

# Networking Digital Documents

*Christiane Sibille*

## **Introduction**

Humanities scholars working on web-based digital research projects often share common objectives: firstly, they want to identify entities in the sources, secondly, they want to connect the same entities within their whole corpus, and thirdly, they want to share and connect their results with other projects. In the latter point this is not only in order to gain a better visibility, but also to contribute to new forms of intellectual exchange and production of knowledge.

This article will outline possibilities for linking information about historical persons by focusing on initiatives that aim to clearly identify persons and their names in a standardized way. While so-called authority control is counted among the core competences of librarians and information scientists, the theoretical reflection and application of standardized data is rather new in the humanities, yet highly relevant. In the context of ideas about linked open data (Berners-Lee 2006) and to avoid the production of data cemeteries, the humanities have to deal with these questions to integrate their highly specialized and high-quality data into virtual networks. To outline these opportunities, the article is divided into two parts. The first reflects on the possibilities of connecting digital edition projects, and the second sketches the development of important initiatives for the identification of persons and names.

I argue that as linked data, digital editions become able to cross disciplinary borders and will connect formerly distinctive fields and areas with each other. We can achieve this by creating our data in a form that will open up active and passive connectivity. The crucial point is that the digital availability of editions means much more than a specific form of a print which is now available as a file on your electronic device. The crucial innovation does not happen on the level of

the document itself, but on the level of the part of an edition which was formerly entitled “index”. In the digital world, the index has come to life. Categories such as persons, topics and places now communicate with one another and are connected to similar groups in the World Wide Web. In the end, this form of digital information network can produce a global perspective on all kinds of sources. It is therefore crucial, not so much for technical reasons as for scientific reasons, to understand how indices initiate or avoid the form and content of this kind of global discussion.

In recent years, literature has reflected on the theoretical and practical implications of digital editions (Pierazzo 2015; Sahle 2013). This article uses the term “digital document” to describe a digitized historical document that was not born digital, but is now selected based on a scholarly decision, scanned and enriched with additional data, at least a defined set of metadata, but also transcriptions of the text. From this it follows that these documents have at least three levels: the physical document, the scanned facsimile of the physical document, and—on a third level—data that was created by scholars to describe the document/scanned image in a way that seemed to be useful in their research context. The first level, the physical document, was the object for generations of scholars during the last two centuries. As for the second level, questions of scanning and preserving digitized objects are today mainly discussed in the field of archival and heritage studies. It is on the third level, where questions about the content of the document, its description and preservation meet, that both groups, scholars and information technologists, should talk to and learn from each other, because the successful solution of possible issues and problems requires this specific kind of interdisciplinary dialogue.

The analogue production of scholarly works in the context of diplomatic studies developed varied approaches to make information on historical source material available. In the German-speaking countries, there were the great “full text” editions like the *Monumenta Germaniae Historica* (2016), regesta like the *Regesta Imperii* (2016), or even editions of selected key documents on certain political questions or events (Institut für Zeitgeschichte 2016). Transferred to the digital area, we can recognize the various techniques in the ways scholars explore the possibilities to add additional data to digitized documents. Even though Extensible Markup Language (XML), following the standard of the *Text Encoding Initiative* (TEI), has become the de-facto standard for the transcription of documents, there are various other

ways to combine the parts of a digital document into a scholarly meaningful collection.

The *Diplomatic Documents of Switzerland* (2016), one of the earliest digital projects in the context of diplomatic documents, regularly puts about 2500 documents, covering a three year period, online. For each of these documents researchers have identified every person, organization and geographical name that is mentioned in the documents. But since the text of the scanned document itself is usually readable without problems, no transcription is provided. However, about 200 of these documents, especially those that summarize the main developments within the respective period, are manually transcribed and regularly published in printed volumes.

Although the success of TEI can partly be explained by its ability to incorporate the demands of a broad range of approaches, it would be utopian to hope for a common standard to handle digital documents, their transcriptions and metadata, especially in the context of digital humanities, which deal with highly heterogeneous source material (Schmidt 2014; TEI 2016).

Instead of arguing about which standard is superior, it seems more constructive to discuss possibilities for exchanging and connecting the digital results of our work and, by doing this, to define what could be our common interests and needs (Hodel/Sibille forthcoming 2018).

