falling into the highest quartile of cell values (Table. 1).

	Parameter weight: Site: 1 Water: 1 Roads: 1 (Figure 4a)	Parameter weight: Site: 3; Water: 2 Roads: 1 (figure. 4b)	Parameter weight: Site: 3; Water: 1; Roads: 2 (figure. 4c)
Sqm data range	18–32	27–41	24–41
Total non-null sqm	15018	1420475	1430878
1° quartile (highest)	529 (3.5%)	322727 (22.7%)	581069 (40.6%)
2° quartile	745 (4.9%)	846210 (59.5%)	565072 (39.4%)
3° quartile	9344 (62.2%)	231943 (16.3%)	265940 (18.5%)
4° quartile (lower)	4400 (29.2%)	19595 (1.3%)	18797 (1.3%)

Table 1: Features of different output maps compared to the Roman centuriation: they are shown in square meters, the entity of centuriae falling into the cells of different parameters weight, and the relative cell values. Map c has the largest space on the highest quartile cells.

Conclusions

Predictive archaeology is a wide and discussed research domain, and it is always hard to present a new analysis tool without getting into a complex discussion on its problems and perspectives (Hodder and Orton 1976, Kamermans *et al.* 2009, Lieskovsk *et al.* 2013). At the same time, virtual archaeology needs tools for shaping ancient landscapes according to a serious – if not realistic – and plausible approach. This work aims to present a module to investigate ancient landscape and try to check areas with higher and lower agricultural potential, according to a spatial analysis approach which is well rooted in archaeological literature (Higgs and Vita-Finzi 1970; Flannery 1976; Gaffney and Stancic 1991; Hunt 1992). The interpretation of such results, usable both for purposes of predictive archaeology and for virtual reconstruction of reliable landscapes, are strictly tied to wider archaeological considerations. As a matter of fact, relations among ancient settlements, paths, water-streams and cultivated land, may be complex, interrelated, and may have as well different ways in terms of cause-effect chains. Moreover, this kind of analysis may be strongly affected by data resolution. The *r.countryman* script may be considered as a tool useful to find analogies among data, according to an heuristic approach in which classification is simply an explorative strategy, and the responsibility of interpretation is always in the hand of the researcher's creativity (Palombini 2001).

r.countryman is available for download at: https://www.palombini.net/sw/country/ Documentation and tutorial are available as well, although currently in progress. The author hopes the topic may be discussed by the Community and is available to reach any kind of suggestions or critics.

Acknowledgments

The creation of this tool represents an ideal follow-up of a work started and conceived together with Antonia Arnoldus-Huydzendveld († 2018) and is dedicated to her memory.

Bibliography

Ammerman, A.J. and L.L. Cavalli Sforza 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Cambridge: Cambridge University Press.

Arnoldus-Huydzendveld, A. and E. Pozzuto 2009. Il castello nel contesto territoriale. In C. Citter (ed.) *Dieci anni di ricerca a Castel di Pietra: edizione degli scavi, 1997-2007*: 1-49. Borgo S. Lorenzo: All'Insegna del Giglio.

Arnoldus-Huydzendveld, A., D.F. Marco Di Ioia, A. Palombini, E. Pietroni, V. Sanna and S. Zanni 2012. The Virtual museum of the Tiber Valley Project. *Virtual Archaeology Review* (https://www.researchgate.net/publication/307780388_The_Virtual_Museum_of_the_Tiber_Valley_Project : accessed 12/7/2021).

Arnoldus-Huyzendveld, A., A. Palombini, E. Pietroni and V. Sanna 2013. Verso una metodologia condivisa per l'analisi del paesaggio antico: il progetto 'Valle del Tevere'. In M. Serlorenzi (ed.) *ARCHEOFOSS 2012, free, libre and open-source software e open format nei processi di ricerca archeologica*: 104–11. Archeologia e Calcolatori 4. (https://www.researchgate.net/publication/294582163_Verso_una_metodologia_condivisa_per_l%27analisi_del_paesaggio_antico_il_progetto_Valle_del_Tevere : accessed 12/7/2021).

Cambi, F. 2003. *Archeologia dei paesaggi antichi: fonti e diagnostica.* Roma: Carocci.

Cambi, F. 2004. Le campagne di *Falerii* e di Capena dopo la romanizzazione. In H. Patterson (ed.) *Bridging the Tiber: Approaches to Regional Archaeology in the Middle Tiber Valley*: 75–101. Rome: British School at Rome.

Cambi, F. and N. Terrenato 2000. *Archeologia dei paesaggi.* Roma: Carocci.

Carbonara, A. and G. Messineo 1994. *Via Tiberina.* Roma: Istituto Poligrafico dello Stato.

Carta dell'Agro. 1990. *Carta Storica, Archeologica, Monumentale e Paesistica del Suburbio e dell'Agro Romano*. Roma.

Castagnoli, F., P. Longo, M. Cancellieri, M.P Muzzioli, G. Gazzetti, E. Stanco, P. Aureli and M. Incitti 1985. La centuriazione nel Lazio. Resti di divisioni agrarie nel territorio dell'odierno Lazio. In R. Bussi and V. Vandelli (eds) *Misurare la terra: centuriazione e coloni nel mondo romano. Città, agricoltura e commercio: materiali da Roma e dal suburbio*: 38–60. Modena: Edizioni Panini.

Childe, V.G. 1958. *The Dawn of European Civilization*. London: Routledge.

Clarke, D.L. 1977. *Spatial Archaeology*. Boston: Academic Press.

De Franceschini, M. 2005. *Ville dell'Agro Romano.* Roma: L'Erma di Bretschneider.

Di Gennaro, F. and A. Guidi 2009. Ragioni e regioni di un Cambiamento culturale: modi e tempi della formazione dei centri protourbani nella valle del Tevere e nel Lazio meridionale. *Scienze dell'antichità* 15: 429–445.

Di Gennaro, F. and A. Guidi 2010. Lo stato delle anime come mezzo per la ricostruzione della popolazione dei villaggi preistorici. In *Arqueología della Población. Actas VI Coloquio Internacional Arqueología Espacial, Arqueología Espacial 28*: 351–358. Teruel: Seminario de Arqueología y Etnología Turolense.

Fanini, B. 2020. Applications of a compact session encoding for immersive WebVR/XR analytics. In M. Lo Turco *et al.* (eds) Pavia *Digital & Documentation – Digital strategies for Cultural Heritage. Vol. 2*. Pavia: Pavia University Press.

Fanini, B., D. Ferdani and E. Demetrescu 2021. Temporal Lensing: An Interactive and Scalable Technique for Web3D/WebXR Applications in Cultural Heritage. *Heritage* 4.2: 710–724.

Fanini, B., S. Pescarin and A. Palombini 2019. A cloud-based architecture for processing and dissemination of 3D landscapes online. *Digital Applications in Archaeology and Cultural Heritage*, 14. DOI: https://doi.org/10.1016/j.daach.2019.e00100.

Flannery, K. 1976. *The Early Mesoamerican Village.* New York: Academic Press.
Gaffney, V. and Z. Stancic 1991. *GIS approaches to regional analysis: a case study of the island of Hvar.* Ljubljana: Univerza v Ljubljani.
Harris, M. 1968. *L'evoluzione del pensiero antropologico. Una storia della teoria della cultura.* Bologna: Il Mulino.
Higgs, E. and C.V. Finzi 1970. Prehistoric Economy in the Mount Carmel Area of Palestine: Site Catchment Analysis. In *Proceedings of the Prehistoric Society* 36: 1–37.
Hodder, I. and C. Orton (eds) 1976. *Spatial Analysis in Archaeology.* Cambridge: Cambridge University Press.
Hofierka, J. and M. Suri 2004. A New GIS-based Solar Radiation Model and Its Application to Photovoltaic Assessments. In *Transactions in GIS* 8: 175–190.
Hofierka, J. 1997. Direct solar radiation modelling within an open GIS environment. In *Proceedings of JEC-GI'97 conference in Vienna, Austria.* Amsterdam: IOS Press.
Hunt, E.D. 1992. Upgrading Site-Catchment Analyses with the Use of GIS: Investigating the Settlement Patterns of Horticulturalists. *World Archaeology* 24.2: 283–309. http://www.jstor.org/stable/124829 (accessed: 12/7/ 2021.)
Kamermans, H., M. Van Leusen and P. Verhagen (eds) 2009. *Archaeological prediction and risk management.* Leiden: Leiden University Press.
Lieskovský, T., R. Ďuračiová and L. Karell 2013. Selected Mathematical Principles of Archaeological Predictive Models Creation and Validation in the GIS Environment. *Interdisciplinaria Archaeologica, Natural Sciences in Archaeology,* http://www.iansa.eu/papers/IANSA-2013-02-lieskovsky.pdf (accessed: 12/7/2021).
Neteler, M. and H. Mitasova 2002. *Open-Source GIS: A GRASS GIS Approach.* New York: Kluwer Academic Publishers.
Palombini, A. 2001. Miti e pregiudizi nell'uso dei sistemi classificatori in archeologia. *Rassegna di Archeologia preistorica e protostorica* 2001: 127–144.
Patitucci Uggeri, S. (ed.) 2007. *Archeologia del paesaggio medievale. Studi in memoria di Riccardo Francovich.* Firenze: All'Insegna del Giglio.
Pietroni, E., A. Palombini, A. Arnoldus-Huyzendveld, M. Di Ioia and V. Sanna 2013. Tiber Valley Virtual Museum: 3D landscape reconstruction in the orientalising period, North of Rome: A methodological approach proposal. In *Digital Heritage International Congress (DigitalHeritage).* Piscataway : IEEE, 223–230.
Renfrew, C. 1973. *Before Civilization: The Radiocarbon Revolution And Prehistoric Europe.* Cambridge: Cambridge University Press.
Sternini, M. 2004. *La romanizzazione della Sabina tiberina*. Bari: Edipuglia.
Tiffany, J.A. and L.R. Abbot 1982. Site-Catchment Analysis: Application to Iowa Archaeology. *Journal of Field Archaeology* 9.3: 312–322.

Little Minions and SPARQL Unicorns as tools for archaeology

Timo Homburg, Florian Thiery

Timo Homburg, Hoschschule Mainz; timo.homburg@hs-mainz.de
Correspondence: qgisplugin@sparqlunicorn.link
Florian Thiery, Römisch-Germanisches Zentralmuseum; mail@fthiery.de
Correspondence: qgisplugin@sparqlunicorn.link

Abstract
We introduce the SPARQLing Unicorn QGIS Plugin as a tool for querying and converting Linked Open Data (LOD) resources and making them accessible using QGIS. The plugin enables QGIS to access LOD resources for the first time, and can be used to include LOD in QGIS projects to enrich geospatial data sets with information gained from the Semantic Web and to prepare geospatial data sets for the publication as LOD resources. We illustrate these functions using examples from an archaeological context, and show how spatial LOD can be published and made accessible using LOD browser implementations.

Keywords: LINKED OPEN DATA; WIKIDATA; SPARQL.

Introduction

In their daily work, archaeologists rely on a wide variety of tools which help to solve particular tasks. Many such tools consist of self-made scripts or home-grown software, which are often of use to other researchers as well. We call them *Little Minions* as referred to by the *Computer Applications and Quantitative Methods in Archaeology* (CAA) working group on *Little Minions*. Apart from tools, researchers have the possibility of sharing research data online in order to further the scientific discourse about their particular topic of research. While this is, in general, a good development it also poses new challenges. Research data needs to be findable, accessible and at best provided in data formats which can be processed by at least the anticipated target user group. Research data which is not accessible in this manner is what we call an *unknown modern data dragon*. This is where the *CAA SIG on Semantics and LOUD in Archaeology* comes into play, which aims to promote the ideas behind LOD (Thiery *et al*. 2019). In this paper, we would like to show that these shortcomings can be overcome by applying linked data and *Semantic Web* technologies and how user-friendly access to this ever-growing wealth of *Semantic Web* data can be given to the archaeological community with the so-called *SPARQL Unicorn tool framework*. Finally, we present the *SPARQLing Unicorn QGIS Plugin* as one example of a tool providing user-friendly access to *Linked Data* in the GIS domain.

Linked Data

Wuttke (2019) shows that areas which are unknown to the map creator were described in ancient times by the phrase *Hic sunt dracones* (engl. *here be dragons*). Today, the web gives geodesists the possibility of sharing their geo data and enables them to participate in the scientific and political discourse. However, much of this shared data is not findable or accessible, thus resulting in *modern unknown data dragons*. Often these *data dragons* lack connections to other data sets, i.e. they are not interoperable and can therefore lack usefulness, reusability or usability. To overcome these shortcomings, a set of techniques, standards and recommendations can be used: Semantic Web and Linked Open Data (LOD), the FAIR principles (Wilkinson *et al.* 2016) and LOUD.[1] Sir Tim Berners-Lee introduced the concept of the semantic web by using the ideas of open-data, semantically described resources and links, as well as usable (machine-readable) interfaces (standardised in the Resource Description Framework: RDF[2]) and applications for creating a Giant Global Graph (Thiery *et al.* 2019). 'The Semantic Web isn't just about putting data on the web. It is about making links so that a person or machine can explore the web of data.' (Berners-Lee 2006). A five-star rating system of openness[3] was introduced to rate linked data, i.e. 'Linked Open Data (LOD) is linked data which is released under an open license.' (Berners-Lee 2006). Furthermore, LOD must be usable for scientists and programmers in order to take full advantage of the full power of LOD. Following the LOUD principles makes LOD even more FAIR.

Linked Data toolset for the cultural heritage domain

For use in archaeology, researchers can rely on a set of tools which allows them to access linked data resources. We can divide them into two sections: (i) tools and (ii) data hubs. Famous linked data tools (section i) in the cultural heritage (CH) domain were developed by members of the Pelagios Network,[4] which connects researchers, scientists and curators to link and explore the history of places. This includes the Recogito[5] online platform for collaborative document annotation, which provides a personal workspace for uploading, collecting and organising source materials – texts, images and tabular data — for collaborative annotation and interpretation. Another LOD search engine is Peripleo,[6] where it is possible to access diverse data about ancient places and physical objects, such as archaeological finds, data about people, time periods, and even geo-tagged literature, or data transcribed from historic maps. The Linked Pasts Network (Grossner *et al.* 2017) hosts some smaller initiatives, e.g. the Linked Pipes[7] idea to store a catalogue of LOD tools and pipelines on GitHub, using RDF and a specialised ontology, which is managed by the Linked Pasts Community (Thiery 2020). The basis of these tools are well-known domain specific LOD hubs (section ii), e.g. Pleiades (Simon *et al.* 2016) as a geospatial hub, Nomisma (Gruber *et al.* 2011) as an ancient coin hub, Kerameikos[8] (Gruber, Gondek and Smith 2019) as an ancient ceramics hub, as well as Wikidata as a community driven data hub.

SPARQL Unicorn

The idea for a SPARQL Unicorn was born at the CAA conference in Kraków, Poland. At a networking event interesting matters came to light: very few databases are made free and

[1] https://linked.art/loud (accessed 01/07/2021).
[2] https://www.w3.org/RDF/ (accessed 01/07/2021).
[3] https://5stardata.info/en/ (accessed 01/07/2021).
[4] https://pelagios.org/ (accessed 01/07/2021).
[5] https://recogito.pelagios.org/ (accessed 01/07/2021).
[6] https://peripleo.pelagios.org/ (accessed 01/07/2021).
[7] http://linkedpipes.xyz/ (accessed 01/07/2021).
[8] http://kerameikos.org/ (accessed 01/07/2021).

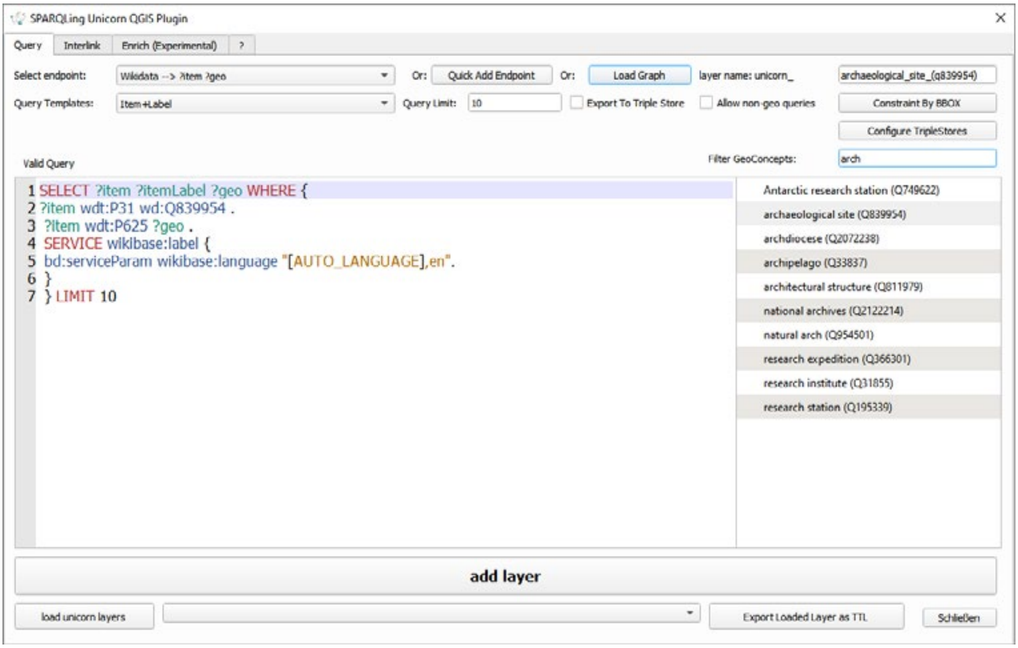


Figure 1: The SPARQLing Unicorn QGIS Plugin (T. Homburg/F. Thiery CC BY 4.0).

openly available and accessible, and even fewer are linked into the Linked Open Data Cloud. This creates challenges for comparative analyses of records across multiple data sets. As mentioned in section 3, Wikidata could be an alternative as a data source and linking hub based on the SPARQL-query-language. But there is a lack of user-friendly, easy to use, FOSS tools, especially for LOD technologies and repositories as well as for Wikidata itself. The SPARQL Unicorn is an idea for a user-friendly series of tools for researchers who are working with Wikidata and other related linked data repositories. SPARQL Unicorn's aim is to help researchers from the humanities or geospatial domains in using the community driven data from Wikidata and to make it accessible to those without expertise in *LOD* or the SPARQL query language (Seaborne and Prud'hommeaux 2006). The Research Squirrel Engineers network and the SPARQLing Unicorn QGIS Plugin working group brings this idea to life and proposes the SPARQL Unicorn principles (Thiery *et al.* 2020), which aim to create LOD and design tools for researchers to use such data without any knowledge of Semantic Web technologies.

SPARQLing Unicorn QGIS Plugin

As shown in section 3, there is a lack of FLOS GIS Tools for LOD in the archaeological and cultural heritage domain. The SPARQLing Unicorn QGIS Plugin[9] (Homburg and Thiery 2020) allows accessing linked geo spatial data resources and makes them usable in a user-friendly way using the open-source Software QGIS. The QGIS plugin, called sparqlunicorn,[10] (GNU

[9] https://github.com/sparqlunicorn/sparqlunicornGoesGIS (accessed 01/07/2021).
[10] https://plugins.qgis.org/plugins/sparqlunicorn (accessed 01/07/2021).

General Public License v2.0[11]) can be installed from within QGIS (by activating experimental plugins for installation) or downloaded from the central QGIS repository. The SPARQLing Unicorn QGIS Plugin allows the execution of Linked Data queries in (Geo)SPARQL to selected triplestores and geo-enabled SPARQL endpoints (Figure 1), and thus prepares the results of the queries in QGIS for the geo-community. The plugin currently offers three functions: (A) simplified querying of Semantic Web data sources (B), enrichment of geodata and (C) transformation of GeoJSON to RDF data.

Example Data

To illustrate how the SPARQLing Unicorn QGIS Plugin can be used in an archaeological context we first present the sample data with which we work throughout this publication.[12] This consists of two data sets. The set includes forts, fortlets and watchtowers along the Upper Germanic *Limes*. The second data set includes a public road network in Roman Britain. In the following subsections, we will examine how the SPARQLing Unicorn QGIS Plugin can help to make working with these data sets more meaningful by applying core functions of the plugin.

limestownpard	limestown	limesobject	ID	limescategory
Orlen	Taunusstein	Zugmantel	1	Kastell

Table 1: Limes data set (example).[13]

The *Limes* data set (Table 1) is an excerpt of data published by the *Deutsche Limeskomission*[14] consisting of a categorisation of the *Limes* data object, an object ID, a town to which the *Limes* object is associated, a description of the object and, if present, a sub-categorisation of the town, e.g. its district.

city	group	name	county	grid	roman_name
Carlisle	2	luguvalium	Cumbria	NY3956	LVGVVALIVM

Table 2: Trade network data set (example).[15]

The trade network data set (Table 2) is an excerpt of *Itinerarium provinciarum Antonini Augusti*[16] regarding Great Britain, a famous itinerary, a register of the stations and distances along various roads. Seemingly based on official documents, possibly from a survey carried out under the Roman emperor Caracalla, it describes the roads of the Roman Empire (Chrisholm 1911). In particular, the data consists of a given modern city name, a group id, a modern county name, a grid position according to British National Grid (BNG) and a Roman name.

[11] https://github.com/sparqlunicorn/sparqlunicornGoesGIS/blob/master/LICENSE (accessed 01/07/2021).
[12] https://github.com/Research-Squirrel-Engineers/limes (accessed 01/07/2021).
[13] https://github.com/Research-Squirrel-Engineers/limes/blob/master/gj2ttl/ObergermanischRaetischerLimes.geojson (accessed 01/07/2021).
[14] http://www.deutsche-limeskommission.de (accessed 01/07/2021).
[15] https://github.com/Research-Squirrel-Engineers/limes/blob/master/gj2ttl/ItinerariumAntoniniAugusti-IterBrittaniarum.geojson (accessed 01/07/2021).
[16] Based on https://web.archive.org/web/20131216035218/http://www.roman-britain.org/antonine-itinerary-map.htm (accessed 01/07/2021).

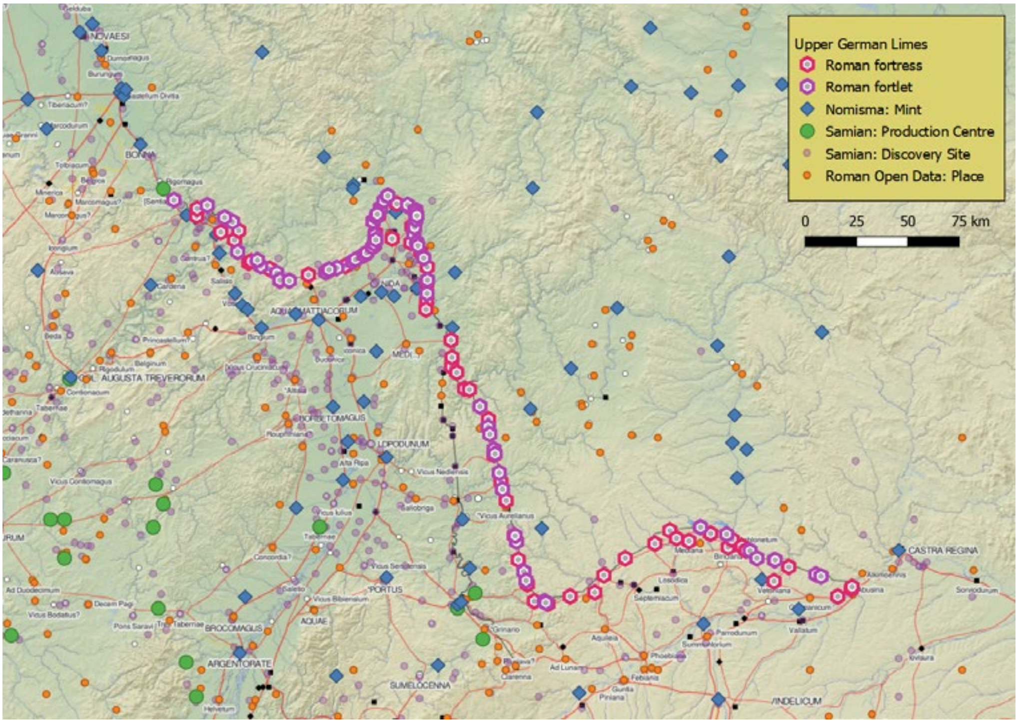


Figure 2: The German Upper Limes combined with data from LOD resources, e.g. Nomisma, Roman Open Data and Linked Open Samian Ware (F. Thiery CC BY 4.0, via Wikimedia Commons https://commons.wikimedia.org/wiki/File:LOD_Upper_German_Limes.png).

Function A: Querying Linked Data

'Function A' is the ability for assisted SPARQL-querying. The SPARQLing Unicorn QGIS Plugin can retrieve geospatial concepts from a given triple store database and may make the geospatial concepts visible and accessible to QGIS. A user may query all attributes of a given geospatial concept using a preconfigured query or – with knowledge of the SPARQL query language – create a customised query for their needs. The results of the SPARQL query are converted to a QGIS vector layer either with or without geometry, depending on the query result. Using this function it is possible to get related data for the data sets described in section 5.1. For example, we could query Mints in Nomisma (cf. SPARQL Query Q1; geometry attributes in red, concept data in green), Roman Open Data and Linked Open Samian Ware repository to find out about Roman mints,[17] the findspots of Roman amphora[18] and Samian Ware production centres/findspots[19] in the vicinity of our data sets. Figure 2 (German Upper *Limes*) and Figure 3 (Roman roads) show our example data (section 5.1) combined with query results obtained by the SPARQLing Unicorn QGIS Plugin from Nomisma, Roman Open Data and Linked Open Samian Ware. The layers for *Limes* points and Roman routes were obtained from a triple store set up to provide this data. A style has been applied to illustrate the result. Finally, 'Function A' provides the possibility of creating SPARQL query variables from already existing

[17] http://nomisma.org/ontology (accessed 01/07/2021).
[18] https://romanopendata.eu/sparql/doc/index.html (accessed 01/07/2021).
[19] https://rgzm.github.io/samian-lod/doc/ (accessed 01/07/2021).

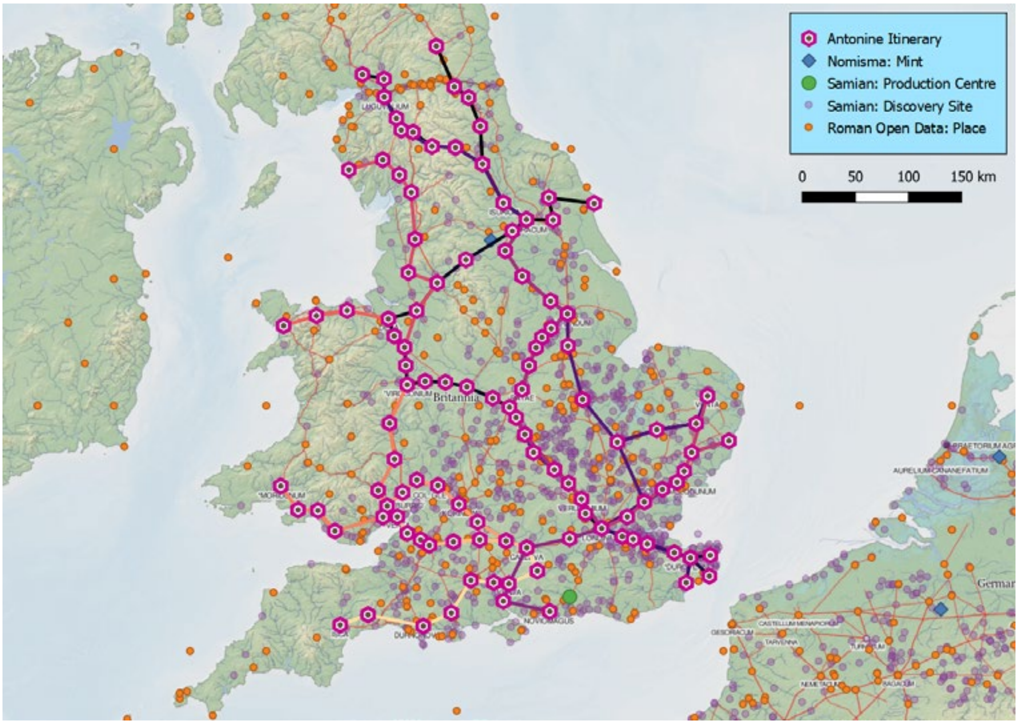


Figure 3: *Itinerarium Antonini* trade network combined with data from LOD resources, e.g. Nomisma, Roman Open Data and Linked Open Samian Ware (F. Thiery CC BY 4.0, via Wikimedia Commons https://commons.wikimedia.org/wiki/File:LOD_Itinerarium_Antonini.png).

QGIS layers. This function allows the querying of specific geometries with a list of matching values represented by the query variable. Furthermore, it is possible to query a non-geo-layer, which can be used as a CSV file to combine geospatial results with non-geospatial data.

SELECT ?item ?lat ?lon WHERE { ?item a <http://nomisma.org/ontology#Mint>. ?item geo:location/wgs84_pos:lat ?lat . ?item geo:location/wgs84_pos:long ?lon . }	(Q1)

Function B: Data enrichment

Function B provides the enrichment of existing *QGIS* vector layers with information present in the Semantic Web. The process begins with a search for appropriate data that is thematically fitting for enrichment and — once appropriate data attributes have been defined — leads to a modified QGIS vector layer with data imported from the Linked Open Data Cloud. We illustrate this with the example of the Upper Germanic *Limes* data set. This data set includes forts and fortlets along the *Limes* but does not include their altitude. This means that no analysis of their history which takes into consideration the altitude of the point at which the *Limes* outpost was built can be taken into account. This information may be relevant when

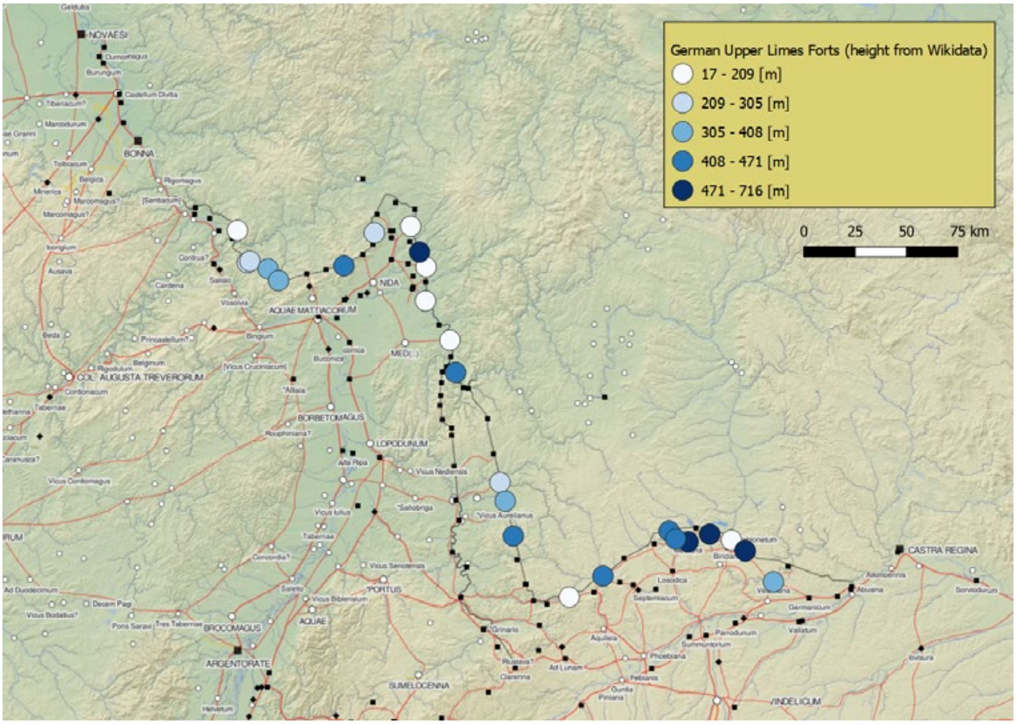


Figure 4: German Upper Limes extended with Wikidata information (elevation above sea level) (F. Thiery CC BY 4.0, via Wikimedia Commons https://commons.wikimedia.org/wiki/File:LOD_Upper_German_Limes_Wikidata_Enriched_Height.png).

attacks on the *Limes* need to be reconstructed. While this information may not be available in linked data either, as an approximation the altitude of the nearest settlement can be used. This information represented by the Wikidata property elevation above sea level (P2044; cf. SPARQL Query Q2 in green) is very often found in the Wikidata knowledge base. We can now ask the SPARQLing Unicorn QGIS Plugin to enrich this information to the given QGIS layer. To achieve this, we need an at best a unique identifier in our data set which can also be found in the knowledge base. In our case this is the ‘limestown’ attribute (cf. VALUES statement in SPARQL query Q2 in orange) and the language German. In addition we narrow down our search result by only asking for municipalities of Germany (Q262166; cf. SPARQL Query Q2 in red). With this information we can ask the SPARQLing Unicorn Plugin to execute a query exemplified by a simplified SPARQL query Q2 in the background to match the elevation above sea level attributes which are displayed in Figure 4. The results show that now the question of where the *Limes* might have been attacked frequently may be approximated.

```
SELECT ?item ?elevation WHERE {
   VALUES ?limestowns { “Idstein”@de, “Rosbach”@de }
   ?item wdt:P31 wd:Q262166 .                              (Q2)
  ?item rdfs:label ?limestowns .
  ?item wdt:P2044 ?elevation . }
```

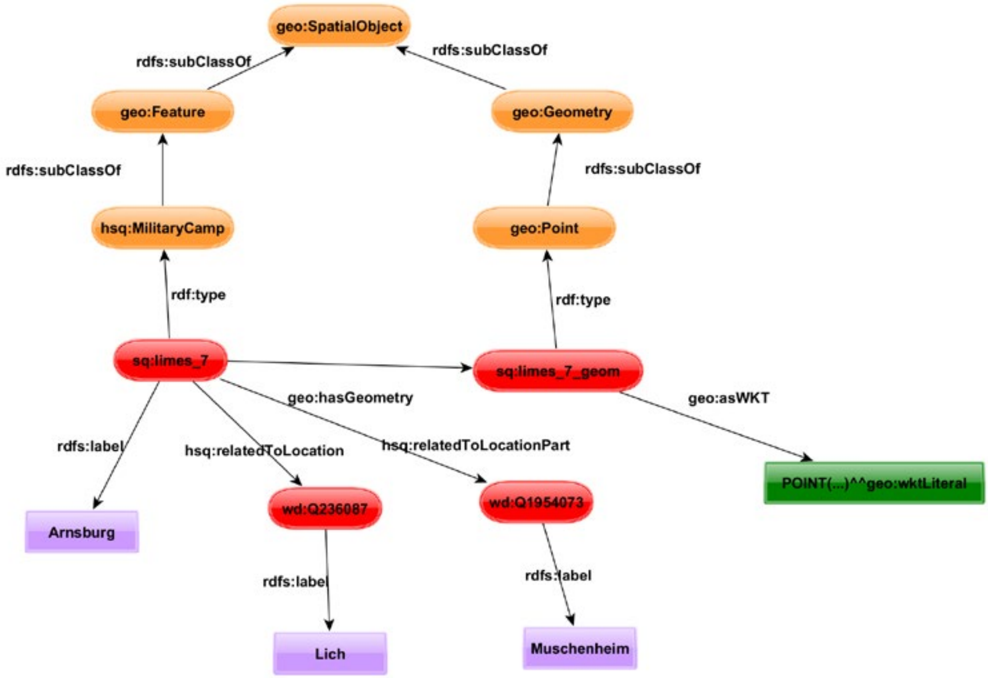


Figure 5: Example of a semantic uplift process (T. Homburg CC BY 4.0).

Function C: Geodata to RDF conversion

Function C provides the possibility of converting a QGIS vector layer to RDF. This can be achieved in a generic automatic conversion process, yielding only a semantic representation of the geometry. An alternative option is for the user to use and/or create a mapping profile which maps QGIS vector layer columns to semantic concepts. To create such a mapping schema, the SPARQLing Unicorn QGIS Plugin offers the possibility of searching for appropriate geospatial concepts, importing and exporting mapping schemas and performing the conversion of a QGIS layer to RDF.

Mapping Schemas to Semantic Web concepts

Mapping schemas map columns of vector layer representations to URIs on the Semantic Web. With this information geospatial data sets can be converted to linked data representations in an automated process with a maximum of information. Currently, creating a mapping schema involves expert knowledge in how to create an RDF graph. In the future, our plugin might make suggestions for mappings or rely on state-of-the-art interlinking tools for the conversion process.

