DARIAH-CH Study Day

October 20, 2022, Università della Svizzera italiana, Mendrisio

Book of Abstracts

Plenary presentations

Reliable Semantic Indexing of Historical Newspapers at Scale: Are We There Yet? Findings from the impresso Project and Future Perspectives
Maud Ehrmann (EPFL/UNIL)

Following the decisive efforts led by libraries to digitize newspaper collections, research initiatives to apply computational methods to historical newspapers at scale have recently multiplied. In this context, the interdisciplinary project '*impresso* - Media Monitoring of the Past' brings together a team of computational linguists, designers and historians to collaborate on the datafication of a multilingual corpus of historical newspapers. The main objectives of the project are to improve text mining tools for historical text, to enrich historical newspapers with automatically generated data, and to integrate such data into historical research workflows by means of a newly developed user interface. Beyond the challenges specific to the different research areas underpinning each of these goals, the question of how best to adapt text mining tools and their use by humanities scholars is at the heart of the *impresso* enterprise. In this talk, I will present the challenges of processing and mining large-scale collections of digitized newspapers, share hands-on experience and discuss our efforts to overcome them, and highlight key priorities for future developments around digitized historical newspapers.

~

From Vision to Application : DARIAH's Services to Enable Researchers in SSH
Edward Gray (DARIAH EU)

Founded as an ERIC in 2014, DARIAH has long held a place in the SSH Research Landscape. Like any research infrastructure, or indeed any type of infrastructure, its work often passes unseen. That does not mean that DARIAH has been inactive, it simply means that the work carried out by the infrastructure focused more on the construction of our international research infrastructure and of the various national consortia, as well as policy and advocacy on behalf of our communities.. DARIAH also served, and serves, as an excellent incubator for European project grants and the creation of project consortia. These important initiatives, however, are rarely seen or appreciated by the individual researcher. However, with the increasing maturity of the infrastructure and notably the SSH Open Marketplace and DARIAH Campus platforms, DARIAH can proudly boast services that appeal to the individual researcher. Joining with the existing, grassroots and researcher-led Working Groups, DARIAH now has multiple means of enabling research in social sciences and humanities in a way that is visible to the individual researcher. Our important unseen infrastructure work continues, and it is joined now by these more visible and tangible initiatives.

~

Project presentations

Towards Computational Historiographical Modeling: Concepts in Art History – Floor Koeleman (UNIL)

In February 2022, the SNSF project "Towards Computational Historiographical Modeling: Corpora and Concepts" was launched. This four-year project aims to lay methodological foundations for digital humanities research. In the field of digital humanities, the phenomena studied are constructed and modeled through data and code. Our goal is to investigate the epistemological implications of imposing this structure on the phenomena under study, specifically in the context of historical research. By analyzing at a meta-level the causal relations that historians establish, we attempt to formalize the scholarly discourse. My sub-project is concerned with the manifestations of concepts in art history. Especially the historical art of painting is about shaping concepts, because it renders the invisible visible: paintings give visual form to abstract ideas. Central to art historical research is the identification, description, and interpretation of these concepts in the visual arts. However, the ideational character of concepts renders them unstable and subject to change over time. As a result, concept driÀ can be observed both at the level of the artworks and the historiography of art. In order to provide insight into the processes of meaning-making, concepts will be understood as models. On the one hand, painted representations of concepts can be regarded as explicit models, albeit complicated by the historical distance. On the other, the use of models in art historiography still remains largely implicit in the absence of overt methodological reflections. The revision and formalization of concepts as models in these contexts will allow them to be tested and compared. This paper focuses on presenting the challenges and discussing possible approaches.

~

Geovistory ecosystem - ORD practices made easily accessible
Francesco Beretta and David Knecht (UNINE, KleioLab)

In the Humanities and Social Sciences (HSS) open research data practices are quite limited. New data management systems are being developed with different approaches and potential to make data reusable. (1) There are virtual research environments that come with rich visualization features but data are kept in silos and information across projects is not interconnected, nor data models aligned with standards, which greatly hampers the reuse potential. (2) Projects use Wikidata as a common knowledge graph to produce linked open data. This solution has the limitation that the data model is not based on established semantic modeling methods but rather on the spontaneous conceptualization of contributors. Hence, the meaning of data is not consistently defined and data reuse is challenging. A new approach was initiated by the Digital History team of "Laboratoire de Recherche Historique Rhône-Alpes" with the symotih.org project and, together with KleioLab, a Swiss startup, further developed into the Geovistory ecosystem (geovistory.org). Geovistory provides two access points. First, a Toolbox, for researchers to produce, curate and analyze data. Second, a freely accessible data publication platform, in line with common Open Research Data requirements and the Linked Open Data vision. The data model is defined in OntoME according to CIDOC CRM (ISO21127) and an ecosystem of domain specific extensions (sdhss.org). Geovistory contributes to a growing knowledge graph for historical research providing open data for academia, GLAM institutions and the larger public. It thus bridges the gap of (a) easy-to-use multi-project research environments and (b) collaboratively curated, reliable socio-economic data to understand global contexts. Geovistory allows dealing with complexities, inconsistencies and contradictions within data. It innovates in the field of information systems for HSS in publishing open data that are interlinked with authority files and expressed with semantically clearly defined, extensible data models.