It is possible to differentiate between at least three forms of shared data. These three forms can be distinguished by the efforts a project is willing to invest in the formalization of data. Firstly, we can make available all our research data and enable other researchers to re-use our results when we have finished a project. For instance, a TEI edition from a historian that is encoded to describe the physical object and the historical content could be reused and enriched by a linguist with a stronger focus on semantic structure. Here the TEI standard is the basis on which other scholars can add an additional layer of information. However, this possible re-use has to be conceptualized from the start of a project (Barbera et al. 2013; Schmidt 2014: 6).

Secondly, we can expose our data, especially our metadata, to meta-engines like *europeana.eu*. *Europeana* is the cultural heritage platform of the European Union. It hosts the metadata of more than 50 million cultural items and connects collections from all over Europe. To share its metadata, a project has to format its data according to specific guidelines, set out by a data exchange scheme (Europeana.eu 2014). This means that there can be two versions of the data. On the

one hand, there are the original data within the project, and on the other hand, the data that meet the standard of the meta-project. The added value lies in the increased visibility. Connected and findable with a large quantity of other data, users can discover content they did not know of and become aware of a wide range of collections and projects. However, to realize the problems of this strategy, it is sufficient to think about the question of multilingualism, different spellings, etc.

The third possibility is to explicitly connect data from a project with authority files or similar entities in other projects. This means the researcher reflects on standardization and the exchange of data as he or she is creating it. While this is still difficult for the assignment of intellectual concepts, it is especially possible with regard to persons, corporate bodies or places. Depending on the amount and quality of data, it can be done supported by automatized named entity recognition (Hooland et al. 2015) or manually, by looking up standardized identifiers for the respective entity (Barbera et al. 2013; Ciotti/Lana/Tomasi 2014; Eide 2014; Iglesia/Göbel 2014).

### **Name Authority Control and Agent Description in the Humanities**

“The library community has been conducting name authority control for over a century” (Niu 2013: 404). When it comes to the standardized management of identities in the context of digital scholarly editions, we can build upon the experiences and the knowledge of librarians, archivists, and information scientists, who have been dealing with this problem for a long time. The following part of the article will focus on initiatives in the domain of standardized information about persons and their names. There are numerous initiatives for other entities, like the *Getty Thesaurus of Geographical Names* (Getty Research Institute 2016), or *GeoNames* (GeoNames 2016), to name but two from the domain of geographical references. In general, we can distinguish between projects that (a) act on a national level, (b) bring together national projects on an international level, and (c) are based on the collection of international projects and separated domain specific initiatives and interests.

The following case studies will represent these three levels. The first will focus on the German Integrated Authority File (Gemeinsame Normdatei, GND), the second on the Virtual International Authority

File (VIAF), and the third on the International Standard Name Identifier (ISNI). A fourth case study, based on the Swiss project *Metagrid*, will discuss domain specific solutions that are able to integrate the data of the aforementioned projects. The following reflections have an intrinsically historical point of view and deal with historical persons. Discussions about researcher IDs, that are important for today's scholarly landscape, are excluded.

### **Gemeinsame Normdatei (GND)**

The context of the creation of the Gemeinsame Normdatei (GND) and its predecessor shows the normative effects of technological changes. The implementation of new cataloguing techniques, especially the digital turn and the increasing use of Online Public Access Catalogues (OPAC), urged the stakeholders to reflect upon the introduction of common tools to ease the transformation of analogue to digital records, a kind of work that had to be done in the same way in almost all libraries. In the context of a retro-conversion project, the German Research Foundation recommended that all supported libraries use the newly created Personennamendatei (PND) (Fabian 1995: 604–605). To overcome the separation of distinct authority files for persons, corporate bodies, subject headings, etc., a project for a new consolidated authority file was launched in the late 2000s, the GND. Digital biographical and historical dictionaries were among the first projects to use the authority files to identify their entities on a second level. This was an important moment: the transformation of the former analogue and static index into an innovative tool for the digital networking of information. In 2005, the German version of *Wikipedia* established a cooperation with the German National Library to link articles about persons with the respective PND/GND (Geipel et al. 2013: 180). Moreover, the German Research Foundation continued its commitment for the adoption of the PND/GND. It recommended the application of PND/GND to all its funded projects and therewith actively supported its implementation outside the narrower field of libraries (Ebnet 2010: 47–51). Based on this broad dissemination, the community developed new tools to exchange information. Based on so-called Beacon files, it became possible for projects to retain their own IDs but be paired with the equivalent GND, using it as a common identifier to link to other projects. The standards of the Beacon format are described on <https://>

gbv.github.io/beaconspec/beacon.html. At the moment (May 2016), about 450 of these Beacon sources provide about 7.6 million distinct identifiers on the website of the *Beacon-Findbuch* (<http://beacon.findbuch.de/seealso/pnd-aks>).