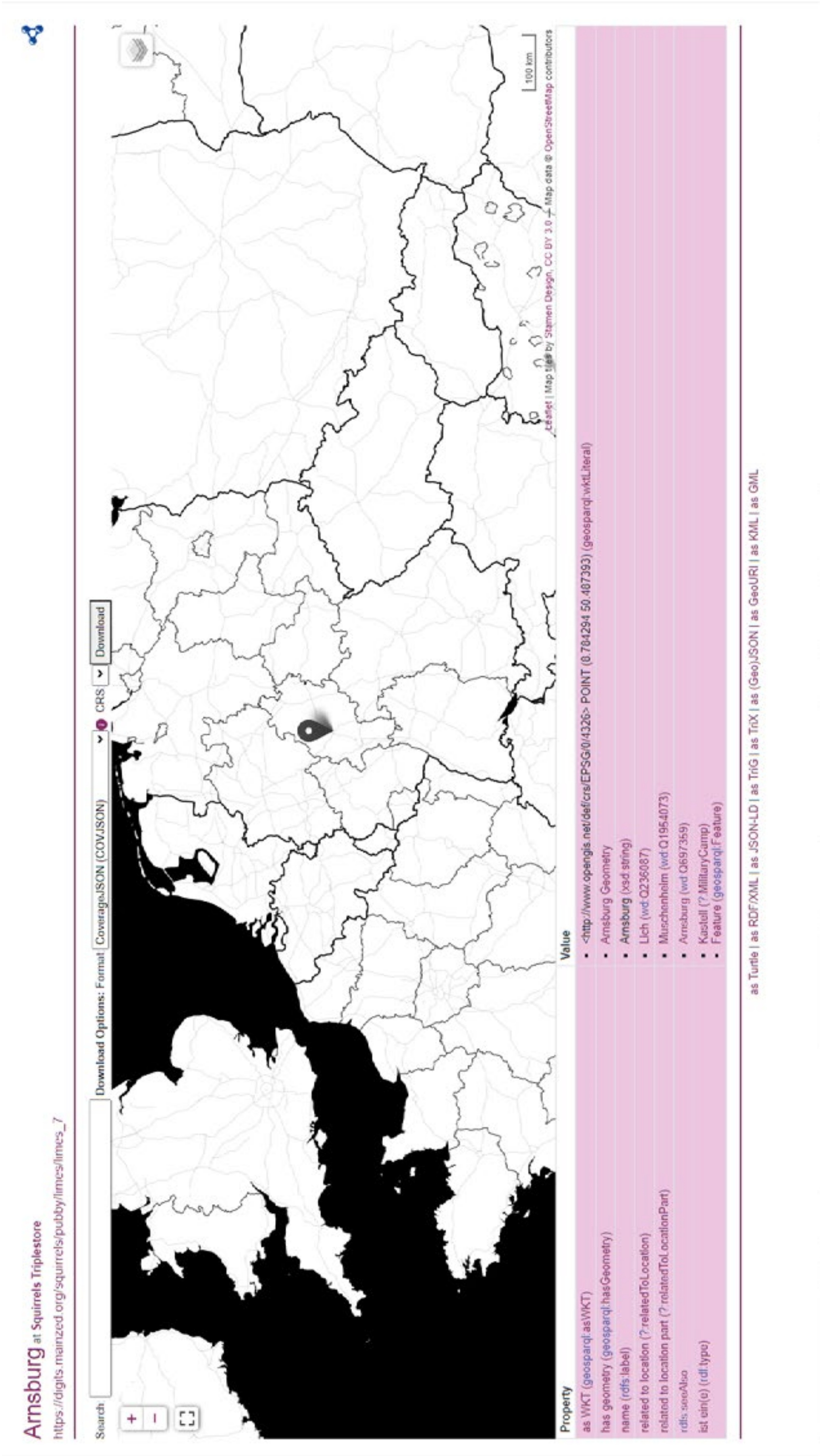


Figure 6: Integrated data shown in the GeoPubby Linked data browser interface (T. Homburg/F. Thiery CC BY 4.0).

Conversion process of geospatial data to RDF

Once a mapping scheme has been created, an automated conversion process from a feature collection to an RDF graph may take place. This conversion process converts every feature to an instance of a given owl:Class, the geometry to a geometry representation in the GeoSPARQL vocabulary[20] and all attached properties to properties in the RDF graph. The result is a graph representing the whole geodata set, which may be serialized in any of the RDF representations supported. Figure 5 shows an example of a semantic uplift process for a row in the *Limes* data set. Each column is mapped to a corresponding value in another knowledge base as specified in the mapping schema (here Wikidata). The geometry column is converted to a WKT literal according to the GeoSPARQL specification.

Using generated RDF data: Hosting and publishing

To be useful, created linked data sets need to be hosted and be made accessible in a SPARQL endpoint web service (Prud'hommeaux and Seaborne 2008). In addition, linked data frontends should ensure the indexing of linked data in search engines and allow users without specific software to browse linked data effortlessly. This requires a web server, including a semantic database and a linked data frontend as a web service. For our example data we use GeoPubby[21] as a linked data frontend for rendering linked geodata. GeoPubby is able to convert into a variety of different geo-formats, is aware of coordinate reference systems and can display vector geometries of all kinds. An example is shown in Figure 6, which displays the fort Arnsburg.[22]

Conclusion

We have presented the SPARQLing Unicorn QGIS Plugin, a plugin which may be used to enrich linked data resources in order to solve, among others, application cases in archaeological contexts. In the future, the plugin should serve two main purposes which we wish to focus on. Firstly, the plugin's user interface should be enabled to execute a maximum of commonly occurring queries without the need to type SPARQL queries in the query input. Secondly, we want to make entering SPARQL queries easier. Methods to achieve this are query autocompletion and further techniques. In general however, we would like the archaeology community to give us feedback about the plugin and foster further developments. The software is designed as an open-source and community project, which means we will be happy if linked data enthusiasts and researchers in the cultural heritage and archaeological domain will help us to develop the plugin further as best as we can, inspired by the SPARQL Unicorn principles.

Acknowledgments

We are grateful to David Wigg-Wolf for English language corrections, and would also like to thank the Research Squirrel Engineers network for their technical and mental support. This research received no external funding. The research is a community work, applied and supported by the Research Squirrel Engineers network.

[20] https://www.ogc.org/standards/geosparql (accessed 01/07/2021).
[21] https://github.com/i3mainz/geopubby (accessed 01/07/2021).
[22] http://lod.squirrel.link/data/limes/limes_7 (accessed 01/07/2021).

Bibliography

Berners-Lee, T. 2006. Linked Data – Design Issues. https://www.w3.org/DesignIssues/LinkedData.html (accessed 01/07/2021).

Chisholm, H. (ed) 1911. *Antonini Itinerarium*. *Encyclopædia Britannica*, 11th edn, Vol. 2. Cambridge: Cambridge University Press.

Grossner, K., T. Hill, R. Simon, R. Light, A. Bosse, G. Bodard, M. Ridge and W. Schmidle 2017. *From* Linking Places to a Linked Pasts Network. https://github.com/LinkedPasts/linkedpasts.github.io/blob/master/assets/lp_whitepaper.pdf (accessed 01/07/2021).

Gruber, E, R. Gondek and T.J. Smith 2019. Kerameikos.Org: Linked Open Data for Greek Pottery. https://doi.org/10.5281/zenodo.3243538.

Homburg, T. and F. Thiery 2021. *SPARQLing Unicorn QGIS Plugin* (version v0.12.2). zenodo. https://doi.org/10.5281/zenodo.3786814.

Prud'hommeaux, E. and A. Seaborne 2008. SPARQL Query Language for RDF. *W3C Recommendation*, January 15.

Thiery, F. 2020. #LinkedPipes – Linked Data Tools and Workflows. https://doi.org/10.5281/zenodo.4311282.

Thiery, F., S.C. Schmidt, T. Homburg and M. Trognitz 2020. The SPARQL Unicorn: An Introduction. https://doi.org/10.5281/zenodo.3742185.

Thiery, F., M. Trognitz, E. Gruber and D. Wigg-Wolf 2019. Hic Sunt Dracones! The Modern Unknown Data Dragons. https://doi.org/10.5281/zenodo.3345711.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3 (March): 160018.

Wuttke, U. 2019. 'Here Be Dragons': Open Access to Research Data in the Humanities. *Ulrikewuttke* (blog). April 9, 2019. https://ulrikewuttke.wordpress.com/2019/04/09/open-data-humanities/ (accessed 01/07/2021).

The ArchAIDE Archive: the open-data policy and management of material covered by copyright

Francesca Anichini, Gabriele Gattiglia

Francesca Anichini, University of Pisa, Department of Civilisation and Forms of Knowledge; francesca.anichini@unipi.it
Gabriele Gattiglia, University of Pisa, Department of Civilisation and Forms of Knowledge; gabriele.gattiglia@unipi.it

Abstract
This paper is focused on one of the less well known aspects of the ArchAIDE project: the open-data policy and management of material covered by copyright. To achieve the correct management of the material that falls under copyright or database protection, the EU directives on Copyright (2001/29/EC) and Database protection (96/9/EC) were analysed. Furthermore, participating in the H2020 open-data pilot, ArchAIDE was committed to creating sustainable outputs where the project held the copyright. This included making the interoperable, multilingual vocabularies, video corpus, 2D and 3D models, and the source code and neural network models created by the project available.

Keywords: IPR MANAGEMENT; COPYRIGHT; OPEN-DATA.

Introduction

The ArchAIDE project (2016–2019) was a European funded project developed by a consortium spread within five countries (Germany, Israel, Italy, Spain and UK). The project built a system for the automatic recognition of pottery with an innovative app for tablets and smartphones (Gualandi, Gattiglia, and Anichini 2021). This goal has been implemented by developing two distinct neural networks for appearance-based and shape-based recognition, which lay in the creation of a digital comparative collection, incorporating existing digital collections, digitised paper catalogues, and multiple photography campaigns. For this reason, the EU directives on copyright (2001/29/EC) and Database protection (96/9/EC) were analysed to achieve the correct management of the material that falls under copyright or database protection.

IPR Management

In the European Union, the copyright is governed by the Directive 2001/29/EC, whereas Directive 96/9/EC governs the Database protection.

The Directive 2001/29/EC of 22 May 2001 on harmonising certain aspects of copyright and related rights in the information society, also known as the InfoSoc Directive (Information Society Directive), implements the WIPO Copyright Treaty and harmonises the aspects of copyright law across Europe. Copyright is the legal right granted to an author, composer, playwright, publisher or distributor to exclusive publication, production, sale, or distribution of creative works with enough originality (individual character) to warrant such a right in the literary, scientific, and artistic domains.[1] Copyright is made of two components: exploitation rights and moral rights.

Exploitation rights are those that can be transferred and licensed and represent the economic value of the copyright. Moral rights are non-transferrable and, in some European legislations, perpetual. They protect the creator and guarantee that attribution must be given. In Europe, this right lasts for 70 years after the death of the latest creator of a published work. In some countries, when the author is not a natural person (e.g. an institution) or the author is anonymous/pseudonymous, this right lasts for 70 years after the first publication. The rights are not absolute but framed by limitations and exceptions to copyright law, including scientific research use. A major limitation on copyright is that it only protects the original expression of ideas and not the underlying ideas themselves (Wright *et al.* 2016: 4).

The Directive 96/9/EC of 11 March 1996 on the legal protection of databases (also known as Database Directive) considers both copyright and *sui generis* right to protect the databases. Article 3.1 provides that 'databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright'. The Directive creates *sui generis* protection against unauthorised use or extraction of the facts in the database, although again, not to the facts themselves. Article 7.1 stipulates that 'Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilisation of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.' *Sui generis* right lasts for 15 years, but each time a database is substantially modified, a new set of rights are created for that database. A *sui generis* database right is considered a property right, comparable to but distinct from copyright, that exists to recognise the investment made in compiling a database, even when this does not involve the creative aspect that is reflected by copyright. *Sui generis* right is created automatically and does not have to be registered to have an effect. The *sui generis* right can be divided into two rights: the extraction right and the re-utilisation right. Owners have the right to object to copying substantial parts of their database, even if data is extracted and reconstructed bit by bit.

The implementation of the ArchAIDE project can be described as scientific research where (i) part of the content is extracted from outside sources (see below), which falls under copyright or database protection; (ii) the content is, when necessary, transformed to fit operational needs; (iii) the content is loaded into a data set or repository; (iv) researchers gain access to the data and analysis tools are applied to the data set; (v) new knowledge is created. As for the outside sources, they can be mainly outlined as: (i) books and journal papers which contain the paper catalogue and are under copyright protection; (ii) database such as the ADS Roman Amphorae digital resource;[2] Ceramalex Hellenistic and Roman Pottery in Alexandria digital resource that is covered by *sui generis* right, whose uses and limitations were stipulated in the ArchAIDE Consortium Agreement.

[1] https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/european-ip-helpdesk/europe-glossary/glossary-c_en (accessed 10/06/2021).
[2] https://archaeologydataservice.ac.uk/archives/view/amphora_ahrb_2005/ (accessed 23/11/2020).

Both Directives provided specific exceptions for scientific research in Article 5.3.a) of the InfoSoc Directive, Article 6.2.b) and Article 9.b) of the Database Directive.

InfoSoc Directive

Article 5.3 a) of the InfoSoc Directive provides for an exception to the right of reproduction (Article 2) and the right of communication to the public (Article 3) when the protected work is used 'for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved'.

As it stands, the text of Article 5.3 a) allows acts of reproduction, communication to the public and distribution consisting in the use of protected works for scientific research. In the Articles or the Recitals of the InfoSoc Directive, there are no indications that the use should be limited to extracts of the protected works (Triaille *et al.* 2014: 59). The Netherlands, Greece and Slovakia did not implement an exception for scientific research in their national legal order. In the absence of a legal provision recognising an exception for scientific research, the researchers were presumably infringing the copyright of the authors of the works used for the scientific research. The United Kingdom law does not use the expression scientific research, considering the insertion of the latter word as redundant. Reproduction is included in all the national provisions, to a greater or lesser extent. Most of the time, this act of reproduction permits digital reproduction, explicitly or not. Communication is referred to in Belgium, France, Luxemburg, and (along with publication) in Italy, making available to the public and transmission in Germany, borrowing in Hungary, summary and quotation, and publication on the Internet in Italy. Acknowledgement (i.e. the indication of the source including the author's name) is required in Belgium, France, Germany, Luxemburg, Hungary, Italy, Denmark, and the United Kingdom (sufficient acknowledgement). Most of those countries provide for this indication 'except when this is impossible' (Triaille *et al.* 2014: 57–58).

Finally, the non-commercial nature of the purpose pursued is implemented in Belgium, Denmark, France, Germany, Hungary, Italy, and the United Kingdom. The non-commercial criterion can be understood as follows: 'Non-Commercial means not primarily intended for or directed towards commercial advantage or monetary compensation' as used in the CC licenses ('Non-Commercial Interpretation – Creative Commons', n.d.).

Beneficiaries of the research exception are not defined because of the adoption of a functional approach based on the activity of doing research. Hungary, Denmark, Italy, Luxemburg, and the United Kingdom have not included in their domestic law a specific beneficiary. In Germany, the exception authorises the reproduction of protected works 'for one's own scientific use' and the making available 'exclusively for a specifically limited circle of persons for their personal scientific research to be made available to the public'.

To summarise, the use of works (data, texts, images, etc.) for ArchAIDE purpose will not infringe authors' exclusive rights if (cumulative conditions):

- The works are used for the sole purpose of scientific research.
- The source, including the author's name, is indicated unless this turns out to be impossible.
- Works are used to the extent justified by the non-commercial purpose to be achieved.
- The use of the works does not conflict with a normal exploitation of the work or other subject/matter and does not unreasonably prejudice the right holder's legitimate interests.

Secondly, Article 5.2 c) of the Infosoc Directive provides for an exception to the reproduction right (Article 2) for certain non-profit making establishments, such as publicly accessible libraries and equivalent institutions, as well as archives: 'c) In respect of specific acts of reproduction made by publicly accessible libraries, educational establishments or museums, or by archives, which are not for direct or indirect economic or commercial advantage.'

Database Directive

As for our purpose, we take into consideration two different articles of the Databases Directive: 6.2.b) related to databases protected by copyright and 9a) for databases protected by sui generis right.

Article 6.2.b) contains the exception to copyright for scientific research, which applies not to the works contained in a database but to the structure of the database itself: 'Member States shall have the option of providing for limitations on the rights set out in Article 5 in the following cases: [...] (b) where there is use for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved.'

This exception has been implemented in Belgium, Spain, the United Kingdom, and Italy. The requirements of indication of the source and 'non-commercial purpose to be achieved' have been taken up in all these Member States, and the exception always applies to the database as a whole (and not just a part of it). Some Member States have sometimes added conditions for the exercise of this exception: in Italy, the user can 'access and visualise' the database without the author's consent, but 'permanent reproduction [...] shall always be subject to the right holder's authorisation'. The Netherlands, Germany, Poland, Luxembourg, Denmark, and Hungary have not implemented the exception for scientific research to the copyright protection of databases; moreover, they make no explicit reference to databases in their copyright legislations (Triaille *et al.* 2014: 69).

In other words, in the event that process involves an act of copying of the structure of the database, this will not infringe the author's rights if the user can prove that:

- The database is used for the sole purpose of scientific research.
- The sources are indicated.
- The database is used to the extent justified by the non-commercial purpose to be achieved.

It is not explicitly stated in the Database Directive that the exception only applies to published databases (Triaille *et al.* 2014: 71). Article 9b) contains the exception for scientific research to the *sui generis* right contained in the Database Directive: 'Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorisation of its maker, extract or re-utilise a substantial part of its contents: [...] (b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved.'

If the database is not freely accessible, users will only be lawful if they can avail themselves of authorisation as granted through the Consortium Agreement. Under Article 9b) of the Database Directive, scientific research does not have to be the sole purpose behind the use of the database. Article 9 provides for an exception to the right of extraction and not to the right of re-utilisation.

The exception to the *sui generis* right for scientific research has been implemented in: Belgium, Spain, the United Kingdom, the Netherlands, France, Germany, Poland, Luxembourg, and Hungary. The requirements of 'indication of the source' and 'non-commercial purpose to be achieved' have been taken up in all these member states and that the exception always applies to a 'substantial part of the database'. As for the copyright, the United Kingdom does not specify that the research must be scientific. Finally, the exception has not been implemented in Italy and Denmark (Triaille *et al.* 2014: 79–81).

The extraction of data, for the purpose of ArchAIDE, will not infringe the database maker's rights when the user is the lawful user of the database and can prove that (cumulative conditions):

- Data are extracted for the purpose of illustration for teaching or scientific research.
- The source is indicated.
- Data are extracted to the extent justified by the non-commercial purpose to be achieved.

Article 9b) of the Database Directive does not include the adjective 'sole' purpose of scientific research. As a result, uses done for non-commercial purposes other than scientific research the exception would still cover, i.e. statistical analysis, etc.

Discussion

Analysing the scientific research exceptions in the InfoSoc Directive and the Database Directive, we may conclude that:

As regards the area of copyright: published works, mentioning the source and the authors' name of the works, can be used to the extent justified by a non-commercial purpose; the use of the structure of published databases can only be used, mentioning the source, to the extent justified by a non-commercial purpose.

As regards the *sui generis* right: databases can be used, even if scientific research is not its sole purpose, mentioning the source and the authors' name of the works, to the extent justified by a non-commercial purpose.

Participating in the H2020 open data pilot, ArchAIDE was committed to creating sustainable outputs where the project held the copyright. Unfortunately, not all the collected data could be disseminated as open data. The research exceptions allowed by the EU Directives do not mean the ArchAIDE project automatically holds the copyright to the newly digitised or remixed data. Negotiation with copyright holders will be necessary for making these data available outside the project. ArchAIDE is able to demonstrate that paper catalogues, once digitised, can be actively re-used, also many years later from the first publication. This opens the possibility of reaching an agreement with publishers and other data providers for making their resources available in new ways, 'with a tangible benefit (seeing their data in use within the app), thus furthering the long-term discourse around making research data open and accessible' (Anichini *et al.* 2020). Instead, data owned by the project, i.e. multilingual vocabularies, videos created by the project, as well as the 2D and 3D models created from the ADS Roman Amphorae digital resource, were made available as open-data for download.[3] The ArchAIDE archive contains 2D vector drawings in SVG format and interactive 3D models navigable through a 3DHOP 3D viewer that can also be downloaded for 3D printing. These models exemplify an excellent standard of best-practice re-use. When the Roman Amphorae digital resource was deposited in 2005, creating automated 2D and 3D models for training a neural network could not have been a use envisioned. As 2D and 3D models were produced for each type included in the digital resource, it was possible to link the two archives, amplifying their mutual usefulness. The multilingual vocabularies were published from the UoY SPARQL endpoint and are also freely available for download and re-use in other Linked Open-data projects focussed on archaeological pottery. Finally, the ArchAIDE archive contains the video corpus created to both documents and promote the project, including the 30-minute ArchAIDE Documentary, created with footage shot over the course of the project. The ArchAIDE video archive represents a unique record of the project and an unusual record of the experience of implementing a European Commission-funded project with partners working across several countries.

It is also hoped that the thousands of photos taken by the project for training the algorithms and currently not available might result in new comparative collections that could be deposited as open research data into the ArchAIDE archive. Still, in many European countries, copyright on cultural heritage is very restrictive and does not allow us to make available the images of potsherds taken by ArchAIDE partners in national and regional collections. Showing the usefulness of these data within the ArchAIDE application might help convince Cultural Heritage national institutions to move towards more open-data policies. Finally, the source code and neural network models are publicly available as open-source in a GitHub repository[4] to allow re-use and future development by other researchers. Although all the data collected by users are, by definition, private and are not published, and all system components are designed to comply with this privacy statement, the system offers the option to publish the data as open data. Sponsoring the open-data philosophy and EU open data pilot, ArchAIDE

[3] https://archaeologydataservice.ac.uk/archives/view/archaide_2019/ (accessed 26/11/2020).
[4] https://github.com/mappaLab/archaide-software (accessed 21/07/2021), https://github.com/mappaLab (accessed 30/11/2020).

suggests to the user to share the data with the community, leaving each user the choice to do that or not.

Acknowledgements

This research was supported by the EU Horizon 2020 grant agreement No. 693548. We thank all the members of the ArchAIDE team (www.archaide.eu). This contribution was conceived by Francesca Anichini (AF) and Gabriele Gattiglia (GG) together. AF wrote 1. Introduction, 2.1 Infosoc Directive. GG wrote 2.2 Database Directive and 3. Discussion. 2. IPR Management was written by the authors together.

Bibliography

Anichini, F., F. Banterle, J. Buxeda i Garrigós, M. Callieri, N. Dershowitz, N. Dubbini, D. Lucendo Diaz, T. Evans, G. Gattiglia, K. Green, M.L. Gualandi, M.A. Hervas, B. Itkin, M. Madrid i Fernandez, E.M. Gascón, M. Remmy, J. Richards, R. Scopigno, L. Vila, L. Wolf, H. Wright and M. Zallocco 2020. Developing the ArchAIDE Application: A Digital Workflow for Identifying, Organising and Sharing Archaeological Pottery Using Automated Image Recognition. *Internet Archaeology*. 18 March 2020. https://doi.org/10.11141/ia.52.7.

Gualandi, M.L., G. Gattiglia and F. Anichini 2021. An Open System for Collection and Automatic Recognition of Pottery through Neural Network Algorithms. *Heritage* 4 (1): 140–59. https://doi.org/10.3390/heritage4010008 (accessed 13/10/2021).

Traille J.-P., J. de Meeûs d'Argenteuil and A. de Francquen 2014. *Study on the legal framework of text and data mining (TDM)*. http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf (accessed 13/10/2021) https://doi.org 10.2780/1475.

Wright, H., K. Green, F. Schäfer, H.S. Hollander and H.D. Tjalsma 2016. *Archaeology and Intellectual Property Rights: CARARE – Project Deliverable*. https://pure.knaw.nl/portal/en/publications/archaeology-and-intellectual-property-rights-carare-project-deliv (accessed 13/10/2021).

Supplementary Materials

Open Data is available online at https://archaeologydataservice.ac.uk/archives/view/archaide_2019/index.cfm (accessed 13/10/2021).

SITAR: a new open-data infrastructure for a public archaeology of Rome

Mirella Serlorenzi, Ascanio D'Andrea, Riccardo Montalbano

Mirella Serlorenzi , Ministero della Cultura, SSABAP – SITAR Scientific Director;
mirella.serlorenzi@beniculturali.it
Ascanio D'Andrea, Data Manager & GIS expert; ascaniodandrea@gmail.com
Riccardo Montalbano, GIS expert; ricca.montalbano@gmail.com

Abstract

SITAR is a project by the Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma that has changed the relationship between the institution and citizens as regards the knowledge, protection, enhancement and communication of the city's archaeological heritage. The need to open up to external users was taken into account when structuring the technological platform, so that the work carried out by the Soprintendenza could benefit the various communities that make up our society. SITAR – awarded with the Archaeological Heritage Prize 2021 by the European Archaeology Association (EAA) – has become a platform that provides open-data under a CC-BY-SA 4.0 license, characterized by a participatory approach thanks to which different types of users can interact with the institution, directly intervening in the process of creating and reviewing the data.

Keywords: ARCHEOSITAR PROJECT; ARCHAEOLOGICAL OPEN DATA CC-BY-SA 4.0; *LOD.*

SITAR for a public archaeology of Rome

The health emergency has highlighted the lack of data digitization projects and major infrastructure for their management and communication within public institutions. This delay can be ascribed to a lack of investment in strategic sectors, such as research, technological development and Cultural Heritage. The gravity of the situation can be seen in the failure to consider the extent to which the absence of investments leads to stagnation in Italy's cultural and economic growth, given the collective utility and social function of Cultural Heritage.

One of the problems to be resolved is the absence of infrastructure capable of guaranteeing long-lasting data preservation strategies (*infra*) on the one hand, and on the other the potential to create specific interoperability protocols, creating new shared, open and accessible knowledge without any limitations. We therefore need to rethink the concept of Cultural Heritage, so that archaeology is experienced as an integral part of a local area, whose significance may be of interest, above all, to the communities that have inherited it. New technologies are the tool with which public sector institutions can disseminate data, information and contents extensively, beginning that effort to democratize our heritage that represents both its constitutional fulfilment and the most effective way of safeguarding it.

In this context, already in 2007 the *Soprintendenza Archeologica di Roma* decided to create an archaeological information system: SITAR. The system has evolved over time to become a complex digital platform, structured to host the data on Rome's extraordinary archaeological heritage. SITAR took its first steps in a fruitful phase of discussion, starting with its participation in two ministerial joint committees that had opened a debate between different institutions, connecting the academic world to that of safeguarding and creating the conditions for adopting shared guidelines that serve the needs and objectives of research, protection and enhancement of heritage (Sassatelli 2011).

The decade that separates us from that debut and the long road travelled demonstrate that SITAR is now among the longest-lived and locally well-rooted platforms; unfortunately, very few digital projects developed over the last decade still exist and operate today. The overview presented in the latest edition of ArcheoFOSS highlighted the difficulties of the medium and long-term survival of projects to digitize and manage archaeological data at various levels (see in this volume Bogdani, Sciacca, table. 3). The burden of keeping a complex system such as SITAR alive over time is onerous and depends on factors that are not always predictable or subject to the will of those who operate and manage it. These include the medium and long-term planning of the objectives to be pursued and the communities to be reached:

- Continuous updating of the technological infrastructure, involving constant evolutionary maintenance with the related costs.
- Data maintenance that also includes storage, where possible prioritizing reliable public partners who can guarantee security and the appropriate dissemination of data.
- The availability of annual grants that may not remain constant or be lower than expected, with the consequent forced rescheduling of planned activities.
- A robust scientific management that oversees the general planning process, considering the variables that arise over time, but simultaneously able to pursue the objectives and the agreed development lines with determination, even in the absence of the necessary conditions.
- The support of a specialized working group, made up of different professionals, who share the project philosophy.
- An awareness that staff numbers are subject to variation, with operational repercussions.
- Contact with universities and the world of freelance professionals, to guarantee the dissemination of the contents and logical of the platform and training on its use.

An information system is thus a living organism that must adapt to the times and change in line with contingent circumstances. Given the difficulties that may arise in the long term, it is advisable to evaluate the existence of well-established projects already in place before embarking on new ones of a similar type. Interoperability protocols now make it possible to create a dialogue between projects of a very different nature. Therefore what is really needed is the launch of collaboration policies between subjects who share fundamental assumptions, such as the provision of open-data and the awareness that the circular reuse of information generates new knowledge.

The following pages focus on the latest developments of the SITAR platform, which have made it possible to meet some of the objectives underlying the project from its early design stages. The solutions available today have made it possible to enhance the collaborative aspects and improve the processes of acquisition and use of SITAR data. Furthermore, a new website has been created,[1] now also available in English, which provides information on the project at various levels and comprises helpful user support tools, such as tutorials and manuals.

The purpose has always been to provide open data and libre contents to the community, with the aim of opening up a dialogue on fundamental issues such as those related to the protection of Cultural Heritage, the quality of suburban areas, the participatory development of policies for the city's growth (Serlorenzi 2011; Serlorenzi *et al.* 2020). Special users that SITAR has targeted since the beginning include freelance archaeologists working daily in the field on behalf of the *Soprintendenza*. The activities carried out generate archaeological knowledge in the form of scientific documentation, which must be processed and standardized and whose essential data must be made public as quickly as possible to benefit not only the planning of new public works but research in general. In this regard, it was essential to consider the intended social significance of such a complex project managed by a public body (Serlorenzi *et al.* 2020; Figure 5), how to create a dialogue between the institution and users, and the original contributions that may arise from this exchange (Cerami *et al.* 2020).

Here, therefore, it is unnecessary to recall the general outlines of the project;[2] instead, we will illustrate the latest changes that have transformed SITAR into a technologically advanced platform, allowing for the dissemination of high-quality knowledge and improved governance of the local area and its Cultural Heritage, thanks to the involvement of civil society. This approach is also sanctioned by the Faro Convention, which sees the active participation of the community as a new resource, fundamental for the conservation and enhancement of heritage.

M.S.

Three pillar-based infrastructures: data, IT, users

The temporal element in the development and implementation of a public administration Information System is undoubtedly a foundational and structurally congenital value. On the one hand, it may favour a controlled evolution of functions and the reformulation of objectives, whilst on the other it may act as a technological brake given the constant adaptation that such systems require. SITAR is a clear example of longevity both in terms of design and data implementation, having enjoyed an operational continuity that has allowed for a constant renewal over the past decade. Starting from 2018, on the wave of a drive to reformulate the strategic objectives and thus expand the user base, we have embarked upon a significant re-engineering of the system. Four main objectives were taken into consideration and used to guide the process:

- *Scalability*, tailoring the resources on the basis of the actual needs, both in terms of the number of users and the delivery of services.
- *Performance*, increasing the processing capacity with new systems for indexing and request/response management, ensuring an interaction within milliseconds.

[1] https://www.archeositarproject.it/ (accessed 07/08/2021).
[2] For the history of the project, see Serlorenzi 2011; Serlorenzi et al. 2012; Serlorenzi 2013; Serlorenzi and Jovine 2013; Serlorenzi and Leoni 2015; Serlorenzi and Jovine 2017; Serlorenzi, Lamonaca and Picciola 2018; Serlorenzi et al. 2020.

Figure 1: Example of life-cycle of the three main Sitar platform components.

- *Maintenance*, improving scheduled maintenance operations and bug-fixes, guaranteeing the potential for immediate intervention on the individual components of the platform.
- *Interoperability*, with the creation of a data-independent technological system to ensure maximum communication with other existing platforms.

The creation of such a fully 'on-demand' system has greatly optimized the delivery of open-data services, from the outset at the heart of the entire platform. The potential to add/modify 'technologies' and the related exposure of public APIs independent of the data model ensures the system will have a long lifecycle that can easily adapt to the needs dictated by technological evolution. In this regard, it is important to stress that the IT component is just one of the three pillars on which the system is based. At the same level of importance, both quantitative and qualitative, are the 'data' component and that relating to the user experience and more generally to the reference communities (Figure 1).

This is a sound type of structure for SITAR, but more generally it can be considered extensible to all those information systems in the field of Cultural Heritage that aspire to become endemic operational tools within the state sector (and beyond). This is a system of design and operational forces that must necessarily 'work' in unison to ensure full efficiency:

- *The data component:* SITAR reaps the rewards of the innovative, shared and consolidated design of the data structure, at the level of both the conceptual and logical model of the database. The ongoing funding, though not always homogeneous in terms of resources, has made it possible to carry out constant and onerous data entry over time. This process has ensured, and continues to ensure, that the system remains constantly up to date on the state of scientific archaeological knowledge on Rome, a factor that underlies many of the decision-making/administrative and documentation standardization processes within the institution in its relationship with users, whether they are ordinary citizens or archaeological professionals.
- *The IT component:* As described below, the technological investment was tailored to the medium and long-term objectives and not to satisfy specific contingent needs. Thanks in part to prior experience, this is ensuring a lasting lifecycle for the platform. The creation of an on-demand system, with a clear separation of the data structure from the application technologies, now makes it possible to keep maintenance interventions to a minimum. This in turn, albeit with a higher initial investment in terms of both resources and development hours, means that the working group can focus on testing new features to expand the range of services. Once again, the IT component should be seen as a supporting 'pillar' of the structure that, once established, acts as a driving

force for the planned implementations, namely the development of a 3D GIS and the creation of tools for assessing the archaeological potential of the city of Rome.

- *The community component:* This term describes everything related to the user experience and more generally the needs of the project's reference 'communities'. SITAR responds to the need to bring to light Rome's vast documentary archaeological heritage and make it accessible to the public. Over the years there have been discussions, internal to the project and shared through the intensive dissemination of the results, about 'what' and 'who' the project's target audience should be; these discussions have entrenched the belief in the fundamental role played by attention to the system 'outputs'. Creating a tool that satisfies all audiences is impossible, but it is certainly feasible to create one that allows all audiences to use data regardless of individual needs, guaranteeing the possibility of choosing between differentiated services (public APIs, geo-services, direct download or a simple WebGIS interface). The next development will be to further implement the interaction with users, in this case with the specific class of 'insiders', allowing those working in the archaeological sector (companies, individual professionals, etc.) to upload data into the system. This will make it possible to further ensure the continuous updating of the information going beyond the initial objectives of an albeit complex digital archive of Rome.

In this light, whilst on the one hand, in terms of continuity of action, SITAR's longevity among national and international archaeological information systems may be surprising, on the other it allows us to identify some key factors that lay under projects in this sector, also in light of the substantial funds that will shortly be made available for the digitization of Cultural Heritage: a shared conceptual and logical modeling, distributed server technologies, data structures independent of specific products, mapping, analysis and interaction with user communities of reference. An integrated approach to technology and human resources in which to invest in the medium and long term to create endemic and sustainable systems within the Public Administration.

A.D.A.

SITAR cloud-based infrastructure, web application, open-data and open-data services

The SITAR re-engineering started in 2018 and is still in progress. It also led to a general restructuring of the IT infrastructure. Since 2013 the development and implementation of new web applications, combined with a wish to provide additional services and data, had prompted the transition from a self-hosted infrastructure, based on servers located at the *Soprintendenza* Data Centre to one based on virtual machines residing on a server infrastructure managed by the GARR Consortium. The latest evolution (2018–2020) entailed the creation of a new and wholly cloud-based infrastructure, hosted on the GARR network servers and managed through *OpenStack* open-source cloud technology. The GARR IAAS cloud helped in containing the server management and maintenance costs, and offered moreover a guarantee of the long-lasting preservation of the documentary heritage acquired and processed (Fresa and Justrell 2014).

It is important to stress that the new SITAR microservices architecture – based on the separation of the individual components, residing on dedicated machines as shown in the diagram in

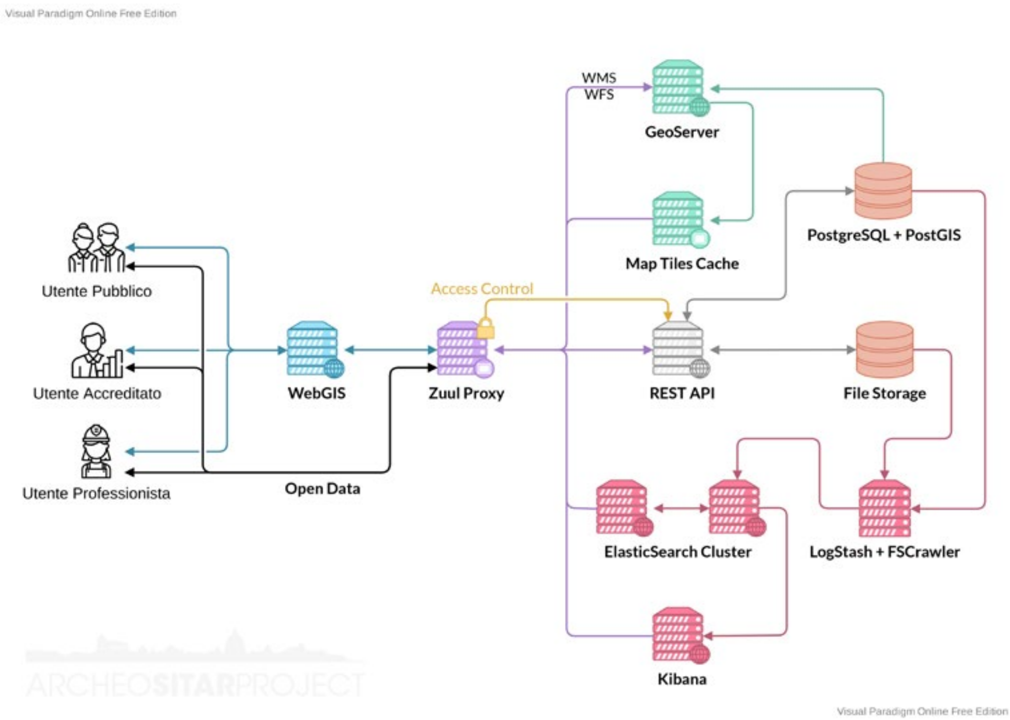


Figure 2: Server and services schema.

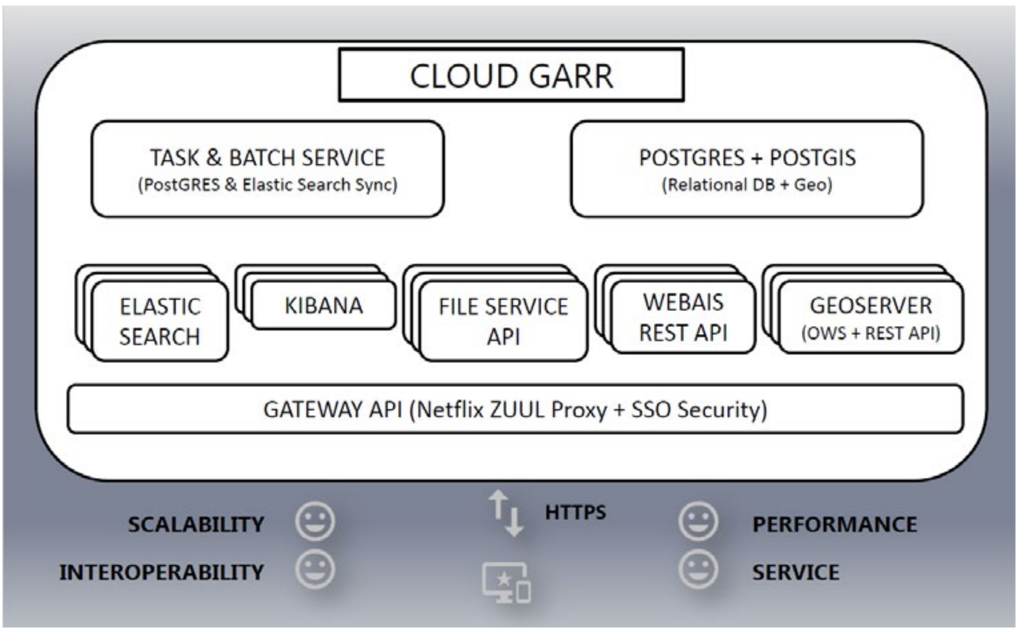


Figure 3: SITAR service scalability.