~

Images in Global Circulation: A Multi-scalar Approach
Nicola Carboni and Béatrice Joyeux-Prunel (UNIGE)

Pictures are the tangible traces of our visual world. They are the materialization of a classifiable, datable, and localizable object; a vocabulary of expressions that, when examined on a large scale, becomes key to understanding changes in visuality over time and space. This is precisely the goal of Visual Contagions, a project at the University of Geneva that aims to use computational analysis to understand what images circulate and how some of them become iconic. To achieve this goal we have developed a novel multi-scalar analysis that combines computational exploration of visual similarities with spatiotemporal classification. The method is applied to a global and diachronic corpus of illustrated periodicals, using algorithms to compare images extracted from the pages and group them into image-types: Vectors of visually similar images. The components of the image-types are then interconnected using their spatiotemporal properties and examined using a combination of digital and traditional art historical methods. The initial result will lead to a panoramic study of globalization through images, encompassing both their geography and their structural and conjunctural logics of circulation. This study will help to make clear the nature of the iconic, how it manifests itself and how it circulates. Using the outcome from the distant reading, the identified iconic features will be further explored drawing from traditional historical methods, thus making evident the aesthetic, social and historical reasons for the emergence and diffusion of the iconic. The presentation will showcase the first results of the project, focusing on the analysis of a large global corpus of images (120 countries – 6 million pictures) from illustrated magazines published between 1890 and 1990, and showcasing what circulated most, when, how, and through what type of media.

~

Atlante Calvino. Literature and Visualization
Valeria Cavalloro, Virginia Giustetto and Margherita Parigini (UNIGE)

The project *Atlante Calvino. Literature and Visualization*, funded by the Swiss National Science Foundation and directed by professor Francesca Serra, conducts extensive experimentation on the interactions, limits, and mutual enrichment of the two fields of literary criticism and data visualization. In order to explore how these two areas of expertise could advance each other, the project brought together a team of literary scholars (Unité d'italien, Université de Genève) and information designers (DensityDesign Lab, Politecnico di Milano), who worked alongside to develop an interactive web platform. This platform allows viewers to discover a set of nine interactive visualizations organized in three *itineraries*, centered around three major issues of literary interest in the contest of Italo Calvino's narrative writings: the formalization of the *doubt*, the exploration of *space*, the construction of the *plot*. Each itinerary confronts progressively more complex aspects of literary analysis, always referring to their overarching main issue. The peculiarity of the project lies not only in the visual elaborations but also in the *criteria* and methods underlying the previous moment of data mining. A choice that resulted in what we have called *human data mining*, in which the process of extracting desired data from the texts is transferred from automated software to human interpreters who can bypass the limitation of current AIs in identifying abstract and elusive literary features. All the final visualizations are the result of an intensive stage of intermedial and interdisciplinary translation, through which the findings of the reading phase were *made visible* by a thorough process of conversion from raw literary observations into defined, finite, and organized visual elements. After a general presentation on the structure of the *Atlante Calvino*, we will focus on one of the nine visualizations, which will work as a case study on the type of visual artefact that can be produced by a similar critical approach.

~

Untangling the cordel

Elina Leblanc and Pauline Jacsont (UNIGE)

The project *Untangling the cordel* (2020-2023) is dedicated to the study of a collection of 19th century Spanish chapbooks preserved at the University Library of Geneva (Leblanc and Carta 2021). Chapbooks have existed since the beginning of printing. Sold in the streets, they relate fictitious or real events, songs, plays or religious writings. Even if their topics are numerous, their editorial form is consistent. They are composed of a few pages, in in-quarto, arranged in columns and decorated with woodcuts (Botrel 2001; Gomis and Botrel 2019, 127–30). Due to their specific layout and the diversity of their content, chapbooks are complex documents that involve many levels of interpretation (book and art history, literature, linguistics, sociology, etc.). To represent this diversity and allow varied analysis, we have elaborated a digital library (DL) with digital scholarly editions (DSE) of chapbooks, and a catalog of woodcuts. Regarding the DSE, we transcribed the texts with different HTR tools. Then, we converted the results in XML-TEI via XSLT. The encoded texts have been stored in an eXist-DB XML database, and the images in the IIIF server of the Geneva University. Finally, we have published the DSE with the TEI-Publisher web application. The woodcuts catalog has also been encoded in XML-TEI. The main distinctive features of this catalog are 1) a list of similar woodcuts within our corpus, that have been detected with the VISE visual search software, and 2) information about the chapbook they come from. Through this workflow, the DL offers services that stress different aspects of chapbooks, i.e. document, text or image, and their close relationships: parallel consultation of transcription and facsimiles, comparison of documents with Mirador, full-text and thematic searches, links between text and woodcuts. Later we will provide further services through a work on named entities and the detection of similar woodcuts in a larger corpus.

