

SUSTAINABILITY AND BEST PRACTICES FOR LINKED DATA HERITAGE RESOURCES: SOME CASE STUDIES FROM SWEDEN

Agnieszka Backman and Marcus Smith*

digital infrastructure projects and scholarship are becoming increasingly significant, and digital projects within the humanities are becoming the norm rather than the exception. At the time of writing, in 2020 and 2021, this trend was thrown into sharp relief when the Covid-19 pandemic forced most scholars to work from home and digital access to cultural heritage went from being desirable to essential. In this chapter we describe and argue for best-practice approaches to digital humanities data involving or in relation to linked open data (LOD), which uses standardized definitions, design principles, ontologies, and open licences to create opportunities for reuse across datasets. It furthermore allows scholars to interrogate data from multiple sources to answer research questions in ways that were not previously possible, and provides the opportunity to enrich or compare datasets with one another.

A digital project should not simply be an analogue project but done with computers; it has its own challenges, but also offers new opportunities. One difference is long-term access: “Digital work does not, and in fact cannot, valorize durability in the same way as conventional modes of scholarship because digital workspaces are connected to the ever-changing ecosystem of the Internet.”¹ In the humanities, there is an expectation for

* Agnieszka Backman, Uppsala University, got her PhD in Scandinavian languages in 2017 from Uppsala University on the dissertation *The Materiality of the Manuscript: Studies in Codex Holmiensis D 3, the Old Swedish Multitext Manuscript Fru Elins bok*. From 2017 to 2019 she worked on the “Norse Perception of the World” project, helping develop a digital platform for interdisciplinary research. From 2019 to 2021 she was a Wallenberg Postdoctoral Fellow at Stanford University, working on a project on the materialities of medieval manuscripts in digital repositories by using social semiotics to investigate the affordances of digitized manuscripts. ORCID iD: <https://orcid.org/0000-0002-9233-0407>.

Marcus Smith, Swedish National Heritage Board, read archaeology (BA) and archaeological information systems (MSc) at the University of York, and Viking and early medieval studies (Magister) at Uppsala University. He is chair of the Swedish chapter of Computer Applications and Quantitative Methods in Archaeology (CAA) and has previously worked at the Archaeology Data Service and the Council for British Archaeology. Since 2012 he has been based at the Swedish National Heritage Board in Visby, where he works with the SOCH linked open data aggregation platform on digital archaeological fieldwork documentation and the digitization of runic inscriptions. Email: marcus.smith@raa.se; ORCID iD: <https://orcid.org/0000-0002-7438-5498>.

1 Morreale, “Medieval Digital Humanities,” 5.

the output to be relevant and preserved for decades or even hundreds of years, such as in the form of archived books in libraries. In contrast, the output of digital humanities projects can often be radically different from conventional scholarship. Although humanities research is likely to result in, for example, articles or monographs, it's equally likely that the result of a digital humanities project will be an interactive online resource, digital media, or a digital dataset, which brings with it new challenges of sustainability.

Part of the problem is funding. The current system for funding is not set up for digital projects: the assumption is that a project will run and at the end produce a deliverable, not a resource with ongoing upkeep costs, such as hosting or maintenance. To really take advantage of open datasets this needs to be resolved, but it can mean different things for different kinds of projects, and this chapter will help in navigating between them. The system being what it is, it is nonetheless somewhat possible to successfully plan for or accommodate a digital project.

This chapter deals with a number of concepts and acronyms that may be unfamiliar to readers unacquainted with linked open data; a glossary is provided at the end of the chapter to explain many of these terms, with references for further reading.

The Case for Linked Open Data

The principles of linked open data and the semantic web were first articulated by Sir Tim Berners-Lee as a logical extension of the World Wide Web.² They need not be recounted in detail here but, in essence, instead of a disparate web of human-readable linked *documents*, linked data consists of a semantic web of machine-readable linked *data*, where by “semantic” we mean that edges (links) between nodes (resources) describe the nature of the relation they represent, and where IRIs³ are used not only as unique identifiers for records and digital objects but also, ideally, as resolvable addresses on the World Wide Web.⁴

In much the same way that HTML is the standard of interchange used to describe the web of documents we're familiar with, the Resource Description Framework (RDF) is the model used to describe the web of data. RDF statements take the form of simple three-part statements, *subject* → *predicate* → *object*, in which a thing—the *subject*, typically identified by its IRI or following on from another statement—is described using attributes or properties—the *predicate*, given using an IRI—that take either string values, an IRI, or a reference to the subject of another such triple: the *object*. All

2 Tim Berners-Lee, “Linked Data.”

3 Internationalized resource identifiers, a generalization of URIs (uniform resource identifiers). These are unique machine-readable identifiers that include—but are not limited to—the resolvable URLs most Internet users are familiar with. See Duerst and Suignard, “RFC 3987” (IRIs), Berners-Lee, Fielding, and Masinter, “RFC 2396” (URIs), and URI Planning Interest Group, “URIs, URLs, and URNs,” for the (lack of) distinction between URIs and URLs.

4 “Resolvable” in the sense that such identifiers take the form of valid locations on the global Internet and return structured data when accessed.

assertions in the web of linked data, from the abstract to the nuanced, are built using this simple, yet powerful, pattern. Linked data services can be queried both singly and in a federated fashion—combining results across multiple systems simultaneously—using the SPARQL query language, which acts as a common application programming interface (API) for RDF resources.⁵

Open—in the sense of openly licensed—data need not necessarily be *linked*, however, and Berners-Lee has proposed a five-star model describing various levels of open data, from unstructured documents published on the web (one star), through structured data published on the web, using open formats, using linkable IRIs as identifiers, to fully linked open data that links itself to other resources (five stars).⁶

From the perspective of researchers, publishing and making use of linked open data offer a number of benefits. At a basic level, publishing as linked open data provides an opportunity to present and highlight data in an accessible, standardized, and reusable way—data that may otherwise be obscured in articles or languish sealed in appendices in traditional publications. The key way in which linked data publication differs from other solutions to this problem, such as tabular data, data papers, or online journals, is that, by its very nature, it places a dataset within a broader context on the semantic web, and allows it to be easily augmented by, compared with, and combined with other related datasets.

In traditional, “analogue” humanities research, datasets are seldom given centre stage, so even an ephemeral resource with properly archived data is a big step up in research transparency. It is also in line with the perspective of open science,⁷ FAIR data,⁸ and “open research that advocates for exposing the research process rather than hiding the mess behind authoritative results,”⁹ and will also be a source for future studies of research processes, investigating our time’s preconceived notions and biases.

The strengths of linked open data are interoperability, interconnectedness, and reusability. By publishing data on the web using persistent dereferenceable identifiers, it is easier to cite, check, and reuse. By describing data with the help of shared vocabularies—expressed as, for example, SKOS¹⁰ or RDFS¹¹—and according to the frameworks of established ontologies—expressed as OWL.¹² By making use of

5 Harris, Seaborne, and Prud’hommeaux, “SPARQL 1.1 Query Language.”

6 Tim Berners-Lee, “Linked Data.”

7 See, for instance, the Center for Open Science, www.cos.io (accessed May 11, 2021), and European Commission, “Open Science.”

8 Findable, Accessible, Interoperable, Reusable; see www.go-fair.org/fair-principles (accessed May 11, 2021).

9 Rockwell, Day, Yu, and Engel, “Burying Dead Projects,” para. 34.

10 Miles and Bechhofer, “SKOS.”

11 RDF Schema; see Brickley and Guha, “RDF Schema 1.1.”

12 Web Ontology Language; see Motik, Patel-Schneider, and Parsia, “OWL 2 Web Ontology Language.”

vocabularies and authority files already in established use, projects can ensure that the data is compatible with others' data within the same field¹³ and that attributes are decoupled from identifiers (that is, that attribute values and terms derived from them are not used as identifiers). By linking a project's data resources against other published data, it may enrich not only its data but also others'.

