Automated Processing and Exploitation of Heraldic Data from European Archives and the necessary prerequisites for a data-driven historical research

Torsten Hiltmann, Philipp Schneider

Abstract: The paper addresses the issue of reusing data from historical archives (and GLAM institutions in general) in data-driven research projects by presenting a catalogue of supporting factors. These factors center around the FAIR principles and how archives and other GLAM institutions can support research by implementing them in their data services. Mainly, historical data should be made accessible through APIs, be describable through its historical context, and to be as interoperable and reusable as possible. These preconditions for using archival data in data-driven historical research are presented by using the example of the research project "Coats of arms in practice". It aims to study the development and usage of heraldry as a tool of visual communication in the Middle Ages and the Early Modern Period. It employs a data-first approach by integrating data of coats of arms as well as the historical contexts of sources in which they were used into a single Knowledge Graph, built with Semantic Web Technologies. The coats of arms themselves will be described with the help of a specific ontology. Image detection methods based on Machine Learning are used to detect (and describe) coats of arms in image data of historical sources that have not yet been described. This paper focuses on the reuse of archival data from a research perspective. We would like to address the preconditions archival data and GLAM data in general has to meet from the point of view of data-driven research – especially when such research draws on data from multiple institutions. But first, we will introduce our project to make it more transparent, where we come from and in what way we are working with archival data.

The research project "Coats of arms in practice"

Called "Coat of arms in practice",[1] our research project deals with coats of arms as means of communication and thus, as sources of cultural history. Coats of arms were a system of signs that originated in the High Middle Ages and were since then widely adapted as a way to mark persons, groups, institutions and abstract ideas.[2] They were used in various contexts – in manuscripts, on charters, seals, ceiling paintings, tapestries, grave stones, clothes, furniture, tableware and on other objects of all kinds.[3]

The first phase of the project focused on the way coats of arms functioned as means of communication and how they gained the success they had as such in the late

[1] For more information on the current phase of the project, see "Digital Heraldry. Cooperation between Historians and Computer Scientists." n.d. Accessed October 28, 2021. http://www.digitalheraldry.org/.

[2] Hiltmann, Torsten. 2011. "Potentialities and Limitations of Medieval Armorials as Historical Source. The Representation of Hierarchy and Princely Rank in Late Medieval Collections of Arms in France and Germany." In *Princely Rank in Late Medieval Europe. Trodden Paths and Promising Avenues*, edited by Thorsten Huthwelker, 1:157--198. Rank. Politisch-Soziale Ordnungen Im Mittelalterlichen Europa. Ostfildern, p. 160.

[3] Paravicini, Werner. 1998. "Gruppe Und Person. Repräsentation Durch Wappen Im Späteren Mittelalter." In *Die Repräsentation Der Gruppen. Texte - Bilder - Objekte*, edited by Otto Gerhard Oexle and Andrea von Hülsen-Esch, 141:327--390. Veröffentlichungen Des Max-Planck-Instituts Für Geschichte. Göttingen.

Middle ages and early modern times.[4] The second phase of our project is now set to analyze the concrete development of coats of arms as a means of communication. We are interested in how they have become more complex over time and how the range of possibilities for their use and form has become more and more differentiated.

For this purpose, we are developing an infrastructure, built on various digital approaches. This infrastructure will integrate heraldic data from multiple historical sources while – most importantly – also keeping the information about the historical context of these sources. This will allow us to study not only coats of arms as images themselves, but to better understand how and when they were used in different concrete contexts, to compare them and their usage, and to understand them in their development over time.