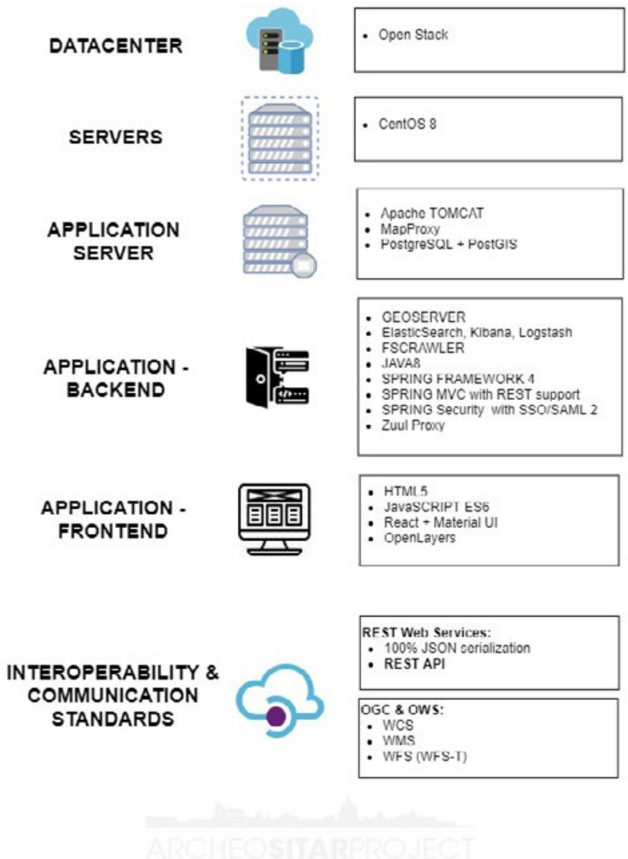


Figure 4: SITAR Tech Stack.

Figure 2 offers countless advantages. The services are distributed independently, resulting in easier maintenance, which can be limited to individual parts of the application. In terms of performance, the interactions between components are more streamlined, the horizontal and vertical scalability can be modulated on demand (Figure 3), and finally the automation of the development phases can be managed more easily, thanks to the implementation of CI/CD (Continuous Integration/Continuous Development) strategies, increasing the frequency and speed of testing and the release of new features. In general, adopting a paradigm based on a distributed cloud-based architecture hosted by an institutional Data Center and the use of open-source solutions (Figure 4), not only helps in being sustainable in economic terms, but also allows for the replicability of the system, which may in the future be cloned and transferred to support academic projects or public sector bodies that share its methodological and conceptual structure.[3]

Compared to the previous application, available online until 2019, the most important novelty of the new SITAR is the merger and integration of the three original *webAIS*, *SIGEDO* and *Ambiente*

[3] A code refactoring is currently under way, aimed at making the application and all its services available to the SOttoSiena project. This is the first step in a long roadmap, launched but not completed, that aims at creating an application that can be made available to other bodies and state institutions free of charge and on the basis of specific memoranda of understanding.

Tutela applications within a single environment. Today the WebGIS has a totally renewed interface, rendered more intuitive and enriched with tools that allow users to customize the map layers, search for entities on a geographical basis, filter individual excavation or find, explore their contents, carry out simple measurement operations (areas and distances), download and freely use spatial data and the related attributes in the most commonly used open formats. To allow users to browse the SITAR's rich corpus of documents, a Digital Library was created, using the open-source *ELK* suite (ElasticSearch, Logstash, Kibana). Thanks to the powerful search and indexing engine, users can perform full text and fuzzy searches on the records for the four SITAR entities and on the individual documents linked to them, consisting in texts in pdf format and images in various formats. This is possible thanks to an OCR system that identifies the texts contained within the static documents, which are subsequently indexed. Special filters allow users to refine the results obtained, for example by selecting the logical entity to be searched or the file type (pdf, image, ZIP, etc.).

A need seen as central from the early design stages, but that has now finally been implemented, is the collaborative aspect of the platform. Users now have a new set of tools thanks to which they can actively interact with the database, triggering a participatory procedure aimed at improving the dataset. Through the *request for change* users can send requests to update the data already in the system, report new work in the territory of Rome that has not yet been registered – potentially attaching photographs, videos, documents – or other types of evidence relating to the scope of SITAR and the *Soprintendenza*, about which a user deems it necessary to inform the offices in charge. These reports are subjected to an internal validation process, after which the data is acquired and published.

The new tools implemented also include the 'booklet' function, which allows registered users to create their own selection of SITAR data consisting of individual excavations and/or finds ('bookmarks'), making up a personal library that can also be shared. In terms of data sharing, on the one hand it is possible to publish the records for all entities with a public link on the main social media platforms, while on the other hand, registering with the RSS feed service allows users to receive updated information on the excavations for which updates have been requested, without the need to constantly check them on the WebGIS platform.

A more advanced level of dynamic interaction with the SITAR database is represented by the possibility to transfer the data-entry process to the professional archaeologists who work daily in the city's construction sites. Thanks to the implementation of dedicated back-end tools, to be released in the next months, the SITAR office will be entrusted with the task of identifying the study areas (OI), while the completion of the detailed information, the creation of the records on Archaeological Partition (PA) and the uploading of the accompanying documents will be outsourced to the external professionals, who will be entrusted with the operational privileges for their creation and updating exclusively for the excavations on which they are working. A double mandatory validation system is required for data to be published. A first formal semi-automated verification ensures that the delivery meets prerequisites in terms of georeferencing, attached documents, formats, etc. A second scientific validation of the data is entrusted to the staff members. Once fully operational, this procedure will guarantee real-time updates on developments in the territory of Rome, but above all it will allow the system to self-feed, thus freeing up useful resources to complete the acquisition and processing of all the vast quantity of archive data, still in progress.

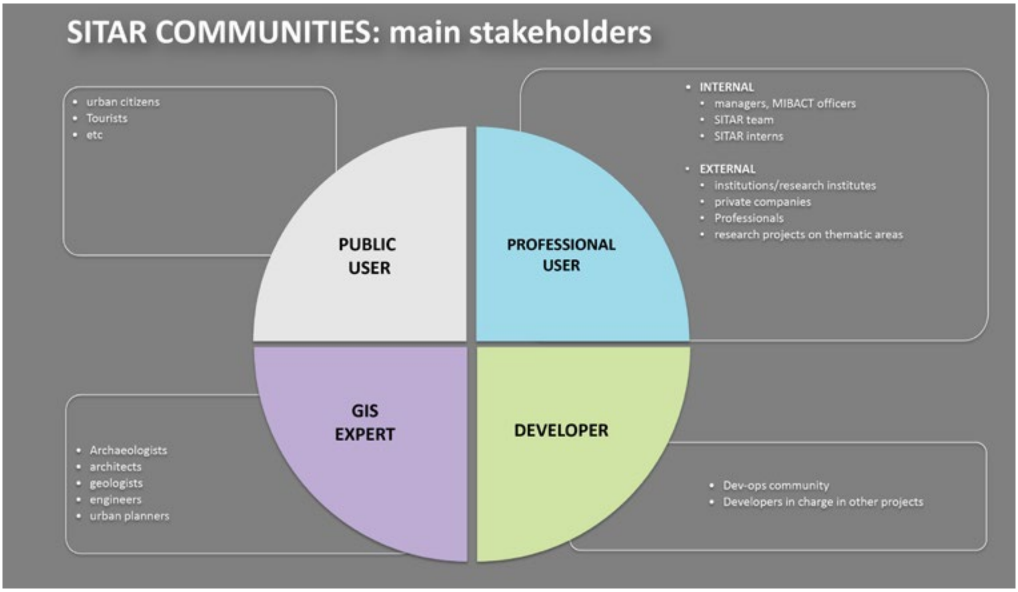


Figure 5: SITAR communities: the main stakeholders.

The SITAR data model has now been recognized on the national and international level, described in numerous publications, aligned with CIDOC CRM through the CRMarcheo extension and shared with the scientific community in linked data format (RDF), through the XLM language.[4] Here, therefore, greater emphasis will be placed on the renewed attention to open-data services, which are now the preferred tool for sharing data. As already mentioned, SITAR has for some time been pursuing a data publication and dissemination strategy in line with the FAIR principles (Serlorenzi 2018). Following the recent developments, SITAR can finally contact its communities of reference (Figure 5) through a help toolkit that facilitates data acquisition by users (Figure 6).

The new application includes a function to export map extracts relating to vector datasets, thanks to which users — using their monitor as a framing area — can freely download the spatial data and with the related attributes with a CC-BY-SA 4.0 license, choosing the format that best suits their needs. The main open GIS formats are available, vector types (Shapefile, GML2, GML3, KML, SVG), raster types (GeoTIFF) and textual types (CSV, GeoJSON). The SITAR platform also delivers geo-data using WMS and WFS protocols, and is thus aligned with the main operating standards and regulations defined by the OGC. At present, only attached documents (scientific reports, photographs, graphic documentation, in general the so-called raw data) are excluded from public use and restricted to *Soprintendenza* staff, who can download and use them if data is related to the area they administer. A geographically-based ownership system has been developed, entailing a dynamic association between excavations and staff through the areas of which they are in charge. The system is able to dynamically assign the new excavations to the officers in charge, allowing them to download the attached documents. By contrast, public users can view a list of available documents, which can be

[4] https://www.archeositarproject.it/wp-content/uploads/2020/10/Mapping-SITAR-towards-CIDOC-CRM-final_12.10.2016-1-1.pdf (accessed 01/07/2021).

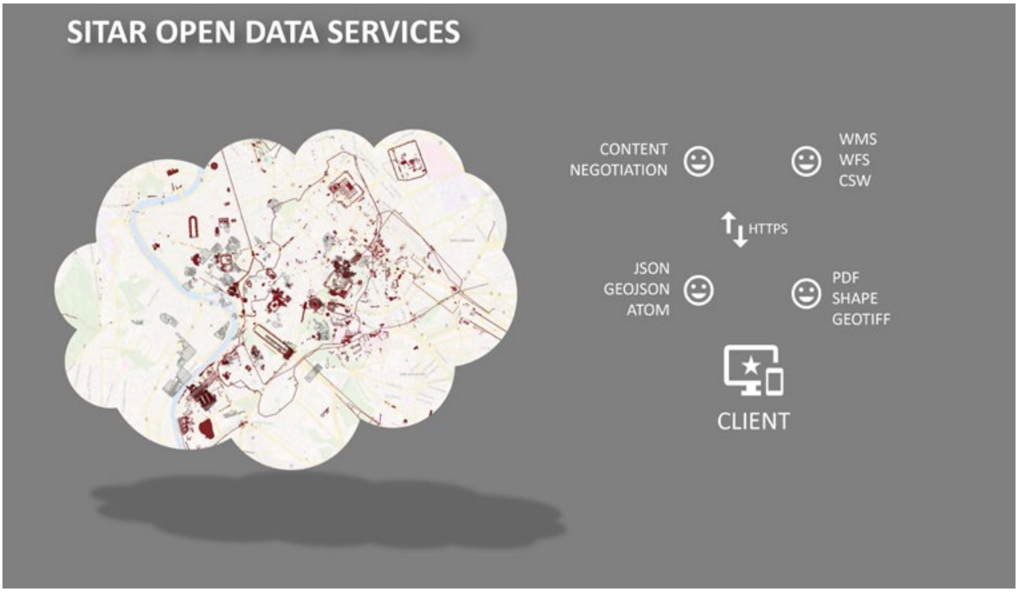


Figure 6: SITAR open data services.

requested to the officers in charge through the integrated request for download procedure. A special management panel allows users to make the request, monitor its status (accepted, in process, rejected) and — upon acceptance — finally download the document.

The need to establish institutional synergies with other public institutions and projects that produce and/or implement datasets in the Cultural Heritage sector required a significant effort in terms of interoperability. Thanks to the latest implementation, SITAR provides a series of specific public APIs to make REST calls both to the Elastic Search engine and to the REST API component.[5] In both cases, the data is returned in JSON format. This technological renewal allows for a further level of data and information services exchange, both inbound and outbound. SITAR will thus be able to receive data from external sources and to configure and provide services in support of other infrastructure and/or projects based on the same interoperability standards.[6] Another stage of the path, now ten years old, has therefore been reached, thanks to which SITAR has become a *trait d'union* between the Public Administration, professionals working on the ground and citizens, making its data available to the various users and communities of reference, who with differentiated skills produce or use data relating to Cultural Heritage.

R.M.

[5] https://www.archeositarproject.it/piattaforma/open-data/ (accessed 01/7/2021).
[6] The most recent experiment is a memorandum of understanding that provides for the exchange of data, via API, with the Forma Romae project curated by the Sovrintendenza Capitolina ai Beni Culturali.

Bibliography

Cerami, F.R., M.L. Scaduto and A. De Tommasi (eds) 2020. *I bacini culturali e la progettazione sociale orientata all'Heritage Making. Tra politiche giovanili, innovazione sociale, diversità culturale*. Firenze: All'Insegna del Giglio.

Fresa, A. and B. Justrell 2015. *A Roadmap for preservation for digital cultural heritage content*. Roma, ICCU - DCH-RP Project. Rome.

Sassatelli, G. 2011. La Seconda Commissione ministeriale per la formazione di un Sistema Informativo Territoriale Archeologico Nazionale. In M. Serlorenzi (ed.) *SITAR. Sistema Informativo Territoriale Archeologico di Roma. Atti del I Convegno (Roma, Palazzo Massimo, 2010)*: 99–101. Roma: Iuno Edizioni.

Serlorenzi, M. 2011 SITAR. Sistema Informativo Archeologico di Roma. In M. Serlorenzi (ed.) *SITAR. Sistema Informativo Territoriale Archeologico di Roma. Atti del I Convegno (Roma, Palazzo Massimo, 2010)*: 9–29. Roma: Iuno Edizioni.

Serlorenzi, M. (ed.) 2013. *ARCHEOFOSS Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11-13 giugno 2012)*. Archeologia e Calcolatori Supplemento 4. Roma: All'Insegna del Giglio.

Serlorenzi, M. 2018. Accessibilità e diffusione del dato archeologico: l'esperienza del SITAR. In M. Arizza *et al.* (eds) *I dati archeologici. Accessibilità, proprietà, disseminazione (Roma – CNR, 23 maggio 2017). Archeologia e Calcolatori* 29: 31-40.

Serlorenzi, M. and J. Jovine 2013. In M. Serlorenzi and I. Jovine (eds) *SITAR. Sistema Informativo Archeologico di Roma. Potenziale archeologico, pianificazione territoriale e rappresentazione pubblica dei dati. Atti del II convegno (Roma, 9 novembre 2011)*. Roma: Iuno Edizioni.

Serlorenzi, M. and I. Jovine 2017. *Introduzione*. In M. Serlorenzi and I. Jovine (eds) *Atti del IV convegno (Roma, 14 ottobre 2015)*. Archeologia e Calcolatori Supplemento 9, 9-18. Roma: All'Insegna del Giglio.

Serlorenzi, M. and G. Leoni. 2015. SITAR. Sistema Informativo Archeologico di Roma. Il SITAR nella Rete della Ricerca italiana. Verso la conoscenza archeologica condivisa. In M. Serlorenzi and G. Leoni (eds) *Atti del III convegno (Roma, 23-24 maggio 2013)*. Archeologia e Calcolatori Supplemento 7, 199-208. Roma: All'Insegna del Giglio.

Serlorenzi, M., F. Lamonaca, S. Picciola and C. Cordone 2012. Il Sistema Informativo Territoriale Archeologico di Roma: SITAR. *Archeologia e Calcolatori* 23: 31–50.

Serlorenzi, M., F. Lamonaca and S. Picciola 2018. The SITAR Project: Web Platform for Archaeological Knowledge Sharing. In V. Apaydin (ed.) *Shared Knowledge, Shared Power Engaging Local and Indigenous Heritage*: 125–127. Cham: Springer.

Serlorenzi, M., G. Leoni, F. Lamonaca and S. Picciola 2020. Il SITAR e le comunità degli utenti: un'infrastruttura culturale al servizio del patrimonio archeologico di Roma. In F.R. Cerami *et al.* (eds) *I bacini culturali e la progettazione sociale orientata all'Heritage Making. Tra politiche giovanili, innovazione sociale, diversità culturale*: 201–212. Palermo: All'Insegna del Giglio.

SPARQLing Publication of Irish ᚛ᚑᚌᚐᚋ᚜ – Ogham Stones as LOD

Florian Thiery, Sophie C. Schmidt, Timo Homburg

Florian Thiery, Römisch-Germanisches Zentralmuseum; mail@fthiery.de
Sophie C. Schmidt, Free University Berlin; scschmidt@zedat.fu-berlin.de
Correspondence: ft_scs@squirrel.link
Timo Homburg, Hoschschule Mainz; timo.homburg@hs-mainz.de

Abstract

Linked Open Data (LOD) is used to interlink data within the WWW semantically, enables citizen science and improves data sharing practices for academia. We discuss the semantic modelling and publication strategies in Wikidata and within our own ontology. Using a bespoke ontology gives more freedom for modelling, but requires high technical specialisation. Wikidata facilitates citizen science, but the data author loses data sovereignty. We will exemplify this using Ogham data, which is the main subject of the Ogi Ogham Project as well as the Irish ᚛ᚑᚌᚐᚋ᚜ Stones in the Wikimedia Universe project.

Keywords: OGHAM; LINKED OPEN DATA; WIKIDATA; OPEN SCIENCE; OPEN SCIENCE FELLOWS PROGRAM; ONTOLOGY; RDF; DATA MODELLING.

Introduction

In this paper we discuss two ways to model and provide archaeological data in the Semantic Web. The Ogi Ogham Project (Thiery, Schmidt and Homburg 2020a) currently being undertaken by a working group of the network Research Squirrel Engineers on the digitisation and provision of Irish Ogham Stones in Wikidata as Linked Open Data (LOD) is used as an example. Ogham Stones are an archaeological and linguistic source on Proto and Early Irish societies for which several separate databases exist. Shown here are two ways to model information on Ogham Stones as LOD: within the Open Data hub Wikidata and with a bespoke self-developed data model. Both workflows are described, compared and evaluated. They are suitable for different aims and approaches: Wikidata offers possibilities for Citizen Science and the opening of research processes, whereas the hosting of an own data model offers data sovereignty and bespoke ontologies

Ogham

Ogham Stones (Figure 1) are monoliths inscribed with the Ogham (᚛ᚑᚌᚐᚋ᚜) script, erected in Ireland and the western part of Britain (Wales, Scotland, Cornwall, Devon, Isle of Man) between the 4th – 6th centuries AD. The letters consist of strokes that are written at differing length and angles along (often natural) ridges of the stone and run up from lower left and back down on the right hand side (McManus 1997: 47). The standard work on Ogham inscriptions is the *Corpus Inscriptionum Insularum Celticarum* (Macalister 1945; 1949). In it, Macalister establishes the widely used numbering scheme CIIC and describes two different

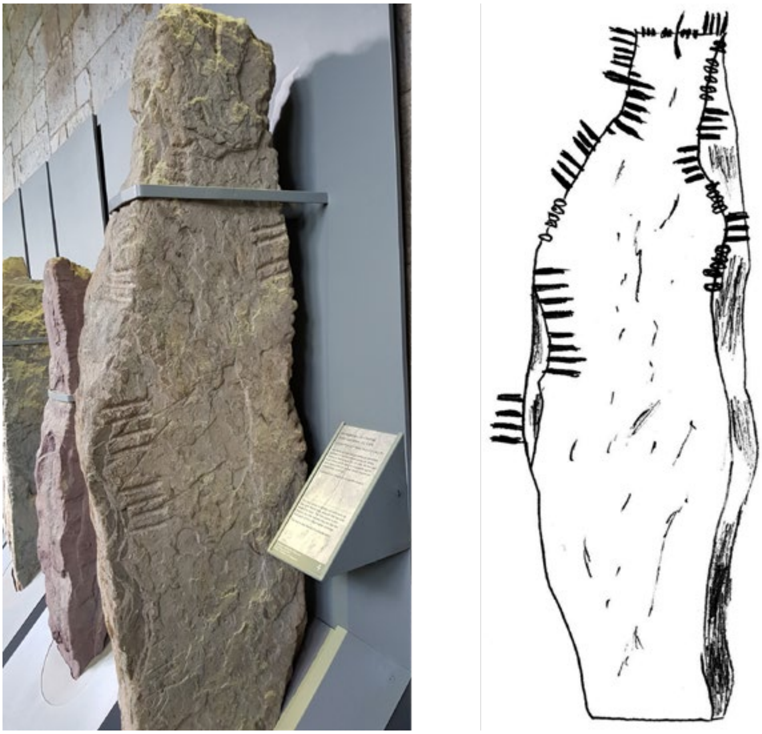


Figure 1: Ogham Stone CIIC 81, left: stone in the Stone Corridor at University College Cork (UCC); right: drawing of CIIC 81 (Florian Thiery CC BY 4.0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:CIIC81_UCC_Drawing.png).

types of words: formula words, mostly denoting personal relationships, such as son of or son of the tribe and nomenclature words that form names. The names are given in the genitive case so that a stone seems to be dedicated to a person. It is unclear though whether the stones were grave markers or denoted land ownership. Nonetheless they offer insights into Proto- and Early Irish communities. Next to the CIIC, several sources for the study of Ogham Stones exist. Next to a number of print publications, databases from different projects exist. For the Ogi Ogham Project, the CIIC, the Celtic Inscribed Stones Project[1] (CISP) and the 3D Ogham Project[2] are especially important. The CISP Project published their database as an online resource in 2001. Though it is not open (and it does not possess an API), Kris Lockyear was kind enough to make their MS Access Database available to us as well as giving us permission to use their data. The 3D Ogham Project is open and offers an increasing number of 3D scans of Ogham Stones as 3D PDFs, OBJ files and EPIDOC files with annotated stone inscriptions for download under a CC BY-NC-SA 3.0 Ireland License. These databases are thus available but not linked with each other, and combining the information is therefore cumbersome. The Ogi Ogham Project aims to create Linked Open Data from the differing databases. It was created as a side project by members of the Research Squirrel Network in 2019 and is now funded by the

[1] https://www.ucl.ac.uk/archaeology/cisp/database (accessed 01/07/2021).
[2] https://ogham.celt.dias.ie (accessed 01/07/2021).

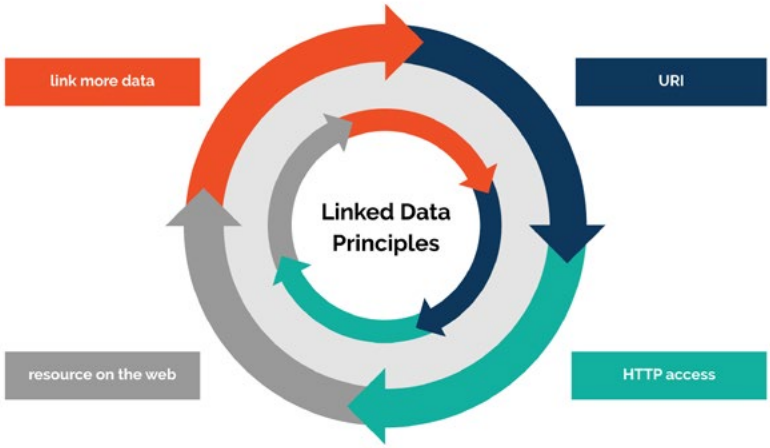


Figure 2: Linked Data Principles (Florian Thiery CC BY 4.0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Linked_Data_Principles.png).

Wikimedia Deutschland Open Science Fellows Program within the project Irish ᚑᚌᚆᚐᚋ Stones in the Wikimedia Universe[3] (Thiery 2021a).

The world wide web (WWW) gives researchers and amateur archaeologists (Citizen Science) the possibility of sharing their research data and thus enables both communities to participate in scientific discourse. Unfortunately, data is often not findable, accessible, interoperable and reusable – FAIR (Wilkinson *et al*. 2016) – which results in modern unknown data dragons (Thiery *et al*. 2019). These data dragons do not have links to related data, are not interoperable and lack reusability and usability. To overcome these shortcomings, we can use the Semantic Web, Linked Open Data (LOD) and Linked Open Usable Data (LOUD). LOD is used to interlink data within the WWW in a semantic way. Sir Tim Berners-Lee introduced the Linked Data life cycle in four rules (Figure 2):

> 'The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data. Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow.' (Berners-Lee 2006)

However, publishing data with unique identifiers on the WWW (as HTTP URIs) and using the LOD principles is of little use unless the data is programmatically usable, meaning, for example, accessible via an API and employing a well-documented ontology that describes the semantic data model for the user community. Therefore, Sanderson (2019) created the *Linked Open Usable Data* (LOUD) principles (Figure 3). All of these principles can be combined as

[3] http://ogham.squirrel.link (accessed 01/07/2021).

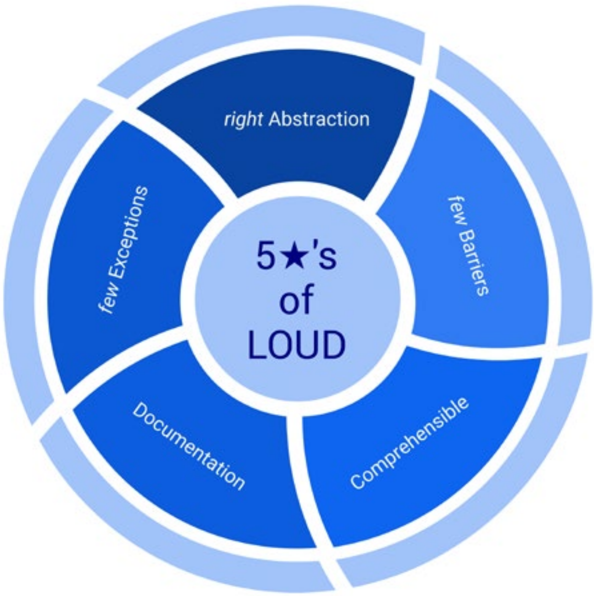


Figure 3: 5 Star Linked Open Data (Florian Thiery CC BY 4.0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:5_Star_LOUD.png).

shown by Thiery (2019) in the *Sphere 7 Data Model*. For archaeology, Leif Isaksen described the first applications of the Semantic Web and Linked Data in his PhD (Isaksen 2011). Since then, several LOD initiatives in Ancient Studies like Pelagios Commons and Linked Pasts have been established. To uncover the archaeological unknown data dragons, the *Computer Applications and Quantitative Methods in Archaeology* (CAA) *Special Interest Group on Semantics and LOUD in Archaeology* (SIG-DataDragon) was founded in 2019.

Linked Open Ogham Data

LOD is a recommendation for the storage of semantically modelled archaeological and cultural heritage data. There are several possibilities for hosting this modelled and published LOD: (A) create your own ontology, formulate a triple creation process (semantic uplift) and host the triplestore with a SPARQL endpoint, or (B) use established community-based systems, e.g. Wikidata. Strategy A (section 4.2) is based on an ontology with publication using a triplestore. Strategy B (section 4.1) uses a community-based approach: Wikidata. Wikidata allows for attribution, linking to other entities and the specification of provenance and sources. In the following sections we will introduce the two different strategies and discuss the pros and cons for different ways of representing Ogham stones as LOD (section 4.3). Overall, the ogham.link[4] platform will serve as an LOD Hub for Ogham Research.

[4] http://ogham.link (accessed 01/07/2021).

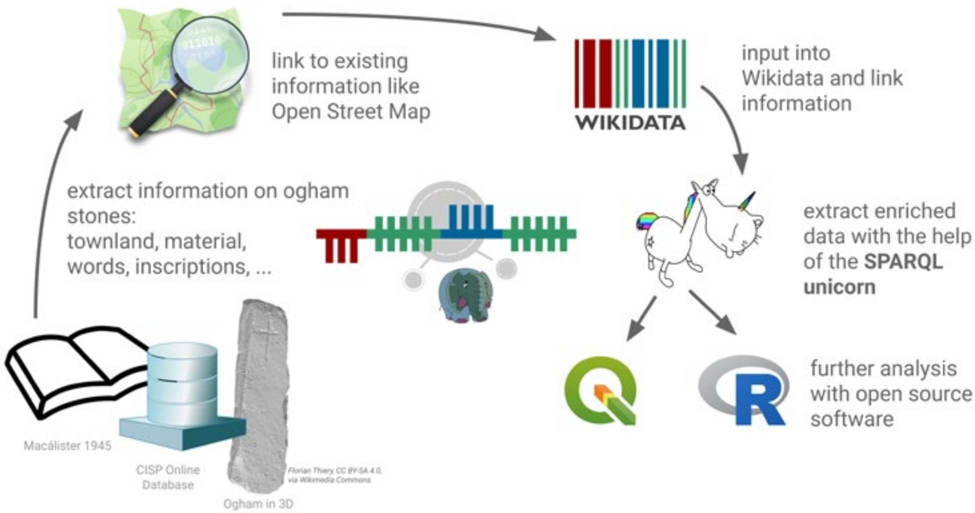


Figure 4: The Wikidata Ogham Workflow (Florian Thiery CC BY 4.0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Ogham_Wikidata_Workflow.png).

Ogham in Wikidata

Wikidata is the community-based knowledge hub in the Wikimedia Universe. The main workflow for the integration of Ogham in Wikidata[5] can be described as shown in Figure 4. In a first step, we extract information on Ogham Stones, e. g. townland, material, dimensions, and inscriptions, from books (Macalister 1945), the CISP online database and the Ogham in 3D project, and store them in CSV files on GitHub.[6] This is followed by the linking to existing information (e.g. Open Street Map; many locations are already present in Wikidata), especially in case of the geographical location (e. g. county, barony, electoral district and townland), resulting in a semi-automatic Wikidata input process using Open Source Software, e.g. Open Refine[7] and Quick Statements.[8] The raw data (Thiery 2021b), as well as the Wikidata mapping schemes (Thiery 2020), are published on Github under CC BY 4.0 license. The current semantic modelling structure in Wikidata for Ogham stones is as follows:

- P31 (instance of) Ogham stone (Q2016147) [https://www.wikidata.org/wiki/Q2016147]
- P361 (part of) Ogham Stones (Q67978809) [https://www.wikidata.org/wiki/Q67978809]
- P361 (part of) Ogi-Ogam Project (Q70873595) [https://www.wikidata.org/wiki/Q70873595]
- P17 (country) Ireland (Q27) [https://www.wikidata.org/wiki/Q27]
- P625 (coordinate location) derived from the Ogham site [https://www.wikidata.org/wiki/Property:P625]

[5] https://w.wiki/HRe (accessed 01/07/2021) under CC0 license.
[6] https://github.com/FellowsFreiesWissen/Ogham (accessed 01/07/2021).
[7] Open Refine is licensed under the BSD 3-Clause 'New' or 'Revised' License, see https://github.com/OpenRefine/OpenRefine/blob/master/LICENSE.txt (accessed 01/07/2021).
[8] QuickStatements is released under the GPL-3.0 License, see https://github.com/magnusmanske/quickstatements (accessed 01/07/2021).

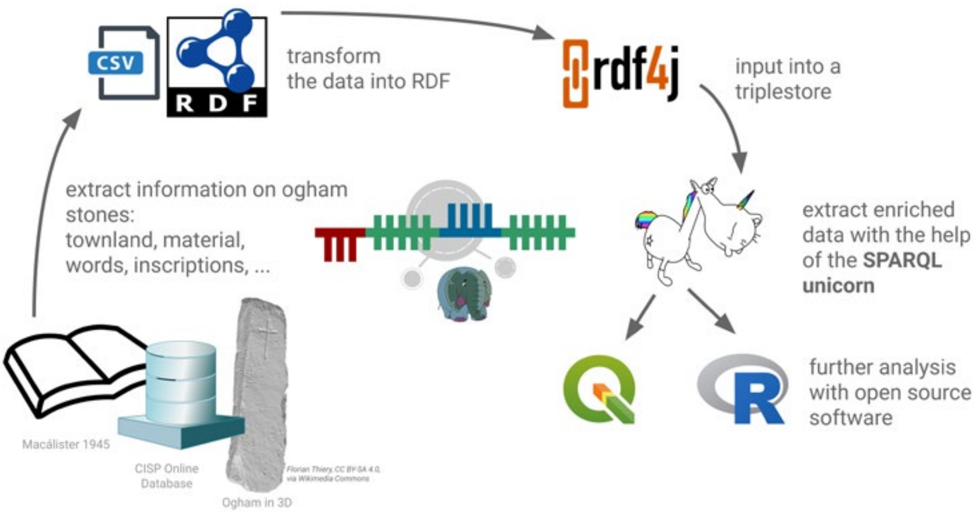


Figure 5: The RDF Ogham Workflow (Florian Thiery CC BY 4.0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Ogham_RDF_Workflow.png).

- P189 (location of discovery) county and Ogham site [https://www.wikidata.org/wiki/Property:P189]
- P1684 (inscription) [https://www.wikidata.org/wiki/Property:P1684]
- P1545 (series ordinal) [https://www.wikidata.org/wiki/Property:P1545]
- P186 (material used) [https://www.wikidata.org/wiki/Property:P186]
- P195 (collection) [http://www.wikidata.org/prop/direct/P195]
- P2043 [https://www.wikidata.org/wiki/Property:P2043]/P2049 [https://www.wikidata.org/wiki/Property:P2049] / P2048 [https://www.wikidata.org/wiki/Property:P2048] (length, width, height)
- P6568 (inscription mentions) [https://www.wikidata.org/wiki/Property:P6568]
- P5816 (state of conservation) [https://www.wikidata.org/wiki/Property:P5816]
- P18 (image) [https://www.wikidata.org/wiki/Property:P18]
- P2888 (exact match) [http://www.wikidata.org/prop/direct/P2888]
- P1382 (partially coincident with) [https://www.wikidata.org/wiki/Property:P1382]

The reference information, i.e. where the information is derived from, will be modelled as reference. Wikidata contains a number of Ogham related entities,[9] including the words written upon the stones. MacManus (1997) offers a large number of different words which were found on Ogham stones. Ogham inscriptions can contain formula words like MAQI (ᚋᚐᚊᚔ → son) or MUCOI (ᚋᚒᚉᚑᚔ → tribe/sept) and nomenclature words, describing names, that reveal details of early Gaelic society, e.g. CUNA (ᚉᚒᚅᚐ → wolf/hound) or CATTU (ᚉᚐᚈᚈᚒ → battle). A list of formula and nomenclature words according to MacManus (1997), together with translations, references and variants, is published on GitHub[10] under CC BY 4.0 license.

[9] https://github.com/ogi-ogham/ogham-wikidata (accessed 01/07/2021).
[10] https://github.com/ogi-ogham/oghamextractor/blob/master/words/words.csv (accessed 01/07/2021).

Self-hosted Ogham LOD

In comparison to a modelling approach in a community hub like Wikidata, it is also possible to create a bespoke ontology. This results in a slightly different workflow to that shown in section 4.1 (Figure 5). The first step is the same, we extract information on Ogham Stones, from books and online databases and store them in CSV files on GitHub. Parallel to this, we create our own ontology using Protegé[11] for the representation of Ogham data according to established modelling approaches and ontologies. We then transform the CSV files according to this ontology by using Python scripts into Resource Description Framework files (RDF) (Thiery 2021b). RDF data can be imported into a triplestore, for which we used RDF4J[12] in combination with the LOD publication frontend GeoPubby[13] (Homburg 2021). The Ogham ontology (Thiery 2021c) in version 0.4 (CC BY 4.0 license) consists of a set of classes which are based on well-known ontologies and vocabularies in the LOD world: CIDOC Conceptual Reference Model (CIDOC-CRM), Simple Knowledge Organization System (SKOS), Friend of a Friend (FOAF), Pleiades, GeoSPARQL; for provenance information the PROV-O ontology is used. Each entity is a E1 CRM Entity[14] as a subclass of a prov:Entity. CIDOC CRM entities are split into two major categories: skos:Concept[15] and E24 Physical Man-Made Thing.[16] E24 Man Made Things are E25 Man-Made Features[17] like Inscriptions or E22 Man-Made-Objects[18] like Ogham Stones. SKOS Concepts are divided into four CIDOC CRM classes: E39 Actor, E53 Place,[19] E55 Type[20] and E5 Events.[21] Each Actor Entity (e.g. a foaf:Group[22] such as Ogham Tribes; a foaf:Person[23] such as an Ogham Person) is modelled as a skos:Concept[24] in order to make a connection to thesaurus modelling approaches. Each Place Entity (e.g. a State such as Ireland) is a GeoSPARQL Spatial Object, as well as a Pleiades Place Concept, in order to link to other (ancient) gazetteers and to provide the possibility of creating standardised GeoSPARQL geometry features. Each Type Entity (typically collected in SKOS based thesauri, e.g. Sandstone) is a E55 Type[25] and skos:Concept to simplify the linkage to other thesauri.