## Virtual Authority File (VIAF)

The Virtual International Authority File (VIAF) collects and connects the authority files from numerous national libraries, specialized libraries and other data contributors. Today, it embodies not only information about persons and corporate bodies, but also works, expressions, meetings and geographical names (Angjeli/Mac Ewan/Boulet 2014: 2–3). The content of VIAF is connected to so-called clusters. These clusters contain entities that, based on an algorithmic calculation, are the same. This fully automated process reduces the workload that has to be invested in the curation of data. However, this approach also reduces the persistence of the dataset. If new or updated data are loaded into VIAF, it is possible that the mathematical plausibility for certain connections decreases and clusters will be merged or split. More generally, that means that a VIAF ID only addresses a specific cluster, but the content of this cluster, i.e. the entities that seem to be the same, are not stable (*ibid.*: 7).

While the VIAF clusters as assemblages of data provided by separate partners can be merged or split off, the data of each partner stay unchanged. That means that errors that are within the authority files are displayed as they have been sent to VIAF and it is not possible to change them by VIAF. They have to be corrected by the original data provider itself (*ibid.*: 10).

The plates below exemplify the functioning of VIAF. Plate 1 shows the standard view of a person entry with information about providing countries and institutions. Here it is the cluster with ID 32065073, that represents information about the American diplomat Elihu Root. Plate 2 represents a dataset that is not yet merged with ID 32065073, probably because another entity from the same provider is already connected to the same cluster. The Resource Description Framework (RDF) representation of another dataset—representing Harold Alexander—in the GND (plate 3)—as a typical example of a national VIAF provider—indicates all local modifications. In this case, it indicates that GND ID 171958489 was merged with GND ID 12257365X. This information is automatically corrected in VIAF (plate 4) and a redirect

VIAF  
Virtual International Authority File

**Suchen**

Feld auswählen:  Index auswählen:  Suchbegriffe:

Root, Elihu, 1845-1937  
Elihu Root  
VIAF ID:32065073 (Person)  
Permalink:<http://viaf.org/viaf/32065073>  
ISBN: 0000\_0000\_8111\_1300

Vorzugsbezeichnungen

- 100 0\_1a Elihu.Root
- 200 \_1a Root, Elihu, 1845-1937
- 100 1\_1a Root, Elihu, 1845-1937
- 100 1\_1a Root, Elihu, 1845-1937
- 100 1\_1a Root, Elihu, 1845-1937
- 100 1\_1a Root, Elihu, 1845-1937
- 100 1\_1a Root, Elihu, 1845-1937

Plate 1: VIAF-Cluster of ID 32065073 (Elihu Root) with information about the data-providing countries and institutions (<https://viaf.org/viaf/32065073>).

VIAF  
Virtual International Authority File

**Suchen**

Feld auswählen:  Index auswählen:  Suchbegriffe:

Root, Elihu  
VIAF ID:310569019 (Person)  
Permalink:<http://viaf.org/viaf/310569019>

Vorzugsbezeichnungen

- 100 1\_1a Root, Elihu

Plate 2: Dataset that is not yet merged with other data (<http://viaf.org/viaf/310569019>).

```
<http://d-nb.info/gnd/12257365X> a gndo:DifferentiatedPerson ;
foaf:page <http://de.wikipedia.org/wiki/Harold_Alexander%2C1._Earl_Alexander_of_Tunis> ;
owl:sameAs <http://dbpedia.org/resource/Harold_Alexander,_1st_Earl_Alexander_of_Tunis> , <http://viaf.org/viaf/79137182> ;
gndo:gndIdentifier "12257365X" ;
gndo:oldAuthorityNumber "(DE-588)125839723" ;
owl:sameAs <http://d-nb.info/gnd/125839723> ;
dabt:deprecatedUri "http://d-nb.info/gnd/125839723" ;
gndo:oldAuthorityNumber "(DE-588a)125839723" ;
owl:sameAs <http://d-nb.info/gnd/125839723> ;
dabt:deprecatedUri "http://d-nb.info/gnd/125839723" ;
gndo:oldAuthorityNumber "(DE-588)1087768055" ;
owl:sameAs <http://d-nb.info/gnd/1087768055> ;
dabt:deprecatedUri "http://d-nb.info/gnd/1087768055" ;
gndo:oldAuthorityNumber "(DE-588)171958489" , "(DE-588a)171958489" , "(DE-588a)12257365X" , "(DE-588c)4617904-5" ;
```