Whereas vocabularies define the terms that can be used to describe things, ontologies define what can be described, and the inferences that can be made from those descriptions. Common useful vocabularies within the heritage sector include *inter alia* the Getty Art & Architecture Thesaurus (AAT);¹⁴ PeriodO¹⁵ for spatio-temporal chronologies; sites and monuments types of thesauri, such as those at Heritage Data;¹⁶ VIAF¹⁷ for authors; and GeoNames¹⁸ (and relevant national authorities) for places. Linking to generic and well-known meta-authorities such as Wikidata¹⁹ is also useful. Common ontologies in use include the Dublin Core (DC)²⁰ minimal set of metadata attributes, the Friend-of-a-Friend (FOAF)²¹ ontology for describing people and institutions, the CIDOC-CRM²² conceptual reference model for heritage assets, and the Europeana Data Model (EDM)²³ used by the international cultural heritage aggregator Europeana.²⁴ In (increasingly rare) cases when a project or institution finds that its domain or terms are not covered by existing established vocabularies or ontologies, the project or institution may decide to create its own, or to augment an existing resource.

This is all very well in theory, but when implementing a linked data strategy there are practical considerations to bear in mind: not just things such as choice of ontology and vocabularies, but a change in mindset. Interoperability becomes a logical extension of linked infrastructures but policy, standards, and business models often determine interoperability more than the technology as such.²⁵ There is also a paradigm shift when describing items with RDF, from a focus on the item itself to its relationships or

13 Or, if choosing to publish a bespoke vocabulary or ontology, ensure that *it* allows comparison to others'.

14 See <http://vocab.getty.edu> (accessed May 11, 2021).

15 See <https://perio.do> (accessed May 11, 2021).

16 See <https://heritagedata.org/live/schemes> (accessed May 11, 2021).

17 See <http://viaf.org> (accessed May 11, 2021).

18 See www.geonames.org (accessed May 11, 2021).

19 See www.wikidata.org (accessed May 11, 2021).

20 See <http://purl.org/dc/elements/1.1> and <http://purl.org/dc/terms> (both accessed May 11, 2021).

21 See <http://xmlns.com/foaf/0.1> (accessed May 11, 2021).

22 See <http://cidoc-crm.org/cidoc-crm> (accessed May 11, 2021).

23 See www.europeana.eu/schemas/edm (accessed May 11, 2021).

24 See www.europeana.eu (accessed May 11, 2021).

25 Thylstrup, *The Politics of Mass Digitization*, 68.

entities,²⁶ which can be helpful in investigating the data from new angles. By describing items in terms of not only their attributes but their relationships to other items (and the records describing them), they are placed within a wider context beyond a single dataset. It is also important to distinguish between the physical item itself and the digital document(s) that describe(s) it; often, in an application of RDF, both will need their own IRIs with their own attributes (e.g., “date created” will likely be different for each) and to be related to one another. To ensure that data are still returned when IRIs for physical objects are resolved, one can use redirects or IRI fragment identifiers.²⁷

Sustainability

There is a paradox between the popular axiom that the Internet is forever and the reality that, if no one is paying and doing the upkeep, what has been put online will become unreadable or even unavailable.²⁸ The phenomenon of link rot—of previously valid URLs no longer resolving or generating an error—is one that is all too familiar to many, not least within academic citations of online work. Furthermore, the costs of producing, publishing, and maintaining access to digital resources—particularly LOD—are often not borne by those who benefit most from them. That is, it is not the owners of data who are in greatest need of the information being published in an open, accessible format, but their time and resources are needed to make it happen. Without adequate planning for permanent identifiers, ongoing management, platform migration, digital preservation, etc. the half-life of online digital resources can be short. In fact, long-term funding for the maintenance of projects is a contributing factor to digital work disappearing.²⁹

Sustainability for a digital humanities project relies on a variety of circumstances that might not be immediately evident to a newcomer to digital humanities work. As linked open data are especially dependent on stable IRIs to facilitate future reuse, the long-term accessibility of an LOD project is one of the cornerstones everything else depends on. Every time URLs change, links are broken, and hours of work will be needed to restore access from the other pages on the web that link to that resource. What exactly that sustainability is will differ depending on project aims, so, given the short-term nature of project funding, ensuring long-term access to data can pose a challenge. Research institutions are often reluctant to commit to publishing data—much less maintaining a digital platform—for an indefinite period after a project has ended and funding has run out. Other options, such as depositing with an accredited digital archive,

26 Smith-Yoshimura, “Summarizing Project Passage Experiences.”

27 See the `httpRange-14` problem and its resolution: Fielding, “[`httpRange-14`] Resolved.” The `httpRange-14` problem refers to the question of how services should correctly respond when a client requests `http` IRIs for physical or otherwise non-digital things for which no electronic representation exists, and how the identifier for the thing itself relates to identifiers for digital documents describing the thing. See “TAG Issues List: `httpRange-14`.”

28 See Chun, “The Enduring Ephemeral,” 169.

29 Morreale, “Medieval Digital Humanities,” 9.

may have to be considered. Accredited digital archives for research-produced heritage data exist in a number of countries, such as Digital Antiquity³⁰ in the United States, the Archaeology Data Service (ADS)³¹ in the United Kingdom, DANS³² in the Netherlands, and the Swedish National Data Service (SND)³³ in Sweden. Accreditation is conferred by independent bodies, such as Core Trust Seal,³⁴ nestor,³⁵ and the Digital Preservation Coalition (DPC).³⁶

Different Options

There is a difference between long-term access to data and a website (whether it is called platform, portal, resource, or something else) being created to present the digital project research output. If the goal is an up-and-running infrastructure, it will need dedicated resources in the form of maintenance and server costs. The bland word “maintenance” hides a lot of work: systems administration, updating code to eliminate security issues, optimizing search algorithms, and remaining compatible with contemporary hardware, software, and technical standards. The Internet of ten years ago was and looked very different from how it is today, and it will most likely change as much in the next ten years. An infrastructure, especially if it is intended to connect to other platforms or social media, will need to be prepared to work even harder to stay the “same” in the eyes of the users. Major changes will also in all probability require someone from the research side able to make overarching decisions for the resource, as updates might impact the functionality of the site or changes to the connecting platforms necessitate changes to the project site.³⁷

There are different options for how to address sustainability. The Celtic Inscribed Stones Project (CISP) from University College London³⁸ is an example of a comparatively early (non-LOD) online digital heritage project. Published in 2001, it is a queryable database of early medieval Celtic epigraphic inscriptions. Although the interface appears dated by modern standards, the resource was still published under its original URL twenty years later (i.e., at the time of writing). The project creators planned for sustainability, however, and when it was completed the dataset was also deposited with the Archaeology Data Service³⁹ for long-term preservation. Consequently, when the

³⁰ See www.digitalantiquity.org (accessed May 11, 2021).

³¹ See <http://ads.ahds.ac.uk> (accessed May 11, 2021).

³² See <https://dans.knaw.nl> (accessed May 11, 2021).

³³ See <https://snd.gu.se> (accessed May 11, 2021).

³⁴ See www.coretrustseal.org (accessed May 11, 2021).

³⁵ See www.langzeitarchivierung.de (accessed May 11, 2021).

³⁶ See www.dpconline.org (accessed May 11, 2021).

³⁷ Morreale, “Medieval Digital Humanities,” 8.

³⁸ See <https://dx.doi.org/10.5284/1000215> (accessed May 11, 2021).

³⁹ See <http://ads.ahds.ac.uk> (accessed May 11, 2021).

original platform inevitably stops working in the fullness of time, the dataset will still be available from the ADS archive via its DOI⁴⁰ (see “Technical Long-Term Availability/IRIs” section below for more on DOIs).⁴¹

If there is a way to reuse an existing platform it is often better than building a new one, as resources may be pooled for upkeep and more stakeholders can help preserve the platform. If the new project collaborates with an infrastructure already connected to a community, it might also get an inbuilt audience of users from the start, either a scholarly community or the general public. Building entirely upon an existing platform has the potential drawback, however, of reduced flexibility: it might be necessary to adjust the project’s data model to fit that of the resource, and accept the limitations of not having the entities needed. If there is a good match between the existing model and a project’s needs, though, it can be a preferable solution. Additionally, by putting the project output in someone else’s hands there is also a risk of potentially losing control of what happens to it. This fear is often strong when it comes to open-source projects that anyone can edit, such as Wikidata,⁴² especially for those not already part of the Wiki community. While it might not be advisable to save copies of project data only in a crowdsourced environment, the advantages of being part of a community working around the same data issues and for preservation are substantial.