The most prolific and relevant types of sources for us are medieval manuscripts, seals, and heraldic mural paintings. These sources can provide us with such context, albeit to varying degrees. To list just a few important points:

1. Seals with coats of arms can usually be more easily assigned to a specific person or institution through its inscription. If the seal is attached to a document, the time frame, when the coat of arms on the seal was being used, can be inferred from the dating of the document. On the other hand, seals are monochrom and thus don't provide any information on the tinctures of the coats of arms.[5]
2. In manuscripts, coats of arms are colored. However, here it is more difficult to date coats of arms and to attribute them to their bearers since heraldry in manuscripts is often represented without a specific context and may have been copied from older examples[6].
3. Murals with coats of arms, finally, are, for their part, clearly associated with a specific place. It is therefore easier to deduce from these sources where a coat of arms was used and who probably saw it. Here, however, the identification of the individual coats of arms is particularly difficult, as corresponding captions are often missing[7].

Since coats of arms were used on a huge variety of sources, we have multiple instances of the same coat of arms. One single coat of arms may, for example, be found at the same time on seals, but also e.g. in manuscripts, on charters, mural paintings, stained glas or on tapestries.[8] Therefore, one instance of a coat of arms

[4] Hiltmann, Torsten. 2019. "Zwischen Grundwissenschaft, Kulturgeschichte Und Digitalen Methoden. Zum Aktuellen Stand Der Heraldik." *Archiv Für Diplomatik* 65: 287--319.

[5] Drös, Harald. 1988. "Wappen Und Stand." In *Codex Manesse. Katalog Zur Ausstellung Vom 12. Juni Bis 4. September 1988, Universitätsbibliothek Heidelberg*, edited by Elmar Mittler and Wilfried Werner, 30:127--152. Heidelberger Bibliotheksschriften. Heidelberg.

[6] Hofman, Elmar. 2021. *Armorials in Medieval Manuscripts. Collections of Coats of Arms as Means of Communication and Historical Sources in France and the Holy Roman Empire (13th - Early 16th Centuries)*. Heraldic Studies. Ostfildern.

[7] Seixas, Miguel Metelo de, and Torsten Hiltmann. 2020. "Heraldic Decor as a Research Problem. The 'Sala Dos Brasões' of the Sintra Palace and Heraldry in Medieval and Early Modern State Rooms." In *Heraldry in Medieval and Early Modern State Rooms*, edited by Torsten Hiltmann and Miguel Metelo de Seixas, 3:11--28. Heraldic Studies. Ostfildern.

[8] Hablot, Laurent. 2019. *Manuel de Héraldique Emblématique Médiévale*. Tours.

on one particular source is always connected to other instances of the same coat of arms on other sources.

Figure 1: Examples of sources and objects depicting heraldry

This means, we are interested in multiple information: the image of a historical source - like a seal with a coat of arms -, its provenance (e.g. *which* document was it part of), and the context of its provenance, e.g. when was it used, by whom was it used, and in what other contexts did it appear.

The goal of our project is to build a digital infrastructure that covers all these different aspects. By using Machine Learning and Semantic Web Technologies, we aim to detect, link and integrate heraldic data from multiple sources, and enrich it through Linked Data. This will allow us to study them with qualitative as well as quantitative methods on a large scale, based on supplementary data on their temporal and historical context. Aside from studying the development of coats of arms and in which contexts they were being used, such an infrastructure can also help in better understanding the different historical objects and sources, we are capturing with our approach. With the help of enriched data about a coat of arms, used in a source, it is possible to do more inferences regarding the time the source was created or used and by whom this was done. In other words: such data could help greatly in providing external evidence to date and contextualize historical sources.

This approach needs the integration of data from multiple data repositories, provided by a variety of institutions across Europe. This includes databases describing and identifying coats of arms themselves.[9] But also other databases that are not dedicated to heraldry as such, but that provide data on the various types of historical sources bearing heraldic imagery – like manuscripts, seals, charters, glass paintings

[9] See e.g. “Medieval Armorials.” n.d. Accessed October 28, 2021. http://www.armorial.dk/ with over 30,000 descriptions of coats of arms in armorials, but also “ArmmA. ARmorial Monumental du Moyen-Âge.” n.d. ArmmA. Accessed May 13, 2021. https://armma.saprat.fr/; “Stemmario. Wappen Florentiner Familien, Kirchen, Bruderschaften Und Hospitäler — Portal.” n.d. Accessed October 28, 2021. http://stemmario.khi.fi.it/.

etc.[10] In all of these sources, coats of arms can be found**.** However, although these sources can be located via catalogues and databases, this does not mean that the coats of arms depicted in these sources are also somehow recorded and thus findable for us. Therefore, as a first step, all sources that could contain heraldry, or their images, must be recorded and processed further**.**.