~

Bullinger Digital – The Transformation and Expansion of an Analogue Edition into the Digital Age

Philip B. Ströbel (UZH)

Heinrich Bullinger (1504-1575) was an important multiplier of the ideas of the Reformation in Switzerland and Europe, as evidenced by his correspondence of 12,000 letters. 3,100 letters had already been manually transcribed, edited, and published in printed form in efforts that spanned three decades. Another 5,400 letters had been transcribed by various scholars and are available as electronic texts. Our project, Bullinger Digital, makes the letter collection accessible to researchers. The aim is to provide images of manuscripts, transcriptions (from printed editions and the electronic versions), and translations on an online platform that allows efficient searches through the collection and facilitates research in many domains.

To transform an analogue edition into a digital one and expand it, we conducted the following steps:

1. Acquisition of metadata: We extracted and corrected the information from the digitized images of the manually created index for non-edited letters from the 1960s to build a metadata database with the help of a citizen science initiative.
2. Data formats: We received the already edited letters and scholarly transcriptions in various formats (TUSTEP, HTML, PDF, or TXT files). We extracted the metadata, summaries, texts, and almost 100,000 footnotes from the edited letters and integrated this information into the metadata database. We enriched the database further with texts from the non-edited scholarly transcriptions and stored everything in an XML format.
3. Scan verification: We verified whether the scans delivered by the Zurich State Archives and the Zurich Central Library matched the metadata with the help of a dedicated scan verification interface.
4. Handwritten text recognition (HTR): We trained an HTR engine using the already transcribed Bullinger letters as training data for the letters for which no transcription is available.

5. Machine translation: We developed a machine translation system that translates the contents of letters written in Latin to modern German for the ease of access to foreign-language text passages.

~

Graph-based Digital Edition: Interim Report from the Semper Project
Elena Chestnova (USI)

Graphs have been postulated for some time as an alternative to TEI XML in digital editions. The advantages they offer are well known: greater flexibility in representation of real-world textual phenomena and variants, greater ease of integrating LOD. Practical applications to date have included small-scale examples and applications with partial use of graphs (e.g. for representing metadata or annotations). But how does a graph-based digital edition actually work out if we decide to represent the whole of the text as a graph? The edition of Gottfried Semper's manuscripts www.semper-edition.ch has done just that. It is a medium-scale (ca. 15'000 pages) digital edition of manuscripts and print text, produced using a hybrid workflow and using a Neo4j database as the basis for its final productive environment. The edition is currently half-way through its projected duration and this presentation will give insight into its experience of working with graphs so far.

~

hallerNet: methods and perspectives of a data-centric edition platform
Lukas Heinzmann and Christian Forney (UNIBE)

hallerNet makes accessible knowledge networks in Switzerland in its European exchange during the transition from a Republic of Letters to a Scientific Community (1700-1850). It is based at the University of Bern and carried by the «Albrecht-von-Haller-Stiftung». The platform was developed from 2016 to 2019, when a large number of metadata from two completed research projects has been converted from a relational database to an XML/TEI-based system. Starting from the Bernese polymath Albrecht von Haller (1708-1777) as central actor and name giver for the platform, hallerNet strives for a Swiss-wide perspective. As a data-centric edition platform, it combines extensive and highly interlinked metadata with digital editions. Currently, it comprises about 100'000 objects (e.g. people, publications, letters or plants) which are linked in various ways via authority files, Metagrid, correspSearch and others. *hallerNet* is the foundation for an ongoing edition project (2018-2023) funded by the Swiss National Science Foundation that aims at editing 6,000 (of about 17,000) surviving letters from Haller's correspondence as well as all of his reviews (9,000) published in the «Göttingische Gelehrte Anzeigen». This venture is enabled by an editing environment which is made up of several components: a customized Oxygen XML Editor with forms and schema validation, a SOLR-powered search, an IIIF Server to deliver high-quality images, and a backend built on the open-source software XSLWeb. Just as crucial as the technical environment are conventions and the documentation which are tracked comfortably in Markdown (Git and codiMd). In our presentation we would like to introduce some workflows applied within the project regarding transcription and encoding of text as well as data processing and analysis. Also, we would like to highlight visions for the future of hallerNet, in the direction of a more generic and Swiss-oriented platform.

