



Teaching Linked Open Data using Bibliographic Metadata

RESEARCH PAPER

TERHI NURMIKKO-FULLER 

ubiquity press

ABSTRACT

This paper describes LD4DH, the *Linked Data for Digital Humanities: Publishing, Querying, and Linking on the Semantic Web* workshop at the Digital Humanities Oxford Summer School. It includes a description of the general structure of the workshop, how it has changed over the course of the last seven years, between 2015 and 2021, and evaluates the differences between in-person delivery in 2018–2019 and the online mode in 2020–2021. Discussion is centred on the description of the data as well as the illustration of the processes, methods, and software used throughout the workshop. The paper concludes with a summary of participant evaluation, and reflects on the opportunities and challenges of teaching Linked Open Data to a mixed cohort of predominantly Humanities researchers and professionals from the cultural heritage sector.

CORRESPONDING AUTHOR:

Terhi Nurmikko-Fuller

Centre for Digital Humanities
Research, Australian National
University, Canberra, Australia
terhi.nurmikko-fuller@anu.edu.au

KEYWORDS:

Linked Open Data;
bibliographic metadata;
pedagogy; participant
evaluations

TO CITE THIS ARTICLE:

Nurmikko-Fuller, T. (2022).
Teaching Linked Open Data
using Bibliographic Metadata.
*Journal of Open Humanities
Data*, 8: 6, pp. 1–11. DOI:
<https://doi.org/10.5334/johd.60>

1 INTRODUCTION

The *Linked Data for Digital Humanities: Publishing, Querying, and Linking on the Semantic Web* (henceforth, LD4DH¹) workshop has formed part of the proceedings of the Digital Humanities Oxford Summer School (DHOxSS) since 2012. I am an alumna of the workshop myself, having attended it as a participant in its first iteration. At that time in my academic career, LD4DH was a space to acquire essential practical skills for implementing Linked Open Data (LOD). Having since become the convener and tutor of the same workshop, it has now become an annual highlight, an opportunity to fully immerse myself in the methodology, to discuss research, and to engage with diverse groups of researchers, academics, and GLAM (galleries, libraries, archives, and museums) sector professionals.

The workshop aims to provide participants with an understanding of the theories behind LOD as an information publication paradigm, and then build on that foundation with practical and hands-on activities. This deliberate pedagogical structure is designed to reflect the insight that neither the use and implementation of digital methods, nor the critical evaluation of the projects and platforms that have been developed using those methods can be taught exclusively in abstract terms (Brier, 2012). In recognition of the role of collaboration and co-authoring in digital humanities (DH) research (Needham & Haas, 2019), workshop participants are encouraged to work together and communicate openly as a group.

Since 2015, I have taught LD4DH with John Pybus and Graham Klyne, both from the Oxford e-Research Centre at the University of Oxford. The secret to our successful delivery of the workshop has not rested only on our friendship, or our common interest in LOD, but also because of our differences in interests, expertise, and academic backgrounds. This diversity within the tutor group enables us to discuss each topic from different perspectives. There is no guarantee of unanimous agreement and that gives the learners access to a greater diversity of ideas. We can thus more confidently cater for the needs and intellectual preferences of diverse cohorts.

In recognition of the challenges of the course, and the role that a pleasant and supportive learning environment can play in successful information and skills-acquisition (Imlawi & Gregg, 2014), there has been a deliberate attempt to create a jovial and friendly atmosphere. Humour is used to promote openness between the teachers and the learners, and to make acronyms and concepts more memorable. An examples of this is theme-specific clothing, such as a pair of golden trousers worn in homage to the query language SPARQL² (pronounced “sparkle”) and a skirt with owls, in reference to the Web Ontology Language (OWL).³

2 STRUCTURE OF THE WEEK

LD4DH is a five-day workshop. Each day follows the same three-session pattern, with a different topic: a 90-minute theory session followed by a two-hour practical one. The final hour of the day is a lecture by a guest speaker discussing the use of LOD in their area of research, often their own project (**Table 1**). Although there has been some flexibility to the list of speakers, in most years the topics covered numismatics (Prof Andrew Meadows, University of Oxford, speaking about *Nomisma.org*⁴), digital musicology (Dr Kevin Page, University of Oxford, reporting on a range of projects), and digital libraries (Prof Stephen Downie, University of Illinois Urbana-Champaign, summarising on the work of the HathiTrust Research Center⁵), as well as geolocation and digital mapping (Dr Valeria Vitale, now at the Alan Turing Institute in London, and Chair of the Pelagios Network⁶). Other colleagues who have contributed to LD4DH include Dr Daniel Bangert (talking about the JazzCats project⁷), Dr Paula Granados García, (from the Open University, who gave a summary of the experience of an alumna of the workshop), as well as Dr Athanasios Velios and

¹ The acronym is the name of the Slack channel (<https://ld4dh-dhoxss.slack.com>) and the Twitter hashtag (#LD4DH) for this workshop.

² <https://www.w3.org/TR/rdf-sparql-query/>.

³ <https://www.w3.org/OWL/>.

⁴ <http://nomisma.org/>.

⁵ <https://www.hathitrust.org/htrc>.

⁶ <https://pelagios.org/>.

⁷ <http://jazzcats.cdhr.anu.edu.au/>.