Find and detect coats of arms in image data with Machine Learning

To mine for coats of arms in image data, we use image detection methods based on Machine Learning. Up to this point, we trained two models to detect coats of arms, both using the architecture *yolov4*.[11] Initially, we only used digitizations of medieval manuscripts as training data, where the coat of arms are depicted with colors. To create the necessary training data, we have mainly used armorials, which contain large numbers of heraldic depictions.[12] This way we were able to establish a substantial ground truth. Furthermore, in order to include material from outside the armorials in our ground truth, we used an iterative training process, where we added new training data in batches of a thousand randomized manuscript pages from our pool of medieval manuscripts at each step. Each iteration included retraining the model with the newly incorporated training data and manually checking and correcting the results, in order to add them to the ground truth. This allowed us to look out for errors and biases of our Machine Learning model and correct them early in the training.

All in all we trained 10 different detection classes to capture heraldry in its various forms of representation, e.g. coats of arms on shields as the most common method of representation, but also on banners, pennants or clothes. Overall, we achieved a F1-score[13] of 90 percent. A peculiarity, we could correct with our iterative training method, was the fact that our detector also classified stamps that were put in the manuscripts by librarians as (historical) coats of arms. We were able to circumvent this issue by creating a new detection class for these stamps, thus teaching our model to distinguish between them and other images of coats of arms.

Our next goal was to also detect coats of arms on seals. It quickly became apparent that it was not possible to simply transfer the model trained on the manuscripts to this new image domain. An attempt to do so failed and only achieved a F1-score of less than 50 percent. This is probably due to the fact that coats of arms on seals are not coloured and, more importantly, are represented here with a different technique

[10] See e.g. “Base Numérique Des Sceaux Conservés En France | SIGILLA.” n.d. Accessed October 28, 2021. http://www.sigilla.org/; “Corpus Vitrearum Deutschland.” n.d. Accessed October 28, 2021. https://corpusvitrearum.de/; “Literatur Und Wandmalerei. Erscheinungsformen ‘höfischer’ Kultur Und Ihre Träger Im Mittelalter.” n.d. Accessed May 13, 2021. http://wandmalereien.imareal.sbg.ac.at/; “Monasterium.Net.” n.d. Accessed October 28, 2021. https://www.monasterium.net/mom/home.

[11] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. “YOLOv4: Optimal Speed and Accuracy of Object Detection.” *ArXiv:2004.10934 [Cs, Eess]*. http://arxiv.org/abs/2004.10934.

[12] Hofman, Elmar. 2021. *Armorials in Medieval Manuscripts. Collections of Coats of Arms as Means of Communication and Historical Sources in France and the Holy Roman Empire (13th - Early 16th Centuries)*. Heraldic Studies. Ostfildern.

[13] Flach, Peter, and Meelis Kull. 2015. “Precision-Recall-Gain Curves: PR Analysis Done Right.” In Advances in Neural Information Processing Systems, edited by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, 28:838–46.

in a different material. It is known that current neural networks are biased towards texture.[14] Therefore, for the time being, we have decided to create a separate ground truth for heraldic representations on seals and to train a new specific model.

Through this process, we are able to automatically find coats of arms in large amounts of image data. Since we are using a Machine Learning architecture doing image detection rather than classification, we also get the exact positions of the retrieved coats of arms in the form of coordinates. This allows us to analyze the visual representation of coats of arms with respect to their size, positioning and their ratio to each other, but also to extract the images of single coats of arms and process them further.

