

N University of Applied Sciences and Arts Northwestern Switzerland

DARIAH-CH Study Day 2024

Basel Academy of Art and Design FHNW

November 22, 2024

Data pitches: Book of abstracts

Organising committee:

Linda Ludwig (FHNW) Tabea Lurk (FHNW) Cristina Grisot (DARIAH-CH) Rita Gautschy (DaSCH, DARIAH-CH) Tobias Hodel (UNIBE)

Table of contents

- 1. Adrian Demleitner (UNIBE/HKB) Labours of Love
- 2. Giacomo Alliata (EPFL) FAIR and CARE Principles in AI-driven Workflows: Distant Listening of the Radio Télévision Suisse
- 3. Aurea Dominguez (FHNW) Voices in Wax: Balancing FAIR and CARE Principles in the Context of Early Sound Recording Technologies
- 4. Dhruva Gowda Storz (eM+, EPFL) How FAIR and CARE principles can help shape the digital restoration and augmentation of historical stereoscopic photographs
- 5. Guillemette Legrand (HGK-FHNW) Climate Cosmograms
- 6. Francesco Beretta (UNINE) Experiences and Challenges of FAIR Data Reuse
- 7. Dominic Studer (FHNW) Herausforderungen im Datenmanagement und der Nachnutzung von Unterrichtsvideografie
- 8. Felipe Castelblanco, Ayënan Quinchoa Juajibioy (FHNW) *Plants_Intelligence: Learning Like a Plant*
- 9. Tsz-Kin Chau (EPFL) FAIR and CARE consideration for the Digital Scholarly Edition of the Panorama of the Battle of Murten (DIAGRAM)
- 10. Victoria G. D. Landau (UNIBAS) Papyrus Fragments as an Example for Distributed Ancient Heritage Object Data

Structuring labours of love

Adrian Demleitner (HKB BFH/Digital Humanities University of Bern) adrian.demleitner@hkb.bfh.ch

While the broader historical trajectory of video game graphics development is generally understood, there has been a notable lack of comprehensive, factual resources to systematically explore and analyse this history. To address this gap, I assembled the Video Game History Screenshot (VHS) dataset, to provide a map of early video game graphics. The VHS dataset comprises approximately 113,000 images sourced from roughly 4,300 games spanning the period from 1950 to 1990. The images are enriched with basic metadata attributes. This dataset draws exclusively from the visual material of MobyGames, complemented by contextualising metadata from Wikidata. Other comparable sources of such scope are non-existent.

This pitch presents the methodological details of assembling the VHS dataset, exploring the challenges posed by its inherent biases, and reflecting on the interplay between FAIR and CARE in the context of utilising MobyGames as a primary source and creating a structured dataset. It is crucial to acknowledge that the preservation of video game history has largely been unpaid labour, primarily undertaken by amateur communities. This reliance on volunteer efforts introduces inherent biases and limitations to the dataset. MobyGames, like Wikipedia, faces challenges stemming from a contributor base that is predominantly from the Global North, white, and male. Furthermore, without overarching curatorial guidelines, the content of the MobyGames archive is largely shaped by the interests and preferences of its community members. Further, access to the MobyGames platform is restricted due to its ownership by a private company. These factors complicate efforts for inquiries and indicate that the VHS dataset inevitably carries biases that are difficult to fully mitigate or quantify.

The utilisation of MobyGames as a primary source for the VHS dataset also raises ethical considerations regarding scientific inquiries that rely on community-assembled materials, particularly when the subject is deeply meaningful to that community. While the VHS dataset can adhere to FAIR principles, the infrastructural and governance differences between academic research and community-driven platforms necessitate a re-evaluation of its approach to CARE principles. While the dataset aims to unlock valuable information from a "walled garden," the act of extracting, copying, and restructuring the community's labour of love into a scientific dataset can be perceived as a form of epistemic violence, potentially disconnecting the data from its original context and the passionate individuals who meticulously curated it.

This presentation opens critical discussions about data collection, curation, and interpretation in the context of digital cultural heritage and born-digital research objects. It highlights the need for more diverse and inclusive approaches to documenting and studying video game history, as well as the importance of developing ethical frameworks for utilising community-driven archives in academic research.

