The Pelagios Network: Collaboration as a Community of Practice

Barker, Elton

elton.barker@open.ac.uk
The Open University, United Kingdom

Gheldof, Tom

Tom.Gheldof@KULeuven.be
KU Leuven, Belgium

Gordin, Shai

shygordin@gmail.com
Ariel University, Israel

Lewis, Orly

orly.lewis@mail.huji.ac.il
Hebrew University of Jerusalem

Nury, Elisa

Elisa.Nury@sib.swiss
SIB Swiss Institute of Bioinformatics, Switzerland

Vitale, Valeria

v.vitale@sheffield.ac.uk
University of Sheffield, United Kingdom

Simon, Rainer

rainer@rainersimon.io
rainersimon.io

McDonough, Katherine

kmcdonough@turing.ac.uk
The Alan Turing Institute

Chen, Anne

achen@bard.edu
Bard College, USA

Williams, Miranda

miranda.williams@classics.ox.ac.uk
Oxford University

Almohamad, Adnan

almohamad@bbk.ac.uk
Bard College Kingston, New York

Middle, Sarah

s.middle@nms.ac.uk
National Museums Scotland, United Kingdom

Hay, Duncan

dh418@sussex.ac.uk
University of Sussex

Butterworth, Alex

A.Butterworth@sussex.ac.uk
University of Sussex

In this panel we discuss the Pelagios Network (https://pelagios.org/), a decade-long collaboration that has been developing methods, tools and communities for linking Humanities resources online.

From its establishment in 2011, Pelagios has used collaboration as the means to transform the practice of Linked Open Data (LOD). In contrast to other initiatives that have promoted complex, if robust, ontologies for producing LOD, Pelagios phases 1-2 developed — in collaboration with popular ancient Mediterranean data providers — a lightweight method of connecting heterogeneous resources based on the semantic annotation of place references and their alignment to gazetteer URIs (Vitale et al. 2021). To extend this method to other disciplines lacking the maturity of such a specialist environment, Pelagios 3-5 developed the annotation tool, Recogito (https://recogito.pelagios.org/), which empowers the domain expert to semantically annotate materials without need for technical know-how (Simon et al. 2017). Finally, Pelagios 6-8 coordinated a "small grant" programme that supported the conceptual articulation of Humanities LOD ("Working Groups") and the development of common tools ("Resource Development Grants"), which succeeded in extending the Pelagios method into knowledge domains from the global South (Kahn et al. 2021). Ultimately, this growing and maturing community has been realised in the foundation of a formal voluntary, and free, organisation: the Pelagios Network Association.

This panel is a demonstration of the decentralised collaborative principle at the heart of Pelagios. The Pelagios Network is formed of equal and interdependent Partners who, via a process of a formally-agreed plan, come together on a regular bi-monthly basis to coordinate work according to thematically-focused Activities — currently: annotation, gazetteers, registry, and visualisation. The five panellists represent Partner projects or institutions who have been contributing to this endeavour in various ways; the opportunities afforded by Pelagios's collaborative framework that we showcase here range from the low-hanging fruit of sharing knowledge to more complex interactions through which Partners are continuing to develop tools and infrastructure. Two speakers discuss the opportunity for collaborative problem-solving (Chen) and the sharing of not only knowledge but also data, best practice, and standards (Ghelfdof). Three speakers discuss the opportunity for developing tools in the round (and thus avoid reinventing the wheel), whether in the adaptation of a common toolset to different projects (Nury), the development of machine-learning techniques to existing software (Vitale), or the application of tools from one domain to another (Gordin). The final speaker (Middle) reflects on the set of continuing challenges facing the widespread adoption of LOD in the Humanities — namely, how rich, heterogeneous datasets, with huge research potential, can be managed and published in a form that is flexible, scalable, interoperable, and sustainable

— and the opportunities afforded by working collectively under the Pelagios umbrella.

Together, we emphasise a ground-up, co-creation approach where collaboration is conceived as a community of practice. Critical in our discussion will be the identification and consideration of core issues, specifically the challenge of addressing and managing a series of tensions between decentralisation and coordination, development and sustainability, and individual needs versus community growth.