Plate 3: When a dataset is merged by the provider—here the GND ID 171958489 was merged with ID 12257365X (<http://d-nb.info/gnd/12257365X>) ...

History of VIAF ID:69064291 (22)		
Record ID	Action	Time
<a href="#">BNF 13012637</a>	add	2009-03-03T12:03:23+00:00
<a href="#">LC n 50038656</a>	add	2009-03-03T12:03:29+00:00
<a href="#">SELIBR 174773</a>	add	2009-03-03T12:03:42+00:00
<a href="#">NKC pna2006370689</a>	add	2009-05-07T09:10:55+00:00
<a href="#">NLA 000035942374</a>	add	2009-08-21T09:42:41+00:00
<a href="#">LAC 0009B6413</a>	add	2010-02-22T06:44:12+00:00
<a href="#">NII DA04077018</a>	delete	2010-11-16T05:07:56+00:00
<a href="#">SUDOC 070313687</a>	add	2011-12-15T07:58:58+00:00
<a href="#">BIBSYS x90953250</a>	delete	2012-11-27T04:00:16+00:00
<a href="#">NTA 073958042</a>	add	2012-12-19T19:11:18+00:00
<a href="#">NLI 000409308</a>	add	2013-04-17T17:27:37+00:00
<a href="#">NLI at 000409308</a>	delete	2013-05-13T19:14:22+00:00
<a href="#">NLP a16723673</a>	add	2013-06-14T19:03:38+00:00
<a href="#">BAV ADV12590677</a>	add	2013-08-19T18:59:48+00:00
<a href="#">ISNI 0000000109125967</a>	add	2013-09-16T16:15:07+00:00
<a href="#">NSK 000452769</a>	add	2013-12-12T18:02:00+00:00
<a href="#">NUKAT n 97002164</a>	add	2014-10-27T16:17:29+00:00
<a href="#">WKPI Harold_Alexander,_1st_Earl_Alexander_of_Tunis</a>	delete	2015-04-14T18:56:54+00:00
<a href="#">WKPI Q39168</a>	add	2015-05-12T00:14:00+00:00
<a href="#">N6 lvt s000075104</a>	add	2015-05-12T00:14:00+00:00
<a href="#">DNB 171958489</a>	delete	2016-07-11T03:44:24.122600+00:00
<a href="#">DNB 12257365X</a>	add	2016-07-11T10:53:42.709355+00:00

Plate 4: ... this information is automatically corrected in VIAF (<http://viaf.org/viaf/69064291>).

is generated. However, it is rather difficult to follow entities that are split off from an existing cluster to another one, especially if the former cluster continues to exist.

By displaying all these data, VIAF offers the opportunity to have an overview of the diversity of information and to detect possible contradictions or errors (*ibid.*: 11–13). Problems that are generated on a local level, especially the existence of entities that cannot be differentiated due to lack of relevant information, such as date of birth or date of death, are thus also transferred to the VIAF level. On a practical level, these characteristics of VIAF should be considered if VIAF IDs are linked or used by a project. Referring to a VIAF ID means referring to a cluster of automatically assembled information from diverse data providers. Especially in the context of other big data projects,



this can cause problems (Wikidata 2016). Projects referring to VIAF should always be aware that these clusters can change their content and that it could be possible that a link has to be checked and updated manually.

## **International Standard Name Identifier (ISNI)**

ISNI, the International Standard Name Identifier starts exactly at this point. As an official international standard and in contrast to VIAF, it has been conceived of to persistently identify persons and organizations. In contrast to VIAF, the wish to establish the ISNI evolved not in the context of library and information science, but was a desideratum of industrial stakeholders to facilitate the rights management for creative works (ISNI 2016). It complements other international norms, such as the ISBN for books or the ISSN for periodicals (Angjeli 2012: 101). ISNI exists of so-called Public Identities that are defined in its accompanying ISO-Standard 27729 as “the identities used publicly by parties involved throughout the media content industries in the creation, production, management and content distribution chains” (Angjeli/Mac Ewan/Boulet 2014: 7).