FAIR and CARE Principles in Al-driven Workflows: Distant Listening of the Radio Télévision Suisse

Giacomo Alliata (EPFL) giacomo.alliata@epfl.ch

The intersection of the FAIR (Wilkinson et al., 2016) and CARE (Carroll et al., 2021) data principles present unique challenges and opportunities for managing large-scale digital archives. Within this context, I offer some reflections on a project I contributed to at the Laboratory for Experimental Museology, EPFL, on the "distant listening" (Clement, 2016) of the Radio Télévision Suisse (RTS), unlocking its rich cultural and historical content for both researchers and the broader public. This contribution will focus on how our work aligns with both FAIR and CARE principles, offering a new model for managing cultural heritage data.

Audiovisual archives such as the RTS repository, containing over 200,000 hours of footage, represent an enormous reservoir of Switzerland's collective memory. However, managing such archives in line with FAIR guidelines remains a challenge. Our project addresses this through the application of machine-learning techniques like speech-to-text conversion, speaker diarization, and named entity recognition (NER), transforming unstructured data into structured, interoperable formats. These methods make the archives more accessible, allowing humans and machines to find and analyse historical data efficiently, thereby supporting the FAIR principles.

The CARE principles, on the other hand, prioritise ethical considerations, particularly when managing data related to marginalised or underrepresented groups. In the context of the RTS archive, our pipeline can extract much more content than what was originally manually tagged by the RTS archivists. Furthermore, by connecting extracted metadata to open databases like Wikidata, we ensure that the derived insights can be shared equitably, promoting inclusive cultural development.

Our AI pipeline operationalizes audiovisual content using Natural Language Processing (NLP) models such as WhisperX (Radford et al., 2023), facilitating "distant listening" to audio recordings and enhancing metadata through large knowledge bases. A key outcome is the development of an interactive Map of Switzerland, visualising the geographical distribution of voices and linguistic diversity captured in the RTS archive. This tool exemplifies how FAIR-compliant workflows not only improve data accessibility and reusability but also contribute to collective cultural understanding. However, FAIR data practices are not without challenges. Ensuring long-term data preservation and interoperability requires robust and standardised communication protocols, particularly when dealing with audiovisual data that often lacks structured metadata. Additionally, while machine-assisted processes accelerate data discoverability, ensuring the ethical handling of data demands the careful application of CARE principles. To address this, our project links all named entities extracted to Wikidata, ensuring that data is not only findable and reusable but also managed with care and respect for its cultural significance.

In conclusion, our AI-driven workflow for the RTS audiovisual archive exemplifies the balancing act between FAIR and CARE principles. By enhancing the accessibility, interoperability, and ethical management of cultural heritage data, we create new possibilities for the Arts and Humanities sector to leverage these archives responsibly and innovatively, contributing to a broader discourse on the future of digital data stewardship in the GLAM (Galleries, Libraries, Archives, and Museums) sector.

References

Carroll, S.R., Herczog, E., Hudson, M. et al. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Sci Data 8, 108 (2021). <u>https://doi.org/10.1038/s41597-021-00892-0</u>

Clement, T. E. (2016). Towards a rationale of audio-text.DHQ: Digital Humanities Quarterly,10(3)

- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. International Conference on Machine Learning, 28492–28518
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Voices in Wax: Balancing FAIR and CARE Principles in the Context of Early Sound Recording Technologies

Áurea Dominguez (FHNW / Musik-Akademie Basel Hochschule für Musik) <u>aurea.dominguez@fhnw.ch</u>

The Voices in Wax project explores the earliest developments in commercially recorded music (1885–1915), focusing on phonograph recording technology from a technical perspective. By combining historical research with hands-on experimentation, the project aims to understand the conditions under which early recordings were made and their profound influence on the music industry. This research is essential for musicology and performance studies, as it provides insights into how these early recordings shaped musical styles and practices during the critical transition from the 19th to the 20th century.