~

Infrastructure pitches

Swiss National Data and Service Center for the Humanities (DaSCH)
Rita Gautschy (DaSCH)

This contribution will introduce the infrastructure and services of the Swiss National Data and Service Center for the Humanitites (DaSCH) for researchers and research ITs. DaSCH develops and operates a FAIR long-term repository and a generic virtual research environment for complex and simple open research data in the humanities in Switzerland, including law and theology. The primary goal of our platform is to guarantee direct access to the research data: it brings your data to life and keeps it alive in the long run. At the same time, it lets you edit, delete and enrich your data, even after it has been archived. Each object within a dataset has its own persistent identifier to allow reliable citability. We set value on interoperability with tools used by the Humanities and Cultural Sciences communities and foster the use of standards. The data is also accessible via an API, which allows computer scientists to collect data in an automated way. Our services for researchers and the community include hands-on training in the use of the DaSCH infrastructure, workshops thematizing frequently asked questions by researchers when writing a data management plan, participation in lectures, or workshops about best practices in the management and (re-)use of qualitative data in humanities research. As the coordinating institution and representative of Switzerland in DARIAH members of DaSCH actively engage in community building within Switzerland and abroad.

~

Text Crunching Center (TCC) at UZH's LiRI
Gerold Schneider (UZH)

This talk introduces the Text Crunching Centre (TCC), also known as NLP group, which is a Computational Linguistics and Digital Humanities service hosted at the Linguistics Research Infrastructure (LiRI) of the University of Zurich. We present a selection of our case studies using text analytics from social, political and historical studies. We show how stylistics, document classification, topic modeling, conceptual maps, and further methods can offer new perspectives. We also discuss opportunities and challenges of our service.

~

Iconoteca database
Silvio Bindella and Elisabetta Zonca (Academy of architecture Library)

Iconoteca went live in May 2020 after two years of study and intensive work to create a digital repository suitable to preserve and disseminate our photographic collection. Biblioteca dell'Accademia holds 100.000 ca. analogic photographic items, most of them rare and unique pieces that are part of archival and bibliographic fonds: they are related to a specific cultural context that we need to valorize through a scientific description of each image. Scientific is archived using international standards for cataloging, applying schemas for structure and syntax of metadata, creating controlled lists, vocabularies and thesaurus. As we decided to run an Italian/English web site – our academic community is international – we choose to rely on Getty Vocabulary Program for the English authorial references; Italian key words are taken from both from Nuovo soggettario Thesaurus and Getty itself, because we are officially contributors for their vocabulary program. We want to take care of our photographic collections, of their materiality, documental value and cultural

significance, using technologies to valorize these aspects and to share them with students, scholars, and curious people, because our task as librarians is to create opportunities to improve knowledge."

~

Knowledge Graphs, Ontologies and Computational Analysis for Cultural Heritage at Swiss Art Research Infrastructure (SARI)
Olya Nicolaeva (UZH)

Despite the rapidly growing number of cultural artifacts being made available from archives, collections, and museums (GLAM) and the strong support for digital research components within the scientific community, researchers still face fundamental challenges in leveraging the full potential of digitized archival resources. The lack of accessibility, of joint ontological foundations and of easy-to-use analytical toolkits makes meaningful and efficient use and interoperability of holding and research data cumbersome both for domain-specialists and computer scientists. Swiss Art Research Infrastructure (SARI) strives to provide comprehensive solutions for data-centered research in the Cultural Heritage domain: through data alignment and data modeling to richly populated knowledge graphs enhanced with reference data and computational analysis results. But how exactly is the path to Linked Open Data being paved? In this practice-based presentation we will present our everyday approaches for ETL (Extract, Transform, Load) processes and computational methods for data enrichment across cultural collections and digital archives.

~

SWIss PipelinE for textual data (SwiPe)
Simon Gabay (UNIGE)

SwiPe (SWIss PipelinE for textual data) Textual data represent an essential part of those produced and used by Swiss researchers, but also by Swiss heritage institutions (museums, archives, libraries, etc.). Digitisation of prints and manuscripts has considerably increased the amount of possibilities regarding corpora creation, digital editions and dataset production. Yet, it has also created new challenges: that of an accelerated transcription of texts (especially for ancient languages and old scripts) using open methods, that of their semantic description (which makes it possible to query, analyze, and exchange the texts) producing interoperable datasets, and that of their Open Access (OA) publication. We propose a complete pipeline for efficient production, management, description, analysis, publication, and distribution of textual data, in compliance with international standards and OA rules. The SwiPe (Swiss Pipeline for textual data) will bring together infrastructure components already developed by different Swiss institutions, and will accept interoperable image formats (IIIF) now published by most heritage institutions as a source. Our project should advance Open Research in the humanities, but also open digital engineering. The location of the complete pipeline on Swiss soil allows the management of sensitive data, as well as the digital empowerment of institutions, scholars and students.