Building a research infrastructure to digitally represent coats of arms

Our next step is to encode the acquired coats of arms in a machine-readable way. Only then can we make individual occurrences comparable to one another and identify which coats of arms were used in which other documents such as other seals, manuscripts or murals. Machine-readability, of course, offers us to do this in a scalable way with large amounts of data. To achieve this, we are currently developing a formalised ontology, which is based on the description of coats of arms by terms of blason.[15] The terms and concept of blason are modeled as classes and relation of the ontology. With this we aim to propose a formalised and standardised way for machine-readable descriptions of heraldic sources in general.

[14] Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. 2019. "ImageNet-Trained CNNs Are Biased towards Texture; Increasing Shape Bias Improves Accuracy and Robustness." *Proceedings of the International Conference on Learning Representations (ICLR);* Hiltmann, Torsten, Sebastian Thiele, and Benjamin Risse. 2020. "Friends with Benfits: Wie Deep-Learning Basierte Bildanalyse Und Kulturhistorische Heraldik Voneinander Profitieren." In *DHd 2020 Spielräume: Digital Humanities Zwischen Modellierung Und Interpretation. Konferenzabstracts*, edited by Christof Schöch, 135–38. Paderborn. https://doi.org/10.5281/zenodo.3666690.

[15] Hiltmann, Torsten, and Thomas Riechert. 2020. "Digital Heraldry. The State of the Art and New Approaches Based on Semantic Web Technologies." In *L'édition En Ligne de Documents d'archives Médiévaux*, 102--125. Turnhout.

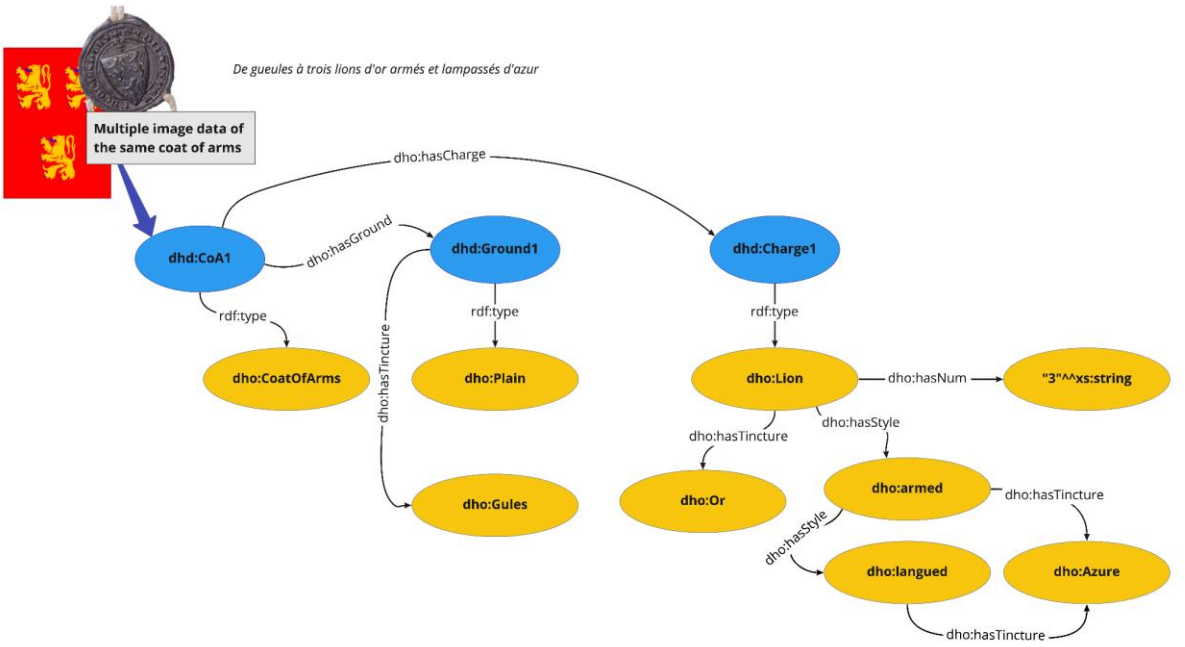

Figure 2: Description of the coat of arms of *Archambaud III de Périgord* with the *Digital Heraldry Ontology*