A key challenge of Voices in Wax is its commitment to using original technology from the early 20th century, including authentic phonographs and wax cylinders. This historically based, hands-on approach is crucial for understanding the technical and acoustic limitations of early recording technologies, as described in historical sources such as publications and laboratory notes. The project will systematically experiment with various parameters, including material properties, machine settings, and recording environments, to establish a standardised framework that future researchers can use to produce comparable and repeatable results. Vocal performances in modern classical, jazz, and historical performance styles will be recorded on wax cylinders.

However, the use of original phonographs and wax cylinders introduces significant preservation and data management challenges. The physical artefacts—fragile wax cylinders and over a century-old phonograph—pose questions about the long-term sustainability of this experimental work. This challenge also intersects with the tension between FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles in data management. While FAIR principles promote the broad availability and reuse of data, CARE emphasises the ethical considerations of working with cultural artefacts. For Voices in Wax, striking a balance between these principles involves careful consideration of how to digitise and share the data generated through the project. Digitising wax cylinder recordings presents unique difficulties, as the process of transferring analog sound to digital formats may compromise the fidelity of the original recordings. How can we best capture the nuances of the original wax cylinder recordings in a digital format that retains their historical and acoustic integrity?

Additionally, the project faces the challenge of developing a standardised method for documenting the recording process itself. Given the experimental nature of the project and the uniqueness of each recording session, making the data interoperable and reusable is complex. At the same time, CARE principles demand that the cultural significance of the recordings is respected, ensuring that they are not merely reduced to technical data points.

Voices in Wax illustrates the intricate interplay between historical preservation, ethical data management, and technological experimentation in digital humanities research. By using original recording technologies and addressing the challenges of preservation and digitization, the project aims to create a framework that supports rigorous scientific inquiry while respecting the cultural and historical context of early sound recordings.

How FAIR and CARE principles can help shape the digital restoration and augmentation of historical stereoscopic photographs Dhruva Gowda Storz Laboratory for Experimental Museology (eM+), EPFL <u>dhruva.gowdastorz@epfl.ch</u>

My PhD research focuses on facilitating access to stereoscopic photographs through contemporary 3D and virtual reality systems. Stereographic photography sparked a global craze lasting nearly a century from the mid-19th to mid-20th century. In the wake of the medium's immense popularity, millions of stereographs documenting important historical events and perspectives survive in archives around the globe today, representing a priceless threedimensional archive of the past. My current research project aims to develop and assemble a number of AI tools that facilitate access to historical stereographs and encourage their dissemination through virtual reality platforms that are descendents of and natural contenders for viewing stereo photographs. The core of my proposed toolkit is an automatic restoration pipeline that improves the viewing comfort for stereographic photographs, some of which are heavily damaged, and prepares them for viewing on various virtual reality platforms. It will also include a series of AI augmentation tools that can be used to automatically 3D-reconstruct, Sonify, and recolor stereographs, and we also plan to create methodologies to link these images to commentaries or other textual information that they may be associated with through automatic NLP techniques. The goal of this toolkit is to enable the automatic processing of large batches of stereographic images, to address the temporal and financial restoration cost faced by large archives who seek to disseminate stereographs today. Due to the scale and ubiguity with which the proposed toolkit is intended to operate, it comes with a unique opportunity to help structure the data it processes in a way that better adheres to FAIR principles. Additionally, the AI tools we are developing, in particular those that modify the historical sources, raise interesting questions in the context of CARE principles.

The toolkit will be designed to work with any dataset of stereo photographs, but we use a collection of 1000+ stereographs from the universal exposition of 1867 as a case study in its development. This dataset alone has raised many questions on how FAIR principles can be incorporated into the toolkit to promote the long-term usability of data processed using the toolkit. To begin with, the dataset is inhomogeneous as it is sourced from a wide variety of archives, many of which use different metadata structures from each other, which raises issues with findability and interoperability within the dataset. Our toolkit presents an opportunity to address this by automatically homogenising or translating various metadata types used to incorporate all facets of FAIR into the data it processes.