A public infrastructure that anyone can edit is not the only valid reason not to choose Wikidata as a primary data deposit. In Project Passage, librarians from US institutions learned to create LOD without needing the technical know-how, in a Wikibase instance with a custom interface.⁴³ Wikibase⁴⁴ software powers Wikidata, but it can also be installed on a user’s own machine to store and manage structured data. One of Project Passage’s goals was to evaluate the software for librarians, but most relevant here were the reasons as to why this was not done directly in Wikidata; these considerations had to do with time, scope, and privacy. As the project planned to be experimenting with new property entities, these might not have been accepted for inclusion in the Wikidata data model in time for use during the project’s limited time span. For that matter, the entities might not be notable enough—i.e., general and widely accepted—for inclusion in Wikidata.⁴⁵ As the data in the project’s use cases could be experimental or provisional, privacy to try things out was also an important concern.⁴⁶

Yet another approach was taken by the Ogi Ogham project and subsequent developments under the fellow program Freies Wissen,⁴⁷ for which a fully linked data

40 For example, CISP, <https://dx.doi.org/10.5284/1000215> (accessed May 11, 2021).

41 See www.doi.org (accessed May 11, 2021).

42 See <https://wikidata.org> (accessed May 11, 2021).

43 Godby et al., *Creating Library Linked Data with Wikibase*.

44 See <https://wikiba.se> (accessed May 11, 2021).

45 Wikidata, “Wikidata:Notability.”

46 Godby et al., *Creating Library Linked Data with Wikibase*, 11.

47 Thiery et al., “Ogi Ogham Project.”

platform for Celtic ogham inscriptions has been created, almost exclusively building upon existing open platforms and datasets. Data about the inscriptions, derived from a number of sources—including the aforementioned CISP database—are stored and structured according to a custom CIDOC-CRM-based ontology and published as linked open data with a SPARQL endpoint. The inscriptions are linked against corresponding entities in Wikidata, allowing the records to be further developed by the community, thereby augmenting the resource. The project is run on Wikiversity, and the code published on GitHub.⁴⁸ In this way the project makes the most of existing infrastructure and increases its reach and impact by embracing open methodologies.

When designing a project, it is important to have a clear idea of what the project's data could be used for, which will have a bearing on platform and sustainability decisions.

Personnel Requirements

In research environments, funding for projects is often short-term and tied to an individual or an institution, and only recently has there been a push from Swedish funding agencies to require sustainability plans. It shouldn't be forgotten that "length of a project's life cycle, from the first creative moments until it goes dark, relies as much on the personal circumstances of those needed to maintain the project"⁴⁹ as the technology used. Often there is an implied assumption of stable employment at one institution, preferably as a (tenured or the equivalent) professor, with the clout and resources to put in the time to keep the project up to date and available. Digital project creators with more precarious employment—short-term, early-career, or alt-ac—will often be expected to move for employment opportunities, leaving digital projects without a champion at their home institutions, and have very little power to dedicate work hours to keeping the resource alive. A model in which volunteering after-hours work to keep digital resources online is an inherently unsustainable, unstable, and unfair practice, impacting people with, for instance, small children or health issues disproportionately.⁵⁰

Another issue for academics is that there is still no standard way for digital projects to be evaluated and peer-reviewed, making digital projects a career gamble. Especially when developing new technologies, this has a tendency to take up the majority of project time, at the cost of scholarship.⁵¹ Depending on the field and the nature of the project, it may be possible for the digital project to be reviewed as a companion to print output, helping with the issue of career advancement.⁵²

⁴⁸ See www.wikiversity.org and <https://github.com> (both accessed May 6, 2021).

⁴⁹ Morreale, "Medieval Digital Humanities," 9.

⁵⁰ See Dombrowski, "The Directory Paradox"; and Rockwell, Day, Yu, and Engel, "Burying Dead Projects," para. 50.

⁵¹ Dombrowski, "What Ever Happened to Project Bamboo?," 327.

⁵² Rockwell, Day, Yu, and Engel, "Burying Dead Projects," para. 48.

Infrastructural Planning

It is easy to be seduced by the possibilities of digital projects to try to be everything to everyone, perhaps especially in the humanities, in which funding opportunities are limited and there is often a feeling that this might be the only chance. After all, adding another field in a database is trivial, but too many tangential add-ons will make the design and workflow messy and the digital resource unwieldy, just as it would for any other sort of project. The first thing to do is to identify a clearly defined research niche and resist the scope creep that will almost certainly beckon. Funding permitting, focusing on research questions and letting them drive the digital methods and tools that are used will also lead to better projects.

Digital humanities projects can often be collaborations between several institutions, but choosing a partner or a collection comes with some epistemic perils. The origins of many digital projects in the cultural heritage sector are either in national collections or at elite institutions, as these are often deemed important for national digitization programs or other funding. National collections, often founded in the nineteenth century to establish the national state, come with national taxonomies and epistemic frameworks, whereas elite institutions can reproduce a class-based bias.⁵³ A consequence of this selection can be exclusion of data from under-represented communities, or foreign-focused or other low-status collections that do not fit a nationalistic narrative, which are either left completely undigitized or under-described in terms of metadata. National borders can easily impede digital humanities projects but digital tools also have a great potential to cross them. An interesting example is the Swedish Caribbean Colonialism project, which studies the Swedish governmental archives from Saint Barthélemy, preserved in the French Colonial Archives.⁵⁴

Applying existing linked data ontologies and vocabularies can come at a price if it means applying a model that is not an exact fit for the data, or is too abstract for the project's needs. Instead, the data model should be designed with the requirements of the project's research in mind, to be used within its own system. This bespoke model can then be mapped to a more generic and consumable LOD model using shared vocabularies and public ontologies for publication or aggregation.

Project Management Resources

For those who are new to digital humanities project planning, DevDH.org⁵⁵ has useful lectures on everything from conceptualizing the project to budgets, management, publicity campaigns, and more. One such aspect that might be new to digital humanities principal investigators is the need to manage bigger and more diverse teams due to the

53 Thylstrup, *The Politics of Mass Digitization*, 25, 52.

54 See www.digarv.se/en/participating-research-projects/svensk-karibisk-kolonialism-1784-1878 (accessed May 15, 2021).

55 Appleford and Guiliano, "DevDH."

fact that programmers often do not have domain knowledge and scholars do not have programming skills.⁵⁶

For documentation, both important internally and for potential users, Read the Docs is a free open option for documenting software that is especially useful if there are to be versions of the documentation, but it requires some technical expertise.⁵⁷ The Digital Documentation Process website offers a process through which digital humanities projects can create catalogue entries with persistent identifiers and a description of the project.⁵⁸ For simple deposits of papers or other material, Humanities Commons is also a non-profit option.⁵⁹

Maintenance and Planning for End of Life

The best time to start planning for the end of a project is at the very beginning. Wrapping up a project is hard work and takes a lot of time, but doing it in a sustainable way that allows for future research should be seen as one of the most important outcomes. If possible, try to hold back funding to use after the project's research objectives are finished so as to have the space to deposit it properly.⁶⁰

The most important question is to decide what makes sense to archive and deposit for the specific project. Examples of aspects to consider archiving are: content, code, process (research and encoding), user experience, project management documentation, deposit process documentation, and draft versions of all previous aspects. Even if user experience is documented, it doesn't mean it needs to be recreatable or emulated; screenshots and narrative descriptions are other valid approaches. Depositing every item on the above list would likely take almost as long to prepare as the project itself, and might be going overboard. Archiving the data with documentation should take priority. It is important to remember that some archiving is better than none if the task becomes too big. As long-term preservation is still more of an idea with unclear guidance, which we don't know the results of yet, it can be prudent to deposit at multiple (well-regarded and -established) repositories.⁶¹

Formats

The long-term preservation of digital data is an active rather than a passive process, requiring periodic integrity checks and interventions, not least among which is that

56 See also Dombrowski's syllabus, "DLCL 205/305: Project Management & Ethical Collaboration for Humanists."

57 See <https://readthedocs.org> (accessed May 7, 2021).

58 Fostano and Morreale, "The Digital Documentation Process."

59 See <https://hcommons.org> (accessed May 7, 2021).

60 Rockwell, Day, Yu, and Engel, "Burying Dead Projects," paras. 10, 22, 25, 31, 50.

61 Rockwell, Day, Yu, and Engel, "Burying Dead Projects," paras. 22, 37–41, 43, 46, 50.

of format migration. File formats change over time: specs are updated, new versions released, new formats launched, and old formats fall out of common use. Often newer software cannot read older file formats. Thus, for all but the very simplest text-based formats, in order for the underlying data to remain accessible and readable, the format it is stored in will need to be periodically updated and the data migrated.