By-products of project development as collaborative opportunity : Arabic vocabulary equivalents in cultural heritage terminology as LOD

Anne Chen [presenter] (Bard College), achen@bard.edu

Miranda Williams (Oxford University), miranda.williams@classics.ox.ac.uk

This paper argues that foregrounding the "struggle" of interpretive decision-making in the lead-up to a target output (publication or otherwise) presents previously unanticipated collaborative opportunities with the potential to benefit future humanities research more broadly. Traditional humanities research has been a rather solitary pursuit: a single scholar collects thematically related content, reads past scholarship, makes editorial or interpretive decisions, and then (hopefully) synthesises the products for publication in an article or book. It is only in the rather mature stages of research — either in the context of a final publication, or else in a conference presentation (often in the lead-up to publication) — that there has traditionally been an opportunity for interfacing with a broader research community. How often have humanities scholars unnecessarily reinvented the wheel or else worked at cross-purposes thanks to this traditional research structure?

With the introduction of digital methodologies into humanities disciplines, one of the not-so-often reflected-upon benefits is the fact that a sense of "strangeness" or destabilisation in the "how-to" of research has necessitated repeated consultation with a broader community long before project completion. The Pelagios Network, with its international membership and online meeting structure, has created a thriving community of practice for in-process collaborative problem-solving. As an illustrative case study, this presentation draws on a collaborative endeavour — the formalisation and unification of independent project efforts in Arabic vocabulary disambiguation for cultural heritage descriptors — an effort conceived thanks to regular network-partner reporting on work-in-progress via the Pelagios Network's linked-data Annotation activity.

Expanding the TM Places gazetteer in the Pelagios (Linked Pasts) Network

Tom Gheldof (KU Leuven), Tom.Gheldof@KULeuven.be

Trismegistos (TM, https://www.trismegistos.org/) has a long-standing collaboration with Pelagios that is exemplary for the importance of sharing data and knowledge across partners, projects, and domains. To scale-up and improve the TM Places gazetteer, we exchanged data with partners in the Pelagios Gazetteer Activity. An initial helpful tool was Peripleo (https://github.com/pelagios/peripleo), a map interface that visualised connections between different projects based on their place references. Not only could users query this LOD graph; it also allowed researchers to easily view errors or disambiguate place names. This resulted in a further expansion of TM Places with links to other gazetteers, beyond our pre-existing collaborations with Pleiades (https://pleiades.stoa.org/) and EDH (https://edh.ub.uni-heidelberg.de/).

As a next step to share geographical data, we launched the TM GeoRelations API (part of the TM Data Services, https://www.trismegistos.org/dataservices/georelations/documentation/), enabling users to query multiple gazetteers. By including links to modern place names, TM Places can also be combined with other gazetteers to form a larger data set of historical place names (e.g., data sets collected in the Pelagios Registry Activity). To further grow TM, the Pelagios Network, as well as the Linked Pasts ecosystem that it has initiated, provides an optimum infrastructure, precisely because it is based on the exchange and development of best practices, standards, and expertise (see https://medium.com/pelagios/from-linking-places-to-a-linked-pasts-network-ebc974f5b342). Specifically, collaboration with Partners in the Registry Activity and further development of PeripleoLite — an adaptable LOD visualisation tool (https://github.com/britishlibrary/peripleo) — will help us continue to ensure the usability and value of TM.

MARK16: visualising manuscripts and their repositories withPeripleo

Elisa Nury [presenter] (SIB Swiss Institute of Bioinformatics), Elisa.Nury@sib.swiss

Claire Clivaz (SIB Swiss Institute of Bioinformatics), Claire.Clivaz@sib.swiss

Silvano Aldà (SIB Swiss Institute of Bioinformatics), Silvano.Alda@sib.swiss

To create a geographic data visualisation for the MARK16 project, the critical question facing us has been: what tool and data format can we use that can be implemented within a short deadline and a limited budget, while still being able to display complex research data? The Pelagios Network is grounding MARK16 into the broader community of practices around Humanities Linked Open Data and Historical GIS.

MARK16 is focused on the last chapter of Mark's Gospel, an enigma in New Testament textual criticism (Clivaz 2019). The Virtual Research Environment created for this five-year SNSF project includes four parts, among which *Dataviz* displays the manuscripts and their repositories on a map. The main part is a *Manuscript Room* which currently holds 57 manuscripts with images and transcriptions (https://mr-mark16.sib.swiss).