The toolkit would also benefit from an examination of how CARE principles could be applied to help shape the AI tools it uses. Although the dataset is not directly about indigenous communities, stereographs were a vehicle for many in the west to travel through images, and as a result many indigenous communities and practices were documented through a colonial lens. The use of automatic AI restoration, recolouring and sonification techniques on such images risks perpetuating bias in the algorithms used and misrepresenting history or altering the authenticity of the historical material, and would benefit from an examination of how such algorithms can be shaped while keeping CARE principles in mind.

Climate Cosmograms

Guillemette Legrand (HGK-FHNW and Ecole des Arts Décoratifs ENSAD-PSL) guillemette.legrand@fhnw.ch

My PhD research examines ways to document and diversify climate-modelling and imaging practices, inquiring into how they become more than scientific knowledge, portraying political and socio-cultural imaginaries of the Earth. This practice-based research starts with the proposal that current climate images are not neutral representations but act as visual proxies of complex, autonomous, and speculative modelling systems that mediates, hierarchises, and politicises climate data. Climate images used within the Intergovernmental Panel on Climate Change (IPCC) reports are produced to inform climate governance and policy-makers, which, I argue, construct imaginaries that reframe people's political and socio-cultural relationship to the Earth. To study climate imaginaries, I define three realities of the Earth that co-exist in the simulation of climate: the Earth as a biochemical substrate, its computational modelling, and finally, the visualisation of this data. By identifying the operations of techno-scientific mediations, I aim to document how Earth becomes world(s) and how the world-imagination of Earth has a recursive effect on Earth itself. The research asks: what are the current practices that simulate and visualise climate and its scenarios? Which model(s), visuality, or simulation(s) of the Earth are required by governances and the broader public to act in the face of the climate crisis? The project hypotheses that using the medium of cosmograms-a cultural artefact incarnating the Earth-as a research lens, I can document, mobilise and diversify climate visuality and imaginaries.

On one end, I document the modelling and visualising infrastructure of climate-earth through software and infrastructure studies with climate models, codes, repositories and infrastructures but also through conversations centred around debugging models with the model maintainers, developers or users. On the other hand, I try to mobilise climate infrastructure through cosmograms as artistic and game practice, looking for other forms of computational and non-computational practices inquiring about climate-earth. I have listed below couple points about the inherent tensions in my research when it comes to organising, documenting, and sharing specific forms of knowledge, material and data:

- I study forms of knowledge-production and data collection that I do not ethically and politically align with. This raises questions about modes of annotating, publishing and communicating these data in specific academic and non-academic contexts.
- In the work of annotation, I am trying to reframe specific scientific material/data through the lens of cosmograms to open up the cosmological/socio-technical dimensions of these images. This opens questions around the further circulation of these images and how to propose counter-forms to inventories of these materials/data while not replicating forms of dominance inherent to world-earth-mapping but instead preserving ways to share and categorise these cosmographic materials as fragments, additions and possibilities for pluriversality.
- I am currently working with a simple climate model used to quickly assess socio-economic climate scenarios and their impact on temperature changes these models use very specific sets of data based on carbon emission and uptake calculation and are processed through a series of fundamental equations. This raises questions about the relationship between input data, computational processes and output of these models.

Experiences and Challenges of FAIR Data Reuse

Francesco Beretta (Université de Neuchâtel) francesco.beretta@unine.ch

As the first sentence of the seminal paper by Mark D. Wilkinson et al. 2016 states, "There is an urgent need to improve the infrastructure supporting the reuse of scholarly data". And further that "the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals". Looking at the field of digital humanities, we can observe that ongoing discussions often focus on the issue of publishing Open Research Data (ORD) according to the FAIR principles, in particular findability (F in FAIR) and accessibility (A). In this data pitch, I'd like to take the perspective from the other side, namely that of data reuse (R) and, from this point of view, raise some questions regarding the publication of research data which are related to the issue of interoperability (I) and the challenge of developing shared, discipline-specific semantics of the data, and the corresponding need for an appropriate research data infrastructure (Beretta 2024).