In order to ensure that the data is readable by a variety of software, not tied to any particular product or company, and to maximize the likelihood that it will still be readable in the future, it is strongly advised to store it (or, at least, to deposit it) using open formats. These are formats whose specifications are open (not proprietary) and can be freely implemented by any software developers. This is generally good practice in any case, but is particularly important in the context of digital archiving, and should be planned for in advance. Naturally, the types of the formats used and their contents should be documented as part of the project's metadata.

Sound general advice on choosing open data formats suitable for preservation may be found at the Library of Congress⁶² and the Digital Preservation Coalition.⁶³ Specific advice on formats and metadata relevant for archaeological fieldwork is provided by the ADS/Digital Antiquity Guides to Good Practice.⁶⁴ Further useful resources and advice on ensuring open access to heritage data, choice of formats, and data harmonization may be found at the ARIADNEplus project.⁶⁵

Licensing

Just as with formats, it's important at an early stage in a project's life cycle to consider the licence the data will be released under. This will become more significant as the project proceeds, and a lot of time and trouble can be saved later by planning ahead.

The issue of the explicit licensing of research outputs is one that academia has often chosen to ignore within the humanities, but, as data sharing becomes more and more commonplace, so too the importance of clearly establishing the terms under which others may access, reuse, transform, augment, and reshare data. Consider that this may require explicit permission if the project is to use copyrighted data produced by others; for example, if a photographer is hired to produce visual documentation, the question of licensing will need to be addressed in their contract.

To facilitate reuse and machine-readability, it is recommended to limit choice of licence/rights statements to well-known existing licences, referenced by their IRIs. For practical purposes, this means Creative Commons (CC)⁶⁶ licences for open works, and

62 Library of Congress, "Sustainability of Digital Formats" and "Library of Congress Recommended Formats Statement: 2020–2021."

63 Digital Preservation Coalition, "File Formats and Standards."

64 Archaeology Data Service/Digital Antiquity, "Guides to Good Practice."

65 See <https://ariadne-infrastructure.eu> (accessed May 11, 2021); see in particular Hollander et al., "PARTHENOS Guidelines."

66 See <https://creativecommons.org> (accessed May 11, 2021).

RightsStatements.org⁶⁷ for more restricted works; in some cases Open Data Commons⁶⁸ licences may also be applicable.⁶⁹

Resist the temptation to use more obscure licences or to invent a bespoke one. First, the chances are good that most needs are met by existing licences and that they are better written than the ones a non-expert could produce on their own. Second, by using well-known licences and rights statements, human and machine agents alike are more likely to already understand the applicable terms under which your material may be used; making a human read the terms of your bespoke licence adds friction to reuse by human agents, and presents a barrier to reuse by machine agents.

It is advisable to license metadata under the Creative Commons Zero (CC0) licence,⁷⁰ which for most practical purposes waives all copyright restrictions. The application of such a liberal licence for metadata greatly simplifies reuse. It is often desirable to use metadata in aggregate contexts in which citation of individual sources and rights holders would be onerous or prohibitive, such as when metadata from multiple sources are combined and transformed as mash-ups.

Open licences for the data themselves are also preferable, and, if a licence as liberal as CC0 (or a Public Domain mark) can also be applied there, then all the better. In practice, though, it is often necessary or desirable—particularly in academic contexts—to require source attribution of data if they are reused; in this case, a Creative Commons Attribution (CC BY) licence should be applied.⁷¹ CC BY licences exist in a few different varieties imposing a few combinations of restrictions on reuse, such as forbidding derivative works (ND), or requiring that any derivative works be licensed under the same terms (SA; so-called “viral licensing”); it is advised that such additional restrictions be avoided in favour of the base CC BY licence when possible.

As though they were a swarm of bees, CC BY-NC-* variants⁷² should be avoided, especially in academic contexts. These licences place a “non-commercial” restriction forbidding reuse involving monetary compensation. For some reason we have yet to fathom, researchers seem to be particularly enamoured of CC BY-NC-* licences, which is unfortunate, because the non-commercial restriction often has unintended consequences while delivering no clear benefit to the rights holder. NC licences are

67 See <https://rightsstatements.org> (accessed May 11, 2021).

68 Open Data Commons, “Open Data Commons Open Database License (ODbL) v1.0.”

69 Please note that CC licences should not be used for software; there are plenty of good free or open-source licences expressly for that purpose. See, for instance, Choose A License, an online tool from GitHub intended to help in choosing an appropriate open licence for software: <https://choosealicense.com> (accessed May 11, 2021).

70 See <http://creativecommons.org/publicdomain/zero/1.0> (accessed May 15, 2021); in cases in which the material being published is old enough that its copyright has already expired, it should be marked as Public Domain: <http://creativecommons.org/publicdomain/mark/1.0> (accessed May 11, 2021).

71 See <http://creativecommons.org/licenses/by/4.0> (accessed May 11, 2021).

72 See <https://creativecommons.org/licenses> (accessed May 11, 2021).

problematic because what counts as “commercial” use is only vaguely defined;⁷³ it is generally legally safer to err on the side of caution, which in practice tends to exclude a number of academic contexts that researchers would often like to explicitly allow. The use of material in monographs, articles, and textbooks if said works are to be sold, and the use of material in courses for which a fee is charged, are just two examples that researchers generally wish to encourage but that NC licences prohibit in practice.

In some cases, it may not be possible for whatever reason to apply an open licence; in such cases, it is still important to apply a rights statement (e.g., all rights reserved) so that it is clear to users that they are *not* licensed to reproduce the data. No matter which licence is used, it is also vital to include the copyright holder as well—either names of authors or the project—so that it can be correctly cited.

Technical Long-Term Availability/IRIs

Permanent resolvable identifiers need not necessarily be IRIs, if the resource itself is not LOD. For publications or non-LOD datasets, a digital object identifier (DOI)⁷⁴ has quickly become the citation standard of choice for digital resources over the past two decades, combining a unique identifier with a resolver to ensure continuity even if the resource is moved or can be accessed as separate instances from multiple sources. Archives, research institutions, and national bodies are able to issue DOIs for digital resources.

These considerations of long-term availability are compounded by the requirements of linked data, in particular that of stable resolvable public identifiers. Can you be sure that your IRIs will survive possible future software migrations? If the data are stored using a domain tied to an institution, can you be sure that that institution’s webmaster won’t decide to reorganize the site structure in a few years and move the material, changing the IRIs? Can you even be sure that the institution and its domain will still exist? Sadly, there are no good solutions to this problem, only alternatives with different drawbacks. Decoupling IRIs from any particular institution and using an independent domain can be beneficial in the long term, and offer a degree of flexibility and control, especially for projects that are not strictly tied to a particular organization. The trade-off is that the project is then responsible for maintaining and renewing that domain.

If a third-party service is chosen to manage and maintain permanent IRIs, it can use redirects (which also have to be maintained, and may not be any more reliable than the research institution) or rely on non-dereferenceable identifiers such as URNs or DOIs (which work well for citations, but must be run through a transient resolver service in order to arrive at the digital resource they identify).⁷⁵

73 The Creative Commons do provide a rough guide: <https://creativecommons.org/faq/#does-my-use-violate-the-noncommercial-clause-of-the-licenses> and https://wiki.creativecommons.org/wiki/NonCommercial_interpretation (both accessed May 11, 2021).

74 See www.doi.org (accessed May 11, 2021).

75 For more on planning for permanent identifiers, see Sauermann and Cyganiak, “Cool URIs for the Semantic Web”; and Farias Lóscio, Burle, and Calegari, “Data on the Web Best Practices.”

Case Studies

This section delves into two Sweden-based projects, chosen as case studies on how work with or in relation to linked open data and sustainability has been carried out. Both are infrastructures, but Swedish Open Cultural Heritage (SOCH)⁷⁶ is planned as a long-term national aggregation platform that links to Europeana, while Norse World⁷⁷ is the result of a three-year project with a specific research profile.

SOCH⁷⁸

Swedish Open Cultural Heritage (SOCH; “K-samsök” in Swedish) is an aggregation and search platform for Swedish cultural heritage data. The platform is administered by the Swedish National Heritage Board⁷⁹ on behalf of almost eighty participating cultural heritage institutions (data partners), which contribute metadata from their collections.