This basic principle is important in order to understand the functioning of ISNI. It implies that if a person has more than one public identity, e.g. if that person uses a pseudonym, all known public identities become a separate ISNI ID. However, related identities can be linked. To reach the goal of a reliable and persistent management of identifiers appropriate to an international standard, ISNI actively works with its data. This administration of content influences the way data are handled. While in VIAF, the data from the providers remain stable, ISNI has decided to actively maintain and modify it. In addition, so-called undifferentiated data that do not provide enough information to clearly identify a person are not assigned with an ISNI ID.

Historical projects planning to use ISNI as a possible reference should be aware that its focus is on the contemporary media content industry. There is no information available about the effort the project spends on the maintenance of historical data. Moreover, since ISNI is nevertheless integrated in the continued exchange of data and information, it can also contain wrong datasets. To address this problem, ISNI has its own messaging system to data providers to inform them about “assignments, merges, splits and deletions” (ibid.: 14).

## ***Metagrid***

While GND and VIAF have evolved in the context of libraries and ISNI serves first and foremost as a commercial tool, *Metagrid* was developed directly within the humanities community based on the first-hand experiences of daily work with historical information about persons. It is a project of the Swiss Academy of Humanities and Social Sciences for the online networking of humanities resources. It is implemented by the *Diplomatic Documents of Switzerland*, with the support of the *Historical Dictionary of Switzerland*. Conceived of as a web service, *Metagrid* makes it possible to set up, administrate and analyze links between identical entities from different websites and databases. In contrast to the aforementioned projects, which primarily aim at the creation of a normative dataset (GND), the collection and managing of existing authority files (VIAF) or the creation of persistent identifiers (ISNI), *Metagrid* wants to facilitate the handling of links between research projects and galleries, libraries, archives, museums (GLAM) and institutions. Traditional link connections from A → B operate on a unilateral basis only and it is the user of web page A alone who benefits from the connection. *Metagrid* generates multilateral link connections between A and B, through which the user is able to switch back and forth between web pages A and B with the help of a widget.

The *Metagrid* webservice collects the specific URL and the distinct ID of a person in a central archive. Whenever a match is detected, the entity's corresponding IDs from both participating projects are equated. Comparable to VIAF, every participating project retains full control of its own data, because at its core, *Metagrid* stores only the information about corresponding IDs. With this approach, the *Metagrid* partners can benefit from the specific knowledge of the other participating projects and avoid the multiplication of the same work done by every project.

The additional value of *Metagrid* (plate 5) can best be explained with a practical example. In the database of the *Diplomatic Documents of Switzerland* (dodis.ch), we can find some basic information about Elihu Root, because he is mentioned in some documents (plate 6). Thanks to *Metagrid* it is possible to link up with the website of the Office of the Historian of the U.S. Department of State, that provides detailed information about his biography (plate 7). You see that he was Secretary of State from July 1905 until January 1909 and that he was awarded the Nobel Peace Prize in 1912 (several years after he was Secretary of State) for his work on international arbitration. From

**metagrid.ch**

Metagrid search

Elihu Root

About 2 results

Elihu Root 1845-01-01 1937-01-01 (Concordance #8661)

[Permalink](#) [JSON](#)

The following partners have information about the person.

- Dodis [Root Elihu 1845-01-01 1937-01-01](#)
- Lonsea [Root Elihu 1845-02-15 1937-02-07](#)
- Wikipedia [Root Elihu 1845-05-05 1937-07-07](#)
- U.S.-Chiefs of Mission (FRUS) [Root Elihu 1845-01-01 1937-01-01](#)

Plate 5: Metagrid search is an interface to search all data in Metagrid ([http://www.metagrid.ch/metagrid\\_search/#/concordance/8661.html](http://www.metagrid.ch/metagrid_search/#/concordance/8661.html)).