I will briefly present three use cases related to my teaching and research. (1) The reuse of a dataset presented in a data paper (Bell & Nehrbass 2022) (2) Answering research questions in a history seminar using Wikidata (Cachelin 2024) (3) Collaboratively producing reusable information about university history in the Geovistory community project "Academic Education and Careers" (https://www.geovistory.org/project/1483135). Based on these use cases, I will discuss the importance of developing a common semantics (as initiated in the sdhss.org ontology ecosystem) to make the data easily understandable and reusable (the I in FAIR), and of publishing the data after providing its enrichment and alignment with authority files (the R in FAIR). I'll then address the issue of the skills needed by researchers in the ETL process and data analysis to engage in this endeavour. Finally, I'll stress the importance of ORD, ideally published as Linked Open Data, which is the vision at the origin of the ontome.net, geovistory.org and dataforhumanities.org projects.

References

- Bell Arvid et Nehrbass Alexander. 2022. « Mapping Pre-Westphalia Europe: Military Alliances, Political Settlements, and Alignments During Armed Conflicts, 1528 to 1648: Social and Economic History », Research Data Journal for the Humanities and Social Sciences 7 (1), 12.12.2022, pp. 1-12 https://doi.org/10.1163/24523666-bja10028
- Francesco Beretta. 2024. « Données liées ouvertes et référentiels public : un changement de paradigme pour la recherche en sciences humaines et sociales », Arabesques (112), 2024, pp. 26-27, <u>https://doi.org/10.35562/arabesques.3820</u>
- Cécile Cachelin. 2024. Le champ de la mode vestimentaire dans le monde, du 19e au 21e siècle, GitHub: <u>https://github.com/czeacach/fashion_designers/wiki</u> (CC-BY-SA-4.0)
- Wilkinson Mark D., Dumontier Michel, Aalbersberg IJsbrand Jan et al. 2016. « The FAIR Guiding Principles for scientific data management and stewardship », Scientific Data 3 (1), 15.03.2016, https://doi.org/10.1038/sdata.2016.18

Herausforderungen im Datenmanagement und der Nachnutzung von Unterrichtsvideografie Dominic Studer (PH FHNW)

dominic.studer@fhnw.ch

Im vom SNF finanzierten Projekt untersuchten wir die Fragestellung, inwiefern eine globale Perspektive im Geschichtsunterricht umsetzbar ist und wie diese die beteiligten Lehrer:innen als auch Schüler:innen im Unterricht wahrnehmen. Die Lehrer:innen entwickelten in einem partizipativen Setting eigene Unterrichtseinheiten mit einer globalgeschichtlichen Perspektive. Die so entwickelten sechs Lektionen wurden dann aus jeweils zwei Perspektiven, derjenigen der Schüler:innen und derjenigen der Lehrpersonen, videografiert. Mit drei Schüler:innen und der Lehrperson wurden im Anschluss an die letzte Lektion ein «Stimulated Recall»-Interview geführt, wobei Ausschnitte aus der Unterrichtsvideografie abgespielt wurden und die interviewte Person dazu möglichst frei ihre Gedanken äussern sollte. Diese Interviews sind die Hauptdaten des Forschungsprojekts und wurden inhaltsanalytisch ausgewertet.

Im Pitch möchte ich auf Herausforderungen im Datenmanagement und der Nachnutzung der Videodaten eingehen. Bezüglich des Datenmanagements stellte vor allem die Speicherung der grossen Videodaten eine Herausforderung dar. Während der Datenerhebung musste ein Kompromiss zwischen einer möglichst hohen Videoqualität und einer bewältigbaren Dateigrösse gefunden werden, wobei die Qualität priorisiert wurde. Bis zum Projektende entstanden so ungefähr 200 GB an Videodaten. Da es sich um sensible Daten handelt, konnten diese nicht einfach auf dem regulären Arbeitslaufwerk der Hochschule gespeichert werden, weil dort keine zufriedenstellende Rechteverwaltung möglich war. Für die Projektlaufzeit sind wir deshalb auf Switch-Drive ausgewichen, was auch die Regelung der Zugangsrechte durch das Projektteam selbst ermöglichte.