At this moment, the description of heraldic imagery, based on our ontology, still has to be done manually. In the long run, this may be done at least semi-automatically as well. To achieve this, we are working together with computer scientists on various approaches to segment the detected coats of arms and record them in their individual parts. The goal here is to retrieve all pixels from an heraldic image that belong to a certain *Charge* (or figure) or to a certain geometric pattern depicted on the coat of arms. As the detection classes for the semantic segmentation[16] serve the terms of blason, defined through our ontology. This part of our infrastructure is done in collaboration with computer scientists from the *Technische Universität Berlin* and the *University of Münster*.

At this point we can detect and localise individual coats of arms on seals and in manuscripts. There is, of course, a central challenge we need to address here: How do we get all this data? Our sources are not provided by a single archive, library, or data repository, since heraldry, as a form of visual communication, was a European-wide phenomenon.[17] This makes cross-border approaches that capture data from different European Institutions from Malta and Portugal to Lithuania, and Hungary strictly necessary. Therefore, we have to acquire data from a great institutional and geographical variety of providers of cultural heritage objects. With regard to the amount of data we are dealing with, the acquisition of our data should be done as automatically as possible.

FAIR principles as a precondition for data-driven research

[16] Li, Biao, Yong Shi, Zhiquan Qi, and Zhensong Chen. 2018. "A Survey on Semantic Segmentation." In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1233–40. https://doi.org/10.1109/ICDMW.2018.00176.

[17] Hablot, Laurent. 2019. *Manuel de Héraldique Emblématique Médiévale*. Tours, p. 15.

In the second part of the paper, we would therefore like to concentrate further on the preconditions that may support data-driven projects such as ours, i.e. working with distributed sources and getting access to them in the first place. In essence, these preconditions can be described by the FAIR principles[18]: data should be findable, accessible, interoperable, and reusable. This applies for data provided by archives, but also by other GLAM institutions. We will illustrate the relevance of these principles to our research project.

Findability in general means that data can be found through a globally unique persistent identifier and through its metadata, which describes the data itself. In the case of historical sources this should refer to the context of the source.[19] There are two ways of understanding what the "context" of a source actually means.[20] From an archival perspective, this is mostly understood as the *respect des fonds*, which is the context of provenance of a single historical source.[21] This is what most archives describe with their metadata.

To this, however, we would add the historical researcher's perspective of historical context. This would in general be metadata about known entities that existed independently from the source itself, but that are linked in some way to the existence of the source. In concrete terms, this would be agents (e.g. individual persons, groups of persons, institutions) who created or reused the source, places where the source was created, reused or applied to, but also the relevant circumstances, where a source was created, in the shape of events and timespans.[22] This kind of metadata is a premise for historians to actually infer new knowledge from sources. If one wants to study how coats of arms were used as a means for visual communication, as it is the case in our project, the simple evidence of a coat of arms does not suffice – rather, we need to know how, when and by whom a coat of arms was used, which extends to the questions of who created the object at which time, where the coat of arms is depicted on. This is well established in historical research as the external evidence of source criticism. However, we want to emphasize that this applies to

[18] "FAIR Principles." n.d. GO FAIR. Accessed September 10, 2021. https://www.go-fair.org/fair-principles/.