~

Poster presentations

Space Heroes, an Interactive Installation as Example of the Embodied Transcription of an Audiovisual Archive

Giacomo Alliata (UNIL/EPFL)

Space Heroes is an interactive installation developed to explore the Claude Nicollier Video Archive, a collection of videos about the only Swiss astronaut - Claude Nicollier. The application is designed for the Linear Navigator, a 4k touch screen mounted on a twelve-meter rail, mapping the timeline of videos to the physical range of motion of the screen. This embodied metaphor offers users the ability to walk along the memories of Nicollier, augmenting their experience with respect to a traditional web-based interface. Participants transduce the archive into a virtual landscape of possibilities through their embodied interactions with the system. To be presented in a museum exhibition, this installation encourages social interactions between visitors, where one participant interacts with the system and acts as a performer for the other persons. Astrophysics concepts and data are leveraged to ground the installation in its setting of space exploration. This poster will outline the design process and frame Space Heroes in terms of its engaging paradigm of embodied interaction that operationalizes the archive into a mediated experience. It emphasizes the benefits of browsing rather than searching, responding to the need for innovative frameworks to explore audiovisual collections, and leveraging the idea of embodied understanding. The process of interaction is discussed to define how viewers can create their own narratives and dis-cover this valuable collection through the embodied metaphor of walking along Nicollier's memories, afforded by the Linear Navigator.

~

Back and Forth from Boundary Objects to IIIF Resources: The Recipes of a Community-driven Initiative Specifying Standards

Julien Raemy (UNIBAS)

The exchange of digital objects and their associated metadata is simplified when these meet established standards, but the capture of all the (meta)information is still very much in tension, at the limits of resources, knowledge and indeed the underlying capabilities of given standards. These limitations can be translated into what Susan Leigh Star defines as residual categories and consequently the generation of boundary objects. The question of these non-standardized residuals within the cultural heritage and digital humanities fields is an iterative identification issue that institutions and individuals have sought to mitigate. Take for instance resources conforming to the application programming interfaces (APIs) of the International Image Interoperability Framework (IIIF). These are JSON-LD serialized objects duly specified and vetted by a community trying to break institutional silos. Of particular interest are resources compliant with the IIIF Presentation API which underpins what a Manifest is, i.e. a description of the structure and properties of an object which can be interpreted by a client and displayed to end-users, typically disseminated openly on the Web. While on the surface or theoretically these IIIF Manifests, which are also referred to as *compound objects*, could be quite the antithesis of boundary objects, it remains to be seen to what extent IIIF-compatible resources revolve around residual categories and also from what point onward these Manifests have been at some point or revert to boundary objects, whether they are not well-structured or simply because the architecture and constituent software serving and interpreting them do not operate correctly.

~

Open Science for Art, Design and Music (OS-ADM)

Chiara Barbieri, Iolanda Pensa, Erzsébet Tóth-Czifra, Marta Pucciarelli, Candeloro Jean-Pierre, Suzanna Marazza and Anna Picco Schwendener (Swiss schools of Art, Design and Music)

According to the 'Open Access action plan 2021–2024' compiled by swissuniversities, the outputs of all publicly funded research projects in Switzerland will be released in Open Access by 2024. However, when it comes to Open Access the disciplinary fields of Arts, Design and Music present a series of complex issues. Specific challenges include the reuse and distribution of third-party content under copyright and the production of a wide range of multimedia outputs. This situation makes the implementation of Open Access in these disciplinary fields particularly complex. The research project Open Science for Art, Design and Music (OS-ADM) aims at supporting the implementation of the 'Open Access action plan 2021–2024'. Led by SUPSI and funded by swissuniversities, it involves a network of Swiss schools of Art, Design and Music – HES-SO (ECAL, Lausanne; HEAD – Genève; EDHEA, Valais), ZHdK (Zurich), HSLU Hochschule Luzern – Design & Kunst (Lucerne), BFH (Bern) and FHNW (Basel) – in collaboration with national and international stakeholders – CCDigitalLaw (USI, Lugano) and DARIAH-EU, among others. Over a three-year period, the research team produces guidelines, workflow charts, templates and solutions for a selection of case studies (2022), organizes training (2023), and promotes and negotiates Open Access among national and international publishers (2024). At a local level, the different schools contribute a series of case studies, implement the guidelines, organize training, and include Open Access within their students' curricula. The project goes beyond generic Open Science advocacy as it addresses shared, domain-specific needs emerging from case studies such as obtaining agreements for the reuse of cultural heritage data, sharing rich multimedia resources with shared ownership, providing Open Access to arts monographs retrospectively, advocating data citation practices and Data Management Planning, or helping research teams throughout the project life cycle to work with sharing in mind.

~

Teach historians how to design a data story: Stadt.Geschichte.Basel

Cristina Münch-Wildisen, Nico Görlich and Moritz Mähr (UNIBAS)

Stadt.Geschichte.Basel, more than 70 historians explore the history of Basel from the Celts to the present. The focus is on current and little-researched topics such as the industrial and commercial history of the 19th and 20th centuries or the history of migration. The city's history is not viewed in isolation, but is regionally and internationally interwoven on an economic, political and cultural level. It is published in 10 printed volumes. In addition, selected aspects will be presented as Data Stories. To make this possible, data must be collected, secured and processed. Researchers need to be trained in how to tell a story with data and use digital tools. A field report on the technical, organizational, and social challenges of teaching data literacy from the research data management team.