The SOCH platform harvests metadata records from these museums, archives, and historic environment registers, and publishes them as linked open data. The records are assigned persistent IRIs, mapped to a common RDF-based data model, indexed, and made queryable via a web API. At the time of writing SOCH harvests and indexes just under 9 million records from 174 different datasets. SOCH includes records covering both tangible and intangible heritage, including protected sites and monuments, historic buildings, artefacts and small finds, photographs and drawings, sound and video recordings, and documents and literature, as well as historic personages, events, and more. As an aggregator and publisher of linked data, it is not only the metadata records themselves that are significant but also the links describing how they relate to one another in a machine-readable manner.

The development of SOCH began in 2008, and the platform went live in 2010 as part of a mandate from the Swedish government, with earmarked funding until 2016; since that time it has been funded from the Swedish National Heritage Board’s annual budget, and is now considered to be an important part of furthering the agency’s core goals of ensuring that cultural heritage is preserved, used, and developed.

All the metadata in SOCH are licensed under CC0, and records for media such as images must include rights statements for the media itself. As the Swedish national aggregator for Europeana, SOCH provides metadata further downstream to that international aggregator. SOCH also makes the records available via an open API, which makes the records searchable, can provide aggregated statistics, and makes it possible to easily find linked data relations

76 Swedish Open Cultural Heritage is operated by the Swedish National Heritage Board (Riksantikvarieämbetet); see the “K-samsök” search function at <http://kulturarvsdata.se> (accessed May 11, 2021).

77 Petrulevich et al., “Norse World.” See also Petrulevich and Skovgaard Boeck, this volume.

78 For more in-depth coverage of SOCH in English, including an overview of its history, data model, strengths, weaknesses, and plans for future development, see Smith, “Linked Open Data and Aggregation Infrastructure.”

79 See www.raa.se (accessed May 7, 2021).

connecting records. For example, a photograph might “depict” an artefact that was “found at” an ancient monument, “documented by” a report that “shares an author with” another report ... and so on. A separate but parallel index allows the creation of additional relations by human and machine users as user-generated content (UGC).

SOCH itself offers only a technical interface for machine agents: the web API takes queries as URL query strings and returns RDF/XML, JSON-LD, or “plain” XML depending on the nature of the request; dereferencing SOCH IRIs on kulturarvsdata.se returns either the record’s RDF/XML or JSON-LD for machine agents, or a redirect to the record’s source at the providing institution for human agents. The intention is to encourage third parties to develop applications using the SOCH data and API, which many have done.⁸⁰ Nevertheless, the Swedish National Heritage Board also provides a web interface to SOCH, Kringla,⁸¹ which allows users to search, browse, and view records in SOCH and the links between them.

At the time SOCH was developed, few non-commercial linked data platforms were well proven, the principles of RDF were not widely understood in the Swedish cultural heritage sector, and linked data ontologies for heritage data had not been widely implemented. As a consequence, a number of design decisions were made when developing the platform and its data model that, with hindsight, would not be repeated if it were done today. Among these, and of particular note in the context of sustainability, are the fact that SOCH uses its own bespoke RDF data model, and how it mints persistent IRIs.

SOCH uses an RDF-based data model that was created for it specifically, including a variety of attributes and relation types for describing cultural heritage data, as well as vocabulary authorities for things such as item types, context types, etc. While some attributes and relations correspond to or directly use terms from well-known ontologies such as OWL, FOAF, Dublin Core, and CIDOC-CRM, the overwhelming majority were created from whole cloth when the platform was developed. This poses a problem for interoperability with other linked data resources, since the SOCH data model is not used other than in the SOCH platform; with the exception of the OWL/FOAF/DC-based attributes, clients not created explicitly with SOCH in mind aren’t able to make sense of the data model’s semantics.

In order to provide heritage data to Europeana as part of the platform’s role as national aggregator for Sweden, however, data from SOCH must first be mapped to the Europeana Data Model (EDM). Although EDM is in some respects a less nuanced model than SOCH’s, as a shared model for cultural heritage data across all Europeana’s partner institutions it is more widely known and implemented, and even sees some use outside Europeana. This compensates somewhat for the limited interoperability of the SOCH data model, as the same data are also available from Europeana as EDM. With this in

80 See www.raa.se/hitta-information/k-samsok/att-anvanda-k-samsok/exempel-pa-anvandning-av-k-samsok (accessed May 15, 2021) for a short list of examples.

81 See www.kringla.nu (accessed May 15, 2021).

mind, it is likely that a future version of the SOCH data model will move toward using simple EDM internally, complemented by support for CIDOC-CRM with archaeological extensions (CRM_{archaeo})⁸² to allow for more detailed metadata descriptions where necessary, and standard RDFS, OWL, DC, FOAF, etc. terms for lowest-common-denominator interoperability.

SOCH IRIs live on the kulturarvsdata.se domain, rather than something under the Swedish National Heritage Board's own domain, such as data.raa.se. This reinforces SOCH's function as a sector-wide aggregator, not serving just a single organization, but it also reflects the pragmatic truth that, as a government agency, the heritage board could theoretically at some future date be restructured or dissolved, or the mandate for SOCH assigned to another agency.

Data partners in SOCH mint IRIs for their records under the kulturarvsdata.se domain themselves when their data are harvested. The use of this domain rather than one under the Swedish National Heritage Board's raa.se is intended to decouple the IRIs from the organization and thus make them more stable in the long term. Nevertheless, the fact that partner institutions mint the IRIs themselves means that they are ultimately responsible for ensuring that they do not change. This is exacerbated by the unfortunate structure of SOCH IRIs' path component, which includes elements tying it to the providing institution, dataset, or system, and often the local identifier used by that institution or system. This means that, if an artefact changes hands, or—more likely—the institution changes database systems and local identifiers, those parts of the IRI may in fact *not* be permanent. It would have been preferable to avoid this by centrally assigning unique and opaque IRIs from SOCH, although this introduces the problem of tracking changes across harvests.

Latterly, an attempt has been made to mitigate the potential damage that may be caused by unstable IRIs. In addition to identity relations defined as `owl:sameAs`, SOCH now also supports relations describing changed IRIs such as `dc/terms:replaces`. This allows data providers that, for whatever reason, deem changing their IRIs unavoidable to do so. SOCH will redirect requests for the old IRI to the new one that replaces it, and transitively treats relations that apply to one as if they also applied to the other. SOCH also now keeps track of IRIs that have appeared in its index, and, if an IRI is removed during a harvest and lacks a corresponding forwarding address, requests for that IRI now signal that it has been removed returning “410 Gone,” flagging that the IRI is correct but no longer exists, rather than simply “404 Not Found.”

Norse World⁸³

Norse World is the web resource deliverable of the Norse Perception of the World project “for research on worldviews and spatiality in medieval literature from Sweden and Denmark.”⁸⁴

⁸² See www.cidoc-crm.org/crmarchaeo (accessed May 11, 2021).

⁸³ This section is based on the poster presentation by Backman, “Mapping Foreign Place Names.”

⁸⁴ Petrulevich et al., “Norse World”; see also Petrulevich and Skovgaard Boeck, this volume.

It seeks to make it possible to answer questions on how foreign lands were visualized in the Middle Ages by extracting foreign places names and other spatial information from literary medieval texts with an emphasis on spelling and name variants.

The resource consists of a Leaflet-powered map component connected to a bespoke MySQL database with a REST-API with JSON, including GeoJSON and JSON-LD compatibility, as well as a documentation website with term definitions and technical specifications.⁸⁵ Coordinates for the geo-referencing are taken from gazetteers such as GeoNames or the Deutsches Archäologisches Institut's *iDai.gazetteer*.⁸⁶ Although the resource is not linked open data as such, it provides open, structured data in the form of exportable CSV files. There is also a REST-API with machine-readable GeoJSON and JSON-LD links of original forms and standard forms.⁸⁷ Unfortunately, at the time of writing, these links do not produce valid JSON or JSON-LD, but it's actively being worked on as of spring 2023.