[19] Caserio, Matteo, Anna Goy, and Diego Magro. 2017. "Smart Access to Historical Archives Based on Rich Semantic Metadata:" In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 93–100. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0006487000930100; Goy, Annamaria, Davide Colla, Diego Magro, Cristina Accornero, Fabrizio Loreto, and Daniele Paolo Radicioni. 2020. "Building Semantic Metadata for Historical Archives through an Ontology-Driven User Interface." *Journal on Computing and Cultural Heritage* 13 (3): 25:1-25:36. https://doi.org/10.1145/3402440; Sternfeld, Joshua. 2011. "Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization." *The American Archivist* 74 (2): 544–75. https://doi.org/10.17723/aarc.74.2.644851p6gmg432h0.

[20] Society of American Archivists. n.d. "Context." Dictionary of Archives Terminology. Accessed October 28, 2021. https://dictionary.archivists.org/entry/context.html.

[21] Lodolini, Elio. 1995. "Respect des fonds et principe de provenance : histoire, théories, pratiques." *Gazette des archives* 168 (1): 201–12. https://doi.org/10.3406/gazar.1995.4283.

[22] Odebrecht, Carolin. 2022. "Referenzierung Des Digitalen Kulturellen (Text-)Erbes. Digitale Quellenkritik Und Modellierung von Metadaten." Poster presented at the DHd2022 - Kulturen des digitalen Gedächtnisses, Potsdam.

data-driven research with historical sources as well. Up to this point, neither the quantity nor the level of detail of archival metadata is sufficient for these demands.[23]

Another important precondition for data-driven research across multiple data providers would be **Accessibility**. This means – also for historians – accessibility for machines. This can be provided through Application Programming Interfaces, short APIs. Only in this way is it possible to process large amounts of data automatically.

Of course there are other ways to work around the lack of interfaces, like web scraping. In our case, for example, we created a Python script to acquire the images and metadata of seals provided by the French database *Sigilla*[24] through their representation on their website. But this approach has of course many technical disadvantages, aside from legal difficulties and needing permission from every data provider. Firstly, we would have to program a new individual solution for each archive, museum, library, or other database, whose data we want to include in the corpus of historical sources for our research. This is easily prone to errors, since only slight changes by the data provider can lead to situations where the data is not scraped completely. It may be that in such a case the whole pipeline has to be adapted or even recreated.

This is not only time consuming, but also puts the task of homogenization into the hands of the researcher. This means that he or she has to decide, each time, which fields of metadata from two different data providers are equivalent to each other or at least follow more or less similar concepts. That means to homogenize the value of the metadata like it is depicted in figure 3. Here, we have extracts from the metadata sections of three different IIIF manifests, representing three different manuscripts. IIIF is a collection of standardized APIs to present, request, and share structured sequences of images.[25] The project is mainly focusing on the Cultural Heritage domain. In our case, the *IIIF Presentation API*, which defines the structure of a sequence of images, identified by URIs, is important to access image scans of digitized manuscripts provided by different libraries across Europe. Although the *IIIF Presentation API* is also used by the libraries to share metadata about the sources, it is important to note that these are not provided in a standardised way, as presented in our example. Here, the data comes from three different libraries and employs three different ways of dating the manuscripts. These differences are of course no problem for a human, working with the data. However, processing or analyzing this data with digital methods is impossible without thoroughly cleaning the data beforehand.

[23] Although we have of course to refer to the many initiatives working on this challenge. See e.g. Lawrence, Faith, Adam Retter, and Jone Garmendia. 2021. “Project Omega: Modelling an Archive Catalogue to Support Future History.” In . Berlin. https://d4h2020.sciencesconf.org/data/pages/Lawrence_Retter_Garmendia_Project_Omega_2.pdf.

[24] “Base Numérique Des Sceaux Conservés En France | SIGILLA.” n.d. Accessed October 28, 2021. http://www.sigilla.org/.

[25] “IIIF | International Image Interoperability Framework.” n.d. Accessed October 28, 2021. https://iiif.io/.