Pros and cons of the different ways of representing Ogham stones as LOD

As seen in sections 4.1 and 4.2, when creating and including new Linked Data sets, two main strategies exist for how to store and manage new LOD in the Linked Open Data Cloud: (1) create and model the corresponding Linked Data within Wikidata or (2) create your own data model – ontology – and host the data in your own triplestore. Advantages of strategy 2 are that

[11] Protegé is released under the BSD 2-Clause ‘Simplified’ License, see https://github.com/protegeproject/protege-distribution/blob/master/LICENSE (accessed 01/07/2021).
[12] RDF4J is released under the BSD 3-Clause ‘New’ or ‘Revised’ License, see https://github.com/eclipse/rdf4j/blob/main/LICENSE (accessed 01/07/2021).
[13] GeoPubby is licensed under the Apache License 2.0, see https://github.com/i3mainz/geopubby/blob/master/LICENSE.txt (accessed 01/07/2021).
[14] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E1 (accessed 01/07/2021).
[15] https://www.w3.org/2009/08/skos-reference/skos.html#Concept (accessed 01/07/2021).
[16] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E24 (accessed 01/07/2021).
[17] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E25 (accessed 01/07/2021).
[18] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E22 (accessed 01/07/2021).
[19] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E39 (accessed 01/07/2021).
[20] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E55 (accessed 01/07/2021).
[21] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E5 (accessed 01/07/2021).
[22] http://xmlns.com/foaf/spec/#term_Group (accessed 01/07/2021).
[23] http://xmlns.com/foaf/spec/#term_Person (accessed 01/07/2021).
[24] https://www.w3.org/2009/08/skos-reference/skos.html#Concept (accessed 01/07/2021).
[25] http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html#E55 (accessed 01/07/2021).

more freedom is given to the data modelling specialist to semantically model the classes and their relations, as well the fact that it uses various well-known quasi standard ontologies and vocabularies in the field of ancient studies, e.g. CIDOC CRM, GeoSPARQL, SKOS, FOAF, Pleiades, RDFS or Dublin Core. But as a result, the ontology model can be too specific to be included in other repositories (e.g. Wikidata). It needs to be curated by specialists in a specific field. The data sovereignty and full access to their own triplestore and publication server are an advantage, but also a resource investment. Data sovereignty also includes the free choice of a data license; data in Wikidata has to be licensed under CC0. Moreover, a self-hosted knowledge base offers the possibility of executing reasoning rules. These may be applied on a self-hosted knowledge base to infer further knowledge, such as a classification of Ogham stones by relationships. In contrast, Wikidata currently does not provide support for the execution of self-defined reasoning rules. However, in order to not miss out on the advantages of citizen science, a bi-directional backlink to the semantically modelled Ogham data in the self-hosted knowledge base can provide the possibility of extending the self-hosted data with community knowledge created in Wikidata. Adding new content to Wikidata (strategy 1) may require publishing data in a research data repository for public reference. On the one hand, Wikidata usage includes the pros of native embedding of data into the LOD cloud, which is necessary to reach 5 star LOD.[26] It also provides an already existing active community, access to citizen science, and the easy publication of LOD including the use of Wikidata APIs and tools. On the other hand, data and semantic modelling sovereignty no longer exist, because everybody can edit the data and the semantic data modelling structure is limited to the community will. The modelling of provenance information can be done using both strategies, by the PROV-O Provenance Ontology[27] or by adding a reference to the Wikidata statement.

Conclusion

This paper describes two semantic Linked Open Data modelling and publication strategies: Wikidata and self-hosted ontologies. We discuss the pros and cons as well as the challenges of the two possible data publication strategies using the example of Ogham Stones. One of the major challenges and possibilities lies in the balance between enabling citizen science vs. retaining data sovereignty. Data publications enable researchers to do further geospatial or statistical analysis, e.g. using the SPARQLing Unicorn QGIS Plugin[28] and integrate them into the research workflow (Thiery *et al.* 2020b). In the future the data in both repositories will be further expanded, within the Ogi Ogham Project as well as the project Irish ᚑᚌᚆᚐᚋ Stones in the Wikimedia Universe funded by Open Science Fellows Program.

Acknowledgments

We would like to thank Dr Kris Lockyear who made the CISP database available to us. We are grateful to David Wigg-Wolf for English language corrections. We would also like to thank the Research Squirrel Engineers network for their technical and mental support. This research is partly funded by Wikimedia Deutschland e. V. within the Open Science Fellows Programme, project: Irish ᚑᚌᚆᚐᚋ Stones in the Wikimedia Universe.

[26] https://5stardata.info/en/ (accessed 01/07/2021).
[27] https://www.w3.org/TR/prov-o/ (accessed 01/07/2021).
[28] The SPARQLing Unicorn QGIS Plugin is licensed under the GNU General Public License v2.0, see https://github.com/sparqlunicorn/sparqlunicornGoesGIS/blob/master/LICENSE (accessed 01/07/2021).

Bibliography

Berners-Lee, T. 2006. *Linked Data – Design Issues*. https://www.w3.org/DesignIssues/LinkedData.html (accessed 01/07/2021).

Homburg, T. 2021. *i3mainz / GeoPubby. i3mainz*. GitHub. https://github.com/i3mainz/geopubby (accessed 01/07/2021).

Isaksen, L. 2011. *Archaeology and the Semantic Web*. Phd, University of Southampton. https://eprints.soton.ac.uk/206421/ (accessed 01/07/2021).

Macalister, R.A.S. 1945. *Corpus inscriptionum insularum Celticarum*. Vol. I. Dublin: Stationery Office.

Macalister, R.A.S. 1949. *Corpus inscriptionum insularum Celticarum*. Vol II. Dublin: Stationery Office.

MacManus, D. 1997. *A Guide to Ogam*. Maynooth Monographs 4. Maynooth: An Sagart.

Sanderson, R. 2019. LOUD: Linked Open Usable Data. 28. Mai 2019. https://linked.art/loud/ (accessed 01/07/2021).

Thiery, F. 2019. Sphere 7 Data: LOUD and FAIR Data for the Research Community. *Zenodo*. https://doi.org/10.5281/zenodo.2643469.

Thiery, F. 2020. Ogi Ogham / ogham-wikidata. Ogi-Ogham Project. GitHub. https://github.com/ogi-ogham/ogham-wikidata (accessed 01/07/2021).

Thiery, F. 2021a. Irische ᚑᚌᚆᚐᚋ Steine im Wikimedia Universum. *Zenodo*. https://doi.org/10.5281/zenodo.4568092

Thiery, F. 2021b. Fellows Freies Wissen/Ogham. Fellows Freies Wissen. GitHub. https://github.com/FellowsFreiesWissen/Ogham.

Thiery, F. 2021c. Linked Ogham Ontology. *Zenodo*. https://doi.org/10.5281/zenodo.4407344.

Thiery, F., S.C. Schmidt and T. Homburg 2020a. ᚑᚌᚔ Ogham. *Osf*. https://doi.org/10.17605/OSF.IO/REZX2.

Thiery, F., T. Homburg, S.C. Schmidt, M. Trognitz and M. Przybilla 2020b. SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data. *Zenodo*. https://doi.org/10.5281/zenodo.3751769.

Thiery, F., M. Trognitz, E. Gruber and D. Wigg-Wolf. 2019. 'Hic Sunt Dracones!' The Modern Unknown Data Dragons. *Zenodo*. https://doi.org/10.5281/zenodo.3345711.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (März): 160018. https://doi.org/10.1038/sdata.2016.18.

Towards an ontology of the Museum of Archaeology of the University of Catania: from the digitization of the legacy data to the Semantic Web

Nicola Laneri, Rodolfo Brancato, Salvatore Cristofaro, Marianna Figuera, Marianna Nicolosi Asmundo, Daniele Francesco Santamaria, Daria Spampinato

Nicola Laneri, Department of Humanities, University of Catania, via Biblioteca, 4 - 95124 Catania, Italy; nlaneri@unict.it

Rodolfo Brancato, Department of Humanities, University of Catania, via Biblioteca, 4 - 95124 Catania, Italy; rodolfobrancato@unict.it

Marianna Figuera, Department of Humanities, University of Catania, via Biblioteca, 4 - 95124 Catania, Italy; marianna.figuera@unict.it

Salvatore Cristofaro, CNR Institute of Cognitive Sciences and Technologies, Catania, via Gaifami, 18 - 95126 Catania, Italy; salvatore.cristofaro@istc.cnr.it

Daria Spampinato, CNR Institute of Cognitive Sciences and Technologies, Catania, via Gaifami, 18 - 95126 Catania, Italy; daria.spampinato@cnr.it

Marianna Nicolosi Asmundo, Department of Mathematics and Computer Science, University of Catania, Viale A. Doria, 6 - 95125 Catania, Italy; nicolosi@dmi.unict.it,

Daniele Francesco Santamaria, Department of Mathematics and Computer Science, University of Catania, Viale A. Doria, 6 - 95125 Catania, Italy; santamaria@dmi.unict.it

Abstract
This paper presents the results of digitization of the Libertini Collection of the Museum of Archaeology of the University of Catania (MAUC). The digitization project aims at defining forms of long-term preservation of digital data associated with sites and objects of cultural interest exploiting the Linked Open Data (LOD) paradigm and, more specifically, the Web Ontology Language (OWL), that is the standard language for representing web ontologies and digital tools designed for describing resources of various domains of knowledge. Such an approach provides a high level of expressiveness along with automatic reasoning tools, which allows the interested user to obtain a more comprehensive and efficient form of digital research.

Keywords: SEMANTIC WEB; ONTOLOGIES; CULTURAL HERITAGE; ARCHAEOLOGICAL MUSEUM.

Introduction

Collecting and sharing data are the most common practices when dealing with Cultural Heritage that is either kept in closed or open area museums or made available from archaeological excavations. In interpreting such needs, scholars involved in studying, protecting, and disseminating ancient heritage should define coherent strategies for implementing processes of sharable digitization. These strategies should be inclusive and open in order to foresee the guidelines set out, for example, by the future Horizon Europe programs that support and facilitate 'open science practices'. In archaeology, such a practice is still behind the planned research agenda and thus it appears as an intellectual imperative to foresee a research pathway in this direction.

On these premises, we decided to apply an innovative digitizing practice (i.e. ontologies) for the collection of a small-sized University Museum as is the case of the MAUC. The choice of digitizing the museum's catalogue is based on the relevance of the collection of the MAUC, on the quality and consistency of the supporting documentation, on the significant presence of legacy data, and on the cultural significance that the Museum has for the collectivity and for the University of Catania community. In particular, the inability of the public to (physically) visit the museum due to the crisis caused by the Covid-19 has prompted the scholars involved in the collection management to review the traditional forms of exhibition focusing more on aspects related to digitization, and online data sharing and use. This was possible also thanks to the support provided by the students of Archaeology involved in this project aiming at producing a new image of the Museum based on a fluid museum no longer constrained by its physical dimension to enter a new digital perspective. The project aims not only at reconstructing the history of the collection, but also at transforming it into a digital dimension through the analysis, recovery, and republication of the available legacy data. The project will allow one to identify the context of scattered items and to relate the copy of the item with its original form, according to the standard CIDOC Conceptual Reference Model (CRM)[1] and by means of an appropriate mapping of the ontology with the RDF Pleiades vocabulary,[2] according to the LOD paradigm. The first step involved the identification of the entities and their organization into a system that adopts the ICCD standards,[3] with a view to a forthcoming integration with the SIGECWeb.[4] For the development of an effective and expressive ontological model, the preliminary phase is the normalization of the data and their organization into homogeneous and structured collections with adequately refined levels of granularity. Specific terminological vocabularies of reference were determined. They are useful for normalizing data by means of the identification of suitable and consolidated existing vocabularies and the definition of new terminological entities. Finally, the ontological model was defined, representing the semantics of the application domain, and appropriate SWRL (Semantic Web Rule Language) rules were introduced to efficiently infer implicit relevant domain information.

The archaeological collection

The digitization of the archaeological collection of the MAUC explores innovative models of data entry to encourage not only new research perspectives, but also the discovery and interpretation of cultural heritage by a wider audience, including undergraduate students who have been directly involved in the process. All the finds will be made available and reusable online in an open (CC BY-NC-SA 2.5 IT – Creative Commons) and linked data standard following the spirit of a public, shared, and participatory archaeology.

[1] http://www.cidoc-crm.org/ (accessed 20/7/2021).
[2] https://pleiades.stoa.org/ (accessed 20/7/2021).
[3] ICCD (Istituto Centrale per il Catalogo e la Documentazione), http://iccd.beniculturali.it/ (accessed 06/08/2021), is the Italian government Institute that deals with the cataloging, research and training of Italian Cultural Heritage.
[4] SIGECWeb (Sistema Informativo GEnerale del Catalogo Web), at www.sigecweb.beniculturali.it (accessed 06/08/2021), is a web-based platform that manages the entire flow of the cataloging of Italian Cultural Heritage.

Figure 1: MAUC: finds from the collection on permanent display.

The core of the Museum are the 325 finds belonging to the collection donated by Guido Libertini to the University of Catania in 1953 (Figure 1). Publishing online the collection is not only related with the quality of the finds, but also with the consistency of the supporting documentation, such as the catalogue, transmission letters, drawings, photos and archaeometric data analysis report (Tortorici *et al.* 2015). Archaeological collections may carry an impressive amount of data: indeed, the Libertini collection consists of a wide range of archaeological material (pottery, coins, glasses, metals, inscriptions, etc.) from different periods. In fact, the main part of the collection consists of finds from the Greek and Roman periods (i.e. fine and common wares, epigraphs, terracotta figurines, coins, etc.), but also some objects from prehistory as well as from Late Antique and Medieval periods. Among these, there are some noteworthy forgeries (i.e. 78 objects forged in the first half of the 20th century) which, due to the outstanding level of craftsmanship and the use of authentic moulds, deceived many archaeologists and art historians.

The digitization project

Digital representation of archaeological data in open formats may play a crucial role not only in Cultural Heritage preservation, since it offers significant advantages in the dissemination, use, and comprehension of information stored in old reports, archives, collections, and museums catalogues, but also in research (Bogdani 2019: 57–88). Since the aim of the project is to define forms of long-term preservation of digital data associated with sites and objects of cultural interest, we shall use the LOD paradigm and, more specifically, the OWL 2 language, a digital tool designed for the definition, description, integration and sharing of ontologies for various domains of knowledge. Ontologies, in this sense, can put together data that are expressions of the history of the territory, but that have been kept separated due to administrative and bureaucratic issues: indeed, only connected data can be fully exploited, but they must first be suitably collected and digitalized (Brancato *et al.* 2019).

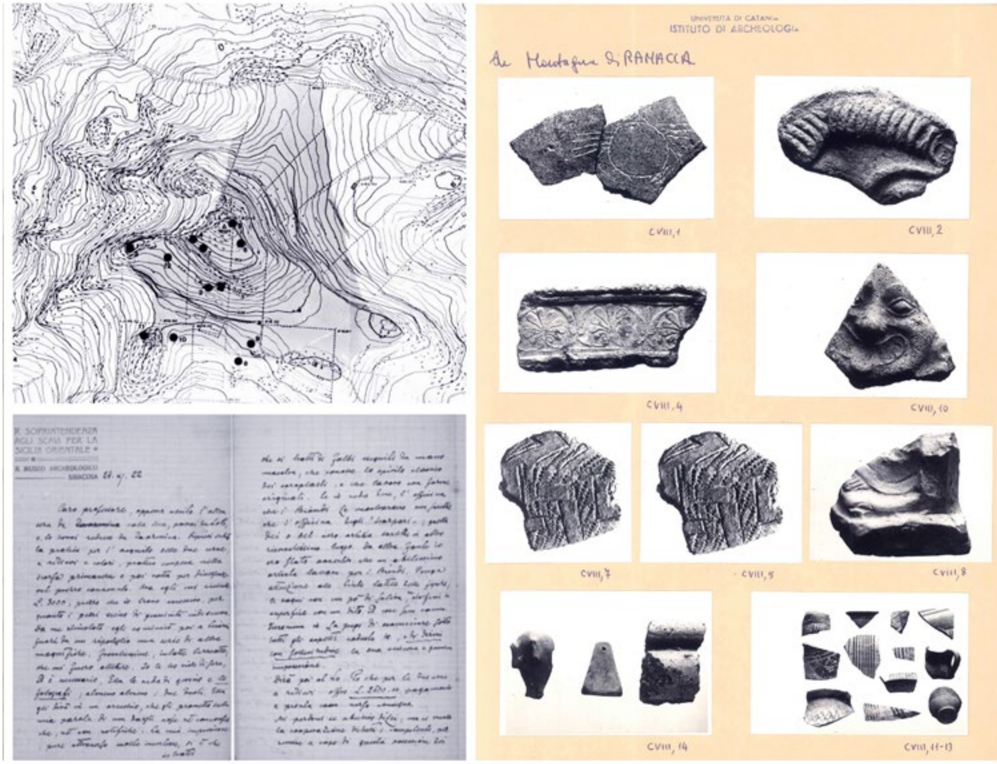
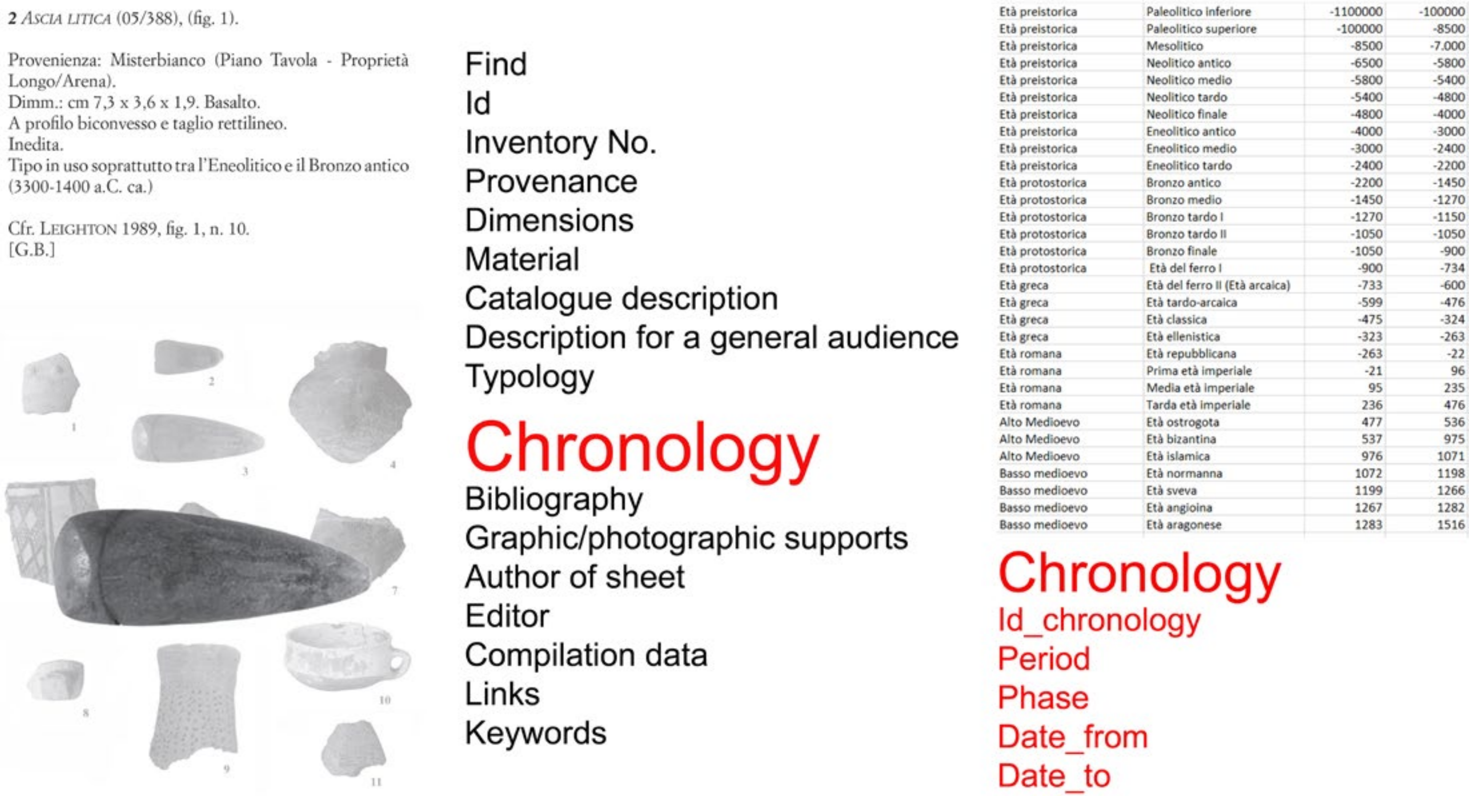


Figure 2. Legacy data available for the digitization process of the MAUC's collection.

The ontological approach adds a high level of expressiveness thanks to automatic reasoning tools, which allow the interested individual to obtain a more complete and comprehensive form of digital research. The project of digitization of the collection started from the mapping of the available legacy data (Figure 2). The use of legacy data is a topic widely considered in archaeology in the last decades (Allison 2008). This process, recently also applied to archaeological landscapes, already involved the archaeological collections of numerous international museums (Terras 2015). Useful archaeographic data (Anichini, Gattiglia 2012) of the MAUC is dispersed in previously published study reports and research papers, in the University historical archive (such as institutional correspondence with museum directors, the first inventory, old photos, etc.), and a recently published scientific catalogue. Therefore, the decision to publish online the content of the Libertini collection is justified by the consistency of the supporting documentation.

The digitization project of the archaeological collection started from the awareness of the plurality of both legacy and digital data archived with obsolete systems. Their analysis pointed out many critical issues, such as the management of limited information, redundancy, and the missing functional organization of data.

To apply data management consistent to the general purposes of the project, which goes beyond this first digitization step, a better choice is to follow the requirement analysis rules.



Età preistorica	Paleolitico inferiore	-1100000	-100000
Età preistorica	Paleolitico superiore	-100000	-8500
Età preistorica	Mesolitico	-8500	-7.000
Età preistorica	Neolitico antico	-6500	-5800
Età preistorica	Neolitico medio	-5800	-5400
Età preistorica	Neolitico tardo	-5400	-4800
Età preistorica	Neolitico finale	-4800	-4000
Età preistorica	Eneolitico antico	-4000	-3000
Età preistorica	Eneolitico medio	-3000	-2400
Età preistorica	Eneolitico tardo	-2400	-2200
Età protostorica	Bronzo antico	-2200	-1450
Età protostorica	Bronzo medio	-1450	-1270
Età protostorica	Bronzo tardo I	-1270	-1150
Età protostorica	Bronzo tardo II	-1050	-1050
Età protostorica	Bronzo finale	-1050	-900
Età protostorica	Età del ferro I	-900	-734
Età greca	Età del ferro II (Età arcaica)	-733	-600
Età greca	Età tardo-arcaica	-599	-476
Età greca	Età classica	-475	-324
Età greca	Età ellenistica	-323	-263
Età romana	Età repubblicana	-263	-22
Età romana	Prima età imperiale	-21	96
Età romana	Media età imperiale	95	235
Età romana	Tarda età imperiale	236	476
Alto Medioevo	Età ostrogota	477	536
Alto Medioevo	Età bizantina	537	975
Alto Medioevo	Età islamica	976	1071
Basso medioevo	Età normanna	1072	1198
Basso medioevo	Età sveva	1199	1266
Basso medioevo	Età angioina	1267	1282
Basso medioevo	Età aragonese	1283	1516

Figure 3. The chronology digitization process: from the legacy data to the standardization of vocabularies.

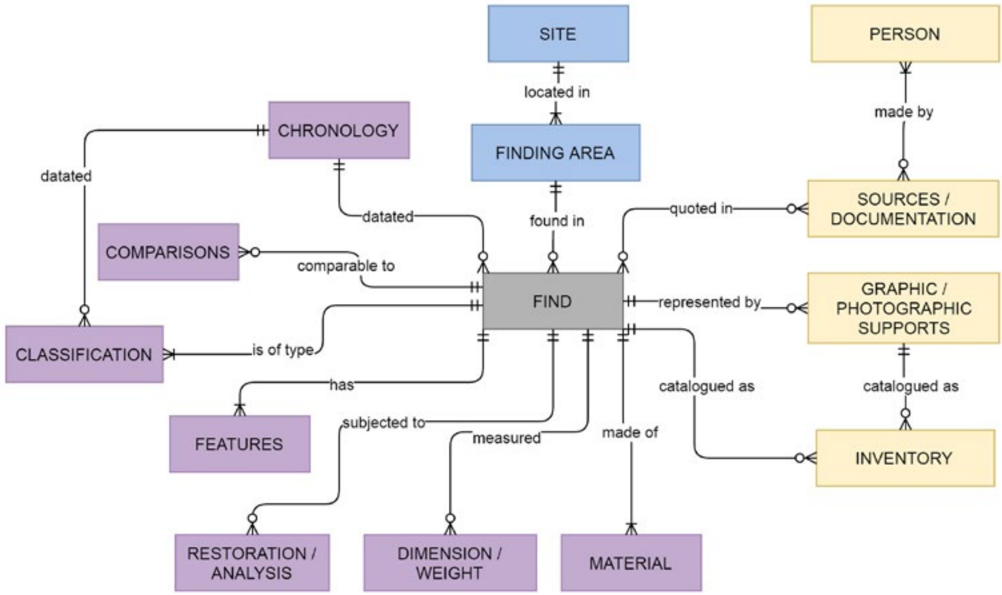


Figure 4. The conceptual model with the entities identified: the main Find entity (in grey) and the others related to their topography (in blue), characteristics (in purple) and documentation (in yellow).

According to the latter, the capabilities of a management system must be identified to guarantee the systematic organization of data and to satisfy standards and stakeholders. The next step is the identification of stakeholders, starting from the domain specialists and the developers involved in the project, to the users of the system belonging to many different levels: scholars, specialists, university students, and the wider audience, including institutions directly or indirectly interested. The possibility to deeply analyse the information, the standardization of vocabularies and the use of ICCD standards were explicitly required, whereas the greater attention to archaeological and archival sources, and the data management of the collection's history were unexpected aspects (Figure 3).

The system has been created to manage information about archaeological finds which differ in chronologies and types, archaeological and archival sources, data related to restoration activities and specialized analysis. It can be further improved by including, for example, information regarding the transfer of ownership of the finds, linked to the history of collecting. The domains have been conceptualized through a formal representation of all relevant entities and their relationships. The conceptual model is based on the prerequisites of completeness, correctness, readability and minimality. The entities which characterize the core of the museum collection have been identified: the central role is played by the archaeological find (i.e. *Find* entity), whereas all the other entities are organized into three main functional data categories related to topography and documentation (Figure 4).

The logical model developed is organized in such a way as to adopt the ICCD standards. Beyond the usual information regarding the background (e.g. chronology, dimensions, materials, discovery areas, sites, etc.), as well as the peculiar features (e.g. material, typology, dimensions, analysis, etc.), particular attention was paid to the documentation (e.g. documentation and archival sources, graphic and photographic supports, etc.). The correct use of the previous documentation involves some methodological problems in the treatment of the legacy data, and their integrity and reuse in a responsible way (Bechhofer *et al.* 2010; Martone 2015; Wilkinson *et al.* 2016).

The data management, in a first phase, is entrusted to a relational database and for the development of an effective and expressive ontological model, the data acquired through museum resources is normalized and organized into homogeneous and structured collections with adequately refined level of granularity.

Data relating to the 325 archaeological finds are digitized and managed by the database through a provisional Web-based interface, implemented thanks to the students in Archaeology of the University of Catania. The next step will consist in the creation of the final web portal that will collect and make available as open formats all the information related with the MUAC.

Relational databases, even though well-assessed tools for organizing and querying information, do not support global and flexible data-integration mechanisms with other sources, and suffer from limited modelling and reasoning capabilities. Moreover, to take advantage of the rich information collected from scholars, it is required a representation system capable of dealing with the archetypal elements of reality. Semantic web offers powerful and well-established methodologies, languages, and tools for knowledge representation systems in which data is published, accessed, and integrated with information from other sources at a global level, thus allowing coherence and dissemination of knowledge.

The process of data normalization requires the identification of suitable and consolidated existing vocabularies and the definition of new terminological entities. For this reason, we exploit the potential of Linked Data to connect archaeological data of the collection of the MAUC with other data sets such as the collection of epigraphs of the Castello Ursino Civic Museum of Catania (Figure 5). To accomplish such a task, we rely on already existing ontologies on inherent domains, such as OntoCeramic 2 (Brancato *et al*. 2019) and EpiONT (Cantone *et al.* 2019), defined according to the most important standard available for the integration of data in cultural heritage and in archaeological contexts. Specifically, OntoCeramic 2, was developed according to the standards of CIDOC CRM, and is conceived to represent and reason on the artificial and natural processes that shaped Sicilian archaeological landscapes, their conformation and topographic information, the distribution of ancient rural sites, and the dynamics of the agrarian organization in Sicily. The ontology models new survey and legacy data on pottery stored in the archives of Heritage Superintendence of Syracuse and Catania, in the Regional Technical Office of Sicily, and in the State Archives of Palermo and Catania. The EpiONT ontology was developed within the EpiCUM project[5] which aims to present and make available through a Web portal the epigraphic *corpus* of the Castello Ursino Civic Museum of Catania encoded in EpiDoc.[6] It provides a means for representing the extraordinarily rich epigraphic heritage preserved by the Castello Ursino Civic Museum and is designed according to the CIDOC CRM standard, making use of the SKOS vocabularies of the EAGLE

[5] EpiCUM (Epigraphs of Castello Ursino Museum) is accessible at http://epicum.istc.cnr.it/ (accessed 20/7/2021).
[6] https://sourceforge.net/p/epidoc/wiki/Home/ (accessed 20/7/2021).

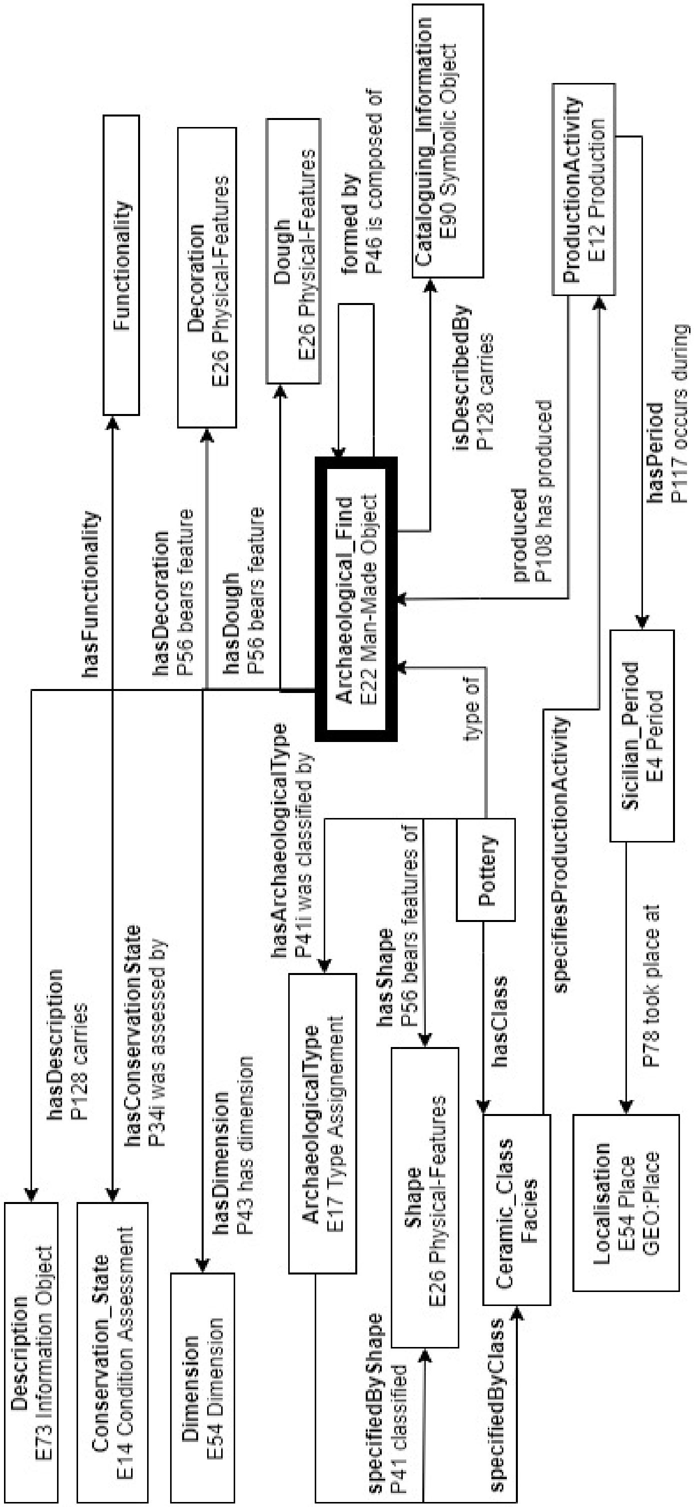


Figure 5. Partial model of the MAUC ontology concerning pottery.

project[7] concerning material, execution technique, type of inscription, and type of support of an epigraph. Currently, the ontology is populated by 580 epigraphs collected in the Castello Ursino Civic Museum and can handle any uncertainty in the origin and place of discovery of the epigraphs.

Finally, the about 21026 records of MUAC currently stored in a relational database will be also available in the JSON-LD format through an OpenLink Virtuoso SPARQL endpoint accessible thanks to a dedicated Web portal implemented in Java Spring and providing access to three main user profiles, namely, a) non-affiliated users that do not need a registration step and read only small amount of data; b) affiliated users that require a standard account with only-read capabilities without limitations; c) accredited users such as scholars that are allowed to publish and modify their own data.

Conclusions

The project is currently not yet completed, whereas the first phase of the digitization process was accomplished in September 2020. This phase of the project was a unique experience also in terms of educational action because Digital Humanities experts, students, and researchers were involved and worked together for a common aim.

The database hosting the Libertini collection of the MAUC will be published online as part of the digital and open-shared practices recently launched by the University of Catania, implemented starting from 2021 by the 'Sistema Museale' of the University of Catania (SIMUA) directed by Germana Barone. The next step is the 3D digitization of each archaeological item, carried on the standards currently in use in other projects. The ontology of the MAUC is also a process of combining and merging well-formed ontology for many ambits. OntoCeramic and EpiOnt are the starting points: all the data will be available in open format and linked to other digitized collections. More specifically, we aim at connecting the ontologies of the MAUC with those already available as is the case of the Epigraphic Database Falsae, that has been recently developed by Gianvito Pio – KDDE Research Group[8] (hosted by the University of Ca' Foscari, Venice) for the fake objects present in numerous museum collections.

In conclusion, the process of digitization of the Libertini collection of the Museum of Archaeology of the University of Catania appears as a fundamental step towards a mandatory objective for scholars of the ancient world that should be to make every archaeological collection available online for the whole community. Such a digitizing process is the only true democratic tool for crossing geographical and scientific boundaries and making data open and available to everyone.

[7] EAGLE. The Europeana network of Ancient Greek and Latin Epigraphy, https://www.eagle-network (accessed 20/7/2021).
[8] http://kdde.di.uniba.it/ (accessed 06/08/2021).

Bibliography

Allison, P. 2008. Dealing with Legacy Data – an introduction. *Internet Archaeology* 24. https://doi.org/10.11141/ia.24.8.

Anichini, F. and G. 2012. #MappaOpenData. From web to society. Archaeological open data testing (Opening the past. Archaeological open data). In M.L. Gualandi (ed.) *MapPapers* 2: 51–56. Roma: Edizioni Nuova Cultura.

Bechhofer, S., I. Buchan, D. De Roure, M. Gamble and C. Goble 2010. Research objects: towards exchange and reuse of digital knowledge. In *The Future of the Web for Collaborative Science (FWCS 2010)*. Springer Nature. https://doi.org/10.1038/npre.2010.4626.1.

Bogdani, J. 2019. *Archeologia e tecnologie di rete: metodi strumenti e risorse digitali*. Roma: Bradypus.

Brancato, R., M. Nicolosi-Asmundo, G. Pagano, D.F. Santamaria and U. Salvatore 2019. Towards an Ontology for Investigating on Archaeological Sicilian Landscapes. In *ODOCH 2019*, *CEUR Workshop Proceedings* 2375: 85–90.

Cantone, D., S. Cristofaro, M. Nicolosi-Asmundo, F. Prado, D.F. Santamaria and D. Spampinato 2019. An EPIDOC Ontological Perspective: the Epigraphs of the Castello Ursino Civic Museum of Catania via CIDOC CRM. *Archeologia e Calcolatori* 30: 139–157.

Martone, M. 2015. FORCE11: building the future for research communications and e-scholarship. *Bioscience* 65.7: 635.

Terras, M. 2015. Cultural Heritage Information: Artefacts and Digitization Technologies. In G. Chowdhury and I. Ruthven (eds) *Cultural Heritage information*: 63–88. London: Facet.

Tortorici, E., G. Buscemi Felici and G. Biondi 2015. *Il Museo di Archeologia dell'Università di Catania. Collezioni Libertini*. Catania: Bonanno Editore.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18.

Fieldnotes for the development and publication of open standards for the vectorisation of archaeologic and architectonic topographic legacy data

Julian Bogdani

Julian Bogdani, Sapienza University of Rome; julian.bogdani@uniroma1.it
Correspondence: julian.bogdani@uniroma1.it

Abstract
The digitisation and vectorisation of archaeological/architectonic graphical legacy documentation is a step that any archaeological project deals with and solves by following its own specific requirements. The article introduces the Simple Vectorisation Protocol (SVP), a GIS-based protocol for acquiring in the digital domain sketches and maps by following a very concise yet rich syntax, able to reverse engineering published and archive data. The result is a multidimensional dataset, ready to be used in specific projects, easy to publish on online repositories and highly re-usable and remixable, while still maintaining a feature-level bibliographic reference to the original source.

Keywords: LEGACY ARCHAEOLOGICAL DATASETS; ARCHAEOLOGICAL DATAFICATION; COLLABORATIVE PROTOCOLS; ONLINE DATA PUBLISHING; GIS; OPEN STANDARDS.