The project is funded by the Swedish Foundation for Humanities and Social Sciences, Riksbankens Jubileumsfond, originally for three years (2017–2020), but this has been extended into 2023. While the funding contract required the hosting institution, the Department of Scandinavian Languages at Uppsala University, Sweden, to guarantee the survival of the project past the funding period for an indeterminate time period, how this would be accomplished has been unclear from the beginning. The project is to be one of the first digital humanities infrastructure projects under Uppsala University's e-administration model for digital operational support, which was still under active development for the duration of the project.⁸⁸

When the funding application was conceived and written in 2016, none of the participants—principal investigator Jonathan Adams, co-investigators Agnieszka Backman and Alexandra Petrulevich⁸⁹—had heard of linked open data, and a relational database was chosen due to earlier experiences. After struggling with public procurement specifications it was instead decided to use a team from Uppsala University IT Services,⁹⁰ who had the relational database know-how necessary for the resource, with domain knowledge provided by the project group. By the time the project participants learned of linked open data as a concept it was too late to change the trajectory of the project, other than to make the resource compatible with LOD. One of the reasons a pivot was not possible was the steep learning curve that creating a linked open data resource

85 Petrulevich, "Infrastructure."

86 See www.geonames.org and <https://gazetteer.dainst.org/app/#!/home> (both accessed May 15, 2021).

87 Backman et al., "REST-API."

88 Ohlsson, "E-administration."

89 Co-investigator Simon Skovgaard Boeck joined the team at project start and additional data entry was carried out by project assistants Jessica Holmlund and Felix Marklund.

90 User experience (UX) designer Jorunn Hartmann, head of group Rasmus Ljungström, and systems developer Andreas Lecerof.

would have entailed. Finding usable ontologies and having to develop an ontology for the onomastic aspects of the project were insurmountable obstacles, mainly because of the limited time available due to the project's three-year duration.

One difficulty had to do with the project's approach to the problem of spatiality in medieval Swedish and Danish texts. Investigating spatiality from a philological and onomastic perspective, the exact spelling of foreign place-names, their textual context and categorization were all intrinsic to the project. The different levels of place-name forms, original form, variant form, and lemma form,⁹¹ which are essential for the project, had no obvious equivalent in the LOD ontologies the project participants had access to. For additional complexity, the project also includes non-names with spatial information, such as inhabitant designations and adjectives describing origin. In comparison, a comparable spatial resource, the Icelandic Saga Map,⁹² has (so far) exclusively place-names from edited texts—that is, normalized place-names—comparable to Norse World's lemma form.

While Norse World is not linked open data, it follows many of the principles of LOD, open science, and open-source programming. The Leaflet library used for mapping is open-source. The data are published under a CC BY licence, allowing for sharing and adaptation. The data are available in (the open) CSV format. The live project data can be downloaded in different configurations: either the whole database or curated by use of search and/or filters. The open licence and providing the data in a non-proprietary, structured format online fulfills the three-star requirement in Berners-Lee's open data model. Additionally, the Norse World resource also uses open resources with stable URLs for the data it reuses, such as GeoNames and Libris.⁹³ The landing pages for detailed views of attestations, works, sources, variant forms, lemma forms, and standard forms are also stable and linkable. As mentioned above, the REST-API is also planned to be compatible with GeoJSON and JSON-LD. In other words, the resource has been made as open and reusable as was possible within the project parameters.

Conclusion

Linked open data provide an opportunity for digital humanities projects to open up their data for potential enrichment and reuse by others than the original stakeholders. This requires the use of standardized vocabularies and ontologies. Many different ontologies exist, from those covering general concepts to very specialized subject domains. Since they are rarely a perfect fit, mapping a project's concepts onto an existing model is preferable to making a bespoke one or trying to squeeze the project's research point of view into too broad or mismatched categories.

91 Petrulevich, Backman, and Adams, "Medieval Macrospace through GIS"; see also Petrulevich, this volume.

92 See <http://sagamap.hi.is/is> (accessed May 15, 2021).

93 See www.geonames.org (accessed May 14, 2021) and <https://libris.kb.se> (accessed May 15, 2021), the National Library of Sweden's catalogue of books in Swedish libraries.

As digital humanities projects often consist of bigger teams than many humanities researchers might be used to, consisting of both domain experts and programmers, a project leader/principal investigator will need to learn about project management, as well as data formats and licences, to ensure optimal project outcomes and sustainability. The two case studies, SOCH and Norse World, showcase different aspects of the LOD spectrum and the sustainability efforts needed due to SOCH being a national aggregator while Norse World is more of a research project.

One of the main points of this chapter has been that not all digital humanities projects are the same, and they will have different needs and paths to sustainability—a concept that might itself have different meanings, depending on the project. In some instances, it might even be more fruitful to view the digital output of projects as ephemeral research results.⁹⁴ Other projects might fail, such as, for example, Project Bamboo, which wasn't able to deliver the digital humanities tools workbench it set out to create but still “brought together scholars, librarians, and technologists at a crucial moment for the emergence of digital humanities”⁹⁵ in the United States; there was value in the journey, even if the destination was not reached. Ideas, connections, and relationships were an intangible outcome of the project and show that success doesn't have to equal something tangible and sustainable.⁹⁶ There needs to be acceptance of both the ephemeral and incompleteness as a part of the digital humanities ecosystem, while still planning for permanence and transparent scholarship.

When planning for the long-term sustainable access of a project, it's important to stress that it's not necessary to guarantee that *everything* remains available in perpetuity, an entire technical platform frozen in amber just as it was created. Indeed, such a goal would be prohibitively costly and likely deliver diminishing returns. Technologies inevitably develop, are superseded, and then are no longer supported, and to attempt to hold back that tide is a fool's errand. The adage goes “Applications age like fish, data like wine,” and this distinction happily also maps to ease of preservation: maintaining access to a usable dataset in the long term requires “only” secure and redundant data storage, regular integrity checks, and the occasional format migration, plus ongoing funding for hosting and systems administration. Yes, the word “only” bears a heavy load in that sentence, but compared to the exponential cost and difficulty of keeping a software-based service up and running in the long term—across component and system updates, protocol changes, components and systems inevitably going through the process of no longer being updated, end of life, deprecation, and ultimately no longer running on contemporary systems, necessitating emulation—maintaining access to a dataset is comparatively easy. In the majority of cases, users don't need the entire experience of using your system to be preserved; they just need to get access to the underlying data in a format they can use, with

94 “What if we were to advocate *ephemerality* of digital resources in those cases where that's a healthy approach that gets scholars where they want to go ... ?” Nowwiskie, “A Skunk in the Library.”

95 Dombrowski, “What Ever Happened to Project Bamboo?,” 335.

96 Dombrowski, “The Directory Paradox.”

its associated metadata. With that in mind, aim to arrange for long-term data preservation (including any identifiers associated with it, so that references to it will still make sense) but assume that any digital platform will have a finite life of no more than a decade or two, after which it will either be retired or replaced. If the data rely on having resolvable public identifiers such as IRIs, it may be worth decoupling these from the application itself if a more sustainable solution is available, so that they may be easily redirected in future.

Digital entropy is an inevitability; it cannot be avoided. Thus, it is important to plan for it from the start: how it will be managed, and how it will be mitigated. Having a project data management plan is essential. Consider how the digital resources the project produces, and the platforms that support them, will be maintained and updated, and for how long. What budget and resources will the project have in the future, and for how long? What is a realistic lifespan for the platform? And, when the platform is ultimately decommissioned, where will the data be archived and how? Will any associated persistent identifiers still resolve?

Glossary

AAT: Getty Art & Architecture Thesaurus; a structured linked data vocabulary for artistic terms, styles, periods, object types, etc. <http://vocab.getty.edu> (accessed May 11, 2021).

ADS: the Archaeology Data Service; an accredited digital repository for heritage data based at the University of York and operational since 1996. <http://ads.ahds.ac.uk> (accessed May 11, 2021).

API: Application Programming Interface; a technical interface that allows programs to interact with a code module, library, or service. A web API uses web technologies to do this.

ARIADNEplus: a continuation and expansion of the earlier ARIADNE project, focused on integrating and harmonizing European digital archaeological research data, and making it accessible and reusable. <https://ariadne-infrastructure.eu> (accessed May 11, 2021).

CIDOC-CRM: the CIDOC Conceptual Reference Model, an event-oriented ontology for cultural heritage with a focus on museum collections. <http://cidoc-crm.org/cidoc-crm> (accessed May 11, 2021).

CRMarchaeo: an extension to the CIDOC-CRM ontology covering concepts and attributes pertaining to archaeological methods, recording, and excavation. www.cidoc-crm.org/crmarchaeo (accessed May 11, 2021).

Creative Commons: a non-profit organization promoting free culture and a set of widely used open attribution-based licences for creative and cultural works. <https://creativecommons.org> (accessed May 11, 2021).