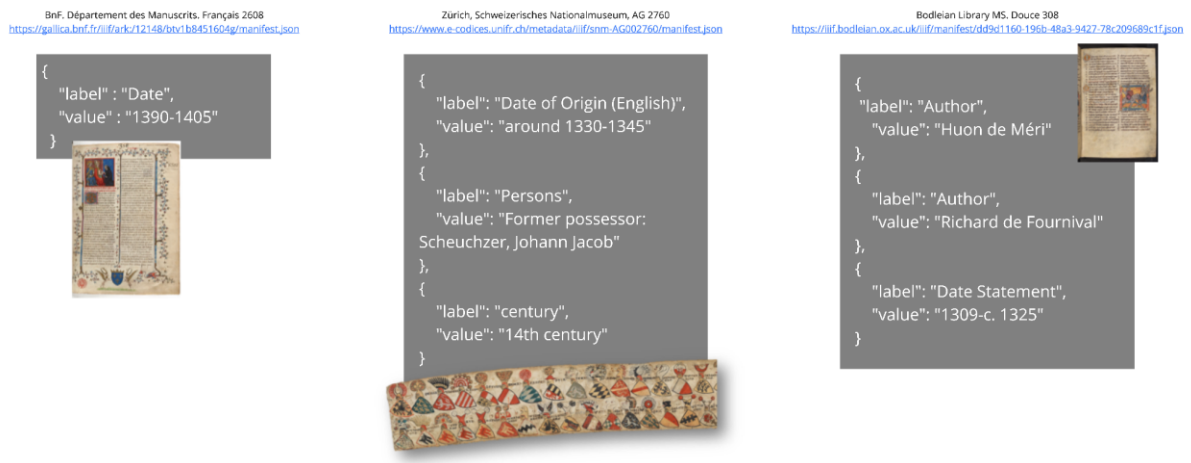

Figure 3: Examples for representing metadata through IIIF manifests

Examples for making historical sources available through standardised APIs exist, but are – especially in the case of archives – not yet widespread. An established service is *Archives Portal Europe*.[26] Here, archival data is made accessible through an API using the standards *Encoded Archival Description, Encoded Archival Context* or the *Metadata Encoding and Transmission Standard*.[27] Europeana would be another example, providing cultural heritage data through its own data model via an API and even through a SPARQL endpoint.[28] To provide image data, IIIF[29] is increasingly accepted as a standard.

An approach to solve this problem, that gained momentum during the last ten years, is the employment of standardized ontologies and vocabularies. This leads us to the topic of **Interoperability**. We have to agree, how we are describing data from the perspective of machine readability. There exists an increasing number of upper-level ontologies. CIDOC CRM[30] is of course increasingly used to describe and contextualize cultural heritage data.[31] But there are also initiatives from the archival domain that are interesting for historical research. A newer standard by the International Council on Archives would be the *Records in Context* ontology, which was developed by the *International Council on Archives*.[32] Although it is not yet

[26] "Archives Portal Europe." n.d. Accessed October 28, 2021. https://www.archivesportaleurope.net/de.
[27] "API - Archives Portal Europe." n.d. Accessed October 28, 2021. https://www.archivesportaleurope.net/de/information-api.
[28] "Europeana SPARQL API." n.d. Europeana Pro. Accessed October 28, 2021. https://pro.europeana.eu/page/sparql.
[29] "IIIF | International Image Interoperability Framework." n.d. Accessed October 28, 2021. https://iiif.io/.
[30] Bekiari, Chryssoula, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. 2021. *Definition of the CIDOC Conceptual Reference Model v7.1.1*. The CIDOC Conceptual Reference Model Special Interest Group.
[31] Beretta, Francesco. 2021. "A Challenge for Historical Research: Making Data FAIR Using a Collaborative Ontology Management Environment (OntoME)." *Semantic Web* 12 (2): 279--294. https://doi.org/10.3233/SW-200416. On the topic of developing interoperable ontologies for historical studies see also the papers from the annual conference of the *Data for History Consortium*; see "Data for History 2020: Modelling Time, Space, Agents - Sciencesconf.Org." n.d. Accessed October 28, 2021. https://d4h2020.sciencesconf.org/.
[32] "Records in Contexts - Conceptual Model | International Council on Archives." n.d. Accessed October 28, 2021. https://www.ica.org/en/records-in-contexts-conceptual-model.