~

DARIAH-EU WG Digital Numismatics & Numismatic Linked Open Data in Switzerland

Rahel C. Ackermann(Swiss Inventory of Coin Finds) and David Wigg-Wolf (Deutsches Archäologisches Institut)

Numismatics is among the best Linked Open Data providers in the Humanities. These initiatives are supported and carried out by the international numismatic community, which consists of institutions and numerous individuals. Swiss Numismatics – and Swiss numismatists – are well integrated into these

international initiatives. Among others, the online portal for Swiss Numismatics OSCAR is one of the pioneers, providing medieval and modern numismatic normdata, the Münzkabinett Winterthur is the Swiss satellite in the ikmk-Verbund, and the Swiss coin finds will be presented in the Cultural Heritage management system Dédalo. This contribution will give an overview of the numismatic LOD initiatives in which Switzerland is an active partner, and will show how we can go further, linking up with other disciplines of the Humanities and their data pools.

~

How to Orient Oneself in Thinking in the Digital Era. A Mixed-Methods Ethnography of Researchers' Navigational Practices on Gallica

Simon Dumas Primbault (UNIL/EPFL)

The birth and development of digital libraries—broadly understood as curated collections of electronic documents accessible online on dedicated platforms with tools for search and consultation—have radically transformed research activities. Consequently, traditional practices observed in physical places of knowledge such as libraries and archives—searching catalogs, browsing through shelving, taking notes—have been supplemented with a series of digital practices developed by researchers to browse websites and databases—searching by keywords, filtering results, navigating through links. The ambition of this project is to reclaim a spatial understanding of digital libraries through the investigation of *navigation* as a fundamental information practice for researchers. Taking Gallica as a case study, this research situated at the crossroads of ethnography, data science, and digital humanities strives to shed light on how researchers "orient themselves" within a digital corpus. Bridging the gap between the *ethnography of the digital*—the qualitative study of scholars' practices through observations and interviews—and *digital ethnography*— the quantitative analysis of navigation traces—, it developed mixed methods combining interviews with topological analysis of navigation paths extracted from server logs. This exploratory study led to promising findings on researcher's navigation within the Dewey classification, on the role of "pivotal literature", and on a first sketch of a variety of "regimes of navigation" clustered based on their topological features. More importantly, this research illustrates a perfect case of heuristic dovetailing between a quantitative approach enhanced by digital tools, and a more traditional qualitative approach. Along the course of this ongoing project, emphasis has been put on a reflexive critique of the nature of the data processed, the implications of every step of the pipeline, as well as their link with actual practices observed and objectified by researchers themselves.

~

Generating Perspectives: Applying Generative Design to critically explore the Atlas of Swiss Folklore

Max Frischknecht (UNIBE)

The poster addresses generative design as a methodology for a critical-visual exploration of the 1940s Atlas of Swiss Folklore. When we look at visualizations, we are often intrigued by the idea that they represent naturally existing phenomena through visual form. This poses a challenge to humanities scholars who interpret context-dependent data that represents complex social interweaving that unwind across time and space. The data beneath the Atlas of Swiss Folklore, for example, is the product of a perennial and multilayered process of collection and analysis. This interpretative framework is often no longer comprehensible in the finished visualizations leading to the impression of unambiguous visual statements. But rather than depicting unique meaning, visualizations should invite the viewer to critically question what is being viewed and the circumstances of its creation. The poster presents Generative Design is a promising approach that utilizes

morphology and iteration to create a plurality of visualizations. Through Generative Design the Atlas of Swiss Folklore and its circumstances of creation can be examined from different perspectives. Generative Design can be used, for example, to visualize the locations of the atlas data collection in different contexts and thus help evaluate which parts of the population the maps of the atlas are representative or not.

~

Natural Language Processing (NLP) driven categorisation and detection of discourse in historical US patents

Bhargav Srinivasa Desikan, Jérôme Baudry, Nicolas Cachereau and Prakhar Gupta (UNIL/EPFL)

Patents have traditionally been used in the history of technology as an indication of the thinking process of the inventors, of the challenges or "reverse salients" they faced, or of the social groups influencing the construction of technology. More recently, historians of science and technology also read them to interpret the way people described technology and how the specific inscriptions of inventions mattered for the justification and operation of the patent system. The digitization of historical patents opens up unique opportunities to assess the feasibility of unsupervised machine learning and natural language methods for such explorations. In this project, we analyze over a million US historical patents from 1830-1930 using a variety of text-based methods, with two major aims: 1) categorizing patents into coherent technical categories, 2) identifying discourses of safety, reflexivity, and environmental concern in technological innovation. We use both frequency-based and context-based methods, and find that a bag of words-based methods such as TDF-IDF and topic modeling do not perform well on semantic categorization due to the linguistic peculiarities of patent specifications. This suggests that a successful approach to categorizing patents would require contextual semantic representations such as Transformers-based methods (e.g. BERT), or static embedding based methods (e.g Word2Vec, Doc2Vec) which have relatively low computational costs but less expressive in some scenarios. We run early experiments using these methods and find that word embedding models are effective in learning semantics from the descriptions of the patents. In this poster, we will describe our early results, as well as exploratory data analysis on this massive historical patents dataset.