Allerdings bleibt nun nach Projektende die Frage offen, wie, wo und ob die Videos langzeitarchiviert werden können. Werden die Videos auf einem hochschulinternen Speicher abgelegt, muss der Datenschutz weiterhin sichergestellt werden, selbst wenn alle Projektmitarbeitenden irgendwann nicht mehr an der Hochschule arbeiten. Hierbei sind die Zuständigkeiten an der Hochschule nicht geklärt. Damit einher geht, dass die Nachnutzung der Videodaten nahezu unmöglich ist. Zwar wären die Videos für viele weitere Fragen der Unterrichtsforschung, auch ausserhalb der Geschichtsdidaktik, interessant, sie können aber wegen des Datenschutzes nicht offen zugänglich gemacht werden. Eine vollständige Anonymisierung der Videos (durch Verpixeln etc.) würde gleichzeitig einen zu grossen Informationsverlust bedeuten und ist deshalb nicht zielführend.

Grundsätzlich führt dies zu einem forschungsethischen Konflikt, vor allem auch, weil die Aufzeichnung der Videos mit einem grossen Aufwand verbunden war, und es deshalb wünschenswert wäre, wenn die Daten auch für weitere Forschung zur Verfügung stehen würden. Selbst wenn nun entschieden werden sollte, dass die Videodaten definitiv gelöscht werden, stellt sich auch hierbei die Frage, wie dieser Prozess möglichst transparent dokumentiert und nachvollziehbar gemacht werden kann.

Plants_Intelligence: Learning Like a Plant

Felipe Castelblanco & Ayënan Quinchoa Juajibioy (Institute Art Gender Nature of the Basel Academy of Art and Design FHNW) <u>felipe.castelblanco@fhnw.ch</u>

My contribution to this project is rooted on the issue of how plants mobilise human assemblies by providing healing, nourishment, and other cognitive capacities that enhance human sensibilities toward plants' agencies in the Andean-Amazon foothills in both Colombia and Peru. Through a practice-based and participatory research process I conduct extensive fieldwork and accompany tropical ecologists, ethnobotanists, agroecology experts, farmers, Indigenous healers, gardeners, and land-defenders to observe how plants make their knowledge intelligible to these actors through plant-based medicine, nourishment, smell, and ingestion. In doing so, I maintain a close collaboration with Indigenous co-investigators, like Ayënan Quinchoa Juajibioy with whom we co-develop and co-produce certain research outcomes that could be better integrated to the CARE Principles for Indigenous Data Governance.

In our work, we use methodologies from documentary film practices and participatory filmmaking. Therefore, in the process we collect many hours of interviews, extensive fieldwork notes in the mode of videos, drawings, diagrams and voice recordings. Some of the challenges I face in handling the data are as follows:

- Identifying a common platform and data-sharing protocol that provides equitable data access to local partners in the Andean-Amazon, particularly where internet connection speeds, low storage capacity devices and the prominence of mobile phones over computers prevail.
- 2. Acknowledge the different degrees of Media literacy and access of my collaborators, which limit the use of certain server-based digital services.
- 3. Annotating video content, commenting and labelling raw or unedited video material so it stands as evidentiary material and can serve to local communities in instances of legal action.
- 4. Developing a file cataloguing and adequate storage system that is easy to transfer once the project is finished and results are returned to local stakeholders in the rainforest.
- 5. Identifying a data stewardship model to highlight the importance of preserving these works once they are returned to the communities or partner institutions.

Finally, my work has brought into focus the need to develop more bilateral avenues for knowledge exchange between Traditional Knowledge and academic expectations over research data. This, in order to enable more room for anecdotal, symbolic, and spiritual descriptions of Indigenous cosmologies in plant-related research. However, while for the academic context my data might be of intellectual interest, for Indigenous stakeholders the same data possesses cultural and symbolic value, thus making even more complicated its uses.

FAIR and CARE consideration for the Digital Scholarly Edition of the Panorama of the Battle of Murten (DIAGRAM)

Tsz-Kin Chau (Laboratory for Experimental Museology eM+, EPFL) <u>tszkin.chau@epfl.ch</u>

The Murten Panorama (1894) by Louis Braun (1836-1916) is an illustrative example of visual heritage characterised by both its vast scale and intricate detail. Depicting the Battle of Murten on June 22, 1476, during the Burgundian Wars fought between the old Swiss Confederacy and Charles the Bold, the Duke of Burgundy, the Murten Panorama stands as both a Swiss national treasure and a global piece of panoramic visual heritage.