Introduction

This contribution has been conceived in the context of PAThs, an ERC Advanced project directed by P. Buzi at Sapienza University of Rome[1] during the activities of the Digital Archaeology Lab at Sapienza.[2] Since the general aims of the project are well known,[3] it is more effective to shift the focus to the specific topic of this article, i.e. the digitizing, georeferencing, vectorizing and processing of archaeologic and architectonic legacy drafts, sketches and maps in order to include these documents in the already publicly available Archaeological Atlas of Coptic Literature.[4] In broad terms, and from the technological point of view, the Atlas has been conceived as a rather lightweight Single Page Application (SPA) written in HTML and JavaScript,[5] pulling structured data from PAThs centralized online relational database and from other custom web mapping services (Figure 1).[6] At present, the Atlas itself is only one

[1] The complete title of the project is Tracking Papyrus and Parchment Paths. An Archaeological Atlas of Coptic Literature. Literary Texts in Their Original Context. Production, Copying, Usage, Dissemination and Storage, project number: 687567, URL: http://paths.uniroma1.it/ (accessed 9/4/2021).
[2] LAD: Laboratorio di Archeologia Digitale, supervised by the author of this paper, http://purl.org/lad (accessed 17/6/2021).
[3] Buzi 2020; Bogdani 2020; Buzi, Berno, and Bogdani 2018; Bogdani 2017.
[4] https://atlas.paths-erc.eu (accessed 9/4/2021), DOI: https://dx.doi.org/10.5281/zenodo.3357946.
[5] The React.js Framework (https://reactjs.org/, accessed 31/3/2021), an open-source (MIT Licensed) project developed and maintained by Facebook Inc., has been used for the coding of the application.
[6] PAThs database has been designed and realized using BraDypUS (DOI: https://dx.doi.org/10.5281/zenodo.4642442) open-source software (A-GPL v3.0 license) that facilitates the creation and usage of relational databases such as MySQL/MariaDB, PostgreSQL or SQLite, with a particular focus on Cultural Heritage use -cases. For the source code: https://dx.doi.org/10.5281/zenodo.4642442. For another use case of a clear separation of data management and data presentation based on BraDypUS, cfr. Bogdani 2016.

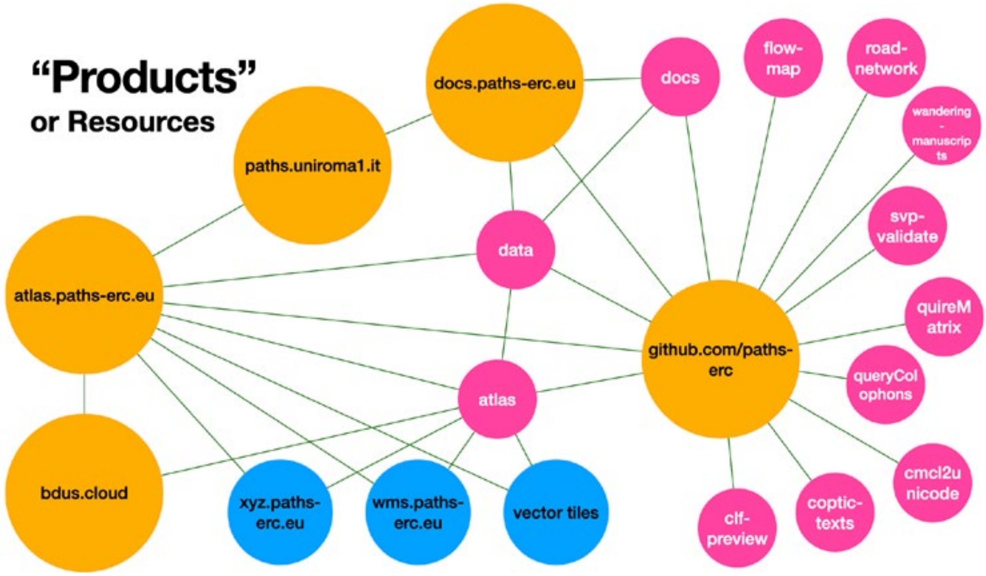


Figure 1: A graphical representation of digital resources powering the Atlas of Coptic Literature.

part of a rather complex net of digital resources, software packages, datasets and web services built to support the research project and released as open-source software and open-access databases.[7] It was important for the PAThs team to share data and algorithms being developed with the broader community of researchers focused on Egypt willing to study, experiment and adapt to their needs these resources, and hopefully also willing to provide feedback or contribute to their further development.

To facilitate data and software reusability, and also to support a complete transparency on the data production and publication processes, a specific documentation and data portal has been made available to the community, where a clear track of updates and new releases is being maintained (Figure 2).[8]

The Simple Vectorisation Protocol (SVP): the general picture

The Simple Vectorisation Protocol (SVP) is meant to provide a very easy way to implement the encoding in vectorial format of architectonic and archaeological drafts, sketches, and maps and to be able to provide, beyond geometry, additional informative and interpretative layers of structured data.

The need for this protocol originated when efforts were made to provide an archaeological context for most of the Late Antique and Medieval Egyptian sites related to the literary manuscript production, usage, and dissemination. It was decided to collect, digitize, georeference and finally vectorize archaeological and architectonic maps of buildings of

[7] All content is available under the terms of CC BY-NC-SA 4.0 (http://creativecommons.org/licenses/by-nc-sa/4.0/). The Atlas itself is released as GNU-AGPL 3.0 License.
[8] https://docs.paths-erc.eu/ (accessed 9/4/2021).

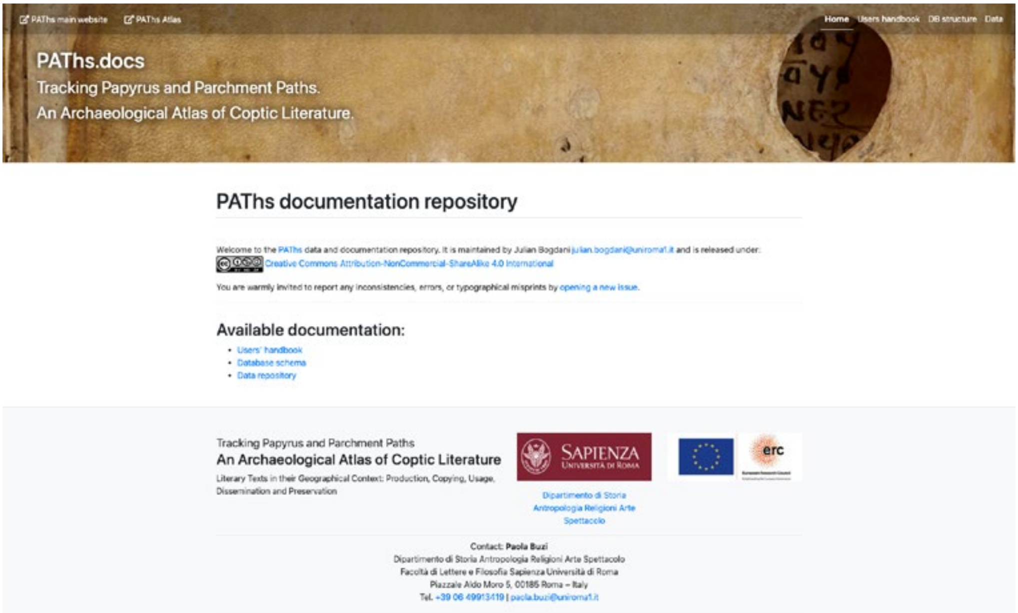


Figure 2: Screenshot of PAThs documentation and data portal, https://docs.paths-erc.eu.

religious destination (chapels, churches, basilicas, and similar) from previously published bibliographical and archival records. The preliminary phases of digitizing and, most importantly, that of contextualizing (i.e. georeferencing), required very attentive work and raised important research questions on the urban and topographical layout and on the diachronic stratification of many Egyptian sites. The detailed documentation and thorough discussion of these steps will be the object of a dedicated work aimed at providing the scholarly community with specific details on the grade of accuracy of the georeferencing process, the archive/bibliographic records involved, and to record and make explicit the not so rare arbitrary decisions that have determined the locations of a specific building inside the urban landscape. Dealing with legacy data often requires discretionary decisions on the interpretation of the available information. These questionable decisions must necessarily be taken to make data processable, but greater transparency must be made available on the *status quaestionis* and process that led to each conclusion. For this reason, the detailed recording of comprehensive *paradata*[9] is a paramount prerequisite for the vectorising process, most importantly in cases when uncertainty and lack of information are greater.

These considerations are also valid for the next step of vectorisation of the legacy raster-based sketches and maps, which beyond uncertainties about precise location, precise scale, and faithfulness of representation, require also great attention to the interpretation of the symbology, and finally to the date when the map was produced. In fact, published maps are to be considered a sort of snapshot of the archaeological situation at the time of the survey, more rarely embedding previous knowledge. Updates or corrections to these maps are never mere additions of physical features or structures, i.e. walls, pavements, doorways, etc., but come always with a deep reconsideration of the archaeological context. It is therefore important

[9] As defined in the London Charter, https://www.londoncharter.org/glossary.html (accessed 9/4/2021).

to precisely annotate the source of information and the relative bibliographic record to fully contextualize each archaeological feature in what we could call the archaeological *continuum* of knowledge production.

It is important to clearly highlight that the idea behind the drafting and implementation of this protocol is not that of obtaining a data model for the thorough representation of space and time in the archaeological domain, aiming at maximising the interoperability, as might be the Star model[10] or the proposed extension of CIDOC Conceptual Reference Model (CRM) to the archaeological excavations, named CIDOC CRM*archaeo*.[11] It is also not an extension of existing standard formats, such as for example GeoJSON-T,[12] developed to add temporal dimension to the GeoJSON format.[13] In particular, GeoJSON-T is of some interest for archaeologists, and more in general for humanists. It was developed to help interoperability for Digital Humanities projects dealing with space *and* time, two tightly interwoven entities. The encoding of time (duration) in addition to space (geometry) has always been problematic and each project has adopted *ad hoc* solutions that have inevitably turned into barriers to data comparison and exchange, i.e. interoperability. The solution proposed by JSON-T is adding a *foreign member* to GeoJSON, the *when* key, where time-related information can be encoded. This information can be one or more timespans, period definitions, labels, durations, or relative chronologies. Most interestingly, a further proposal to add a proper way to encode uncertainty and approximation has been advanced but not yet merged in the main branch.[14] The possibility for having *foreign members* is already present on the GeoJSON format specifications[15] but support for it is optional and depends exclusively on the implementation of the format. While some projects – mainly focused on web mapping and online data publication – are implementing the JSON-T format[16] there is no plan at present to see it adopted broadly on popular open-source GIS software, such as QGIS. This would be an initiative that would greatly ease the life of archaeologists dealing with geographical features through time.

Some conceptual and technical differences exist between SVP and GeoJSON-T that are worth listing, also in the perspective of future integration. While SVP can certainly be serialized and stored as GeoJSON, it does not rely exclusively on this format. Any GIS format able to manage coordinates and attributes can be used to implement SVP, consequently any GIS package can be employed to enter data, with no need to code *ad hoc* software and/or plugins. This implies that the data-structure defined by SVP does not rely on custom *foreign members*, but uniquely on the list of feature properties, their types and the vocabularies used to populate them.

Thus *simple* is a fundamental keyword, and simpleness means a plain structure – a flat list of attributes – that can be easily built (and eventually validated) by archaeologists who feel comfortable with GIS software in their daily workflow but that not necessarily have

[10] For an attempt in this direction, on the basis of ISO Standard 19100, cf. Migliorini, Grossi, and Belussi 2017.
[11] Doerr et al. 2019.
[12] https://github.com/kgeographer/geojson-t (accessed 9/4/2021).
[13] Standard RFC 7946, https://tools.ietf.org/html/rfc7946 (accessed 9/4/2021). In recent times, greater attention has been dedicated to a deeper consideration of data formats for the encoding of archaeological datasets with spatial capabilities, with particular regard to interoperability and standards, cf. D'Andrea et al. 2007; Forte 2019; and the contribution by D'Andrea and Forte in this volume.
[14] https://github.com/kgeographer/geojson-t#possible-extension (accessed 9/4/2021).
[15] Section 6 of RFC 7946, https://tools.ietf.org/html/rfc7946#section-6 (accessed 9/4/2021).
[16] For a list of the projects currently implementing the JSON-T format, cf. the 'Uptake' section of the official repository at https://github.com/kgeographer/geojson-t#uptake (accessed 9/4/2021).

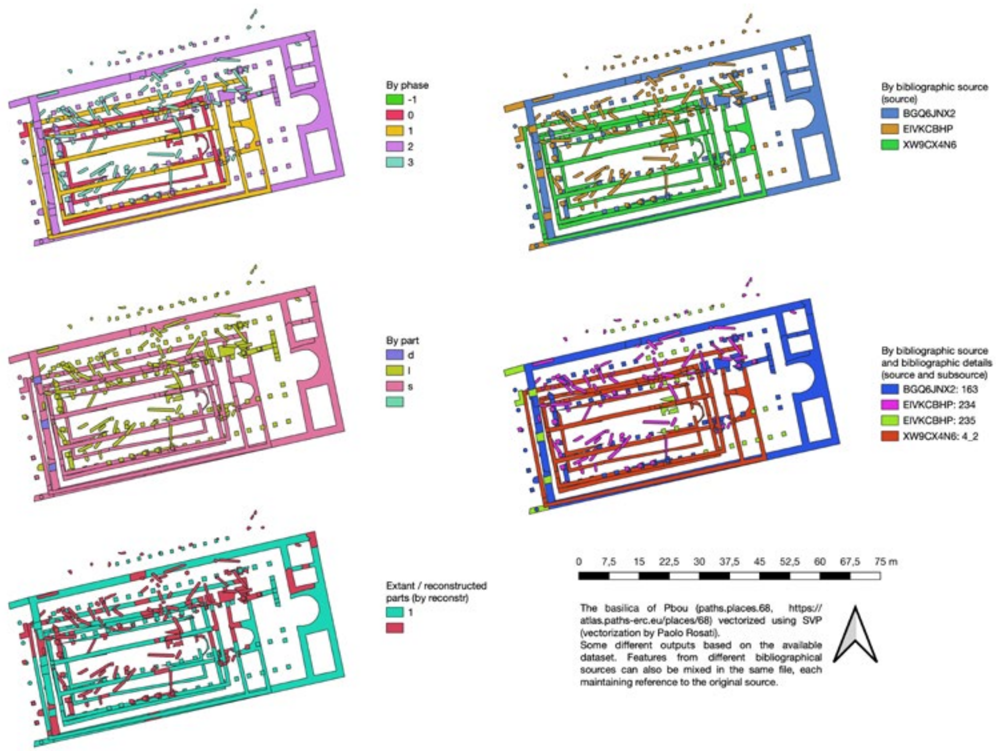


Figure 3: Different representations of the basilica of Pbou (http://paths.uniroma1.it/atlas/places/68) encoded using SVP.

coding experience. Ease of implementation and ease of usage results in consistent datasets embedding archaeological interpretation and providing a clear map, at feature level, of the epistemological process that produced each bit of information and interpretation.

SVP specifications

As already mentioned, SVP consists of a minimal list of attributes suited to extract and digitally encode virtually the full set of information that a paper drawing/sketch/map of an archaeological or architectonic complex conveys (Figure 3). The main effort has been spent on finding an equilibrium between a higher level of abstraction (a minimal set of attributes and vocabularies) and a yet intuitive and user-oriented experience of data-entry. The test-bench for the protocol was a series of GIS classes held at Sapienza University of Rome involving younger BA archaeology students who were partially aware of the graphical conventions of representing architectural and archaeological structures and who were also introduced to the use of GIS-based georeferencing of raster images and basic vectorisation.[17] They were also briefly introduced to the main specifications of the protocol, that were also corrected based on their feedback.

[17] A special note of thanks for their continued support goes to Domizia D'Erasmo and Paolo Rosati.

The attributes by which each feature is described in SVP are the following:

- *place*: a unique identifier of the place or archaeological site where the building represented is located. At PAThs Place ID (https://atlas.paths-erc.eu/places) is filed. URIs also can be used if a linked data output is planned. The usage of identifiers from well-known gazetteers[18] is strongly encouraged.
- *subplace:* eventually the name (ancient or modern) of part of the *place* where the building is located. This is used to provide a more direct (and human readable) topographical location of the represented building.
- *reconstr:* records whether the feature was visible when it was surveyed, or if the surveyor/digitizer reconstructed it based on an archaeological hypothesis. The following values are available:
    - 0: the feature was clearly visible and preserved when surveyed (usually a solid line is used to represent it on paper maps).
    - 1: the feature was not seen by the surveyor, who reconstructed it by hypothesis, based on archaeological analysis or by comparison (usually a dashed or dotted line is used to represent a reconstructive hypothesis on paper maps).
    - 2: as in the previous case the feature was not visible on the ground, and furthermore the original surveyor did not propose a reconstructive hypothesis. The reconstruction was made in the successive phase of the vectorizing. It is clear that SVP supports an active way of digitizing, by keeping a clear track of the source of information.
- *part:* part of the building the feature belongs to, indicating not the function (except for doorways) but relation to the section plane. The following values are available:
    - *null:* in cases when no information is available from the original source
    - *s:* for *sectioned*, used for walls or structures which are cut by the section plane
    - *u:* for *upper projection*, used for features located above the section plane, e.g. roofing, domes, beams, etc.
    - *l:* for *lower projection*, used for features lying below the section plane, e.g. pavements, altars, benches, etc.
    - *d:* for *doorways*, used for thresholds and doorsteps. This is the only exception for an indication of function, which is usually very difficult to recover from mute maps.
- *phase:* an indication of chronology when available. In PAThs an integer is used to indicate relative chronology. Also, URIs can be used, pointing to an external source describing in detail absolute or relative chronologies. The usage of identifiers from shared gazetteers, such as PeriodO[19] is highly recommended if the production of linked data is a priority.
- *lost:* whether the feature is currently preserved or not. Sometimes a feature was visible at the moment of surveying (*reconstr: 1*) but it is not preserved/visible nowadays (*lost: 1*). 0 or 1 values can be used.
- *height:* the maximum height, elevation, or altitude of the single element/feature, if available. In combination with *minHeight* it can be used to graphically and programmatically extrude elements to form a 2.5D representation (Figure 4).

[18] Such as Pleiades (https://pleiades.stoa.org/, accessed 9/4/2021) and World Historical Gazetteer (http://whgazetteer.org/, accessed 9/4/2021).
[19] https://perio.do/ (accessed 9/4/2021).

Figure 4: A screenshot from https://docs.paths-erc.eu/data/demo/#paths.places.16-Elephantine/BGQ6JNX2-186 showing the temple of Knum at Elephantine (blue) and the Late Antique garrison structures (in orange) on left and a 2.5D automatic representation of the structures.

- *minHeight:* the base height, elevation, or altitude of the single feature.
- *scale:* the denominator or the scale (when available) of the original map.
- *source:* unique identifier or URI of the bibliographic/archive record where the original map is published. In PAThs project the identifier of the official Zotero database[20] is being filed for each item.

[20] https://www.zotero.org/groups/2189557/erc-paths/ (accessed 9/4/2021).

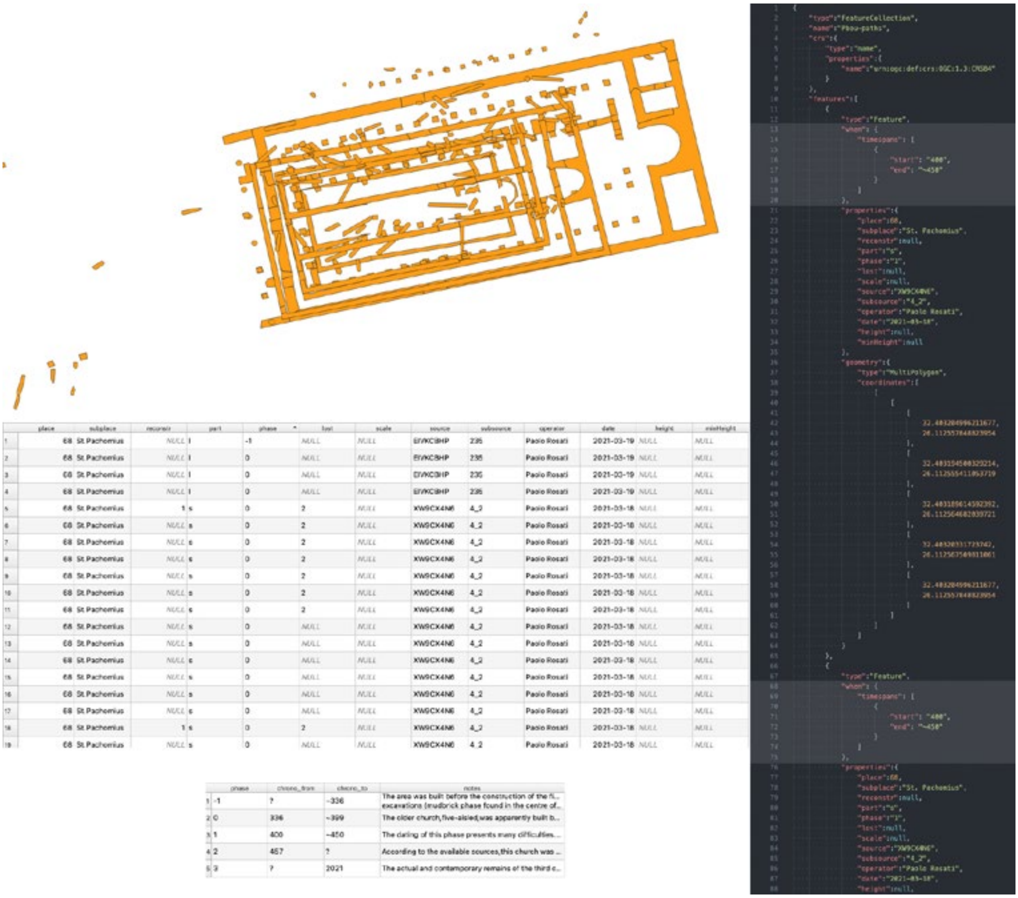


Figure 5: From SVP to GeoJSON-T: in the left-upper part the graphical representation of the basilica of Pbou encoded using SVP; in the left-middle part the SVP attribute list (excerpt); in the left-lower part a flat table mapping phases to absolute chronology; on the right excerpt of the resulting GeoJSON-T, embedding absolute chronology on feature level.

- *subsource:* if needed, further details regarding the bibliographic record, such as page, table, figure numbers, etc., can be provided here.
- *operator:* name, URI or any other string identifying the person who completed the vectorisation.
- *date:* date and time of the vectorisation process.

This structure does not imply strong assumptions on software or on file formats to use, the only requirement being the usage of a GIS format, able to handle geometry and non-geometry attributes. In addition, if a SVP implementation is serialized (or exported) in GeoJSON, it can be easily enhanced into a Linked Data format, by taking advantage of the implicit availability of external links expressed as URIs, as, for instance, for location, bibliographical reference, chronology, etc.[21] The same is true for other extensions, such as the above mentioned GeoJSON-T, that can be easily and programmatically obtained from SVP by *embedding* information on phasing and dating stored, for example, in a parallel database or also in the Linked Open Data Cloud (Figure 5).

While specific tools aimed at enhancing, converting, publishing, and visualising SVP encoded legacy data can be written in any programming language – especially if the dataset is serialized in JSON based formats – two very minimal examples can offer a glimpse on the potentiality and ease of implementation.

The first is a simple web-based demo, hosted inside the PAThs documentation portal,[22] allowing the visualisation about 150 georeferenced and SVP-encoded maps and sketches of churches, basilicas, and other religious buildings of the Late Antique and Medieval Egypt,[23] automatically styled by combining *part* and *phase* attributes (Figure 4, top) with support for 2.5D visualisation based on *height* and *minHeight* attributes.[24]

The second application[25] is a very simple validation tool written in HTML and JavaScript aiming at providing some grade of consistency in the implementation of SVP, with particular focus on the specific implementation of the protocol at PAThs.[26]

Conclusions. Some methodological considerations beyond (or before) practical implementations: digitisation vs. datafication

The digitisation of the archaeological record can be considered as a multi-level process, that too often stops at the very first steps of acquiring in digital format information previously published on paper. This output, be it a raster file or a CAD-based vector file, can be described as *analogical content on digital format*. Simply put, the use we make of these files is no different from the use we have made for a long time of the paper-based originals, reproducing in bytes – by analogy – the appearance of paper. Yet, paper is but a medium that conveys information, and *digitisation* should be intended as the process aimed at acquiring in the digital domain rich, complex, and multi-layered information stored in a rather flat and poor paper format. Datafication[27] should therefore be the fourth step in the archaeological workflow, following (1) digitisation, (2) georeferencing, and (3) vectorisation. SVP can be considered as a theoretical and practical attempt aimed at reverse engineering legacy archaeological geo/topographical data stored on paper support, seeking conciseness and completeness, trying not to introduce new tools or formats, but relying on GIS software that archaeologists have used for decades now. As such, the protocol can be perfectly integrated in workflows that look towards the future of data sharing and publishing, such as the Linked Open Data universe.

[21] This implementation can also benefit from the GeoJSON-LD implementation, https://geojson.org/geojson-ld/ (accessed 9/4/2021), combining GeoJSON and JSON-LD.
[22] https://docs.paths-erc.eu/data/demo/ (accessed 9/4/2021).
[23] The main source being Grossmann 2002.
[24] The demo application uses Leaflet.js (https://leafletjs.com/, accessed 9/4/2021) as web mapping library, Bing and Google maps as base satellite imagery and OSM Buildings (https://osmbuildings.org/, accessed 9/4/2021) as web-viewer for 3D buildings.
[25] Code at: https://github.com/paths-erc/svp-validate (accessed 9/4/2021), GNU A-GLP licensed.
[26] Available at https://paths-erc.eu/svp-validate/ (accessed 9/4/2021).
[27] Anichini *et al.* 2015: 153.

Bibliography

Anichini, F., N. Dubbini, F. Fabiani, G. Gattiglia and M.L. Gualandi 2015. Dall'Open Data Alla Predittività. Nuovi Modi Di Far Crescere l'archeologia Italiana. *Archeologia e Calcolatori* Supplemento 7: 141–154.

Bogdani, J. 2016. Un Archivio Digitale Multidisciplinare per La Gestione e La Conservazione Di Un Patrimonio Culturale a Rischio: Il Progetto Ghazni (Afghanistan). *Archeologia e Calcolatori* Supplement: 236–245.

Bogdani, J. 2017. The Archaeological Atlas of Coptic Literature. A Question of Method. *Vicino Oriente* 21: 59–69.

Bogdani, J. 2020. Defining Methodologies and Protocols for the Use and Re-Use of Archaeological Legacy Data. The Case Study of the Archaeological Atlas of Coptic Literature. In P. Buzi (ed.) *Coptic Literature in Context (4th-13th Cent.). Cultural Landscape, Literary Production, and Manuscript Archaeology. Proceedings of the Third Conference of the ERC Project Tracking Papyrus and Parchment Paths: An Archaeological Atlas of Coptic Literature : Literary Texts in Their Geographical Context ('PAThs')*: 311–324. Percorsi Di Archeologia 5. Roma: Quasar.

Buzi, P. 2020. The Places of Coptic Literary Manuscripts: Real and Imaginary Landscapes. Theoretical Reflections in Guise of Introduction. In *Coptic Literature in Context (4th-13th Cent.). Cultural Landscape, Literary Production, and Manuscript Archaeology. Proceedings of the Third Conference of the ERC Project Tracking Papyrus and Parchment Paths: An Archaeological Atlas of Coptic Literature : Literary Texts in Their Geographical Context ('PAThs')*: 7–26. Percorsi Di Archeologia 5. Rome: Quasar.

Buzi, P., F. Berno and J. Bogdani 2018. The 'PAThs' Project: An Effort to Represent the Physical Dimension of Coptic Literary Production (Third–Eleventh Centuries). *COMSt Bulletin* 4: 47–66. https://doi.org/10.25592/uhhfdm.253.

D'Andrea, A., A. Felicetti, S. Hermon, F. Niccolucci and T. Zoppi 2007. I Linguaggi Standard Del W3C e Gli Strumenti Open Source per La Gestione Dei Dati Archeologici. In R. Bagnara and G. Macchi Janica (eds) *Open Source, Free Software e Open Format Nei Processi Di Ricerca Archeologici: Atti Del I Workshop (Grosseto, 8 Maggio 2006)*: 51–66. Firenze: Centro editoriale toscano.

Doerr, M., A. Felicetti, S. Hermon, G. Hiebel, A. Kritsotaki, A. Masur, K. May, *et al*. 2019. Definition of the CRMarchaeo. An Extension of CIDOC CRM to Support the Archaeological Excavation Process. Proposal for Approval by CIDOC CRM - SIG. Version 1.4.8, February 2019. http://www.cidoc-crm.org/crmarchaeo/ModelVersion/version-1.4.8 (accessed 9/4/2021).

Forte, F. 2019. Informazioni spaziali e strumenti d'indagine semantica sui dati archeologici digitali. *Newsletter di Archeologia CISA* 10: 95–111.

Grossmann, P. 2002. *Christliche Architektur in Ägypten*. Leiden – Boston – Köln: E.J. Brill.

Migliorini, S., P. Grossi and A. Belussi 2017. An Interoperable Spatio-Temporal Model for Archaeological Data Based on ISO Standard 19100. *Journal on Computing and Cultural Heritage* 11 (1): 5:1–5:28. https://doi.org/10.1145/3057929.

Analysis and comparison of open and non-open spatial formats for archaeological research

Andrea D'Andrea, Francesca Forte

Andrea D'Andrea, University of Naples 'L'Orientale', dandrea@unior.it
Francesca Forte, University of Naples 'L'Orientale', fforte@unior.it

Abstract
This contribution attempts an overview on how to manage the spatial information of archaeological data available online in a non-open format and how they can efficiently be reused in different applications to extend the domains of knowledge: a key-role in this process is played by open standards, of which a review of comparison with a non-open format is provided in this study. A practical test has been carried out to demonstrate the benefits of using open formats for the scientific research and dissemination and to concretely verify the advantages of using formats, such as GeoJSON.

Keywords: OPEN SCIENCE; BIG GEOSPATIAL DATA; SPATIAL INFRASTRUCTURE; GEOJSON; OGC.

Introduction

This paper provides a short analysis of the current management and sharing of spatial information in the archaeological field according to the principles of Open Science (OS) and Big Data (BD). Up to now the topic has been treated mainly from the point of view of the implementation of digital infrastructures with scarce attention to standard and format. This issue has already been partially analysed in a previous contribution on comparisons of the most used spatial data in archaeology (Forte 2019: 95–111).

Our attention here focuses on the concept of Big Geospatial Data (BGD)[1] and its relationship with archaeological research. The paper deals with the geo-locational dataset available online and describes a test for re-using accessible archives in a non-open format. Furthermore, it shows how it is possible to integrate archaeological datasets through open systems and web applications, eventually to address the need to manage the growing size of archaeological geospatial datasets. The paper does not provide a solution from among the large availability of open and not-open formats; it offers some suggestions on how to choose the spatial data online to be integrated in the spatial analysis.

[1] BGD are all types of data objects or elements with geographical information unrelated to a specific data model (raster, vector), to a software, or format. Due to volume variety or velocity, BGD overcome the capacity of commonly used spatial systems.

Open Science and Big Geospatial Data

The term 'Open Science' identifies 'an approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology'.[2] One of the goals of research and innovation policy is to share knowledge and data as early as possible in the research process, making publications available in open-access and providing spatial datasets as open as possible; only in this way the scientific research improves its quality, efficiency and responsiveness.

Over the past few years, the growing diffusion of satellite networks and the emerging of the Internet Of Things (IoT) have allowed users to track and correlate the accurate positions of people and objects; furthermore, thanks to the spread of technologies that are becoming more and more accessible, a large amount of Geospatial Data has been produced.[3] Within the flow of data that is collected daily, Geospatial Data occupy a strategic role in the perspective of BD. Therefore, it is possible to use the term 'Big Geospatial Data' to define datasets that include locational information and exceed the capacity of widely available hardware, software, and human resources and require specialized effort to work with: without an appropriate methodology to manage them, the strong potentiality of BD dramatically decreases.

Unfortunately, most of the geospatial data are kept in personal repositories, formalised according to international standards, restricting the dissemination and the enrichment of the scientific knowledge. The adoption of FAIR (Findable, Accessible, Interoperable and Re-usable data) principles[4] and open data should strengthen the diffusion of the results of the scientific research.

The first steps in this direction were made by OGC (Open Geospatial Consortium)[5] in 2017 with the publication of a White Paper[6] aimed at defining more efficient approaches for the treatment of BGD. The goal of the paper was to understand what actions the OGC should undertake for the correct management of spatial data and the improvement of the capabilities of BGD. The White Paper is a survey including open standards needed to guarantee interoperability, efficiency, innovation, and cost effectiveness. No specific recommendation is given, but the paper focuses on the need to adopt an analytical process to better exploit the opportunities offered by BGD.

Big Geospatial Data in archaeology

Geospatial Data in archaeology has an extraordinary importance for the reconstruction of the past, by allowing the identification of territorial patterns. Technological advancements have positively affected archaeology, therefore currently archaeologists create, manage, and share geospatial archaeological data much more than they did before. It is specifically worth noting that today archaeological datasets cannot be defined as BGD, since the volume of data archaeologists work with rarely exceeds the capacity of the available resources. However,

[2] https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science_en (accessed 7/7/2021).
[3] Geospatial Data is any information relating to the relative position of things on the earth's surface.
[4] https://www.go-fair.org/fair-principles/ (accessed 7/7/2021).
[5] https://www.ogc.org/ (accessed 7/7/2021).
[6] http://docs.opengeospatial.org/wp/16-131r2/16-131r2.html (accessed 7/7/2021).

it must be highlighted that although the volume of archaeological data is still manageable, BGD will increase significantly in the future. The debate on these aspects (McCoy 2017: 74–94; Cooper and Green 2016: 271–304; Gattiglia 2015: 113–124) is very active, considering the amount of data constantly produced by rescue archaeology, by public and private organizations that hold excavation concessions, by universities, and by local authorities in charge of safeguarding cultural heritage.

The challenge is to model the growing geospatial datasets before it becomes necessary to find specific solutions for the integration. From this point of view, it is important to identify the best practices for the creation, publication, and preservation of geospatial data also through the adoption of open standard formats. This approach guarantees the reliability of the data, allows their correct divulgation and their valid reuse in different application domains.

The last decade has witnessed a proliferation of standards and formats for the interchange of data: up to now, OGC has approved 69 formats.[7] There is a growing literature on the topic (Previtali and Valente 2019: 17–27; Richards-Rissetto and Landau 2019: 120–135; Carlisle *et al.* 2014), but a specific guideline for orienting researchers among the multitude of formats is still missing; the lack of best practices makes it difficult to encourage the interoperability and the interaction with a broad variety of sources. Notwithstanding, there is a great potential in using standard formats for the online publication of spatial databases.

The state of the art

A key approach for understanding in which way to proceed leads to the analysis of how the current scenario in Italy responds to the process of standardizing the data formats available online (geoportals and digital repositories) and how they can be integrated. Firstly, the institutional experiences in the field of management geospatial data for cultural heritage was examined. An important push for the homogenization of the spatial data came from the Ministry of Cultural Heritage with the publication of the Circular 30/2019.[8] This document describes how Geospatial Data must be represented (both for raster and vector data) within the documentation that must be delivered at the end of excavation, and indications relating to the publication and dissemination of research data. The formats admitted are the ESRI Shapefile,[9] OGC GeoPackage[10] and DXF,[11] all with specific geometry (*multipolygon*) and reference system (WGS 84 – EPSG: 4326).

The ESRI Shapefile is one of the most popular formats used to encode geographic data and even if it is not included in the OGC list, it has become a standard *de facto*. It is worth noting that in the above list, only the OGC Geopackage is a standard approved by OGC. It is one of the most recent standards (2014) designed to store and transfer complex and voluminous data (raster and vectors), without losing the style of vectors among the data elements. The documentation, organised through these *criteria*, will then populate the National Geoportal

[7] https://www.ogc.org/docs/is (accessed 7/7/2021).
[8] http://www.ic_archeo.beniculturali.it/getFile.php?id=494 (accessed 7/7/2021).
[9] https://doc.arcgis.com/it/arcgis-online/reference/shapefiles.htm (accessed 7/7/2021).
[10] https://www.ogc.org/standards/geopackage (accessed 7/7/2021).
[11] https://documentation.help/AutoCAD-DXF/ (accessed 7/7/2021).

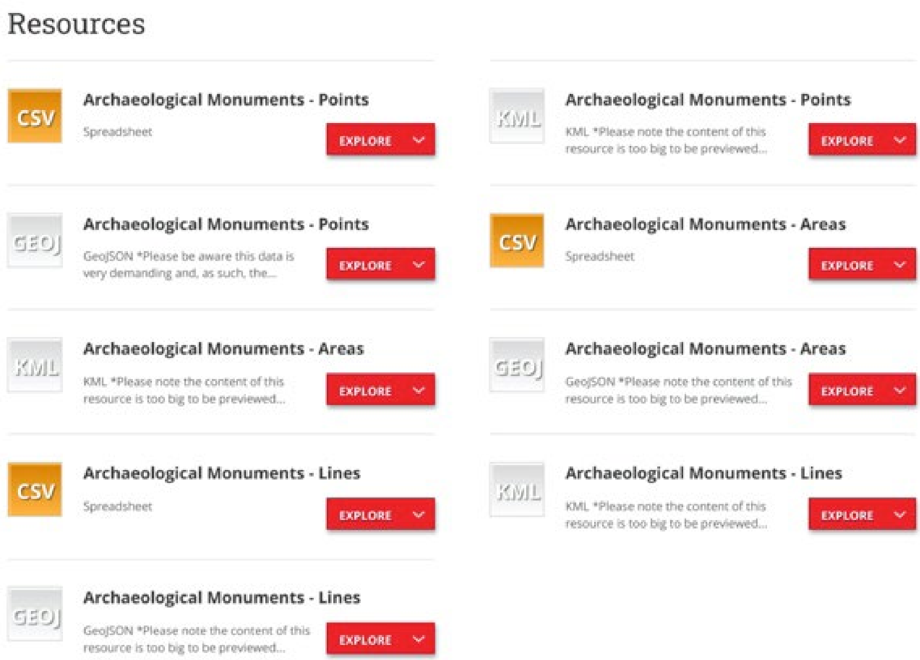


Figure 1: Resources available for 'Archaeological Monuments' (https://data.yorkopendata.org/dataset/archaeological-monuments, accessed 13/10/2021).

for Archaeology,[12] which will collect data already partially structured. At present, the platform hosts some local, regional, or national portals dealing with spatial data. The list, particularly useful for finding spatial data already available and accessible, includes some specific archaeological initiatives or thematic sites. Unfortunately, a search engine has still not been developed, and users must look for data in every single archive. Furthermore, there is no available information on licensing and the re-use of data, and on how archaeologists should structure their data to make them completely interoperable.