Diplomatic Documents of Switzerland 1848-1975

**Dodis - Simple search**

Search

**Information about Person [dodis.ch/P34767](http://dodis.ch/P34767)**

 **Root, Elihu**  
\* 15.2.1845, † 7.2.1937

Links [Metagrid.ch](#): [Lonsea](#) [Wikipedia](#) [U.S.-Chiefs of Mission \(FRUS\)](#)

**Mentioned in the documents (6 records found)**

Date	Nr	Type	Subject	Summary	L
14.2.1906	<a href="#">42962</a>	Letter		Russland versucht, die Revision der Genfer Konvention an die Haager Friedenskonferenz zu bringen. Die Politik der amerikanischen Regierung richtet sich in dieser Frage nicht gegen die schweizerischen...	de
11.5.1907	<a href="#">43025</a>	Political report		Ausser einer Rede des Staatssekretärs Elihu Root bot die nationale Friedenskonferenz in Washington wenig Interesse. Die USA befürworten die Schaffung eines permanenten Gerichtshofs im Haag, sind aber...	de

Plate 6: Information about Elihu Root's diplomatic activities in the database dodis.ch (<http://dodis.ch/P34767>).



The screenshot shows the 'Office of the Historian' website. At the top left is the Department of State seal. The main header reads 'OFFICE OF THE HISTORIAN'. A search bar is on the top right. A blue navigation bar contains links: Home, Historical Documents, Department History, Guide to Countries, More Resources, and About Us. Below this, a breadcrumb trail reads 'Home > Department History > People > Elihu Root'. The main heading is 'Biographies of the Secretaries of State: Elihu Root (1845–1937)'. Underneath is an 'Introduction' section with a paragraph about Root's appointment. To the right is a grey box with biographical details: Born 1845, Died 1937, Non-career appointee, State of Residence: New York, and a list of roles starting with 'Secretary of State'. Below the introduction is a 'Rise to Prominence' section with a paragraph about his early career. To the left of this section is a portrait of Elihu Root. Below the portrait is a caption: 'Elihu Root, 38th Secretary of State president of the U.S. Bar.' At the bottom right of the page is a 'DEPARTMENT HISTORY' button.

Plate 7: Elihu Root's biography on the website of the Office of the Historian of the U.S. Department of State (<https://history.state.gov/departmenthistory/people/root-elihu>).



The screenshot shows the 'Lonsea' website profile for Elihu Root. The header includes the Lonsea logo and the name 'Elihu Root'. Below the name is a summary line: '♂ • 15/2/1845 (in Clinton: New York: United States of America) † 7/2/1937 (in New York: United States of America)'. The 'Nationality' is listed as 'US-American'. The 'Roles in Organisations' section lists 'Institut des Hautes Etudes Internationales : Member of a Committee (1921-1929)'. The 'Delegate' section is empty. The 'Residences' section lists 'Clinton: New York: United States of America' and 'New York: United States of America'. The 'Links from Metagrid.ch' section lists 'Dodis', 'Wikipedia', and 'U.S.-Chiefs of Mission (FRUS)'. On the right side, there is a portrait of Elihu Root with a 'Mark for connection search' button. On the left side, there is a navigation menu with categories: Start, Search, Organisations, People, Places, Topics, Connections, Visualize, Network, Organisation Arc, Country Matrix, Charts, Congresses (via VisualEyes), Bibliography, and Contact.

Plate 8: Information about Elihu Root's activities in international organizations in Lonsea (<http://www.lonsea.de/pub/person/456>).

the Office of the Historian's website you can continue to *Lonsea*, the *League of Nations Search Engine*, created by two projects of the Cluster of Excellence *Asia and Europe in a Global Context* of the University of Heidelberg. Focused on the activities of international organizations in the 1920s and 1930s, *Lonsea* indicates that Elihu Root was a member of the Committee of the Institut des Hautes Études Internationales and who his colleagues there were (plate 8).

Expanding to further project partners, like the Swiss National Library, *Metagrid* offers easy access to highly qualitative research results, source material and scholarly literature from and about historical agents.

## Conclusion

The digital turn in the humanities has reached a new level. After a first exploratory stage that resulted in valuable, but mostly separated results, the increased awareness of the benefits of linked data called for new solutions to connect the outcome of digital scholarly efforts. These approaches are supported by a common set of established technologies and standards that serve the very basic needs of digital humanities, such as the utilization of TEI-XML for text encoding, the storage of information in relational databases, or the description of data with RDF. However, the application of a common virtual research environment or rigid standards for dealing with source material appear not to be greatly appreciated in the humanities, with its highly variable and specialized subfields and needs. It is therefore at a meta-level where networking activities are promising and feasible. This meta-level consists of information about persons, corporate bodies, locations, and in some cases also concepts. At this point, humanists enter the classical field of libraries, archives, and information science, a field with long-term experience with the systematic indexing of its holdings, and that has had to deal, as early adopters, with the opportunities and challenges of the digital turn.