MARK16 Dataviz is using the new *PeripleoLite* javascript viewer (Simon et al. 2016, 2022). Our dataset of 36 places linked to 57 manuscripts necessitated adjustments to the faceted search of *PeripleoLite* (https://dataviz-mark16.sib.swiss). To showcase the libraries who participated in MARK16, while also being able to search the manuscripts, we benefited from the help of its creator Rainer Simon. Preparing the visualisation has raised interesting questions for us: what is the physical anchor to the world of a library that exists purely in a digital format? What happens when physical objects are shared in different places, as for the Gothic manuscript Got 1? As we discuss our Dataviz with the Pelagios Visualisation Activity, we are also contributing to the tool co-creation process.

Sustainability through sharing: the many lives of Recogito

Valeria Vitale [presenter] (University of Sheffield), v.vitale@sheffield.ac.uk

Rainer Simon (The Austrian Institute of Technology), aboutgeo@gmail.com

Katherine McDonough (The Alan Turing Institute), kmcdonough@turing.ac.uk

Recogito and Peripleo, the online tools developed by Pelagios, have proven highly popular in Humanities research and pedagogy (Dunn and Vitale 2021). However, when core Pelagios funding ended in 2019, the question of how to maintain these tools or develop others was critical for the community that formed the Pelagios Network (Kahn et al. 2021). The path to achieving long-term self-sustainability has been to encourage as many open customisations as possible, specifically to:

- generate a library of reusable "modules";
- foster a community of developers;
- inspire researchers about the potential of Recogito and Peripleo beyond their main versions.

As shown by projects like MARK16, this "diffused strategy" has generated a number of new applications sharing a common methodology. Here, we focus on the *Machines Reading Maps* project (Chiang et al. 2021). MRM aims to apply machine learning (ML) and semantic technologies to identify, gather, and make available text on maps. Recogito was initially proposed as a means to create gold-standard annotations to evaluate the ML algorithm, but we quickly realised that the potential was much higher, and that Recogito could become a powerful way to investigate maps from a semiotic perspective. The new Recogito has been informed by MRM research questions, and has in turn shaped the research. The MRM-specific version of Recogito has already been adopted by other projects, both within and outside academia, where new applications in community mapping are being explored, including volunteer transcription of historical maps in collaboration with the National Library of Scotland. (See https://github.com/machines-reading-maps/Tutorials-Newsletters/wiki/Edinburgh-Map-Transcription-Event:-Annotation-Guidelines-for-the-Public.)

Machine learning of ancient scripts and languages through collaboration

Shai Gordin (Ariel University), shygordin@gmail.com

Orly Lewis (Hebrew University of Jerusalem), orly.lewis@mail.huji.ac.il

One of the central characteristics of Digital Humanities is the multitude of projects that involve intense collaboration across disciplines. It is much less common to find collaborative research networks like Pelagios. In this talk we consider how Pelagios's collaborative model has the potential to revolutionise the way in which we computationally study ancient languages.

Originally established to bring together scholars around the common theme of historical places and spaces, Pelagios has evolved into an interdisciplinary virtual space for those who study the ancient and pre-modern world from a multitude of computational perspectives. One of its main tools, Recogito, has shown great promise in the collaborative labelling of data for machine learning purposes, such as in the Machines Reading Maps project. From the Near East to the Mediterranean, texts and inscriptions in cuneiform and ancient alphabets share similar challenges not only in their digitization — which requires specialised optical character recognition models — but also in the philological nature of their study. Among the largest corpora, those of Greek and Akkadian share similar types of genres and reconstruction methods. But how available are these sources for advanced methods like machine learning models? Can we use knowledge from one language to improve these models for better results in the other, also called transfer learning? Can we handle the bias in models trained on limited or fragmentary corpora? These are just some of the challenges going forward that we will be addressing through shared annotation and corpus building under the umbrella of the Pelagios network.

Identifying community requirements for Linked Humanities Data infrastructures

Sarah Middle (National Museums Scotland), s.middle@nms.ac.uk

Duncan Hay (University of Sussex), dh418@sussex.ac.uk

Alex Butterworth (University of Sussex), A.Butterworth@sussex.ac.uk

It is likely that the application of LOD approaches to the Humanities will result in large, rich, heterogeneous datasets with huge research potential, but in turn raise a considerable challenge: how can these datasets be managed and published in forms that are flexible, scalable, interoperable and, critically, usable and sustainable? Several infrastructures exist that aim to address this issue, but each come with challenges of their own, including the balance between the need for off-the-shelf front-end solutions and for APIs that can support custom user interfaces, including explorable data visualisations.