The awareness that archaeologists, above all for those not super-technical, have to face some problems when they deal with spatial data, leads to such questions as: among all these standards which could be the best to use? How are they going to ensure the effective interoperability of their data with others? How can we shape our knowledge in the perspective of BGD?

Searching for answers based on consolidated frameworks, it is important to look at two useful English experiences: York Open Data[13] and the Archaeology Data Service[14] have represented in recent years a reference point for open data and public sharing of knowledge in archaeology.

[12] The project, still ongoing, is aimed at the creation of an online service for the archiving, research and knowledge of data relating to the Italian archaeological heritage (http://www.ic_archeo.beniculturali.it/it/222/il-geoportale-nazionale-per-l-archeologia-gna, accessed 7/7/2021).
[13] https://www.yorkopendata.org/ (accessed 7/7/2021).
[14] https://archaeologydataservice.ac.uk/ (accessed 7/7/2021).

Data Type	Preferred File Format	Accepted File Format	Metadata Template Download Type	Example
Collection-level Metadata			Microsoft Word Open Office Document	
3D Models, Visualisation, and Virtual Reality	**Virtual Reality Modelling Language** .vrml **Wavefront OBJ File** .obj (+ .mtl + jpg textures)	**Adobe Portable Document Format (3D)** .pdf This is accepted for dissemination purposes only, it is not suitable for preservation of 3D data. **STL** .stl	Microsoft Excel Open Office Spreadsheet	
GIS (Geodatabases)		**Delimited text and ESRI Shapefile** .csv + .shp **GeoJSON** .geojson	Microsoft Excel Open Office Spreadsheet csv	GIS PDF
GIS (Raster)	**Geo-referenced TIF Image** .tif (+ world file: .tfw) or GeoTIFF	**ERDAS Imagine files** .img (+ .rrd, aux.xml, img.xml) **ESRI GRID ascii** .asc/.grd **ESRI GRID binary** .adf **JPG World** .jpg + .jpw (.rrd, .aux, .xml) **Keyhole Markup Language** .kml **PNG World** .png + .pgw (.rrd,.aux,.xml)	Microsoft Excel Open Office Spreadsheet csv	GIS PDF

Figure 2: Table of all file types accepted for GIS by ADS (https://guides.archaeologydataservice.ac.uk/, accessed 13/10/2021).

The city of York offers several downloadable datasets about the city, e.g. urban assets, the environment, archaeological monuments, events. The resources are divided by the geometry, easily downloadable through three non–proprietary open formats: CSV, KML and GeoJSON (Figure 1). It is possible to download the datasets with no legal restrictions on their use. It is also possible to easily read and understand what information each dataset contains, by the simple use of a text editor, and the data are both human readable and machine-parseable. These data can be re-used with no problems relating to technical specifications.

The cutting-edge ADS project has updated (July 2020) its guideline for depositors with a list of all file types and formats they accept, providing also a template for the required metadata associated with each data-type. ESRI Shapefile and CSV or the GeoJSON file are admitted for GIS-data (Figure 2).

A practical example

It is noteworthy that there is a common thread that connects the experiences mentioned above, identifiable in the GeoJSON format,[15] a format largely used to encode data for browser-based web applications. It is useful to compare it with another widely used format, the ESRI Shapefile, by a practical example of the analysis and management of an archaeological dataset, and to demonstrate the benefits of using open formats for the scientific research and dissemination. The aim was to concretely verify the advantages of using a GeoJSON file instead of an ESRI Shapefile through a simulation of a survey on restrictions in urban development planning, using data available online. The pipeline was the following:

[15] https://geojson.org/ (accessed 7/7/2021). GeoJSON has become a standard (RFC 7946) in 2014, although the specification was finalized in June 2008.

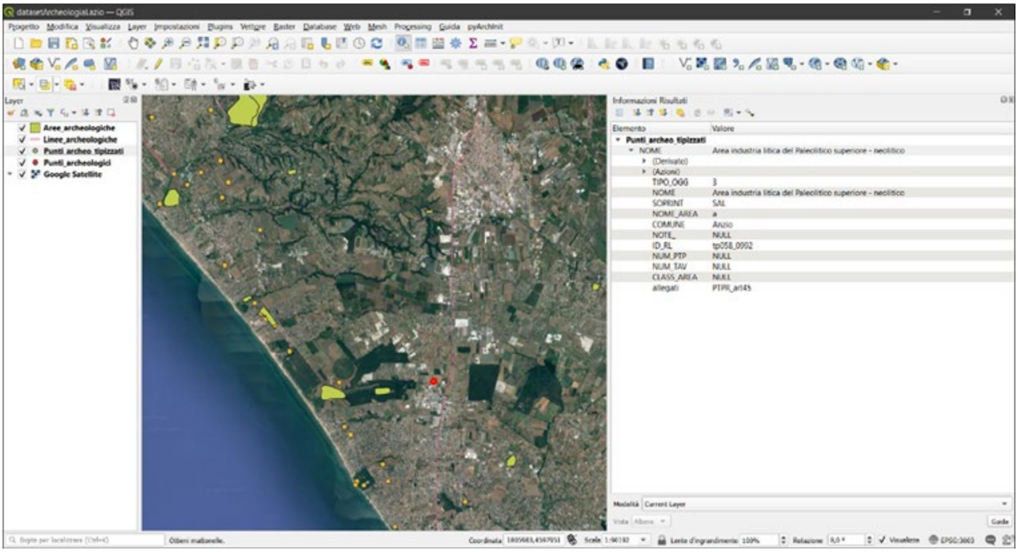


Figure 3: Different geometries and attribute tables of archaeological area in Lazio Region, shown in QGIS interface.

- Downloading a set in Shapefile format from a geoportal.
- Loading the dataset into QGIS.
- Exporting it in GeoJSON.
- Comparing the two datasets encoded in SHP and GeoJSON formats.

Uploading the GeoJSON file into a web and a desktop application and analyse if the whole process, from a non-open format to an open one, worked out.

The dataset for archaeological areas in the Lazio territory was downloaded by the Open Data portal of Lazio Region,[16] in the section Piano Territoriale Paesistico Regionale (PTPR), made of points, lines and polygons in shapefile format. The data were uploaded into the open-source GIS desktop application QGIS 3.10,[17] to visualize the geometries and attribute tables linked to each resource (Figure 3).

The next phase involved the conversion of the original Shapefiles in a standard and open format, i.e. GeoJSON, to simplify the management of data. The original data of the Shapefile were projected in ED50/33N (EPSG:23033) coordinate reference system, and were necessarily reprojected into WGS84 (EPSG:4326), according to the syntax of GeoJSON.[18] The subsequent comparison between the two data sets (Figure 4) has clearly shown:

[16] http://dati.lazio.it/catalog/it/dataset?tags=PTPR (accessed 7/7/2021).
[17] https://www.qgis.org (accessed 7/7/2021).
[18] An 'Internet Standards Track Document' made by the Internet Engineering Task Force (IETF) provides additional information about the CRS for the GeoJSON Format (https://datatracker.ietf.org/doc/html/rfc7946#section-4, accessed 7/7/2021).

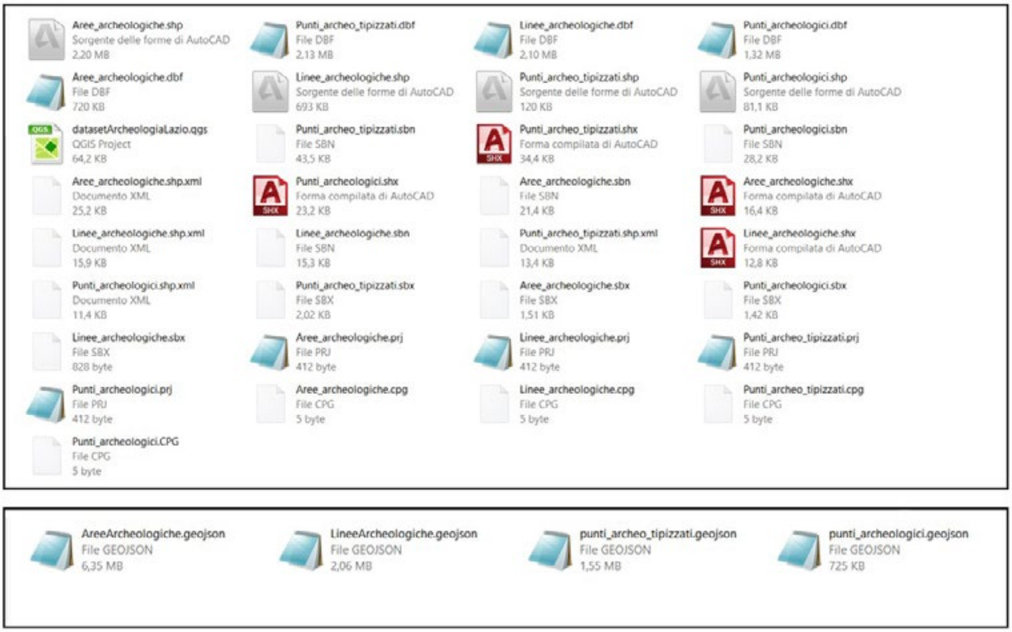


Figure 4: Comparison between the two datasets in shapefile (on top) and GeoJSON (down).

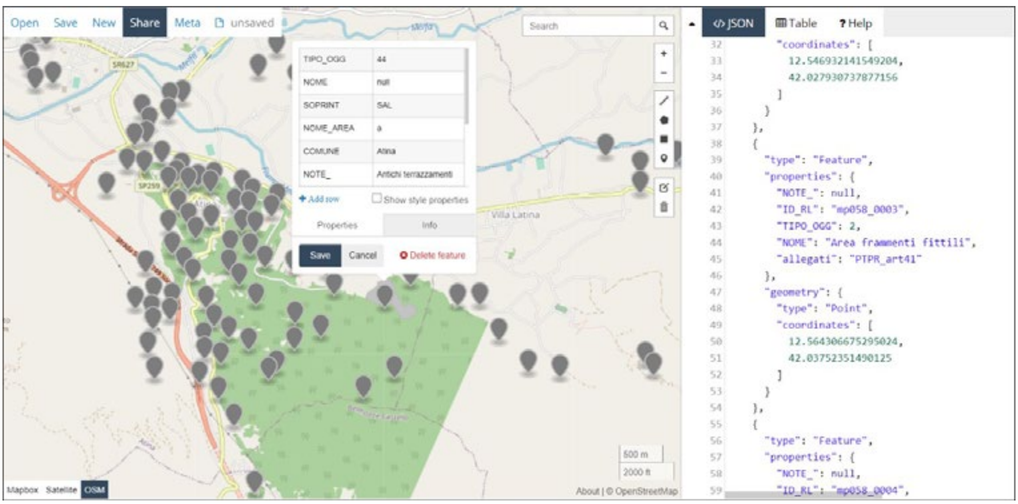


Figure 5: The dataset uploaded on geojson.io.

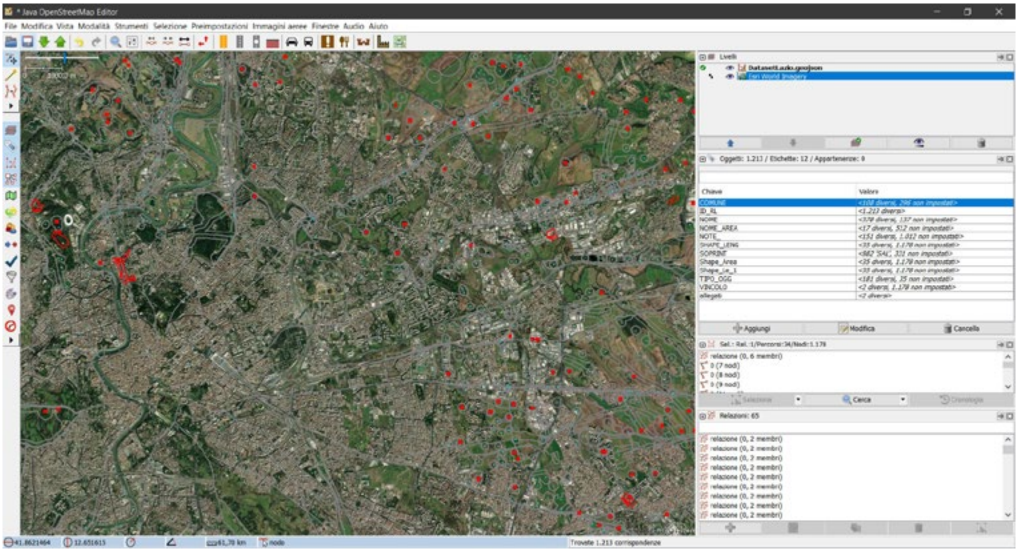


Figure 6: The GeoJSON file uploaded into OpenStreetMap Editor.

- Shapefile is a multifile format, it uses at least three files (*.shp, *.dbf, *.shx), making it impossible to share just one file; furthermore only one geometry type (point, line, polygon and others) is supported by a single file, and mixed geometry features are not possible. As a result, 33 files are needed to describe the four geometries of archaeological features of Regione Lazio. Users could zip all the files into one archive and unzip them on the other end of the distribution chain, but this is error-prone and limits interoperability. In addition, there is no way to describe topological relations in the format or to store 3D data with textures or parametric objects.
- GeoJSON data, exported through the specific QGIS function, has just four files to describe the four different geometries; the advantages of having the same data but with a smaller number of files is unarguable, even though less complexity makes it easier to store and exchange data, thus promoting interoperability. It can handle complex vector data features and build complex hierarchical data models. GeoJSON is recommended as a shapefile replacement for data interchange, particularly for web services, such as the ones used in this procedure.

To check the functionality of the tested datasets, it was necessary to deal with a single file containing all the geometries. To achieve this purpose, an open web tool was used for editing GeoJSON data on the internet: geojson.io.[19] It enables editing raw GeoJSON through a map interface, therefore it has been used to merge the four geometric GeoJSON files, thus combining the 7,363 resources; this stage confirmed that in the fusion process all the linked attributes have been preserved (Figure 5). It has also been integrated with GitHub,[20] allowing users to edit their data in Git repositories and GitHub Gists, making data accessible and downloadable.

[19] http://geojson.io/ (accessed 7/7/2021). This is an open-source software under the permissive MIT license. That license applies to the software and not the data created with geojson.io (http://geojson.io/about.html).
[20] https://github.com/mapbox/geojson.io (accessed 7/7/2021).

A useful tool to upload, visualize, edit, and query different types of data is the open Desktop Application *Java OpenStreetMap Editor*.[21] The merged GeoJSON file was uploaded to JOSM Editor to check that all the attributes and relations had been preserved without losing any part of the information. With this tool it has been possible to search any single information independently from the geometries[22] (Figure 6).

Conclusion

An approach such as the one illustrated in this paper, with the coding of spatial data structured through shared and common standards approved by the scientific community, fits perfectly into the theme of the production and consumption of BGD, an issue that is going to be faced soon within the scope of the implementation of web infrastructures able to integrate resources and open-data. According to a document approved by W3C and OGC in March 2021[23] on the best practices for the publication of spatial data on the Web, there are only two reference geometry formats widely used in the geospatial and Web communities: GML for 3D data and GeoJSON for 2D geometries. As GeoJSON is serialized in JSON, this format is ready for browser-based Web applications.

One of the challenges to face is to understand how to preserve the topological information of the spatial data in relation to web mapping; an extension called TopoJSON[24] has been developed to express the geospatial topology encoded in GeoJSON, but a more in-depth analysis is necessary to understands how it fits with GeoJSON format.

In this context, we focused on the advantages of using specific formats to encode 2D data for web applications; however, further investigation is needed on 3D spatial data: an in-depth study for its excellent prospects is the standard open format Geopackage, the data container based on SQLite, which allows users to manage vectors, rasters, attributes, topology, and, thanks to its extensions, even 3D data. This is the case of the 3D tiles[25] format, even if it is not yet an official extension to the GeoPackage Encoding Standard, but the future is promising: 3D Tiles is an open specification for sharing, visualizing, fusing, interacting with, and analysing massive heterogeneous 3D geospatial content across desktop, web, and mobile applications.

It is also worth highlighting the importance of formats such as GeoJSON and Geopackage, which, differently from Shapefile, are transparent, encouraging interoperability. GeoJSON allows the easy reuse of spatial data, undoubtedly opening new opportunities for integrating data within the scenario of the semantic web; Geopackage does the groundwork for the management of 3D data with spatial coherence.

[21] https://josm.openstreetmap.de/ (accessed 7/7/2021).
[22] Differently, in QGIS the single GeoJSON file, once uploaded in the software, was divided in three layers, corresponding to the three geometries described in the file.
[23] https://w3c.github.io/sdw/bp/ (accessed 7/7/2021).
[24] https://github.com/topojson/topojson-specification (accessed 7/7/2021).
[25] http://www.opengis.net/doc/CS/3DTiles/1.0 (accessed 7/7/2021).

Bibliography

Carlisle, P.K., I. Avramides, A. Dalgity and D. Myers 2014. The Arches Heritage Inventory and Management System: a standards-based approach to the management of cultural heritage information. In *Proceedings of the CIDOC Conference Access and Understanding – Networking in the Digital Era, Dresden, Germany, 6–11 September 2014*. (http://archesproject.org/wp-content/uploads/2014/10/I-1_Carlisle_Dalgity_et-al_paper.pdf, accessed 13/10/2021).

Cooper, A. and C. Green 2016. Embracing the Complexities of 'Big Data' in Archaeology: the Case of the English Landscape and Identities Project. *Journal of Archaeological Method and Theory* 23: 271–304 (https://doi.org/10.1007/s10816-015-9240-4).

Forte, F. 2019. Informazioni spaziali e strumenti d'indagine semantica sui dati archeologici digitali. *Newsletter di Archeologia CISA* 10: 95–111.

Gattiglia, G. 2015. Think big about data: Archaeology and the Big Data challenge. *Archäologische Informationen* 38: 113-124 (https://doi.org/10.11588/ai.2015.1.26155).

Hiebel, G., G. Goldenberg, C. Grutsch, K. Hanke and M. Staudt 2020. FAIR data for prehistoric mining archaeology. *International Journal on Digital Libraries* (https://doi.org/10.1007/s00799-020-00282-8).

McCoy, M.D. 2017. Geospatial Big Data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science* 84: 74–94.

Previtali, M. and R. Valente 2019. Archaeological documentation and data sharing: digital surveying and open data approach applied to archaeological fieldworks. *Virtual Archaeology Review* 10 (20): 17–27 (https://doi.org/10.4995/var.2019.10377).

Richards-Rissetto, K. and H. Landau 2019. Digitally Mediated Practices of Geospatial Archaeological Data: Transformation, Integration, & Interpretation. *Journal of Computer Applications in Archaeology* 2 (1): 120–135 (https://doi.org/10.5334/jcaa.30).

Open Data, Open Knowledge, Open Science: The new research group at the Institute of Heritage Science (CNR)

Alessandra Caravale, Alessandra Piergrossi, Irene Rossi

Alessandra Caravale, CNR-Istituto di Scienze del Patrimonio Culturale (ISPC); alessandra.caravale@cnr.it

Alessandra Piergrossi, CNR-Istituto di Scienze del Patrimonio Culturale (ISPC); alessandra.piergrossi@cnr.it

Irene Rossi, CNR-Istituto di Scienze del Patrimonio Culturale (ISPC); irene.rossi@cnr.it

Abstract

The Open Data, Open Knowledge, Open Science research group at the CNR-ISPC, including archaeologists, philologists, mathematicians, and computer scientists who share editorial work around the open-access journal Archeologia e Calcolatori, has addressed most of its activities to monitoring and studying the use of digital data collections, and open-access policies in the humanities field. The paper illustrates some recent works, with a focus on the adhesion of Archeologia e Calcolatori to the FAIR principles and on the Linked Data annotation of the journal's texts.

Keywords: OPEN SCIENCE; FAIR PRINCIPLES; ARCHAEOLOGICAL COMPUTING.

The CNR-ISPC research group and *Archeologia e Calcolatori*

An interdisciplinary group – including archaeologists, philologists, computer scientists and mathematicians – has gathered over time around the open-access international peer reviewed journal *Archeologia e Calcolatori* (*A&C*).[1] *A&C* was born in 1990 as an observatory of theoretical and methodological aspects of computing and information technology applied to archaeology. Until the early 2000s the journal represented the only regular update on archaeological computing, together with the annual publication of the Proceedings of the CAA Conferences.[2] Since 2005, *A&C* has joined the Open Archives Initiative (Barchesi 2005), and an electronic version of the journal is published under the Creative Commons License BY-NC-ND 4.0, along with the hard-copy publication (Caravale and Piergrossi 2012; Caravale 2020). At present, some other journals related to this research topic use the Web as their means of publication. Recent examples at the European level are *Virtual Archeology Review*,[3] promoted in Spain by the Universitat Politècnica de València since 2010; *Digital Applications in Archeology and Cultural Heritage*,[4] established by Bernard Frischer in 2014; *Archéologies numériques-Digital Archeology*,[5] included in the French ISTE OpenScience since 2017, and the *Journal of Computer*

[1] http://www.archcalc.cnr.it/ (accessed 14/07/2021).
[2] https://caa-international.org/ (accessed 14/07/2021).
[3] https://polipapers.upv.es/index.php/var (accessed 14/07/2021).
[4] https://www.journals.elsevier.com/digital-applications-in-archaeology-and-cultural-heritage (accessed 14/07/2021).
[5] http://www.openscience.fr/Archeologies-numeriques (accessed 14/07/2021).

Applications in Archeology,[6] promoted by the CAA Association in the same year (Moscati 2019b: 26).

The work of *A&C* has not been limited to the scientific and editorial activity on the articles to be published in the journal, but has had broader scopes, aiming at monitoring and studying the use of digital data collections and the open-access policies in the humanities field.[7] The journal has therefore played a double role: on the one hand, it has been an international aggregator of projects, ideas, and reflections about the application of information technology in archeology and human sciences; on the other hand, the editorial team has built over time a laboratory focused on open-data, open-science, and open-digital resources in the archaeological sector (Moscati 2019a).

This research group has recently found adequate space within the Institute of Heritage Science of the CNR, a new hub that brought together four CNR institutes conducting research on cultural heritage under different aspects, comprising humanities, historical processes, experimental sciences, and technological applications. Because of this interdisciplinarity, the Institute is currently compiling a registry of competences in Heritage Science using semantic taxonomy techniques, which will be based on the analysis of projects and publications in bibliometric as well as non-bibliometric sectors (Miliani *et al.* in print). Among its various objectives, the Institute has the mission of supporting innovation in the knowledge, management, conservation, and enhancement of cultural heritage. In line with the CNR policies, this vision of participatory science has been pursued by our Institute since its inception, as shown by the involvement in digital infrastructures linked to Heritage Science, such as E-RIHS[8] and Iperion HS.[9]

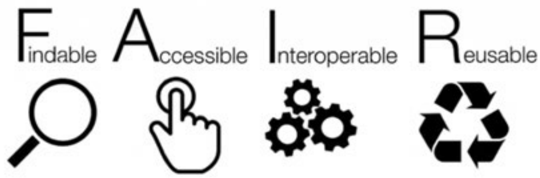
The case-study of *A&C* demonstrates how, for an electronic scientific journal, the adherence to the open access policy and the effort to promote it, are necessarily connected with a theoretical reflection aimed at exploring novel solutions to share scientific data, and with their practical experimentation. The theme is topical in this moment when Open Science is an international priority.

A.C.

FAIR principles

In 2016, a new set of principles was designed to define how to produce research data and what researchers should expect from contemporary data resources (Wilkinson *et al.* 2016). Findability, Accessibility, Interoperability and Reusability (FAIR, Figure 1) summarize and formalize the procedures to enable reuse of data by academies, research institutions and scholarly publishers.[10] Those principles had somehow already implicitly formed the base of all the Open Access and Open Science movements over the years. They are not prescriptions,

[6] https://journal.caa-international.org/ (accessed 14/07/2021).
[7] https://www.ispc.cnr.it/it_it/2020/05/14/gruppo-open-data/ (accessed 14/07/2021).
[8] http://www.e-rihs.eu/ (accessed 14/07/2021).
[9] http://www.iperionch.eu/ (accessed 14/07/2021).
[10] https://www.go-fair.org/fair-principles/ (accessed 14/07/2021).



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Figure 1: The FAIR guiding principles (Wilkinson *et al.* 2016).

but behavioral guidelines and best practices, which should improve the transparency and maximize the availability and reproducibility of research data and all related processing steps.

To make data easily findable and accessible, the first step is to provide every research object with appropriate metadata, possibly standardized and well-established within relevant disciplines: they provide a formal, shared, schematic way of representing knowledge through a common language (see, as a good starting point, the Metadata Standards Directory[11]). Persistent identifiers (PID), i.e. long-lasting links to the data such as DOIs, are another feature to improve findability in the long term. For access to be trusted over time, digital data, or 'digital objects', should be provided with a persistent identifier and a stable URL, so that they can be located even if their actual location on the Internet changes. PIDs can identify many different entities, from natively digital objects (documents, data, software) to physical objects (people, samples), to conceptual entities (organizations, projects). Storing them in disciplinary discovery portals or institutional repositories is another fundamental operation to ensure good data management and retrieval.

[11] https://rd-alliance.github.io/metadata-directory (accessed 14/07/2021).

Interoperability means that data sets, as well as software, can be integrated with other data and software; they can be readable by machines and verifiable using common standards and open-data formats.

Reusability implies both the creation of a clear documentation, which provides context and makes data understandable to others, and the definition of who can reuse the research, defining the licenses that state the conditions under which data (and their related metadata) can be reused.

The idea that scientific data should be FAIR is increasingly endorsed by scientific institutions. As advocated by the European Commission[12] and recommended by UNESCO,[13] these principles now play a decisive role in outlining strategies and valuable tools for the research community. The digital infrastructures that give access to scientific resources and store data promote and support the adoption of FAIR principles, as their remit is to create new opportunities and enhance exchange between disciplines and researchers. Leading by example, the European Commission has introduced the obligation of FAIR Data Management Plans (DMPs) in the context of Horizon 2020, the EU Framework Programme for research and innovation in the period 2014–2020 and over. Horizon Europe, the successor of Horizon 2020, is being conceived to support Open Science at all levels.[14]

At the same time, researchers are encouraged to share their data openly, but as much as FAIR data does not necessarily mean open data, obviously openness helps the outcome of fairness: 'as open as possible, as closed as necessary' state the H2020 Program Guidelines on FAIR Data Management (European Commission. Directorate-General for Research & Innovation. Version 3.0. 26 July 2016). 'FAIR only speaks to the need to describe a process for accessing discovered data; a requirement to openly and richly describe the context within which those data were generated and to enable evaluation of its utility ... They do, however, require clarity and transparency around the conditions governing access and reuse.' (Mons *et al.* 2017: 51–52).

The challenge nowadays, especially in the field of Humanities and Cultural Heritage (and even more so in Archaeology where some research processes, such as excavations, are impossible to repeat, given the variety of practices, demands and dataset available), is to refine and expand their applicability, by defining controlled vocabularies and ontologies, suitable common quality data format and metadata standards, as solicited by the infrastructures supporting the dissemination and storage of scholarly data (Piergrossi 2020).

Obviously, each discipline requires different approaches and strategies according to the different types and use of data. A consequence of this heterogeneity is the need to find connections across multiple datasets and to agree on publishing policies, shared terminologies and metadata specifications that are central to interoperability and reuse.

[12] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 14/07/2021).
[13] https://unesdoc.unesco.org/ark:/48223/pf0000374409.page=10; https://www.openaire.eu/openaire-input-on-unesco-s-open-science-consultation (accessed 14/07/2021).
[14] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf (accessed 14/07/2021); https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf (accessed 14/07/2021); see also Schouppe and Burgelman 2018.

What is in it for the researchers? There is a long list of advantages that can be summarized in better research output; more value to publicly funded research; improvement of the peer review process, but above all, transparency, high integrity, and high quality of the research data.

Similarly, 'FAIR' knowledge and good data management are fostered by high quality digital publications that can facilitate the ongoing process of discovery, evaluation, and reuse. *A&C* is an example of such best practices as its focal points are full open-access, high standards guaranteed by peer review processes and ethical research rigor.

A.P.

Archeologia e Calcolatori and the FAIR principles

The commitment to Open Science and the early compliance with the FAIR principles also resulted in specific operational choices regarding the management and dissemination of *A&C*'s bibliographic resources (Piergrossi and Rossi 2019). The journal has been regularly updating the indexes of its publications on its website, where resources can be found both by consulting the chronological index of the volumes and through a dynamic search functionality.[15] Precisely to facilitate the retrieval and sharing of the contents in the ever-increasing number of bibliographic resources on the Web, the journal adopted the Digital Object Identifiers (DOI) system. Conceived in the 1990s in the sector of digital publishing, DOIs are codes that allow users to identify univocally and persistently the digital objects, and to easily find them on the Web, thanks to the localization of the services associated with the content.[16] Their use is proving more and more effective in scholarly publishing inspired by the principles of open-access. The registration of DOIs for *A&C*'s resources started in 2016, following the signature of an agreement between the CNR and the European Agency for DOI registration.[17]

Even before the practice of registering DOIs spread, *A&C*'s resources were provided with Uniform Resource Identifiers (URI) in the repository of the scientific open-access literature SOLAR, created in 2006 at the CNR Central Library.[18] This early choice of depositing metadata and full texts of the articles in SOLAR was crucial, not only as regards the findability principle, but also for accessibility reasons. To comply with this second FAIR principle, the journal's website had also been configured as a place for the archival and dissemination of the publication's contents. The full texts of the open-access articles, for a total of over 15,000 pages, are available in PDF format and can be downloaded via a link in the webpage describing each resource. However, given the volatility of the digital product and the ever-increasing mobility of the Web, ensuring stable access to digital resources is a concrete issue, and long-term repositories – like Zenodo[19] – have been largely promoted in recent years by national and supranational policies. SOLAR was one of the first initiatives in this direction.

[15] http://www.archcalc.cnr.it/ (accessed 14/07/2021).
[16] https://www.doi.org/ (accessed 14/07/2021).
[17] https://www.medra.org/ (accessed 14/07/2021).
[18] http://eprints.bice.rm.cnr.it/ (accessed 14/07/2021).
[19] https://zenodo.org/ (accessed 14/07/2021).

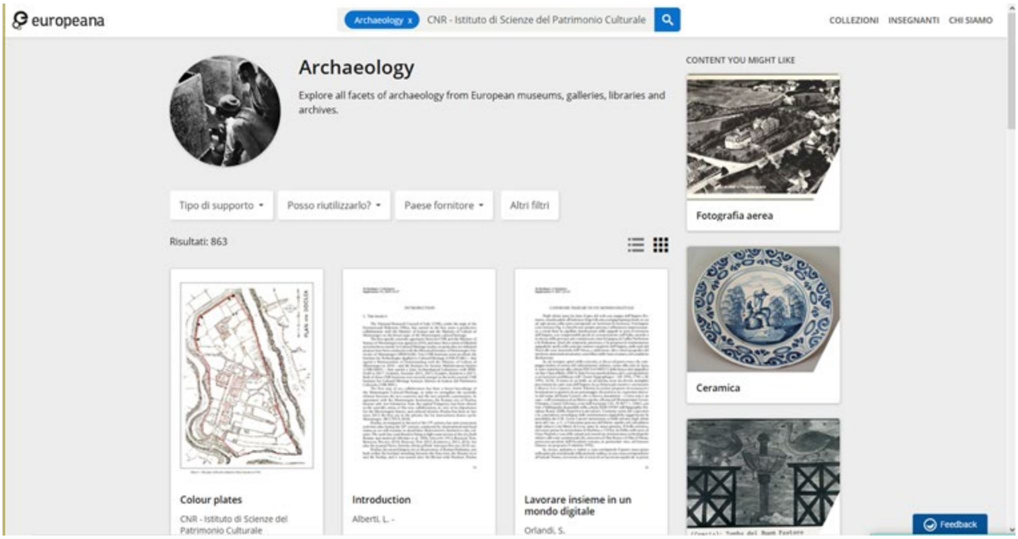


Figure 2: *A&C* in Europeana (https://www.europeana.eu/it).

A key to each of the abovementioned advancements towards findability and accessibility of *A&C*'s resources was the provision of their descriptive metadata. These have been recorded in the database of the journal, which has been implemented over two decades both in content and in technology, pursuing the path of Open Science. *A&C*'s database currently stores the metadata of more than 1,000 publications, structured according to a simple model. This is based on three main entities, *Article*, *Volume*, and *Author*, and implements the Simple Dublin Core as the basic set for its attributes.[20] Their relations return the complexity of the bibliographical resource's description, which is refined by further *ad hoc* attributes that are useful for the study and classification of the articles' contents conducted by our group.

The choice of complying with the standards for metadata creation paralleled the early adhesion of the journal to the Open Archives Initiative (OAI)[21] since 2005 and the provision of *A&C*'s records according to the *oai_dc* metadata format in an OAI-PMH repository, to enable data interoperability.[22] The repository has been recently renewed, with the inclusion of a further metadata format, the *qdc* (Qualified Dublin Core). Moreover, for the integration with the Open Access Infrastructure for Research in Europe,[23] in 2020 the data exposed in the *oai_dc* format were aligned with the OpenAIRE Guidelines for Literature Repositories, version 3,[24] and domain relevant attributes, such as those regarding access rights, identifiers and resource type, were referred to standard vocabularies. As a result, OpenAIRE now harvests more than 1,000 of *A&C*'s resources, providing further services on them.[25]

[20] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (accessed 14/07/2021).
[21] http://www.openarchives.org/ (accessed 14/07/2021).
[22] http://www.archcalc.cnr.it/oai/aec_oaipmh2.php (accessed 14/07/2021).
[23] https://www.openaire.eu/ (accessed 14/07/2021).
[24] https://guidelines.openaire.eu/en/latest/literature/index_guidelines-lit_v3.html (accessed 14/07/2021).
[25] https://explore.openaire.eu/search/dataprovider?datasourceId=openaire____::f3a35cc23d629e74e8d5fe0d8bca1684 (accessed 14/07/2021).

To keep pace with the evolution of best practices and guidelines for the provision and reuse of research resources and relevant metadata, *A&C*'s repository is being constantly enriched with further digital contents, such as metadata records of images and 3D models published in the journal articles, and metadata publication formats, such as *oai_datacite* and *oai_openaire* (Rossi and Paraciani 2021).

This effort towards standardization and interoperability has eased the provision of *A&C*'s bibliographic metadata to several initiatives aggregating cultural and scientific digital content. Thanks to an agreement between the then Istituto di Studi sul Mediterraneo Antico, now converged into the Institute of Heritage Science, and the Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche (ICCU), *A&C* was the first scientific journal to provide its set of data to the Portal of the Italian Culture,[26] in 2017. The scale of provision of content to aggregators was broadened to the European level in 2020: thanks to a renewed collaboration with the ICCU, the journal has contributed more than 900 resources to the Europeana digital library,[27] in the frame of the Europeana Archaeology project (Figure 2).

All those portals, including OpenAIRE, display the metadata of the records but also provide access to the PDF files of the publications stored in the journal's archive, thus supporting in their turn the dissemination and accessibility of *A&C*'s resources, and boosting opportunities for data reuse. For instance, providing to OpenAIRE links to the PDF file for each article allows the portal to retrieve the full texts of the journal, which are further processed for analyses through text mining tools, in order to extract bibliographic references and project information, compute subject classifications, etc.

I.R.

Linked Open Data

Aiming at sharing and reusing our scientific resources online, we also tested the application of Web mapping strategies on the journal texts, according to the criteria offered by the open-source tool Recogito.[28] This is a free online annotation tool and publishing platform developed as part of the Pelagios 3 research project, used mainly for mapping toponyms mentioned in ancient literature. Starting from digitized documents and basing on some specific gazetteers, Recogito allows users to identify the geographical items cited in the texts and to place them on a map.

To carry on the work on the articles of *A&C*, it was decided to experiment with the mapping according to two different criteria: chronological and thematic. The former considered the articles published from 2014 to 2018, while the latter examined a selection of texts published since the first volume of the journal in 1990, where Etruscan toponyms are mentioned (Figure 3). As far as it concerns the 2014–2018 articles, the text mapping shows a wide distribution of the mentioned toponyms; the sites are found throughout the Mediterranean area and beyond. Sites in the Italian peninsula are certainly predominant, though many are also

[26] http://www.culturaitalia.it (accessed 14/07/2021).
[27] https://www.europeana.eu/ (accessed 14/07/2021).
[28] https://recogito.pelagios.org/ (accessed 14/07/2021).