DOI: digital object identifier; an international standard for unique persistent identifiers for digital objects, or an instance of such an identifier. DOIs are not in themselves dereferenceable but, rather, must go through a resolver service. www.doi.org (accessed May 11, 2021).

- Dublin Core: a set of fifteen fundamental metadata properties (later expanded) widely used and accepted as a baseline for metadata. <http://purl.org/dc/elements/1.1>; <http://purl.org/dc/terms> (both accessed May 11, 2021).
- EDM: the Europeana Data Model, a data model for cultural heritage data used by the European cultural heritage aggregator Europeana. www.europeana.eu/schemas/edm (accessed May 11, 2021).
- Europeana: a European digital cultural heritage aggregator operational since 2008 and with collections of data from more than 3,000 cultural heritage institutions, operated under the auspices of the European Commission. www.europeana.eu (accessed May 11, 2021).
- FAIR: a set of principles for Findable, Accessible, Interoperable, and Reusable digital data, complementary to linked open data and movements such as Free Software, but with a focus on academia and scientific data. www.go-fair.org/fair-principles (accessed May 11, 2021).
- FOAF: Friend-Of-A-Friend; a linked data schema for describing people, organizations, and social networks. One of the earliest linked data ontologies; it is in widespread use. <http://xmlns.com/foaf/0.1> (accessed May 11, 2021).
- GeoNames: an authority for geolocated place-names and geographical divisions (countries, provinces, cities, etc). www.geonames.org (accessed May 11, 2021).
- GitHub: online platform for collaborative open software development, based on the *git* distributed version control system. <https://github.com> (accessed May 11, 2021).
- Heritage Data: a set of linked data SKOS vocabularies and thesauri from UK heritage bodies, including monument-type thesauri. <https://heritagedata.org/live/schemes> (accessed May 11, 2021).
- JSON-LD: a variant of the JavaScript Object Notation data format with support for linked data; a serialization format of RDF; see Kellogg et al.
- Kringla: a web application built on SOCH allowing users to search, browse, and view records in SOCH, including associated media, links between records, and out to the broader web, etc. www.kringla.nu (accessed May 11, 2021).
- Norse World: a research project and digital resource mapping exonyms found in the Old East Norse corpus in context. See Backman, “Mapping Foreign Place Names”; Petrulevich, Backman, and Adams, “Medieval Macrospace through GIS”; Petrulevich and Skovgaard Boeck, this volume; and Petrulevich, Backman, Adams, Skovgaard Boeck, Holmlund, Marklund, Hartmann, Lecerof, and Ljungström. www.uu.se/en/research/infrastructure/NorseWorld (accessed May 11, 2021).
- Open Data Commons: an open licence for databases and structured datasets. <https://opendatacommons.org/licenses/odbl/1-0> (accessed May 11, 2021).
- OWL: Web Ontology Language; an RDF ontology for describing RDF ontologies; see Motik, Patel-Schneider, and Parsia, “OWL 2 Web Ontology Language.”
- PeriodO: a linked open data gazetteer for historical and archaeological chronological periods, bounded both temporally and geographically. <https://perio.do> (accessed May 11, 2021).

RDF: Resource Description Framework, the meta-model for linked open data and the semantic web, comprised of three-part statements called triples; see Schreiber and Raimond, “RDF 1.1 Primer.”

RDFS: RDF Schema, a set of RDF types and predicates for describing data schemas and vocabularies; see Brickley and Guha, “RDF Schema 1.1.”

RightsStatements.org: a set of rights statements for describing commonly occurring non-open-rights situations—e.g., full copyright/all rights reserved, initiated by Europeana and the Digital Public Library of America, and now operated by a larger consortium of national bodies. <https://rightsstatements.org> (accessed May 11, 2021).

SKOS: Simple Knowledge Organization System; a set of RDF types and predicates for describing multi-hierarchical structured vocabularies as linked data; see Miles and Bechhofer, “SKOS.”

SOCH: Swedish Open Cultural Heritage (Swedish: “K-samsök”); the Swedish national cultural heritage aggregator and linked open data platform, operated by the Swedish National Heritage Board. <http://kulturarvsdata.se> (accessed May 11, 2021).

SPARQL: SPARQL Protocol and RDF Query Language, an SQL-like query language for RDF data on the semantic web based around matching patterns of triples. In contrast to REST-based web APIs, which will vary from system to system, SPARQL is a common interface for querying any set of RDF data, and queries may be federated over multiple systems to combine datasets; see Harris, Seaborne, and Prud’hommeaux, “SPARQL 1.1 Query Language.”

UGC Hub: SOCH’s platform for user-generated content. <https://ugc.kulturarvsdata.se> (accessed May 11, 2021).

VIAF: Virtual International Authority File, a set of aggregated authorities for bibliographic entities such as authors, works, etc., collected from national libraries across the world. <http://viaf.org> (accessed May 11, 2021).

Wikidata: structured linked open data crowdsourced by Wikimedia. <http://www.wikidata.org> (accessed May 11, 2021).

Wikimedia Commons: open media crowdsourced by Wikimedia. <https://commons.wikimedia.org> (accessed May 11, 2021).

Wikipedia: open knowledge crowdsourced by Wikimedia. <https://www.wikipedia.org> (accessed May 11, 2021).

Bibliography

- Appleford, Simon, and Jennifer Guiliano. “DevDH: Development for the Digital Humanities.” DevDH.org, 2013. <https://devdh.org/about> (accessed May 11, 2021).
- Archaeology Data Service. <http://ads.ahds.ac.uk> (accessed May 11, 2021).
- Archaeology Data Service/Digital Antiquity. “Guides to Good Practice.” 2009. <https://guides.archaeologydataservice.ac.uk> (accessed May 11, 2021).
- Backman, Agnieszka. “Mapping Foreign Place Names in Medieval Swedish and Danish Texts.” Humanities Commons, December 11, 2018. <https://dx.doi.org/10.17613/48gz-5g45> (accessed May 11, 2021).

- Backman, Agnieszka, Jorunn Hartmann, Andreas Lecerof, Rasmus Ljungström, and Alexandra Petrulevich. "REST-API." www.uu.se/en/research/infrastructure/norseworld/infrastucture/rest-api (accessed May 15, 2021).
- Berners-Lee, Tim. "Linked Data." W3C, July 27, 2006. www.w3.org/DesignIssues/LinkedData.html (accessed May 11, 2021).
- Berners-Lee, Tim, Roy T. Fielding, and Larry Masinter. "RFC 2396: Uniform Resource Identifiers (URI): Generic Syntax." RFC Editor, August 1998. <http://tools.ietf.org/html/rfc2396> (accessed May 11, 2021).
- Brickley, Dan, and Ramanathan V. Guha. "RDF Schema 1.1." W3C, February 25, 2014. www.w3.org/TR/rdf-schema (accessed May 11, 2021).
- Celtic Inscribed Stones Project (CISP). <https://dx.doi.org/10.5284/1000215> (accessed May 11, 2021).
- Center for Open Science. www.cos.io (accessed May 15, 2021).
- Chun, Wendy Hui Kyong. "The Enduring Ephemeral, or the Future Is a Memory." *Critical Inquiry* 35 (2008): 148–71.
- Core Trust Seal. www.coretrustseal.org (accessed May 11, 2021).
- Creative Commons. <https://creativecommons.org> (accessed May 11, 2021).
- . "About the Licenses." <https://creativecommons.org/licenses> (accessed May 11, 2021).
- . "Attribution 4.0 International (CC BY 4.0)." <http://creativecommons.org/licenses/by/4.0> (accessed May 11, 2021).
- . "Frequently Asked Questions." <https://creativecommons.org/faq/#does-my-use-violate-the-noncommercial-clause-of-the-licenses> (accessed May 11, 2021).
- . "NonCommercial Interpretation." https://wiki.creativecommons.org/wiki/NonCommercial_interpretation (accessed May 11, 2021).
- DANS (Data Archiving and Networked Services). <https://dans.knaw.nl> (accessed May 11, 2021).
- DIGARV. "Swedish Caribbean Colonialism." www.digarv.se/en/participating-research-projects/svensk-karibisk-kolonialism-1784-1878 (accessed May 15, 2021).
- Digital Antiquity. www.digitalantiquity.org (accessed May 11, 2021).
- Digital Preservation Coalition. www.dpconline.org (accessed May 11, 2021).
- . "File Formats and Standards." 2015. www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards (accessed May 11, 2021).
- Dombrowski, Quinn. "The Directory Paradox." In *People, Practice, Power: Digital Humanities outside the Center*, edited by Anne B. McGrail, Angel David Nieves, and Siobhan Senior, 83–98. Minneapolis: University of Minnesota Press, 2021.
- . "DLCL 205/305: Project Management & Ethical Collaboration for Humanists." GitHub, 2020. <https://github.com/quinnanya/dlcl205> (accessed May 11, 2021).
- . "What Ever Happened to Project Bamboo?" *Literary and Linguistic Computing* 29 (2014): 326–39.
- Duerst, Martin, and Michel Suignard. "RFC 3987: Internationalized Resource Identifiers (IRIs)." RFC Editor, January 2005. <http://tools.ietf.org/html/rfc3987> (accessed May 11, 2021).
- European Commission. "Open Science." https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science_en (accessed May 15, 2021).