As discussed in the case studies above, multiple projects and approaches to handling data about persons exist. Each of these projects has its advantages and disadvantages, and digital humanities projects should decide what fits their needs best. The case of the GND has shown that the support of funding agencies can regulate the integration of data interchange. Moreover, cooperation with external players such

as *Wikipedia* can boost the acceptance and application of standards and technologies. On an international level, VIAF tries to bring together the authority files from national providers and connects them automatically. Here, the language independence of numerical identifiers for persons with a broad variety of possible names is best visible. However, the federative philosophy of VIAF, which does not interfere with the data provided, brings with it a certain instability concerning the content of VIAF clusters. With its explicit goal of persistently identifying persons in a long-term perspective, ISNI tries to overcome this deficit of VIAF, at least for contemporary entities.

From a methodological point of view, we have to be aware that all these initiatives are based in the broader context of libraries. This means that historical persons who had a public visibility in the past, be it for example by writing a book, composing a piece of music or creating a work of art, might be overrepresented. Depending on the research question, it is possible that only a negligible proportion of people is recorded in a national authority file, VIAF or ISNI. In these cases, it is important that projects try to maintain their data as best as possible and to manage their own persistent IDs. The example of *Metagrid* shows that research communities can create their own infrastructures to serve their specific needs. In a technological context, that fosters the idea of linked open data.

The main methodological question as a result is whether a research project wants to refer to a centralized authority to persistently identify a person as exactly this person, or whether a project wants to connect with other projects to gain more visibility and to share the results of its research. In the latter case, established initiatives such as GND, VIAF or ISNI are dealt with like other research projects and connected, if a match is possible. The future will show whether there will still be a need for the central organization of authority control, or whether the possibilities of the semantic web will foster direct connections between similar entities. The best way each project can prepare for this future is to carefully maintain its own data and to be open for new developments.

## Abbreviations

GLAM	Galleries, libraries, archives, museums
GND	Gemeinsame Normdatei (German Integrated Authority File)
ISBN	International Standard Book Number
ISNI	Internationale Standard Name Identifier
ISO	International Organization for Standardization
ISSN	International Standard Serial Number
OPAC	Online Public Access Catalogues
PND	Personennamendatei
RDF	Resource Description Framework
TEI	<i>Text Encoding Initiative</i>
URL	Uniform Resource Locator
VIAF	Virtual International Authority File
XML	Extensible Markup Language

## References

- Angjeli, Anila. 2012. “ISNI: un identifiant passerelle.” *Documentation et bibliothèques* 58: 101–108.
- Angjeli, Anila, Andrew Mac Ewan, and Vincent Boulet. 2014. “ISNI and VIAF: Transforming Ways of Trustfully Consolidating Identities.” Paper presented at IFLA WLIC 2014, 16–22 August 2014, Lyon, France. <<http://library.ifla.org/985/1/086-angjeli-en.pdf>> (Accessed May 10, 2016).
- Barbera, Michele et al. 2013. “Annotating Digital Libraries and Electronic Editions in a Collaborative and Semantic Perspective.” In: *Digital Libraries and Archives: 8th Italian Research Conference, IRCDL 2012, Bari, Italy, February 9–10, 2012, Revised Selected Papers*, ed. by M. Agosti et al. Berlin, Heidelberg: Springer, 45–56.
- Berners-Lee, Tim. 2006. “Linked Data: Design Issues.” <<https://www.w3.org/DesignIssues/LinkedData.html>> (Accessed May 10, 2016).
- Ciotti, Fabio, Maurizio Lana, and Francesca Tomasi. 2014. “TEI, Ontologies, Linked Open Data: Geolat and Beyond.” *Journal of the Text Encoding Initiative* 8. DOI: 10.4000/jtei.1365.
- Diplomatic Documents of Switzerland*. 2016. “dodis.ch.” <<http://dodis.ch/>> (Accessed May 10, 2016).
- Ebneth, Bernhard. 2010. “Vom digitalen Namenregister zum europäischen Biographie-Portal im Internet: Mit einem Bericht aus der Arbeit der Neuen