~

Fantasizing Alternative Pasts in Chinese Online Fiction

Sujie Jin (UZH)

Since the 2000s, writing and reading online fiction has become a cultural phenomenon in the Sinosphere. The number of online fiction users increased from around 162 million in 2009 to over 461 million in 2021. In the meantime, a large number of websites have been established to publish stories and make profits. My study focuses on selected works that have been published on *Jinjiang Literature City*, the most influential female-oriented website with 4.15 million fiction works and 1.78 million registered writers. It aims at exploring how online fiction reflects and affects the ethos of culture situated in the sociological context of changing cultural and natural environments. I use discourse analysis to examine *danmei jiakong* history (DJH). The Chinese phrase *danmei* – also known as boys' love in English – refers to women's imagination of male homosexuality, and *jiakong* means fictive and fictitious. Mainly written by and for women, DJH fantasizes alternative history of China, which contributes to historical narratives that have been dominated by male authors. This presentation will provide a systematic introduction on this genre and conceptualize the interaction between the DJH text and its social context. It will also show how DJH arouses historical nostalgia and offers its readers a strong sense of (imagined) participation in the making of the future world.

~

Grammateus as an innovative publication ?

Elisa Nury, Susan Fogarty, Lavinia Ferretti and Paul Schubert (UNIGE, SIB)

In the grammateus project, we are creating a Virtual Research Environment to present a new way of classifying Greek documentary papyri. This environment comprises a database of papyri, marked up with the standard EpiDoc subset of the TEI. It includes as well the textual research outputs from the project, such as introductory materials, detailed descriptions of papyri by type, and an explanation on the methodology of the classification. The textual research output was deliberately prepared as an online publication so as to fully take advantage of the interactivity with data offered by a web application, in contrast to a printed book. We are thus experimenting with a new model of scholarly writing and publishing. This poster presents the workflow of the project, how we dynamically integrate our data with existing data about documentary papyri, and how we are planning for the long-term sustainability of a fully digital research output.

~

Multimodal Deep Learning in Digital Visual Studies

Eva Cetinic and Dario Negueruela del Castillo (UZH)

Multimodality is inherent to almost all aspects of human perception, communication, and production of information. However, as a phenomenon, multimodality is particularly important for the epistemological, interpretative and creative processes within art and art history. The notion of multimodal transformation became recently computationally operationalized on a meaningful and very convincing level. The field of multimodal deep learning significantly advanced in recent years with the introduction of large pre-trained vision-language models. Those models made it possible to computationally generate semantically aligned textual descriptions of images, or vice versa, to render images corresponding to textual inputs. Such models are usually not easy to analyze or explain. They comprise huge parametric spaces trained on immensely large datasets, composed of data sampled from the Internet, and therefore often integrate and propagate various biases and dominant worldviews. By encoding numerous associations which exist between data items collected at a certain point in time, those models represent synchronic snapshots of collective data traces, embedded in specific technological, historical and cultural paradigms. Despite seemingly reproducing the known problem of cultural framing, their directive biases are inextricable of their intentionality as cultural objects, and can provide useful perspectival shifts and new hybrid possibilities in research and artistic production. The focus of this research project is to analyze how such models can be employed in the context of studying visual art and culture. Research enthusiasm in multimodal deep learning is rapidly expanding. However, its application in the domain of digital humanities is still scarce at the moment. Therefore, this project might be of interest to the DARIAH-CH community, as it addresses questions and methods which are at the forefront of contemporary DH research.

~

Decoding the Expression of Light

Pepe Ballesteros Zapata (UZH)

Art history studies rely firstly on data, whatever the type, which is then interpreted in layers of contextualization aiming to describe the evolution of human artistic production. This project will build upon the conceptual framework of a renewed epistemology for art history brought by contemporary computational models. The analysis of relevant features inherent to visual culture and the study of hidden correlations between them allows the opportunity for renewed interpretations never accessed before. Therefore, the project is focused on the study of two elemental features present in paintings: face expression and light. First, the use of face mesh models allows the abstraction of facial expressions from depicted faces. Second, by means of Intrinsic Image Decomposition (IID) and image processing techniques it is possible to describe light conditions (i,e., direction, intensity, diffuseness).