Having faced these challenges in our use of Arches (https://www.archesproject.org/) for the Tools of Knowledge (https://toolsofknowledge.org/) project database, we sought to harness the knowledge and experience of the Pelagios and Linked Pasts communities to establish requirements for an optimum system. As a foundation for this research, we are organising an activity at the annual Linked Pasts symposium (https://www.seadda.eu/?p=180, December 2022) to generate discussion around the topic, which will take place both asynchronously and in person. Our results will form the basis of a white paper, to be shared with the Pelagios community for further input and refinement.

Our contribution to this panel will take a speculative design approach to our findings, envisaging a potential future system for managing and publishing large, heterogeneous Linked Humanities Data, which meets this community's needs. In addition, we propose to reflect upon our experiences of engaging in community-led, collaborative research, as well as the learning opportunities provided by knowledge sharing among the Pelagios Network.

Bibliography

Chiang, Y., Li, Z., Kim, J., Holmes-Wong, D., McDonough, K., Simon, R., Vitale, V. (2021). *Machines Reading Maps.* https://machines-reading-maps.github.io

Clivaz, C. (2019). The Impact of Digital Research: Thinking about the MARK16 Project. *Open Theology* 5/1 (2019), pp. 1-12. https://doi.org/10.1515/opth-2019-0001

Depauw, M. (2019). Trismegistos: optimizing interoperability for texts from the ancient world. In: *Crossing Experiences in Digital Epigraphy: From Practice to Discipline*. Poland: De Gruyter Open Poland, 2019, pp. 193-20. https://doi.org/10.1515/9783110607208-016

Depauw, M. & Gheldof, T. (2014). Trismegistos. An interdisciplinary platform for ancient world texts and related information. *Theory and Practice of Digital Libraries*: *TPDL 2013 Selected Workshops* (Communications in Computer and Information Science 416), Cham: Springer 2014, pp. 40–52. https://doi.org/10.1007/978-3-319-08425-1_5

Dunn, S., & Vitale, V. (2021). Linking text and maps: Annotation as a critical tool for teaching in the Spatial Humanities. *Literary Geographies*, *7*(2), 292-310. https://www.literarygeographies.net/index.php/LitGeogs/article/view/253

Gheldof, T. & Pietowski, F. (2018). Trismegistos Data Services: Digital tools for ancient world LOD networks. Linked Pasts IV, Dec 2018, Mainz, Germany. https://hal.archives-ouvertes.fr/hal-01955745

Kahn, K., Isaksen, L., Barker, E., Simon, R., de Soto, P., and Vitale, V. (2021). Pelagios – Connecting histories of place. Part II: from community to association. *International Journal of Humanities and Arts Computing* 15(1-2), pp.85-100. https://doi.org/10.3366/ijhac.2021.0263

Monier, M., and Nury, E. (2022). The SNSF MARK16 Project. The first virtual research environment focused on a biblical chapter. *Poster presented at the 2022 DARIAH-CH study day* (Mendrisio, 20.10.22). https://zenodo.org/record/7248296

Simon, R., Isaksen, L., Barker, E., and Cañamares, P. de S. (2016). Peripleo: a tool for exploring heterogeneous data through the dimensions of space and time. *The Code4Lib Journal* 31. https://journal.code4lib.org/articles/11144

Simon, R., Barker, E., Isaksen, L., and Cañamares, P. de S. (2017). Linked Data annotation without the pointy brackets: introducing Recogito 2. *Journal of Map & Geography Libraries* 13(1), pp.111-32. https://doi.org/10.1080/15420353.2017.1307303

Simon, R., Gadd, S., Rees, G., and Isaksen, L. (2022). Peripleo. https://doi.org/10.23636/sd7c-k310

Vitale, V., de Soto, P., R. Simon, Barker, E., Isaksen, L., and Kahn, R. (2021). Pelagios – Connecting histories of place. Part I: methods and tools. *International Journal of Humanities and Arts Computing* 15(1-2), pp. 5-32. https://doi.org/10.3366/ijhac.2021.0260