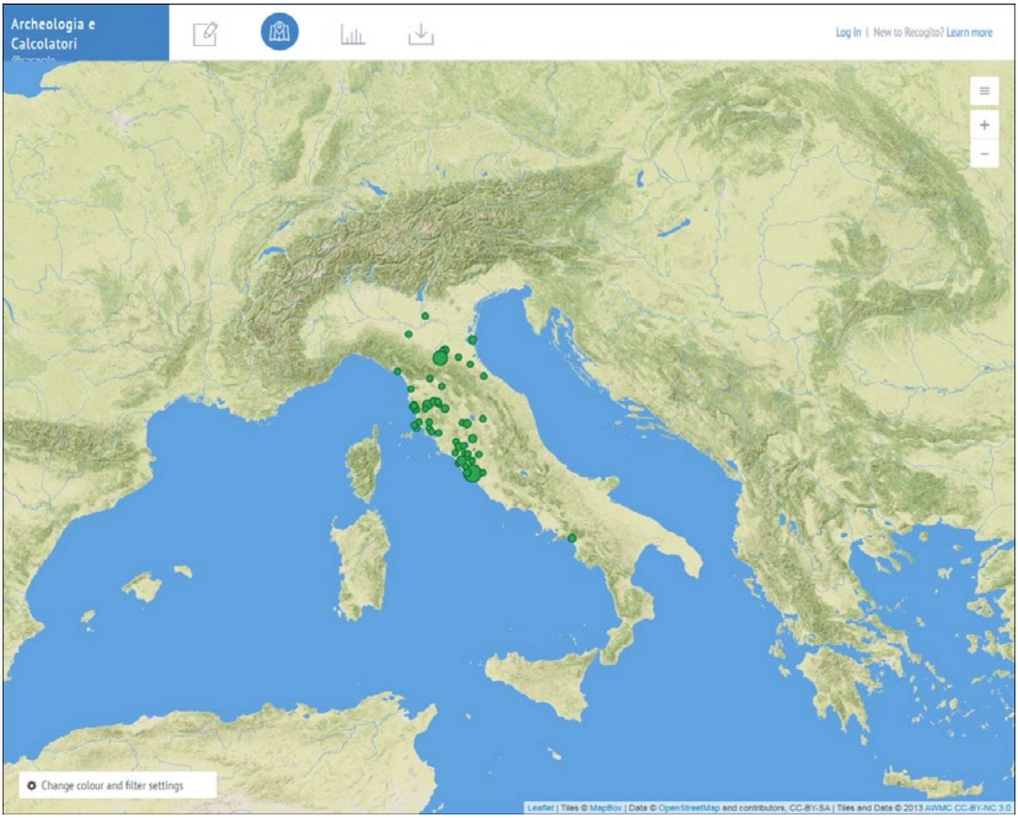


Figure 3: Map showing *A&C* Etruscan sites (Cantone and Caravale 2019).

located in the two major islands of Sicily and Sardinia: in fact, in recent years, numerous archaeological initiatives and projects applying information technologies have been carried out in these regions, and the results of many of them were published in *A&C*. The Near East is also represented with numerous toponyms; in Western Europe, Spanish sites prevail.

Regarding the thematic mapping dedicated to mentions of Etruscan sites, the most attested toponyms are Cerveteri (with over 120 mentions) and Marzabotto (with about 80). Their prevalence is not accidental, because both sites have been subject of particular attention by our journal: the first one, especially in the past years, thanks to the innovative research carried out by our institute as part of the Caere Project; the second one, in more recent times thanks to investigations related to the Kainua Project, aimed at reconstructing the architectural structures of Marzabotto through the most innovative digital technologies.

The theme of linked open data is an important topic in the Italian and international debate related to the dissemination of cultural heritage through the Web. Projects and initiatives increase at the national level (for example those promoted by the ICCU with CulturaItalia and the ICCD with the data.beniculturali.it platform), and at the international level (such as the work of the Getty Research Institute linked to Vocabularies for art, architecture, archival materials, conservation, and bibliographic materials).

Through the mapping made with Recogito, the toponyms mentioned in *A&C* became linked open data and they have been 'connected' with other geographical data through their inclusion within Pelagios Network, a European initiative that links information online through common references to places. So today, if you enter in Pelagios through its geographical search engine Peripleo and search for example Orvieto, you will see a list of some digital repertoires in which the site is mentioned and among these also all the citations of this Etruscan city made in the *A&C* volumes.

A.C.

Conclusions

Based on the theoretical and practical experience of the research group gathered around *A&C*, the main target of the CNR-ISPC research group on open data is the experimentation and promotion of protocols and strategies useful to the enhancement of open and shared scientific resources, mainly publications and datasets, with a focus on the area of the Heritage Science. This is in line with the policies of our Institution, which, in 2012, signed the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities and supports the European Chart for the free circulation of knowledge, and it is consistent with the CNR's latest initiatives aimed at the promotion and dissemination of scientific knowledge, such as the implementation of the CNR ExploRA portal,[29] building on the CNR institutional repository of research products and exposing the metadata of these resources, with an emphasis on open-access publications.

A.C., A.P., I.R.

Acknowledgements

The authors wish to express their gratitude to Paola Moscati and Nicolò Paraciani for their valuable comments on this paper.

Bibliography

Barchesi, C. 2005. Archeologia e Calcolatori: nuove strategie per la diffusione di contenuti in rete sulla base dell'OAI-PMH. *Archeologia e Calcolatori* 16: 225–241. http://www.archcalc.cnr.it/indice/PDF16/Barchesi_16_2005.pdf (accessed 14/07/2021).

Cantone, F. and A. Caravale 2019., Archeologia e Calcolatori. Classificazione geografica e tematica per la condivisione della conoscenza.' In P. Moscati (ed.) *30 anni di Archeologia e Calcolatori. Tra memoria e progettualità*: 93–107. *Archeologia e Calcolatori* 30. https://doi.org/10.19282/ac.30.2019.07.

Caravale, A. 2020. Principi Fair ed editoria elettronica. L'archeologia open di Archeologia e Calcolatori. In F.R. Cerami, M.L. Scaduto and A. De Tommasi (eds) *I bacini culturali e la progettazione sociale orientata all'heritage-making, tra politiche giovanili, innovazione sociale, diversità culturale. Il framework del Progetto ABACUS - Attivazione dei Bacini Culturali Siciliani, alla luce della Convenzione Quadro del Consiglio d'Europa sul valore del Patrimonio culturale per la*

[29] https://publications.cnr.it/ (accessed 14/07/2021).

società: 209–220. Firenze: Edizioni All'Insegna del Giglio. https://www.insegnadelgiglio.it/prodotto/i-bacini-culturali/ (accessed 14/07/2021).

Caravale, A. and A. Piergrossi 2012. Archeologia in rete. Le riviste open access: risorse e prospettive. *Archeologia e Calcolatori* 23: 187–207. http://www.archcalc.cnr.it/indice/PDF23/11_Caravale_Piergrossi.pdf (accessed 14/07/2021).

Miliani C., C. Sfameni, P. Moscati and A. Caravale (in print). La sfida delle competenze per il patrimonio culturale: complementarità, integrazione, interazione (ISPC). In *Relazione sulle Scienze Umane e Sociali – CNR-DSU*. Roma: CNR.

Mons, B., C. Neylon, J. Velterop, M. Dumontier, L. Olavo Bonino da Silva Santos and M.D. Wilkinson 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* 37: 49–56. https://doi.org/10.3233/ISU-170824.

Moscati, P. (ed.) 2019a. *30 anni di Archeologia e Calcolatori. Tra memoria e progettualità*. *Archeologia e Calcolatori* 30: 9–138. http://www.archcalc.cnr.it/journal/idyear.php?IDyear=2019-01-01.

Moscati, P. 2019b. Informatica archeologica e archeologia digitale. Le risposte dalla rete. In P. Moscati (ed.) *30 anni di Archeologia e Calcolatori. Tra memoria e progettualità*. *Archeologia e Calcolatori*: 21–38. *Archeologia e Calcolatori* 30. https://doi.org/10.19282/ac.30.2019.03.

Piergrossi, A. 2020. What is F.A.I.R.? *Archeologia e Calcolatori* 31.1: 253–256. https://doi.org/10.19282/ac.31.1.2020.12.

Piergrossi, A. and I. Rossi 2019., Archeologia e Calcolatori. Accessibilità e diffusione della cultura scientifica.' In P. Moscati (ed.) *30 anni di Archeologia e Calcolatori. Tra memoria e progettualità*. *Archeologia e Calcolatori*: 75–92. *Archeologia e Calcolatori* 30. https://doi.org/10.19282/ac.30.2019.06.

Rossi, I. and N. Paraciani 2021. IT applications to archaeology and the OA diamond journals' challenge. Enhancing access and reuse of textual and visual resources. *Archeologia e Calcolatori* 32.1.

Schouppe, M. and J.-C. Burgelman 2018. Relevance of the EOSC initiative and FAIR principles in the realm of Open Science and implementation phases of the EOSC. In *Data Analytics and Management in Data Intensive Domains. Proceedings of the XX International Conference (Moscow, Russia, 2018)*: 1–4. http://ceur-ws.org/Vol-2277/paper01.pdf (accessed 14/07/2021).

Wilkinson, MD., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18.

FOSS, Open-Data e archeologia: qualche riflessione su passato, presente e prospettive future

Marco Ciurcina, Piergiovanna Grossi

Marco Ciurcina, Studio Legale Ciurcina; ciurcina@studiolegale.it
Piergiovanna Grossi, Università degli Studi di Verona; piergiovanna.grossi@univr.it
Correspondence: piergiovanna.grossi@univr.it

Abstract
Starting from two concepts, common good and digital law, the paper presents some remarks on recent evolution of regulations and technologies that contributed to the process of data opening in the domain of Cultural Heritage. An overview on what is happening in the (rest of) the world is proposed, with reference to some of the most important archaeological open-data projects. After a brief introduction about Italian laws concerning data related to Cultural Heritage, the main projects that have been presented at previous ArcheoFOSS editions are discussed.

Keywords: ARCHAEOLOGICAL OPEN-DATA; ARCHAEOLOGICAL DATABASE; CULTURAL HERITAGE; COMMON GOOD; DIGITAL LAW; FARO CONVENTION.

Prima degli Open Data: il bene comune, il diritto digitale e l'*information technology*

Affrontare il tema degli *open-data* nel 2021 è come discutere della stampa nel XVII secolo: si è giunti a un punto in cui termini e concetti sono ormai noti a tutti, ma in cui ancora non sono consolidate le norme che regolano i diritti di autori ed editori e sono in rapido cambiamento anche le tecnologie che ne consentono la divulgazione massiva e popolare.

L'obiettivo di questo contributo non è essere esaustivi né suggerire considerazioni definitive in un ambito in rapida evoluzione, bensì quello di proporre una breve riflessione sulla situazione attuale, concentrandoci principalmente su due concetti:

- quello di 'bene comune', presente in tutte le epoche e in tutte le civiltà: senza bene comune non è possibile costruire una società e non è possibile il progresso. Il bene comune è un concetto che ha sempre fatto parte della storia umana: basti pensare alle rappresentazioni pittoriche della Preistoria, alla tradizione orale dei 'viaggi di ritorno' di cui fa parte l'Odissea, all'invenzione della stampa, alla nascita del Web;
- quello di 'diritto digitale', che ormai è parte della quotidianità di tutte le persone, indistintamente dal luogo e dal ceto sociale di provenienza, dal livello culturale e dalla conoscenza del diritto digitale.

Il quesito cui vorremmo rispondere è quale sia il collegamento tra questi concetti e quello di 'bene culturale', con il suo sottoinsieme 'bene archeologico'.

Per quanto riguarda il 'diritto digitale', la risposta è scontata. È indubbio, infatti, che dati e contenuti siano ormai trattati quasi esclusivamente in maniera digitale e che la conoscenza dei beni – culturali ed archeologici – non possa più prescindere dall'*Information Technology* (IT). Pertanto, tutto ciò che si interfaccia con l'IT ricade nell'ambito del diritto digitale, compresi appunto i dati e i contenuti relativi ai beni culturali (e archeologici).

Il concetto di 'bene comune' ben si sposa con quello di 'bene culturale, meglio ancora con la definizione anglosassone di 'Cultural Heritage': senza una eredità culturale in grado di accomunare un gruppo di individui non è possibile costruire una civiltà, una società e non è possibile pensare né alla conservazione della stessa né al suo progresso. È del resto, in sintesi, anche quanto sancito dalla Convenzione di Faro, avviata nel 2005 e ratificata da 21 paesi, tra cui l'Italia.[1] Il documento mira a stabilire il valore del patrimonio culturale per la società; tutto il testo, nella sua complessità, rappresenta un'unica forte dichiarazione del legame tra comunità e bene culturale. È significativo citare, a titolo di esempio, l'art. 1, in cui si riconosce 'che il diritto al patrimonio culturale è inerente al diritto a partecipare alla vita culturale, così come definito nella Dichiarazione universale dei diritti dell'uomo', si ribadisce 'una responsabilità individuale e collettiva nei confronti del patrimonio culturale' e si sottolinea, infine, 'che la conservazione del patrimonio culturale, ed il suo uso sostenibile, hanno come obiettivo lo sviluppo umano e la qualità della vita'. L'art. 14 del medesimo documento si ricollega invece al primo concetto qui introdotto, ovvero il diritto digitale, riconoscendo l'importanza dell'*Information Technology* per i beni culturali come strumento efficace in termini di accesso alle informazioni e incoraggiando gli stati membri a supportare standard compatibili e interoperabili.

È dunque importante iniziare il ragionamento partendo dal legame tra l'I*nformation Technology* e il diritto digitale e dalla relazione inscindibile che intercorre tra quest'ultimo, i dati aperti (o *open-data*) e i beni culturali. Come noto, l'informatica degli esordi era caratterizzata dall'utilizzo specialistico di potenti elaboratori concentrati presso poche istituzioni, mentre il primo cambiamento epocale si verifica nel momento in cui il personal computer rende hardware e software più accessibili alle masse. In questo contesto si sviluppano due modelli di software: quello 'proprietario' e quello 'libero', basato quest'ultimo su licenze libere nate negli anni Ottanta del secolo scorso per salvaguardare la conoscenza del codice sorgente (Raymond 2001; Stallman e Gay 2009; Graham 2004). È questo il momento in cui leggi e norme, nate per tutelare gli autori di produzioni intellettuali diffuse su supporti esclusivamente fisici, iniziano a diventare inadeguate.

Tra la fine degli anni Novanta e gli inizi del nuovo millennio si assiste al passaggio massivo agli smartphone. Il software libero è il più utilizzato per i sistemi operativi, ma in ambienti chiusi. Si pensi ad Android, progetto parzialmente *open-source*, o a iOS, parzialmente basato su *kernel* FreeBSD. Entrambi, tuttavia, promuovono l'uso esclusivo di specifici *market store* per le applicazioni, inoltre la licenza iOS è proprietaria. È il periodo in cui inizia il dibattito sull'accessibilità e sulla distribuzione dei dati e in cui sono attive associazioni per la cultura libera, come ad esempio *Open Knowledge Foundation*[2] e *Electronic Frontier Foundation*.[3]

[1] https://www.coe.int/en/web/culture-and-heritage/faro-convention (accesso 23/07/2021).
[2] https://okfn.org/ (accesso 29/06/2021).
[3] https://www.eff.org/ (accesso 29/06/2021).

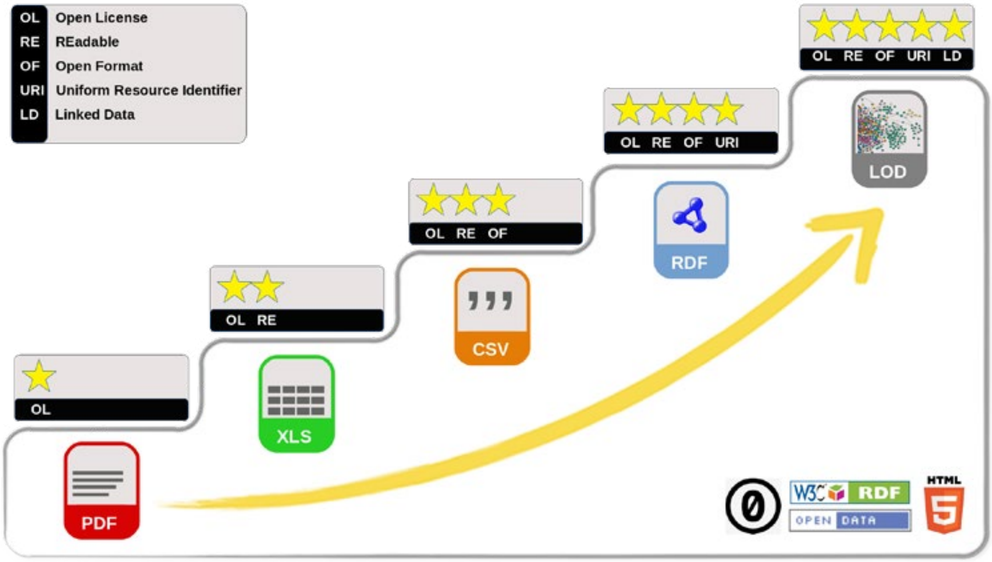


Figura 1: Schema 5 stelle dei formati Open Data. Tratta da: https://5stardata.info/en/ (CC0-Public Domain).

Nonostante l'origine di queste tecnologie risalga ben più indietro nel tempo,[4] è soltanto dopo il 2000 che inizia a farsi strada la nuova prassi del *cloud storage* e *cloud computing*,[5] con i pro (la scalabilità, la delocalizzazione, l'efficienza, l'economia) e i contro (controllo totale da parte del gestore, i rischi per la *privacy*, l'ambiguità sulla proprietà di dati e applicazioni, l'oligopolio dei servizi. Cfr. Stallman 2020; Zuboff 2019).

In conclusione, appare evidente come l'*Information Technology* sia un settore in rapida e continua evoluzione e pertanto all'avanzamento tecnologico deve necessariamente affiancarsi un aggiornamento del diritto digitale a esso connesso.

Fare *open-data*: la situazione attuale e le prospettive future

Il valore del software libero è oggi riconosciuto a livello internazionale, come testimonia il fatto che i principali operatori dell'industria informatica promuovono progetti di software open-source[6]. L'attenzione è attualmente concentrata sull'accessibilità ai dati: formati, licenze, gestione dei dati, servizi che li ospitano ed espongono e relativi termini d'uso. In particolare, le pubbliche amministrazioni e gli enti di ricerca sono focalizzati sugli aspetti, tecnologici e legali, che consentono non solo la conservazione e l'accessibilità ma anche il riuso, senza il quale nella maggior parte dei casi i dati risultano di nessuna utilità. Da questa necessità nascono il concetto *open-data* e i movimenti che lo promuovono.

[4] Si pensi ad ARPANET, la rete di computer istituita negli anni '60 per le comunicazioni e il trasferimento dati interno al ministero della difesa degli USA (Wikipedia, The Free Encyclopedia, s.v. 'ARPANET' 2021).
[5] Solo a titolo di esempio menzioniamo Amazon Web Service, o AWS (https://aws.amazon.com/, accesso 29/06/2021) e Microsoft Azure (https://azure.microsoft.com/, accesso 29/06/2021).
[6] https://opensource.google/ (accesso 29/06/2021), https://opensource.microsoft.com/ (accesso 29/06/2021).

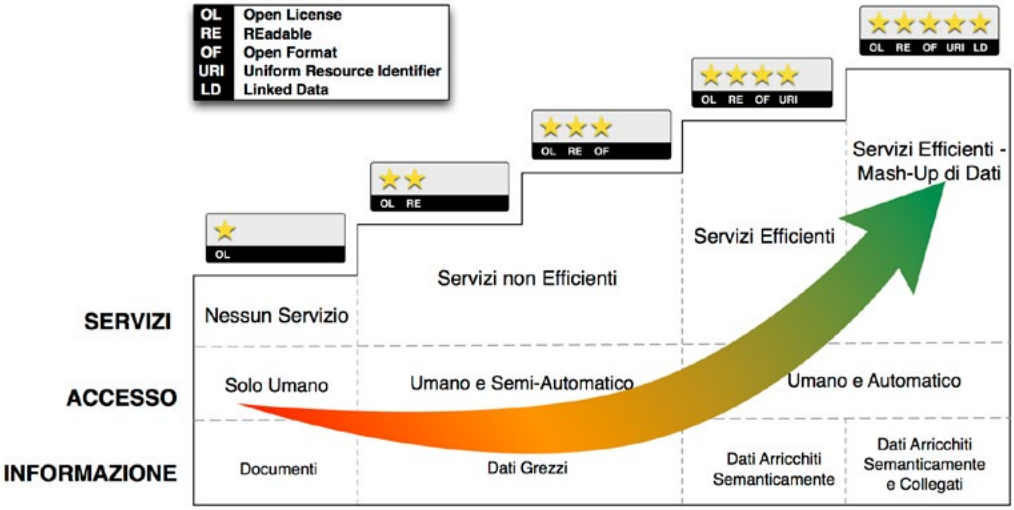


Figura 2: Schema 5 stelle degli Open Data. Tratta da: https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/modellodati.html (accesso 13/10/2021).

Fare *open-data* significa rendere disponibili *set* di dati non solo per la consultazione, ma anche per il riuso, tramite licenze e formati adeguati, che soddisfino i requisiti stabiliti dall'*Open Definition*.[7] È proprio Tim Berners Lee, l'inventore del Web e l'ideatore dei Linked Data (*Wikipedia, The Free Encyclopedia*, s.v. 'Tim Berners Lee' 2021), a fornire una definizione di Linked Data e del diverso grado di apertura degli open-data, attraverso una scala da 1 a 5 stelle:[8]

- 1 stella: dati con licenza libera, ma in formato non processabile;
- 2 stelle: dati con licenza libera in formato processabile;
- 3 stelle: dati con licenza libera in formato processabile e formato aperto;
- 4 stelle: dati con licenza libera con un URI che li identifica univocamente sul Web;
- 5 stelle: dati con licenza libera collegati ad altre informazioni esterne.

Per esempio, un PDF non è un dato processabile e quindi è il livello minimo di *open-data* (Figura 1).

Ma pubblicare *open-data* significa anche utilizzare servizi che garantiscano l'autonomia, la libertà, la conservazione, la divulgazione, di cui si possa controllare la locazione fisica, che utilizzino codice a sorgente aperto e che propongano licenze d'uso rispettose degli utenti (Figura 2).

Le problematiche connesse all'esposizione dei dati *on line* sono dunque numerose. Particolarmente rilevante è ad esempio il rischio, sempre più concreto, che si creino oligopoli nella fornitura di servizi in questo ambito. Il controllo dei dati non va invece sottovalutato, perché implica un potere strategico.

[7] https://opendefinition.org/ (accesso 29/06/2021).
[8] https://5stardata.info/ (accesso 22/07/2021).

I governi hanno, e in futuro ne avranno sempre di più, un ruolo fondamentale nel garantire la libertà e l'autonomia nella gestione dei dati. L'UE sta svolgendo una funzione chiave nell'azione *antitrust* contro i monopoli e nella promozione di politiche relative ai servizi di *cloud* per la Pubblica Amministrazione (cfr. ad esempio l'azione antitrust avviata a settembre 2020 dall'Italia, Reuters Staff 2020).

Anche le comunità non governative e indipendenti potrebbero svolgere un ruolo importante nei prossimi anni, integrando le rispettive attività in modo cooperativo. Solo a titolo di esempio e senza sminuire altri progetti altrettanto importanti, si ricordano *Wikimedia Foundation*[9] e i tanti *chapter* nazionali, come Wikimedia Italia,[10] che già promuovono e sostengono il progetto satellita *OpenStreetMap*.[11] La strategia vincente è sicuramente quella di attuare un piano di forte difesa di standard pubblici, formati e licenze aperte, che sono garanzia di preservazione dei dati nel lungo termine.

Uno sguardo sul (resto del) mondo: dati aperti e archeologia

Nel contesto dei beni culturali ed archeologici la situazione non è molto diversa da quella generale appena descritta. Anche in questo settore il software libero è infatti diventato uno strumento di lavoro a pieno titolo ormai da anni.

Per valutare la situazione sul fronte dei dati aperti, è utile partire da una panoramica a livello internazionale, per verificare quali sono le principali tendenze che si stanno sviluppando. È indubbio che a livello internazionale la necessità di condivisione dei dati e il senso di comunità sono sicuramente molto sentiti.

Si propone di seguito la disamina di alcuni progetti di natura, caratteristiche e provenienze geografiche differenti, ma che condividono la volontà di esporre open-data. Il criterio di valutazione è quello delle 5 stelle già richiamato:

- iDAI.world[12] è un progetto dell'Istituto Archeologico Germanico che espone dati eterogenei, tra cui archivi e biblioteche. I dati del dizionario geografico[13] sono esposti con licenza libera (CC-BY), in formati processabili (RDF, GeoJSON, ecc.); i dati del vocabolario archeologico[14] sono esposti come *linked open-data*, con URI, in formati processabili e con interrogazioni in SPARQL;[15] la licenza non è precisata per questo vocabolario. Seppure non tutte le tipologie di dati possono essere esposte in questi formati (ad esempio i volumi e articoli PDF) e con licenza libera (ad esempio alcuni dati e contenuti collezionati da altre fonti che conservano la licenza d'origine) si può definire un sito che pubblica, ove possibile, open-data 5 stelle.

[9] https://wikimediafoundation.org/ (accesso 29/06/2021).
[10] https://www.wikimedia.it/ (accesso 29/06/2021).
[11] https://www.openstreetmap.org/ (accesso 29/06/2021).
[12] https://idai.world/ (accesso 20/04/2021).
[13] https://gazetteer.dainst.org/ (accesso 20/04/2021).
[14] https://archwort.dainst.org/ (accesso 20/04/2021).
[15] SPARQL Query Language for RDF. W3C Recommendation https://www.w3.org/TR/rdf-sparql-query/ (accesso 22/07/2021); SPARQL 1.1 Overview, W3C Recommendation 21 March 2013 https://www.w3.org/TR/sparql11-overview/ (accesso 22/07/2021).

- Pleiades[16] è un progetto dell'Ancient World Mapping Center (Università del North Carolina), *Stoà Consortium* (nato all'Università del Kentucky e dal 2018 ospitato su server dell'Università di Londra), e Institute for the Study of the Ancient World (Università di New York). È un dizionario geografico costruito comunitariamente e una mappa di luoghi antichi. Pubblica informazioni sui luoghi e sugli spazi antichi, fornendo servizi per ricercare, visualizzare e riutilizzare tali informazioni sotto licenza aperta (CC-BY). Pubblica *linked open-data*, quindi dati aperti a 5 stelle.
- Antiquity À-la-carte,[17] promosso dall'Ancient World Mapping Center, dell'Università del North Carolina, è un progetto di mappatura dell'antichità. I dati sono scaricabili in formato processabile. Non vi sono *linked open-data*. La licenza, *Creative Commons Attribution* (CC-BY), è stata aggiornata nel 2016 (prima era CC-BY-NC). Pubblica dati aperti a 3 stelle.
- Il Getty Institute pubblica diverse risorse con licenza libera, tra cui testi, immagini e *thesauri*. Limitatamente al *thesaurus* di arte e archeologia[18], esso è pubblicato in formato RDF, come *linked open-data*, con licenza ODBL, quindi può considerarsi a tutti gli effetti un open-data a 5 stelle.

A questi grandi progetti se affiancano altri dedicati a tematiche più specifiche, che producono ed espongono on line a pari modo *open-data*. Si ricorda, infine, *Linked Pasts*, un'iniziativa che fa dello sviluppo, pubblicazione e promozione dei *linked data* relativi all'antichità il suo tema principale.[19]

In definitiva, a livello internazionale non mancano esempi virtuosi e la pubblicazione di *open-data* è sempre più diffusa.

Uno sguardo all'Italia: *open-data* e archeologia

Similmente che nel resto del mondo, anche in Italia ormai il software libero è ampiamente utilizzato. La sua fortuna deriva forse più dal fatto che è gratuito piuttosto che libero e non di rado gli aspetti legati alla licenza vengono trascurati, generando una confusione tra *freeware* e *free*, e gratis e libero. Ad ogni modo, il software libero è sicuramente il più adoperato in ambito archeologico.

Sul fronte *open-data*, è doverosa una premessa, seppure sintetica, sulla legislazione italiana. In Italia la pubblicazione dei dati relativi ai beni culturali è soggetta a diversi quadri normativi:

- *In primis* il Codice dei Beni Culturali e del Paesaggio (D. Lgs. 42/2004) che all'art. 107 e 108 regola le riproduzioni, (anche le fotografie), seppure solo dei beni in custodia al ministero e agli enti pubblici regionali e territoriali.
- Il CAD (Codice dell'Amministrazione Digitale, D. Lgs. 82/2005) che regola la pubblicazione dei dati presenti negli archivi delle pubbliche amministrazioni.
- Le norme sul riutilizzo dei documenti nel settore pubblico (D. Lgs. 36/2006, che dovrebbe essere rivista entro il 21 luglio 2021 alla luce della recente direttiva (UE) 2019/1024).

[16] https://pleiades.stoa.org/ (accesso 20/04/2021).
[17] http://awmc.unc.edu/wordpress/alacarte/ (accesso 20/04/2021).
[18] https://www.getty.edu/research/tools/vocabularies/aat/index.html (accesso 21/07/2021).
[19] https://ics.sas.ac.uk/events/linked-pasts-6 (accesso 20/04/2021).

- Le norme sulla privacy (soprattutto il Regolamento (UE) 2016/679 - GDPR), che regola eventuali informazioni relative a persone, come ad esempio proprietà di terreni vincolati.
- Infine, la legge sul diritto d'autore e diritti connessi che regola, tra l'altro, i contenuti e le riproduzioni fotografiche, le banche di dati, ecc. (Legge 633/1941, che potrebbe essere aggiornata con il recepimento della direttiva (UE) 2019/790).

Negli ultimi otto anni i quadri normativi citati hanno subito importanti aggiornamenti (alcuni dei quali ancora in corso), che hanno avviato un processo di apertura e trasparenza nella Pubblica Amministrazione.

Nel Codice dei Beni Culturali e del Paesaggio è stata modificata la norma sulle riproduzioni dei beni culturali: ora è più agevolmente fotografare e divulgare se non vi è scopo di lucro (Ciurcina e Grossi 2019, 2016, 2013a, 2013b). Nel CAD (Codice dell'Amministrazione Digitale) è stato inserito il principio *open by default* per i dati delle pubbliche amministrazioni (cfr. art. 52: 'I dati e i documenti che i soggetti di cui all'articolo 2, comma 2, pubblicano, con qualsiasi modalità, senza l'espressa adozione di una licenza [...] si intendono rilasciati come dati di tipo aperto ai sensi all'articolo 1, comma 1, lettere l-bis) e l-ter) del presente Codice'. La mancanza di indicazioni specifiche definisce dunque sempre un diritto all'uso e riuso dei dati '... da parte di chiunque, anche per finalità commerciali, in formato disaggregato ...' come espresso nel Codice, all'art.1, lettere sopra indicate).

La legislazione italiana ha quindi compiuto negli ultimi anni numerosi passi avanti sul fronte dell'apertura dei dati, come del resto imposto dalle direttive europee. Ma cosa succede, in pratica, nel settore archeologico?

Una sintesi sul panorama FLOSS e open-data, almeno dal 2006 e per come è stata pubblicata nelle conferenze ArcheoFOSS, è quella proposta da J. Bogdani e F. Sciacca in appendice a questo volume. L'analisi conferma una sempre maggiore diffusione del software libero. Ma dimostra anche che al suo ampio utilizzo non corrisponde una altrettanto estesa partecipazione ai progetti di sviluppo. Ad esempio, per i progetti QGIS e Grass GIS, che sono tra i software più usati in ambito archeologico, gli archeologi che contribuiscono attivamente, realizzando attività di sviluppo, traduzione, documentazione, implementazione, debugging, disseminazione e/o diffusione, si contano sulle dita di una mano.

Tornando agli *open-data* archeologici, segue una breve disamina dedicata ai progetti italiani. Lungi dal voler proporre un elenco esaustivo, ci limiteremo ad esporre le proposte presentate nei workshop ArcheoFOSS, in quanto costituiscono una panoramica su quelle esperienze che più di altre sono mirate alla pubblicazione di dati aperti e che più di altre possono considerarsi longeve, testate e consolidate.

- ArcheoSITAR è un progetto di raccolta, sistematizzazione e condivisione della documentazione archeologica di Roma, promosso dalla Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma[20]. Si tratta del primo progetto a tentare un'apertura dei dati archeologici in Italia, strutturando inizialmente dei *web service*

[20] https://www.archeositarproject.it/ (accesso 20/04/2021).

sperimentali per la pubblicazioni di *data set* relativi alle principali classi logiche del sistema informativo (metadati) e iniziando a esporre *data set* archeologici per la consultazione *on line*. Da ottobre 2020 espone i dati in formato processabile con licenza libera (CC-BY-SA 4.0; cfr. il contributo di M. Serlorenzi, A. D'Andrea e R. Montalbano in questo volume), quindi *open-data* a 3 stelle.

- I due progetti del Mappalab dell'Università di Pisa, MappaGIS[21] e MappaOpenData[22] sono nati poco dopo ArcheoSITAR e sono anch'essi pionieristici in Italia in merito all'apertura dei dati archeologici. Per quanto riguarda MappaGIS, i dati non sono scaricabili in formato processabile, la licenza è aperta a seconda del tipo di dati (alcuni dati non hanno licenza libera). Si tratta pertanto di *open-data* a 1 stella. Per quanto riguarda il MOD, alcuni dati sono processabili, ma alcuni dati/contenuti sono in formato non processabile (PDF). I dati processabili hanno licenza libera e sono *open-data* a 3 stelle. Questa attribuzione non si può applicare tuttavia ai documenti con licenza libera in formato PDF (*open-data* a 1 stella). Le relazioni di scavo, inoltre, sono coperte da copyright, non presentano una licenza libera e non si possono pertanto considerare *open-data*.
- I progetti SITAVR e SITARu sono sviluppati presso l'Università di Verona. Il primo, il SITAVR[23] è nato da una collaborazione con ArcheoSITAR, del quale è un derivato in quanto utilizza lo stesso modello dati per il Sistema Informativo Archeologico di Verona; la maggior parte dei dati è attualmente ad accesso riservato, i pochi dati accessibili sono in formato processabile e con licenza libera. Il secondo, il SITARu[24] è invece un piccolo progetto derivato dal SITAVR, realizzato con il supporto finanziario di un imprenditore locale e con l'appoggio della funzionaria archeologa di zona. Si tratta del Sistema Informativo Archeologico del comune di Rubiera (RE). Espone dati articolati di scavi e reperti archeologici, in formato processabile, con licenza libera. I due progetti sono *open-data* a 3 stelle.
- Il progetto Raptor[25] è stato, con ArcheoSITAR e i progetti del MAPPAlab, un'esperienza pionieristica, che si proponeva in origine di sviluppare software *open-source* e di pubblicare i dati in formato aperto e con licenza libera. Oggi, però, il codice sorgente non è disponibile e nessun dato è scaricabile dal portale, l'accesso è solo in consultazione.
- Infine vi sono il Geoportale Nazionale per l'Archeologia (GNA)[26] e il progetto SigecWeb, il sistema di catalogazione ministeriale.[27] Vengono presi in considerazione in quanto il primo è stato oggetto di uno dei workshop dell'edizione di ArcheoFOSS 2020, dove si è anche accennato all'interoperabilità con il secondo.[28] Per quanto riguarda il SigecWeb, non vi è la possibilità di scaricare i dati, l'accesso per il pubblico è solo in consultazione; il Geoportale GNA invece è un'infrastruttura ancora in fase di realizzazione, ma gli annunci sul sito sembrano far ben sperare nell'apertura dei dati che verranno immessi al suo interno.

---

[21] http://www.mappaproject.org/web-gis-2/ (accesso 20/04/2021).
[22] http://www.mappaproject.org/archivio-digitale/ (accesso 20/04/2021).
[23] https://sitavr.scienze.univr.it/ (accesso 20/04/2021).
[24] https://sitavr.scienze.univr.it/sitaru/ (accesso 20/04/2021).
[25] https://www.raptor.beniculturali.it/ (accesso 20/04/2021).
[26] http://www.ic_archeo.beniculturali.it/it/222/il-geoportale-nazionale-per-l-archeologia-gna (accesso 21/07/2021).
[27] http://www.iccd.beniculturali.it/it/sigec-web (accesso 21/07/2021).
[28] https://2020.archeofoss.org/workshops#la-documentazione-archeologica-fra-digitalizzazione-dei-dati-e-dematerializzazione-dei-processi-soluzioni-e-strumenti-per-limplementazione-del-geoportale-nazionale-per-larcheologia-e-per-linteroperabilit%C3%A0-con-il-sistema-generale-del-catalogo-mibact (accesso 21/07/2021).

In sintesi, i dati italiani pubblicati come *open-data* sono davvero pochi. Spesso sono in formati non processabili e spesso hanno licenze che non ne consentono il riuso. In conclusione: anche a livello italiano c'è qualche esempio virtuoso, ma la pubblicazione di Open-Data archeologici in Italia pare meno diffusa che altrove.

Considerazioni conclusive

L'apertura dei dati archeologici in Italia è un processo avviato che ha ancora una lunga strada da percorrere. Un ruolo fondamentale giocheranno il Governo e il Parlamento, perché è con una legislazione adeguata, aggiornata e chiara che si potranno aprire i dati conservati negli archivi.

Una parte importante in questo processo avranno anche le comunità, perché la spinta dal basso, come dimostrano Wikipedia e OpenStreetMap, può incidere (una citazione qui è doverosa al contributo presentato da S. C. Schmidt, F. Thiery in questo stesso volume, un bel connubio tra istituzioni e comunità).

Fondamentale sarà, ad ogni modo, non abbandonare mai la costante e insistente difesa dei formati standard e delle licenze libere. In sintesi, molti processi virtuosi sono già stati avviati e attendono solo di essere migliorati. Si formulano due raccomandazioni:

- a livello tecnico, avvicinarsi il più possibile agli *open-data* a 5 stelle e attivare servizi nazionali o locali di gestione e distribuzione dei dati;
- a livello sociale, fare pressione presso il ministero e le Pubbliche Amministrazioni perché ciò venga fatto.