- Deutschen Biographie.” In: *Jahresbericht 2009*, ed. by Historische Kommission bei der Bayerischen Akademie der Wissenschaften. München, 37–56. <[http://www.historischekommission-muenchen.de/fileadmin/user\\_upload/pdf/jahresberichte/jahresbericht2009.pdf](http://www.historischekommission-muenchen.de/fileadmin/user_upload/pdf/jahresberichte/jahresbericht2009.pdf)> (Accessed May 10, 2016).
- Eide, Øyvind. 2014. “Ontologies, Data Modeling, and TEI.” *Journal of the Text Encoding Initiative* 8. DOI: 10.4000/jtei.1191.
- Europeana.eu. 2014. “Europeana Professional.” <<http://pro.europeana.eu/>> (Accessed May 10, 2016).
- Fabian, Claudia. 1995. “Entwicklung und Aufbau der Personennamendatei in Deutschland: Bericht über Konzeption und Realisierung seit 1989.” *Zeitschrift für Bibliothekswesen und Bibliographie* 42: 605–616.
- Geipel, Markus et al. 2013. “Herausforderung Wissensvernetzung: Impulsgebende Projekte für ein zukünftiges LOD-Konzept der Deutschen Digitalen Bibliothek.” In: *(Open) Linked Data in Bibliotheken*, ed. by P. Danowski and A. Pohl. Berlin, Boston: Walter de Gruyter, 168–185.
- GeoNames. 2016. “GeoNames.” <<http://www.geonames.org/>> (Accessed May 10, 2016).
- Getty Research Institute. 2016. “Getty Thesaurus of Geographic Names.” <<http://www.getty.edu/research/tools/vocabularies/tgn/>> (Accessed May 10, 2016).
- Hodel, Tobias, and Christiane Sibille. Forthcoming. “Die angereicherte Edition: Ideenskizze zur Nutzung des Potentials digitaler Vernetzung (eine Utopie?).” *Itinera: Beihefte zur Schweizerischen Zeitschrift für Geschichte*.
- Hooland, Seth van et al. 2015. “Exploring Entity Recognition and Disambiguation for Cultural Heritage Collections.” *Digital Scholarship in the Humanities* 30: 262–279.
- Iglesia, Martin de la, and Mathias Göbel. 2014. “From Entity Description to Semantic Analysis: The Case of Theodor Fontane’s Notebooks.” *Journal of the Text Encoding Initiative* 8. DOI: 10.4000/jtei.1253
- Institut für Zeitgeschichte. 2016. “Akten zur Auswärtigen Politik der Bundesrepublik Deutschland.” <<http://www.ifz-muenchen.de/publikationen/editionen/ed/edition/akten-zur-auswaertigen-politik-der-bundesrepublik-deutschland/>> (Accessed May 10, 2016).
- Monumenta Germaniae Historica*: “Publikationen: Allgemeines: MGH.” <<http://www.mgh.de/publikationen/publikationen-allgemeines/>> (Accessed May 10, 2016).
- Niu, Jinfang. 2013. “Evolving Landscape in Name Authority Control.” *Cataloging & Classification Quarterly* 51: 404–419.
- Pierazzo, Elena. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. New edition. Farnham, Surrey, UK; Burlington, VT: Ashgate Pub Co.



- Regesta Imperii*. 2016. “Publications: Regesta Imperii.” <<http://www.regesta-imperii.de/en/research/publications.html>> (Accessed May 10, 2016).
- Sahle, Patrick. 2013. *Digitale Editionsformen: Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. 3 vols. [Preprint-Version]. University of Cologne.
- Schmidt, Desmond. 2014. “Towards an Interoperable Digital Scholarly Edition.” *Journal of the Text Encoding Initiative* 7. DOI: 10.4000/jtei.979
- Text Encoding Initiative*. 2016. “TEI: Projects Using the TEI.” <<http://www.tei-c.org/Activities/Projects/index.xml>> (Accessed May 10, 2016).
- Wikidata*. 2016. “Wikidata:Database reports/Constraint violations/P214.” <[https://www.wikidata.org/wiki/Wikidata:Database\\_reports/Constraint\\_violations/P214](https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/P214)> (Accessed, November, 14 2016).