- Farias Lóscio, Bernadette, Caroline Burle, and Newton Calegari. "Data on the Web Best Practices." W3C, January 31, 2017. www.w3.org/TR/dwbp (accessed May 11, 2021).
- Fielding, Roy T. "[httpRange-14] Resolved." W3C, June 18, 2005. <https://lists.w3.org/Archives/Public/www-tag/2005Jun/0039.html> (accessed May 11, 2021).
- Fostano, Katherina, and Laura K. Morreale. "The Digital Documentation Process." Digital Documentation Process. January 31, 2019. <https://digitalhumanitiesddp.com> (accessed May 11, 2021).
- GeoNames. www.geonames.org (accessed May 11, 2021).
- GitHub. "Choose a License." <https://choosealicense.com> (accessed May 11, 2021).
- Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, et al. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. Dublin, OH: OCLC Research, 2019.
- Harris, Steve, Andy Seaborne, and Eric Prud'hommeaux. "SPARQL 1.1 Query Language." W3C, March 21, 2013. www.w3.org/TR/sparql11-query (accessed May 11, 2021).
- Heritage Data. <https://heritagedata.org/live/schemes> (accessed May 11, 2021).
- Hollander, Hella, Francesca Morselli, Frank Uiterwaal, Femmy Admiraal, Thorsten Trippel, and Sara Di Giorgio. Zenodo, August 12, 2019. "PARTHENOS Guidelines to FAIRify Data Management and Make Data Reusable," 2019. <https://dx.doi.org/10.5281/zenodo.3368858> (accessed May 11, 2021).
- Humanities Commons. <https://hcommons.org> (accessed May 11, 2021).
- Icelandic Saga Map. <http://sagamap.hi.is> (accessed May 11, 2021).
- iDai.gazetteer. <https://gazetteer.dainst.org/app/#!/home> (accessed May 15, 2021).
- International DOI Foundation. "Digital Object Identifier System." www.doi.org (accessed May 11, 2021).
- Internet Archaeology. <https://intarch.ac.uk> (accessed May 11, 2021).
- Kellogg, Gregg, Pierre-Antoine Champin, Dave Longley, Manu Sporny, and Markus Lanthaler. "JSON-LD 1.1: A JSON-Based Serialization for Linked Data." W3C, July 16, 2020. www.w3.org/TR/json-ld (accessed May 11, 2021).
- Kungliga biblioteket. "Libris." <http://libris.kb.se> (accessed May 11, 2021).
- Library of Congress. "Library of Congress Recommended Formats Statement: 2020–2021." 2020. www.loc.gov/preservation/resources/rfs (accessed May 11, 2021).
- . "Sustainability of Digital Formats: Planning for Library of Congress Collections." 2019. www.loc.gov/preservation/digital/formats (accessed May 11, 2021).
- Miles, Alistair, and Sean Bechhofer. "SKOS Simple Knowledge Organization System Reference." W3C, August 18, 2009. www.w3.org/TR/skos-reference (accessed May 11, 2021).
- Morreale, Laura. "Medieval Digital Humanities and the Rite of Spring: Thoughts on Performance and Preservation." BodoArXiv, July 18, 2019. <https://dx.doi.org/10.34055/osf.io/7p2t6> (accessed May 11, 2021).
- Motik, Boris, Peter F. Patel-Schneider, and Bijan Parsia. "OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition)." W3C, December 11, 2012. www.w3.org/TR/owl-syntax (accessed May 11, 2021).
- Nestor. www.langzeitarchivierung.de (accessed May 11, 2021).
- Nowviskie, Bethany. "A Skunk in the Library." Nowviskie.org, June 28, 2011. <https://nowviskie.org/2011/a-skunk-in-the-library> (accessed May 11, 2021).
- Ohlsson, Annika. "E-administration." Staff Portal—Uppsala University, 2021. <https://mp.uu.se/en/web/info/vart-uu/e-forvaltning> (accessed May 11, 2021).

- Open Data Commons. "Open Data Commons Open Database License (ODbL) v1.0." <https://opendatacommons.org/licenses/odbl/1-0> (accessed May 11, 2021).
- Petrulevich, Alexandra. "Infrastructure." www.uu.se/en/research/infrastructure/norseworld/infrastructure (accessed May 7, 2021).
- Petrulevich, Alexandra, Agnieszka Backman, and Jonathan Adams. "Medieval Macrospace through GIS: The Norse World Project Approach." *The Cartographic Journal* 57 (2020): 18–27.
- Petrulevich, Alexandra, Agnieszka Backman, Jonathan Adams, Simon Skovgaard Boeck, Jessica Holmlund, Felix Marklund, Jorunn Hartmann, Andreas Lecerof, and Rasmus Ljungström. "Norse World." www.uu.se/en/research/infrastructure/norseworld and <https://norseworld.nordiska.uu.se> (both accessed May 11, 2021).
- Read the Docs. <https://readthedocs.org> (accessed May 11, 2021).
- RightsStatements.org. <https://rightsstatements.org> (accessed May 11, 2021).
- Riksantikvarieämbetet (Swedish National Heritage Board). www.raa.se (accessed May 7, 2021).
- Rockwell, Geoffrey, Shawn Day, Joyce Yu, and Maureen Engel. "Burying Dead Projects: Depositing the Globalization Compendium." *Digital Humanities Quarterly* 8 (2014). www.digitalhumanities.org/dhq/vol/8/2/000179/000179.html (accessed May 15, 2021).
- Sauermann, Leo, and Richard Cyganiak. "Cool URIs for the Semantic Web." W3C, December 3, 2008. www.w3.org/TR/cooluris (accessed May 11, 2021).
- Schreiber, Guus, and Yves Raimond. "RDF 1.1 Primer." W3C, June 24, 2014. www.w3.org/TR/rdf11-primer (accessed May 11, 2021).
- Smith, Marcus. "Linked Open Data and Aggregation Infrastructure in the Cultural Heritage Sector: A Case Study of SOCH, a Linked Data Aggregator for Swedish Open Cultural Heritage." In *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*, edited by Koraljka Golub and Ying-Hsang Liu, 64–85. Abingdon: Routledge, 2022.
- Smith-Yoshimura, Karen. "Summarizing Project Passage Experiences in Creating Library Linked Data in Wikibase (2 of 2)." Hanging Together, July 17, 2019. <https://hangingtogether.org/?p=7398> (accessed May 11, 2021).
- Svensk Nationell Datatjänst. <https://snd.gu.se> (accessed May 11, 2021).
- Thiery, Florian, Sophie C. Schmidt, Timo Homburg, and Jakob Voß. "Ogi Ogham Project." 2020. <http://ogham.link> (accessed May 11, 2021).
- Thylstrup, Nanna Bonde. *The Politics of Mass Digitization*. Cambridge, MA: MIT Press, 2018.
- URI Planning Interest Group, W3C/IETF. "URIs, URLs, and URNs: Clarifications and Recommendations 1.0." W3C, September 21, 2001. www.w3.org/TR/uri-clarification (accessed May 11, 2021).
- W3C. "TAG Issues List: httpRange-14: What Is the Range of the HTTP Dereference Function?" 2002. www.w3.org/2001/tag/issues.html#httpRange-14 (accessed May 11, 2021).
- Wikibase. <https://wikiba.se> (accessed May 11, 2021).
- Wikidata. <https://www.wikidata.org> (accessed May 15, 2021).
- . "Wikidata:Notability." www.wikidata.org/wiki/Wikidata:Notability (accessed May 11, 2021).
- Wikiversity. www.wikiversity.org (accessed May 6, 2021).