The development of a computational model for the study of such features attempts to represent in a mathematical feature space the proportions, light conditions, and colors from depicted faces of particular characters, schools, and iconographies. The interpretation of the generated outcomes will be based on the hypothesis that light creates space, thus the description of light features allows the possibility of revealing compositional patterns across paintings. Light is also present as a sign of knowledge, as illumination has been used as a sign of divine communication. The duality of light and form is present as a metaphor for the soul and the body. If light is used to suit a message rather than to fulfill the requirements of the physical world, can we find similar light conditions present in different narratives? Does the meaning of light have a mathematical representation? Can we find relationships between light features and character's relevance (soul)? Once shedding light on the embedded information hidden under paintings, the project opens up new possibilities for a close understanding of contextualized iconographies, as well as distant viewing approaches to the description of particular motifs and patterns that improve our understanding of the interaction between light and expression.

~

Augmenting art history: a new story of the painted hand

Valentine Bernasconi (UZH)

The use of computational methods within the field of art history is usually restricted to digitization and archiving tools, allowing the preservation of digital collections and their online access to the world. However, over the past decade, many projects have shown the possibility to use computer vision and machine learning methods to automate tasks on large corpuses of paintings and to better outline recurrent patterns. Unfortunately, these works are often overlooked by art historians who have limited access to these complex computational methods and might sometimes perceive them as an intrusion of the scientific methodology to their field. However, computer science has much more to offer than automated data processing and should not only be seen as a new methodological approach, but also as a door to new perceptions of artworks and artistic practice. Through a research project on computational and historical analysis of hands and gestures in early modern time, where different technologies are used to better grasp the many dimensions of the painted hand, the idea of the machine as a new agent in the relationship between paintings and art historians is envisioned. This exploration has many facets, one of which is the creation of an interactive tool to facilitate research on a large corpus of painted hands, directly using the hand gesture of the user as a query. Fostering a new understanding of painted hand gestures through computer vision, animations and enactment, the goal is to propose, through the *eye* of the machine, a new way of seeing.

~

The SNSF MARK16 project. The first virtual research environment focused on a biblical chapter
Mina Monier and Elisa Nury (DH+, SIB)

This poster will present the virtual research environment (VRE) of the five-year SNSF project MARK16, the first VRE focused on a biblical chapter (https://mark16.sib.swiss). The last chapter of Mark is a well-known enigma of New Testament textual criticism (NTTC): at least six different endings have been listed (e.g. Focant, 2016; Clivaz 2019a). We presumed that many useful manuscripts have not yet been studied, and therefore, should be explored. This led to significant primary results that were gradually documented during our research (Monier 2019, 2021 & 2022; Clivaz 2020, 2021). To support the harvest of results, we have created a VRE in four parts (Clivaz, 2019b): the main part of the MARK16 VRE holds 55 items visualized in a Manuscript Room application (https://mr-mark16.sib.swiss), with the code on Github (https://github.com/sib-swiss/dh-mr-mark16). Prepared in collaboration with the New Testament Virtual Manuscript Room (INTF, Münster), it provides folios of Mark 16 from ancient manuscripts in ten ancient languages. More than 20 international colleagues are MARK16 partners, and data have been nominally published in Nakala, the Huma-Num open public repository (https://mark16-snsf-prima-project.nakala.fr). The second part, Interpretations, presents scholarly individual opinions on Mark's endings from the team and some colleagues (https://mark16-etalk.sib.swiss/search.php). It uses the tool eTalk, with the API on Github (https://github.com/sib-swiss/etalk-docker). The third part, Material, presents relevant material from the printed and digital cultures, like printed editions, articles, and also multimedia publications on Mark 16 (https://material-mark16.sib.swiss). The fourth part, forthcoming, Dataviz (https://dataviz-mark16.sib.swiss) is building a geographical map in collaboration with the network Pelagios.

~

The Seven Sages of Rome in Late-Medieval Europe
Rory Critten and Amélie Mc Cormick (UNIL)

This poster reports on a mini project whose purpose was to create a new database of manuscript and early print copies of the *Seven Sages of Rome* and to experiment with visualizations of that database. The work was funded by the Arts Faculty at UNIL and conducted in spring 2022 by a team comprising Rory Critten, Davide Picca, and Amélie Mc Cormick. Digital methods were employed in this instance due to the vastness of the corpus analyzed, which includes several hundred items: after its appearance in French in the twelfth century, the *Seven Sages* quickly spread across western Europe and was translated into all the major vernaculars. Our database combines records of copies of the *Seven Sages of Rome* at www.arlima.net with our own bibliographic research; together with the visualizations it has been compiled with a view to demonstrating afresh the sharing of literary texts across the western continent. The project thus contributes to a broader reaction against nationalizing literary histories of medieval literature. At the same time, difficulties that its programming component encountered re-illuminate the complexity of the data. For example, language is a reliable means of locating the Welsh and English texts quite precisely, but for the French and Latin copies of the *Seven Sages*, other methods of localization and/or visualization must be employed. In this case, the methodological considerations involved in the design of the project are of interest as much as its necessarily provisional outcomes.