Bibliografia

Ciurcina, M. a P. Grossi 2019. Legge 4 agosto 2017 n. 124 e libertà di riproduzione dei beni culturali: cosa cambia. In P. Grossi, S. Costa, A. Jaia, S.G. Malatesta, F. Stasolla (eds) *Atti del XII Workshop Free, Libre e Open Source Software e Open Format nei processi di ricerca archeologica (ArcheoFOSS 2018)*, Roma, 19–22 Febbraio 2018: 447–450. *Archeologia e Calcolatori* 30. https://doi.org/10.19282/ac.30.2019.26.

Ciurcina, M. a P. Grossi 2016. Considerazioni sugli Open Data e i beni culturali e paesaggistici in Italia. Il decreto *Artbonus*: cosa cambia per la riproduzione dei beni culturali? In P. Basso, A. Caravale, P. Grossi (eds) *Atti del IX Workshop Free, Libre e Open Source Software e Open Format nei processi di ricerca archeologica*, Verona, 19–20 Giugno 2014: 35–41. *Archeologia e Calcolatori* Suppl. 8-2016. http://www.archcalc.cnr.it/indice/Suppl_8/06_Ciurcina-Grossi.pdf (accesso 13/10/2021).

Ciurcina, M. a P. Grossi 2013a. Beni culturali: brevi note sui dati e sul loro uso pubblico alla luce delle recenti modifiche legislative. In M. Serlorenzi (ed.) *Atti del VII Workshop Open Source, Free Software e Open Format nei processi di ricerca archeologica, Roma*, Palazzo Massimo, 11–12 Giugno 2012: 35–44. *Archeologia e Calcolatori* Suppl. 4 (2013). http://www.archcalc.cnr.it/indice/Suppl_4/04_Ciurcina_Grossi.pdf (accesso 13/10/2021).

Ciurcina, M. a P. Grossi 2013b. Open Data nella Pubblica Amministrazione: alcune considerazioni sui Beni Culturali. In M. Serlorenzi, I. Jovine (eds) *SITAR Sistema Informativo Territoriale Archeologico di Roma. Potenziale Archeologico, pianificazione territoriale e rappresentazione* *pubblica dei dati*, Atti del II Convegno Roma Palazzo Massimo, 9 Novembre 2011: 47–67. Roma: Iuno.

Graham, P. 2004. *Hackers & painters: big ideas from the computer age.* Sebastopol, CA: O'Reilly.

Raymond, E.S. 2001. *The cathedral and the bazaar: musings on Linux and Open Source by an accidental revolutionary*. Pechino: O'Reilly.

Reuters Staff 2020. Italy antitrust opens inquiry into Google, Apple, Dropbox on cloud computing. https://www.reuters.com/article/us-google-italy-antitrust-idUSKBN25Y0YM (accesso 13/10/2021)

Stallman, R. and G. Joshua. 2009. *Free Software Free Society: Selected Essays;* with introduction by Lawrence Lessig. Boston: Free Software Foundation.

Stallman, R. 2020. Who does that server really serve?. https://www.gnu.org/philosophy/who-does-that-server-really-serve.html.en (accesso 13/10/2021).

Wikipedia, The Free Encyclopedia, s.v. 'ARPANET', https://en.wikipedia.org/w/index.php?title=ARPANET&oldid=1029194011 (accesso 13/10/2021).

Wikipedia, The Free Encyclopedia, s.v. 'Tim Berners Lee', https://en.wikipedia.org/w/index.php?title=Tim_Berners-Lee&oldid=1032059716 (accesso 13/10/2021).

Zuboff, S. 2019. *The age of surveillance capitalism: the fight for a human future at the new frontier of power.* New York: Public Affairs.

An introspective, incomplete, view on the activity of the FLOS community dealing with Archaeology and Cultural Heritage

Julian Bogdani, Federico Sciacca

Julian Bogdani, Sapienza University of Rome; julian.bogdani@uniroma1.it
Correspondence: julian.bogdani@uniroma1.it
Federico Sciacca, Independent researcher; fede__fs@libero.it

Abstract
This short contribution was presented at the demo/poster section of ArcheoFOSS 2020 edition and is aimed at providing a snapshot of the activity of the Italian FLOS community, as it appeared in 2020, through the lens of what has been published in the available conference proceedings. It is inevitably, and by definition, incomplete research. For this reason, the Github repository can be used by anyone willing to update and integrate the available data and actively contribute to improve this self-evaluation attempt.

Keywords: ARCHEOFOSS; OPEN-SOURCE; ARCHAEOLOGY; FREE SOFTWARE; RESEARCH SUSTAINABILITY.

Introduction

The idea behind this contribution was developed during the preparation of the 2020 edition of ArcheoFOSS International Conference with the clear intent to feel the pulse of the community of people and institutions who, since 2006, have gathered and shared projects and ideas in the context of the conference.

The data was gathered by Federico Sciacca and Julian Bogdani and made available with MIT license during the 2020 edition works on a public Github repository.[1] The raw data were also paginated to a form a HTML presentation[2] and finally the single page application was published using the Github Pages.[3] The presentation is rebuilt on each change (commit) on the main repository. It is the firm intention of the authors to update the dataset behind the application as soon as other proceedings will be published. The state of art presented here has been updated for the last time on October 6th, 2020. It is therefore possible, rather desirable, that the state of the data and conclusions presented here will differ in the near future with the state of the presentation in the repository. For this reason, we are looking forward to seeing active contributions, possible corrections or updates to the dataset by the readers of this paper.

[1] https://github.com/jbogdani/af-introspection (accessed 16/6/2021).
[2] Reveal.js (https://revealjs.com/, accessed 16/6/2021) was used to create the Web-based presentation.
[3] https://jbogdani.github.io/af-introspection/ (accessed 16/6/2021).

This is not the first attempt to have an inner look on the activity of the ArcheoFOSS community. In 2013 an article by S. Costa and A. Palombini[4] tried to closely analyse the 'seven-year itch' of the conference, providing a deep insight on topics introduced in the conference, geographical and affiliation distribution of researchers, the growth of the relations and connections between them, being the conference editions the main catalyst. The present article might be considered as a natural follow-up, with a closer focus on the fate of single applications and projects through time and policies aimed at improving collaboration and openness.

J.B.

Methodology and sources

As already revealed in the first paragraph, the principal sources for the dataset are the published volumes of the proceedings of the various editions of ArcheoFOSS. The complete mapping of the conference editions and their publication can be seen in table. 1, par. 4. For each conference edition, the year of the publication, along with full bibliographic reference is provided. Furthermore, the explicit license of the volume is reported.

The publications were analysed to extract information about the different projects that were being presented. Two major typologies of projects were mapped in the bibliography: (1) applications, software packages, plugins, and scripts (Table 2, par. 4) and (2) GIS, webGIS and data portals (Table 3, par. 3). Each project, be it an application or a data-presentation project, was made the object of a Web search aiming at tracing its presence online and at obtaining possible updates on its lifecycle. We are well aware of the fact that the Web search is not a very reliable source of information since projects living in the deep Web, not indexed by search engines or hidden behind virtual private networks might not instantly appear. Yet, since openness does not come without findability, we assume that projects targeting an open community might not have walked many steps towards intentional invisibility. Once more, we warmly invite the reader to correct data vagueness due to the inaccuracy of our research by filing an issue or by requesting a repository merge.

To find the source code for each considered application, we looked for possible links provided as reference in the publications, a relatively rare practice. The main open-source public code repositories, such as GitHub, GitLab, BitBucket, SourceForge, Codeberg, etc. were sounded out. Furthermore, a generic Web search using the most popular search engines was also performed to look for 'ghost' projects.

It was highly problematic to determine whether a project is currently maintained or not, as well. For the aims of this contribution a project was considered 'not maintained' or 'abandoned' when no recent activity is clearly measured in the available source repository. To us 'recent' means a period of at least two years. If the source code is not available, the application is assumed to be not in an active status of development. It is also clear that projects that do not provide explicit reference to the last update have been considered abandoned, even if they might be actively maintained behind the curtains.

[4] Costa and Palombini 2013. La crisi del settimo anno: sette rassegne di progetti open-source per l'archeologia italiana. In M. Serlorenzi (ed.) *ARCHEOFOSS Free, Libre and Open-source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11–13 giugno 2012)*: 27–34. *Archeologia e* Calcolatori Supp. 4. Firenze: All'Insegna del Giglio.

When it comes to measuring the life duration of each project (up to October 2020), the year of the conference where it was presented was considered as the first reference. We are aware of the fact that the development of many projects might have started well before their presentation to the community; if clear evidence of this prior date is available from the text or from the date of the first commit in the publicly available repository, this prior date has been considered.

For applications maintained in git- or svn-based code repositories the total number of commits is available (Figure 7). This information does throw a glimpse on the activity of developers but is not truly representative of the overall life cycle of an application, especially for older projects initially not based on a version control system. Also, different developing techniques (early and frequent releases vs. long-cycle projects) might condition the total number of commits.

In the following sections (par. 3) some basic analysis will be presented in the form of charts. Raw data are presented as tables in par. 4 and a few synthetic conclusions will close the discussion (par. 5).

J.B.

Visual analysis of the data

In the next few pages are presented a series of bar charts displaying some very simple, yet highly informative analyses of the collected data, published in par. 4. For the general discussion and some preliminary observations, see sec. 5.

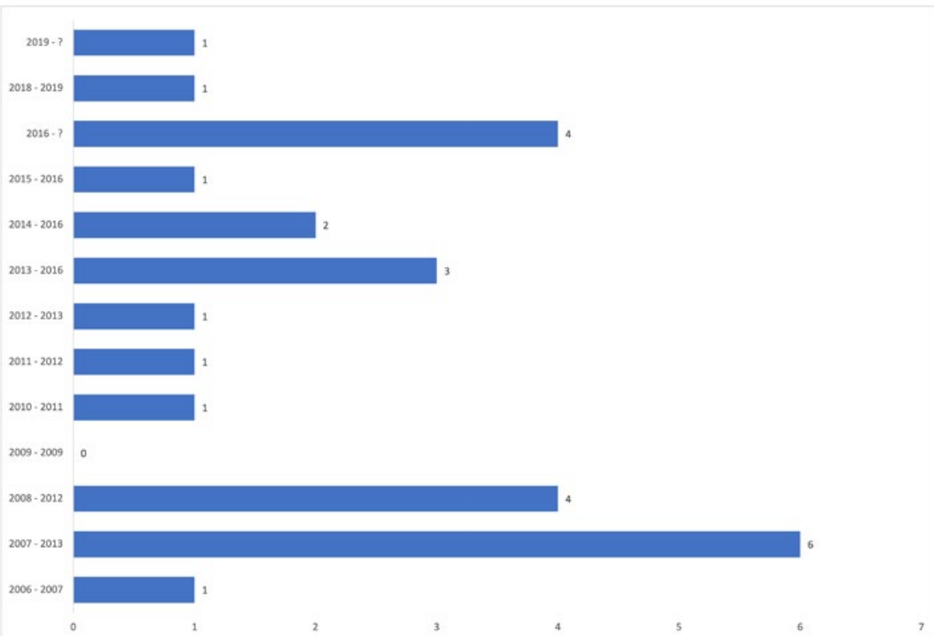


Figure 1: Years elapsed from the conference edition to the publication of its proceedings. Question marks indicate volumes that have never, or have not yet, been published. At the time of writing (2021), the 2019 proceedings are in press.

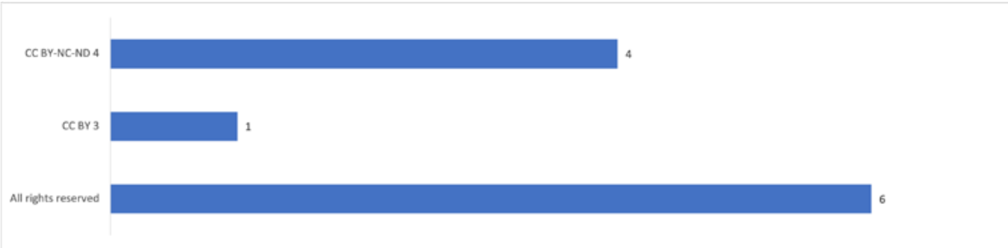


Figure 2: Licensing of the publications.

Figure 3: Availability of the source code of applications, software packages, plugins and scripts presented at ArcheoFOSS conferences.

Figure 4: Explicit licensing of applications, software packages, plugins and scripts presented at ArcheoFOSS conferences.

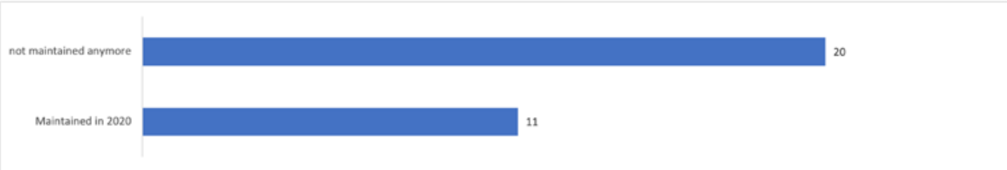


Figure 5: Survival rate of applications, software packages, plugins and scripts presented at ArcheoFOSS conferences.

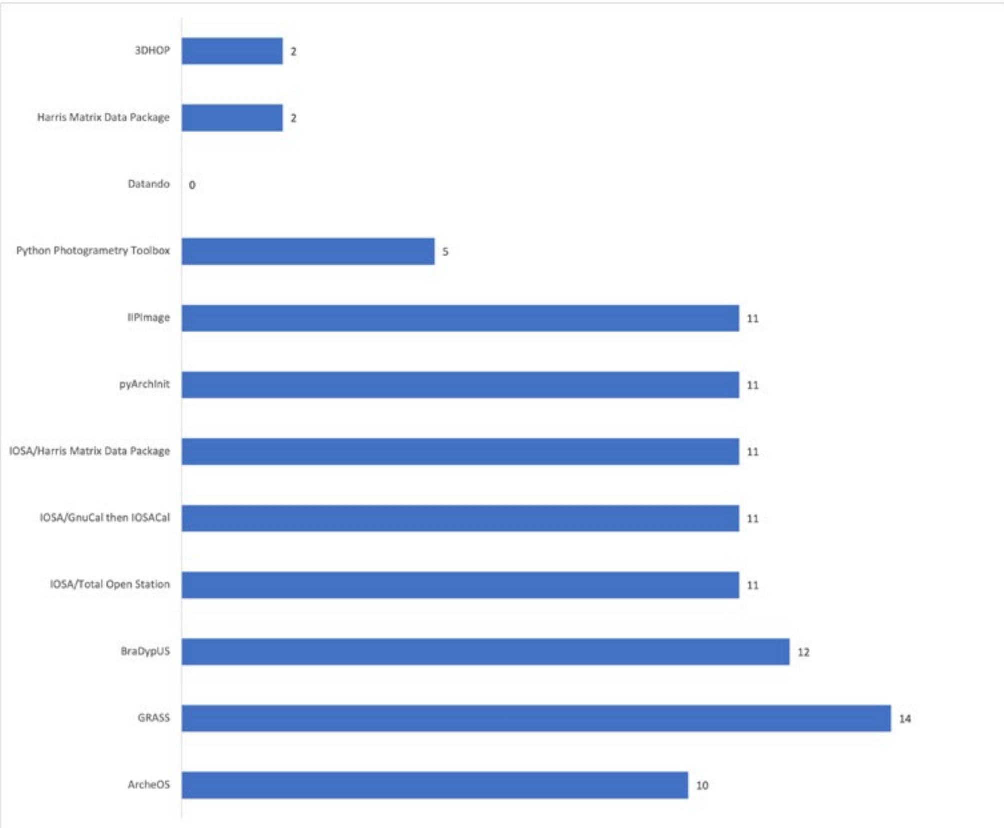


Figure 6: Longevity of applications, software packages, plugins and scripts presented at ArcheoFOSS conferences.

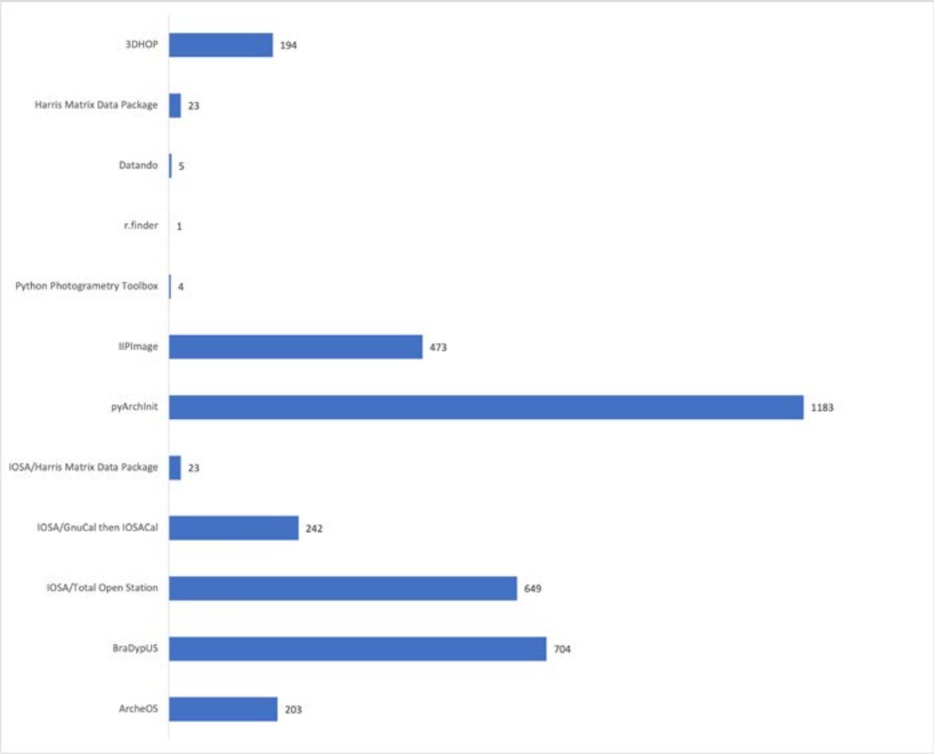


Figure 7: Total number of commits for applications, software packages, plugins and scripts presented at ArcheoFOSS conferences, as available in the public repository. GRASS with more than 20,000 commits has been excluded from the chart.

Figure 8: Online availability of Databases, GIS, webGIS and data portals presented at ArcheoFOSS.

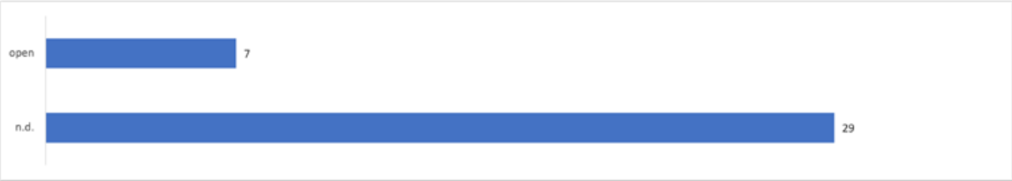


Figure 9: *De facto* data access policy for Databases, GIS, webGIS and data portals presented at ArcheoFOSS.

J.B.

Data sets

Conference year	Publication year	Full bibliographic record	License of the publication
2006	2007	G. Macchi Jánica and R. Bagnara (eds) 2007. *Open-source, free software e open format nei processi di ricerca archeologici atti del 1. Workshop (Grosseto, 8 maggio 2006)*. Firenze: Centro editoriale toscano. ISBN 978-88-7957-260-6	All rights reserved
2007	2013	S. Costa and G.L. Pesce 2013 *Open-source, Free Software e Open Format nei processi di ricerca archeologica 2007*. London: Ubiquity Press. DOI:http://dx.doi.org/10.5334/bae	CC BY 3
2008	2012	L. Bezzi, D. Francisci, P. Grossi and D. Lotto 2012. *Open-source, Free Software e Open Format nei processi di ricerca archeologica. Atti del III Workshop (Padova, 8-9 maggio 2008)*. Roma: Quasar	All rights reserved
2009	2009	C. Paolo, A. Palombini, S. Pescarin (eds) 2009. *ARCHEOFOSS Open-source, Free Software e Open Format nei processi di ricerca* archeologica. *Atti del IV Workshop (Roma, 27-28 aprile 2009)*. Rome: Archeologia e Calcolatori, Supplemento 2	CC BY-NC-ND 4
2010	2011	F. Cantone 2011. *ARCHEOFOSS Open-source, Free Software e Open Format nei processi della ricerca archeologica. Atti del VI Workshop (Napoli, 9–10 giugno 2011)*. Napoli: Naus Editoria	All rights reserved
2012	2013	M. Serlorenzi (ed.) 2013. *ARCHEOFOSS Free, Libre and Open-source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11–13 giugno 2012)*. Rome: Archeologia e Calcolatori, Supplemento 4	CC BY-NC-ND 4
2013	2016	F. Stanco and G. Giovanni (eds) 2016. *Free, libre and open-source software e open format nei processi di ricerca archeologica. VIII Edizione Catania 2013*. Oxford: Archaeopress	All rights reserved
2014	2016	P. Basso, A. Caravale and P. Grossi (eds.) 2016. *ARCHEOFOSS. Free, Libre and Open-source Software e Open Format nei processi di ricerca archeologica. Atti del IX Workshop (Verona, 19–20 giugno 2014)*. Rome: Archeologia e Calcolatori, Supplemento 8	CC BY-NC-ND 4
2015	2016	S. Campana, R. Scopigno, G. Carpentiero and Marianna Cirillo (eds) 2016. *Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*. Oxford: Archaeopress. ISBN 97817849133892	All rights reserved
2016	Never published		
2018	2019	G. Piergiovanna, S. Costa, A. Jaia, S.G. Malatesta, F.R. Stasolla (eds) 2019. *ArcheoFOSS. Free, Libre and Open-source Software e Open Format nei processi di ricerca archeologica, Atti del XII Workshop (Roma, 19–22 febbraio 2018)*. Rome: Archeologia e Calcolatori	CC BY-NC-ND 4
2019	In press		

Table 1: ArcheoFOSS conferences and the publishing of the proceedings, with indication of the license, source: https://github.com/jbogdani/af-introspection/blob/master/data/publications.js (accessed 16/6/2021).

Year	Authors	Application name and description	License	Is it maintained?	Source code available at:	Last commit in	Total number of commits
2006, 2021	A. Bezzi, L. Bezzi, D. Francisci R. Gietl, F. Furnari	ArcheOS Linux distribution targeted to archaeologists. Presented a new release of v.1.1.6 Akhenaton.	-	-	https://github.com/archeos/ArcheOS	2016	203
2006	F. Niccolucci, A. D'Andrea, S. Hermon, T. Zoppi, A. Felicetti	MAD (Managing Archaeological Data) Tool for the management of archaeological data, based on XML and eXist.	-	-	-	-	-
2006	M. Neteler, A. Bezzi, L. Bezzi, R. Gietl, D. Francisci, M. Barton	GRASS Use of GRASS GIS package for archaeology	GNU GPL v2+	yes	https://github.com/OSGeo/grass	2020	20832
2006	F. Cuniolo	ARIS (Archaeological Resource Information System) A multi-user system for the management of the archaeological record.	-	-	-	-	-
2007	L. Calori, C. Camporesi, A. Palombini, S. Pescarin	Osg4Web Software package for 3D navigation and interaction with archaeological landscapes	-	-	-	-	-
2007	D. Lotto, F. Biscani, A. Facchin, S. Tibolla	Knossos DBMS for archaeology	-	-	-	-	-
2007	F. Giudici, A. Macdonnel McLean, A. Palombini	VRC (Visual Resource Collaborative) RDBMS for the online cataloguing and visualisation of Cultural Heritage items	-	-	-	-	-
2008	J. Bogdani, E. Vecchietti	BraDypUS Web-based RDBMS for the archaeological record	MIT	yes	https://github.com/bdus-db/BraDypUS	2020	704

Year	Authors	Application name and description	License	Is it maintained?	Source code available at:	Last commit in	Total number of commits
2009	S. Costa, G.L. Pesce, L. Bianconi	IOSA/Total Open Station A program for downloading and processing survey data from total station devices	GNU GPL v3	yes	https://github.com/totalopenstation/totalopenstation	2020	649
2009	S. Costa, G.L. Pesce, L. Bianconi	IOSA/GnuCal then IOSACal An open-source radiocarbon (14C) calibration software. Written in Python http://c14.iosa.it/	GNU GPL v3	yes	https://codeberg.org/steko/iosacal	2020	242
2009	S. Costa, G.L. Pesce, L. Bianconi	IOSA/Harris Matrix Data Package A Data Package specification for archaeological stratigraphy data following the Harris Matrix convention	GNU GPL v3	yes	https://codeberg.org/steko/harris-matrix-data-package	2020	23
2009	C. Pedelì	ArcheoTRAC Software package for the tracking of the life cycle of Cultural Heritage items	-	-	-	-	-
2009	M. Felicori, M. Gaiani, A. Guidazzoli, M.C. Liguori	Visman Framework based on OpenSceneGraph and wxWidgets for 3D visualisation	-	-	-	-	-
2009, 2012	L. Mandolesi, E. Cocca	pyArchInit QGIS plugin for the management of the archaeological record	GNU GPL v2	yes	https://github.com/pyarchinit/pyarchinit	2020	1183
2009	S. Laurenza, S. Mancuso, A. Costantino	V.I.R.A. (Visualizzatore Interattivo delle Rotte Antiche) Calculation and visualisation of naval routes in the ancient Mediterranean	-	-	-	-	-
2009	D. Pitzalis, R. Pillay	IIPImage Client-server package for the remote visualisation of high-resolution images over the Web	GNU GPL v3	Yes	https://github.com/ruven/iipsrv/	2020	473

Year	Authors	Application name and description	License	Is it maintained?	Source code available at:	Last commit in	Total number of commits
2009	G. De Felice, G. Sibilano, G. Volpe, E. Di Sciascio, R. Mirizzi, G. Piscitelli, E. Tinelli, M. Trizio	IreMaS Web based system for the management of excavation records	-	-	-	-	-
2009	A. Pozzi, P. Salonia	Cloud Manager Software package for the management of point clouds	-	-	-	-	-
2010	D. Gnesi Bartolani, S. Petruzzelli	Extended ICCD Software package for the management of ICC catalogue items	-	-	-	-	-
2011	P. Moulon, A. Bezzi	Python Photogrammetry Toolbox Photogrammetry software that creates 3D models from images	GNU GPL	-	-	-	-
2011	A. Palombini	r.finder GRASS script for predictive archaeology	GNU GPL	-	http://www.palombini.net/sw/finder/	-	-
2012	D. Francisci	ArchaeoSection Application aimed at automatising the documentation of archaeological sections	-	-	-	-	-
2013	D. Gnesi Bartolani, U. Moscatelli	Datando Python library to manage and convert dates in DBMSs	-	-	https://pypi.org/project/datando/	2013	5
2013	M. Monti, G. Arvizzigno, D. Lombardo, G. Maino	AKVIS Retoucher Software package for the digital restoration of photographs	commercial	yes	-	-	-
2014	G. Di Giacomo	GIS Cloud GIS functions distributed on the cloud	-	-	-	-	-
2014	M. Frassine, G. Naponiello, S. De Francesco, A. Asta	Raptor Geo RDBMS for the archaeological record	commercial	yes	-	-	-

Year	Authors	Application name and description	License	Is it maintained?	Source code available at:	Last commit in	Total number of commits
2015	J. Sikora, J. Sroka, J. Tyszkiewicz	Strati5 Mobile application for the creation of Harris Matrix	-	-	-	-	-
2018	S. Costa	Harris Matrix Data Package Harris Matrix Builder	GNU GPL	yes	https://codeberg.org/steko/harris-matrix-data-package	2020	23
2018	A. Vecchione, A. Lureau, M. Callieri	3DHOP 3D Viewer	GNU GPL v3	yes	https://github.com/cnr-isti-vclab/3DHOP	2020	124
2018	D. Bursich	iGIS (Immersive GIS System) Immersive 3D platform	-	-	-	-	-

Table 2: Applications, software packages, plugins and scripts presented at ArcheoFOSS conferences, source: https://github.com/jbogdani/af-introspection/blob/master/data/applications.js (accessed 16/6/2021).

Year of presentation	Authors	Name	URL	Availability	Access policy
2006	S. Alivernini, M. Brovelli, D. Magni	ArchaeoGEW: Pre-Roman archaeological evidences at Spina Verde (Como)	http://geomap.como.polimi.it/	-	-
2006	L. Deravignone, G. Macchi, M. Serragli, A. Vichi	ASFT: Atlas of fortified sites in Tuscany	-	-	-
2006	A. Bonomi, M. Cattani, G. Mantegari, G. Vizzari	WebGIS and portal of the Bronze Age in Northern Italy	-	-	-
2007	M. Baldassarri, G. Naponiello, G. Pagni	WebGIS of Montescudaio in Val di Cecina (PI)	http://www.cediamp.it	-	-
2008	N. Pisu, G. Naponiello	ArcheoTanatOS. WebGIS of the funerary archaeology in Trentino	http://arc-team.homelinux.com/archeotanatos/archeotanatos.php	-	-
2008	M. De Gennaro, L. Zennaro	WebGIS of the Veneto region	-	-	-
2008, 2009	L. Calori, C. Camporesi, A. Negri, A. Palombini, S. Pescarin	Roma città aperta. 3D VR webGIS of the Agro Romano	http://www.virtualrome.it	-	-

Year of presentation	Authors	Name	URL	Availability	Access policy
2008	P. Grossi, F. Pirotti	WebGIS of Montegrotto Terme (Padua)	http://www.ketos.lettere.unipd.it/montegrotto	-	-
2009	G. Di Giacomo, G. Scardozzi	Atlas and Fortifications of Hierapolis	http://antares.ibam.cnr.it/atlante-hierapolis	-	-
2010	L. Endrizzi, M. Frassine, R. Gietl, G. Naponiello, N. Pisu	Castellum Vervassium. WebGIS of S. Martino Vervò (Trentino)	n.d.	-	-
2010	N.M. Mangialardi, M.G. Sibilano	ArchiDB			
2010	E. Scampoli, A. Sahlin	Geodatabase of Florence and Pratomagno	-	-	-
2011, 2012, 2014, 2018	M. Serlorenzi, A. De Tommasi, S. Ruggeri, R. Grassucci, A. Vismara, G. De Rosa, A. Cecchetti, F. Lamonaca, F. Zonetti	SITAR. WebGIS for Rome	https://archeositarproject.it/	yes	open
2011	A. De Stefano, M.G. Sibilano, G. Volpe	La città nascosta. Map of the hypogea in Foggia	-	-	-.
2011	D. Leone, N. Mangialardi	LIBURNA. Database for underwater archaeology	Nd.	-	-.
2012	F. Anichini, G. Gattiglia, M.L. Gualandi, V. Noti	MAPPA: Open-Data container	http://www.mappaproject.org/	yes	open
2012	D. Berardi, A. Ciapetti, M. De Vizia Guerriero, A. Donnini, M. Lorenzini, M.E. Masci, D. Merlitti, S. Norcia, F. Piro, O. Signore	baseculturale.it. Semantic portal for Cultural heritage	http://baseculturale.it	-	-
2013	A. Scianna, S. Gristina, R. Sciortino	3D GIS of the castle of Maredolce, Palermo	-	-	-

Year of presentation	Authors	Name	URL	Availability	Access policy
2013	D. Giusti	Spatial database of the Palaeolithic site of Pirro Nord, Apricena (FG)	-	-	-
2013	M.S. Busana D. Francisci, A.R. Tricomi	Archeologia della lana. A Spatialite database on wool trade in the Roman Cisalpine area	-	-	-
2013	P. Albrizio, F. De Virgilio, G. Panzarino, E. Zambetta	Siponto Aperta: a webGIS of Siponto (Manfredonia)	-	-	-
2013, 2014	B. Bruno, P. Basso, P. Grossi, A. Belussi, S. Migliorini, G. Cavalieri Manasse	SITAVr: webGIS of Verona	https://sitavr.scienze.univr.it/	yes	open
2013	A. Santoriello, A. Rossi, P. Rossi	SIUrBE: urban webGIS of Benevento	-	-	-
2013	L. Gambaro, S. Costa	Archaeological database of Albintimilium, Liguria	-	-	-
2014	C. Alfonso, G. Di Giacomo	Archaeological map of the Protected Marine area of Porto Cesareo	-	-	-
2014	L. Peyronel, D. Bursich, G. Di Giacomo	Draft webGIS of the plain of Erbil, Irachi Kurdistan	-	-	-
2014	L. Michielin, G. Strapazzon	Database of Casa delle Bestie Ferite, Aquileia	-	-	-
2014	D. Rose, E. Ceccaroni	WebGIS of the aqueduct of Alba Fucens	-	-	-
2014	F. Brunet, A. Cosner, G. Naponiello	Le fonti per la storia: webGIS of the Primiero and Vanoi valleys (Trento)	-	-	-
2014	D. Malfitana, G. Cacciaguerra, A. Mazzaglia, S. Barone, V. Noti	OpenCiTy: webGIS of the urban archaeology of the city Catania	-	-	-
2014	G. Andreozzi, G.C. Cianferoni, C. Francini, A. Sahlin, E. Scampoli	ArcheoFi: information archaeological system of Florence	-	-	-
2014	V. Vassallo, N. Kyriacou, S. Hermon, I. Eliades	Draft webGIS of Cyprus	-	-	-

Year of presentation	Authors	Name	URL	Availability	Access policy
2014	J. Bogdani	Ghazni: an open-access digital archive of the Italian Archaeological Mission in Ghazni, Afghanistan, backed by a BraDypUS database	https://ghazni.bradypus.net, https://ghazni.bdus.cloud	yes	open
2014	A. Caravale	Sethlans. Archaeological database of the Faina museum, Orvieto	http://bronzifaina.isma.cnr.it/	yes	open
2018	J. Bogdani	PAThs: an open-access geodatabase and archaeological atlas of Late Antique Egypt, backed by a BraDypUS database	https://atlas.paths-erc.eu	yes	open
2018	S. Mariotti	Percorsi BioGrafici: webGIS and media wiki for the urban archaeology of Monforte San Giorgio (ME)	http://www.percorsibiograficimsg.com/wiki/index.php?title=Pagina_principale	-	-

Table 3: Databases, GIS, webGIS and data portals presented at ArcheoFOSS conferences, source: https://github.com/jbogdani/af-introspection/blob/master/data/webgis.js (accessed 16/6/2021).

F.S., J.B.

Concluding notes

To sum up the previous data visual analyses, it can be said as a conclusion that generally ArcheoFOSS meetings have been promptly published (Figure 1). In most cases (seven) it took only one year for the volume to be published; in one case it was published in the same year of the meeting. Some major difficulties have been encountered in the second edition of the conference (2007) whose proceedings were published only in 2013, and in the case of the 2016 edition that has never been published. As for the 2019 edition, the proceedings are by the time of writing this article under press. While logistic and funding issues might have been the main cause of the delay of publication, it can be observed that the restrictive licensing of the books is due, probably to a scarce attention towards the free and open dissemination of the research, apparently in sharp contrast to the mission of the meetings (Figure 2). An open license (CC BY-NC-ND 4) is available when the proceedings are published as Supplements of the *Archeologia e Calcolatori* journal and only in one case (conference year 2007, published in 2013) it seems to be a conscious choice of the editors (CC BY). Most of the volumes have been published with all rights reserved.

Turning our attention to the single projects presented in the conferences (and published), as for the applications, software packages, plugins and scripts, most of them (18) do not at present expose publicly the source code, in private or third-party repositories (Figure 3). Only 13 projects can be today labeled as open-source, but we must consider that many of the projects that are not maintained anymore might simply have disappeared in the last few years. There is no way to know how many of the missing ones were originally published on the Web, yet it can be argued that published articles rarely cite links, and when they do, they are projects whose code is still available. It is reasonable to assume that the code of the missing projects has never been published online.

Despite the mission of the conference, a very poor attention is dedicated to the licensing of the applications presented in the conferences, as well. The majority of the considered projects are totally reticent on this topic (Figure 4) and focus mainly on the usage of open-source tools, applications or libraries. GNU GPL is the most used FLOS license (ten cases) and MIT is used only once[5]. Unexpectedly, two commercial projects have been presented and at least in one case it was promised by project managers that the first stable release of the product would have been released with a GNU GPL license. Despite the promise, the product is still being maintained but the source code has never been shared with the community. Many of the 18 projects with no clear indication of the license, regard the use of FLOS software and tools, a fact that does not help to understand the forgetfulness of the open license for the final project.

The 'survival rate' represented in Figure 5 also clearly indicates that most of the applications presented in the conference were short-living projects, aimed at resolving very limited tasks or in some cases serving only academic purposes, since no clear track could be found of them except academic articles in scientific journals. Apparently, the push to investigate IT applications to the archaeological domain (and more in general in the domain of Cultural Heritage) has been determined in the past more by academic needs than by market or everyday practice.

The longevity of applications, software packages, plugins and scripts presented in Figure 6, refers to projects whose code is available and is measured in years from the date of their first commit (or from the date of presentation, if prior) to the date of the most recent commit/activity. Most of these projects can boast of being around for more than ten years, clearly demonstrating a huge interest of the community towards them. The count of commits, for applications using a Git-based public repository, gives account of the intensity of development and the general picture confirms the information about the longevity. By far, pyArchinit seems to be the most intensively updated project, probably reflecting the interest of the community in its development.

When it comes to projects about databases, GIS, WebGIS and data portals, the data on availability clearly shows an impressive volatility and instability of the presented projects (Figure 8). Of the vast majority of the considered projects (28/35) no traces could be found online, with the obvious exception of scientific articles published in ArcheoFOSS proceedings and in few other scientific journals. We can assume that some of them might have gone offline and disappeared from the reach of search engines after the scientific (and financial) interest on the project faded, yet many of them probably were never published online, since in many cases no link could be found in the publications. The good news is that the remaining seven projects still available today on the Web are all adopting open policies for the access to data (Figure 9). This overview does not clearly measure the openness of each project, but the contribution by P. Grossi and M. Ciurcina in this volume can help for a first attempt to look closer to data sharing policies.

J.B.

[5] The new Version 4 of the Bradypus project, previously licensed as MIT, is now distributed as GNU AGPL v3.0. As the well-known MIT license does not include a share-alike clause, and forks of a MIT licensed project can be released under a commercial license.