Physically measuring approximately 10 x 100 meters, the panorama was digitised in 2023 at 1,000 dpi and fully processed in June, 2024 under the initiative of the DIAGRAM project, creating a 1.6 Terapixels digital twin. The content within the panorama is exceptionally diverse, encompassing a wide range of named geographical locations, historical characters, heraldic representations, recognized historical events, providing an immense opportunity for linked data annotation which motivate the digital scholarly edition project of the Panorama of the Battle of Murten.

Using linked open data and ontologies as our backbone, we are mindful of the typical challenges in semantic web technologies—such as the 'link rot' of ontologies and datasets after project completion (resulting in inAccessibility), or the obsolescence of dataset repositories (leading to 'unFindability'). At this early stage of the project, we are reviewing our best options for FAIR LOD in 2024 from the perspectives of both a DH research project and a Swiss DH research initiative.

In line with CARE principles, we promote inclusive representation in LOD vocabulary and ontology terms. The Murten Panorama scholarly edition project aligns with UNESCO's Memory of the World nomination program, where panoramas as 18th-19th century historical media will be preserved, promoted, and researched using critical media archaeology. The nationalism and colonialism embedded in the golden age of this medium will be acknowledged, deconstructed in the digital edition through a data-driven approach. We are actively developing our CIDOC-CRM-based data model to disentangle, document, and highlight social discourse found in the panorama. We are enthusiastic and looking forward to exchanging and learning from the data experts.

Papyrus Fragments as an Example for Distributed Ancient Heritage Object Data

Victoria G. D. Landau (Digital Humanities Lab, University of Basel) victoria.landau@unibas.ch

In the digital sphere, ancient papyri are inventoried and presented – ideally in the form of textual metadata and images – in one of two ways: a) in a manner closely reminiscent of print editions, or b) standardised to match other object types in the care of their holding institutions. Comparability across institutions and particularly institution types is therefore a challenge. These are usually GLAM institutions (galleries, libraries, archives, museums) with different specialisations and prioritizations.

The dataset being produced in the context of the speaker's ongoing doctoral research on ancient heritage digitization intends to offer for the first time a directory of known papyrus-holding institutions permitting this comparability worldwide. The data is being assembled at the institutional level, categorising available and unavailable information on the digitization status for the papyrus collections in each institution's care. Papyri have been distributed across more than 39 countries, a historical spread which has been well-documented to a certain degree. This data is partially collected, partially accessible and partially usable, owing to a number of challenges. FAIR:

Findability and accessibility of existing ancient heritage data is hampered by «siloization», in large part due to the nature of funding: financed on either a project-basis or on an institutional basis, interoperability beyond these constraints has not always been automatically encouraged. Standards are slowly being considered and implemented, as seen in a growing use of cultural heritage metadata schemata such as CIDOC CRM and Dublin Core. Cross-linking between projects and platforms is a work-in-progress and is not always done effectively. This pace has impacted the reusability of digitised objects, usually facilitated by newly creating or improving available metadata to be both machine-readable and usable to scholars of the ancient world. CARE:

While the heritage objects in question are ancient, there are living communities in their modern countries of origin that have a vested interest in them and educated scholars that do not have equal opportunity to access and research them. Information related to dating (e.g., periodization), location (e.g., findspots, sites mentioned) and provenance (e.g., ownership history, collection connections) is challenging enough to standardise or harmonise on its own, but this work is also usually conducted and enforced by academics removed from the context in which objects were produced and rediscovered. For items long coveted by commercial antiquities markets, scholarly ethics and responsibility must take center stage to ensure custodianship of both physical and digital versions of the ancient heritage objects is undisputed and future uses are possible open access.

Knowing where the data is (i.e., at institutions holding papyri) and what each collection's digitization status is, offers a first step in enhancing and connecting available data, ideally working towards equitable access for scholars and members of the interested public alike.