widely employed by institutions, this model is promising for the needs of data-driven historical research. Its classes describe historical documents not only in their context of provenance, but also in their historical context (see Findability). Historical documents – or *Records* – are modelled as entities that are created, reused or acquired by an agent. An agent can be a single person, or a group of persons (e.g. an institution). The creation, reuse or acquisition of a *record resource* happened in relation to a place and at a specific time. *Records in Context* allows to model time either as a date or as an event. Such an approach, centering on events and agents and placing sources in their historical context, would make historical data comparable on an abstract level and would therefore be an important precondition for data-driven research that spans multiple data providers. Unlike CIDOC CRM, which was created with objects in museums in mind, *Records in Context* is explicitly designed for an archival context. It has a much more reduced set of classes than CIDOC CRM, thereby making it more easily adaptable to the archival context. Nevertheless, both ontologies use very similar base concepts (e.g. an event-centric approach or a similar definition of *agents* as individuals and groups alike). Thus, mapping between the two should be possible. The widespread usage of an ontology like *Records in Context* would make archival sources interoperable.

As a last aspect, there is **Reusability**. Naturally, everything we talked about up to this point facilitates the reuse of European cultural heritage for historical research. But there are further important obstacles that can hinder researchers in reusing the data provided by archives and other GLAM institutions such as licensing. To work, data-driven research needs data to be open. In this regard, we would like to introduce another example from practice: a database on coats of arms provided by the Tyrolean State Museum. This database is actually the digitization of a handwritten and hand drawn catalogue collecting coats of arms from the region of Tyrol.[33] The catalogue is based on about 30,000 index cards, each containing a reference to a single coat of arms as a drawing. The bearer of the arms is denoted as handwritten text, which is transcribed and provided through the database. Although some of the coats of arms are from the 20th century, the database provides a valuable resource in identifying unknown medieval and early modern coats of arms.

Since the images of coats of arms are linked to families who bore them, this data can also be an important reference for the study of the use of heraldry. Unfortunately, the images are protected by copyright and on the website of the Museum it is indicated that their reuse costs money – 16.50€ per document.

Of course copyright is not an easy issue, especially for GLAM institutions. In short: The Tyrolean State Museum will have reasons to protect their data this way. The database makes a good reference catalogue nonetheless. But when we want to make use of new digital methods for our research, this imposes a high barrier on processing such resources.

To conclude, digital methods enable us with new approaches to research historical sources in archives and make it possible to analyze them in new and innovative

[33] "Tiroler Wappen." n.d. Accessed October 29, 2021. http://wappen.tiroler-landesmuseen.at/login.php.

ways. These pose multiple methodical and technical challenges we have yet to overcome, as we illustrated in the first part of the paper. But aside from these, there are also a few remaining practical challenges, which have to be tackled by GLAM institutions. In summary, the following points would further support the use of archival data in data-driven historical studies:

1. It would be helpful, if historical sources and their metadata would be made available in a formalized and standardized way.
2. This should be done through standardized APIs, so that the data is automatically processable by machines on a large scale.
3. The reusability of metadata would, from the researchers' perspective, be increased, if it would not only describe the provenance of a source, but also contain information about its historical context, in an interoperable format. This is a complex endeavor for the whole of historical scholarship, which cannot be done by a single institution alone – therefore minimal standards describing relations to relevant agents, events, time spans and places would already be helpful, if they were widely employed.
4. And finally, cultural heritage data should be understood as part of our common cultural heritage and thus being as openly available and reusable as possible.

From the perspective of research projects such as ours, which we presented at the beginning, we propose these points as supporting steps that can help to further unlock Europe's digital treasures on a large scale and integrate them into the common history of Europe.