	MONDAY		TUESDAY		WEDNESDAY		THURSDAY		FRIDAY
9:00–10:00	Registration	9:00–10:30	Ontologies (theory)		Producing RDF (theory)		SPARQL (theory)		British Museum (intro, theory)
10:30–12:00	Introduction to LD4DH	11:00–13:00	Ontologies (praxis)		Producing RDF (praxis)		SPARQL (praxis)		British Museum (praxis)
LUNCH									
13:30–15:30	Introduction to LOD	14:30–15:30	Guest Speaker (Musicology)		Guest Speaker (Libraries)		Guest Speaker (Alumna)		Guest Speaker (Museums)
16:00–17:00	Guest Speaker (Numismatics)								

Table 1 Table illustrating the daily structure of the workshop. Each day consists of a theory session, a hands-on session (praxis), and a talk by a guest speaker.

Prof Donna Kurtz, who have spoken on the topic of OxLOD⁸ and its precursor projects, such as CLAROS,⁹ respectively.

The common thread throughout all these diverse projects is the illustration of practical use of LOD in the Humanities. All speakers favour Open licensing for data and software, promote collaboration, and have developed tools that enable users to engage with it without the need to learn programming. These talks strongly support the philosophy of LD4DH, and serve to provide a complementary and enriching context for the learners.

Over the years, LD4DH has undergone several tweaks, rearrangements, and changes. The most recent version has been a response to the COVID-19 crisis, and the move to an entirely online delivery. The workshop became an hour-long lecture on the fundamentals of LOD for a general audience, followed an afternoon hands-on session for those who had opted to enrol in it. The latter is delivered by Dominic Oldman and Diana Tanase, both of the British Museum, using ResearchSpace.¹⁰ At the time of writing, expectations are high for an in-person event in 2022, which would see a return to the pre-COVID-19 mode of delivery of the LD4DH workshop.

3 DATA

Since 2016, the workshop has centred on the data of the ElePHäT¹¹ (Early English Print in HathiTrust, Linked Semantic Worksets Prototype) project.¹² This prototype (which was funded through the Andrew W. Mellon Foundation Workset Creation for Scholarship Analysis project award) combines bibliographic metadata from two very different types of collections: the behemoth HathiTrust Digital Library (HTDL), and the rather more boutique Early English Books Online Text Creation Partnership (EEBO-TCP¹³).

The aim of the ElePHäT project was to see whether two digital library collections (which at a distance appeared to share similarities, but on closer inspection had many idiosyncratic features) could be bridged at the metadata level (Page, Nurmikko-Fuller, Cole, & Downie, 2017). Both the HTDL and EEBO-TCP are aggregators: the EEBO-TCP contains information from some 150 sources – the number of institutions (each of which contain a multitude of collections and sources) that form the HathiTrust is closer to 250. The considerable variation both within and between these two large projects is evident from the metadata.

Although the data for the ElePHäT project thus consisted of two aggregated datasets, for the purposes of the LD4DH workshop, the focus has been exclusively on the data from the EEBO-TCP. The reason for this are two-fold: first, throughout the project, the HTDL data was modelled and provided by the HathiTrust Research Centre (HTRC) team, whilst the EEBO-TCP data was worked on by scholars at Oxford, meaning the team at Oxford had the opportunity to gain

8 <https://www.glam.ox.ac.uk/oxford-linked-open-data-pilot>.

9 <https://eng.ox.ac.uk/claros/>.

10 <http://researchspace.org/>.

11 The project workset viewer is available at <https://eebo.oerc.ox.ac.uk/>.

12 The project has been reported on by Page, K. and Willcox, P. in the 2015 project report, available from https://www.ideals.illinois.edu/bitstream/handle/2142/79017/IEPH%C3%A3T%20final%20report-with_appendix-20150615.pdf.

13 <https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online/>.

familiarity with that dataset, thus making it easier to work with in the context of the workshop. Second, of the HTDL data, 66% remains subject to copyright restrictions, limiting access and use (Jett, Nurmikko-Fuller, Cole, Page, & Downie, 2016). The EEBO-TCP data on the other hand consisted of 25,000 records which became publicly available in 2015.

The EEBP-TCP data has a number of idiosyncracies resulting from the combination of historical data and the processes of aggregation. This manifests in the dataset containing several categories for the same concept, e.g. discrete ID numbers and titles. As the data is derived from historical sources it contains a significant quantity of complex and messy details that do not comfortably sit with modern metadata categories (see for example column “Publisher” in *Table 2*, which provides a sample of the dataset).

AUTHOR	PUBPLACE	PUBLISHER	DATE	IDO	ID1	ID2
Fennor, William	London :	Barnes, and are sold at his shop in S. Dunstans Church-yard in Fleetstreet	1615	A00648	STC 10783	ESTC S114801
Bacon, Francis, 1561-1626	London :	Printed [by Richard Field] for Felix Norton and are to be sold in Pauls Church-yard at the signe of the Parrot	1604	A01003	STC 1111	ESTC S104433
Forser, Edward, 1553?-1630	London :	Printed by B. A[lsop] for Nathaniel Butter, and are to be sold at his shop, at the Pyed Bull, neere Saint Austens Gate	1624	A01075	STC 11189	ESTC S119405
Bacon, Francis, 1561-1626	London :	Printed by I. Okes, for Humphrey Mosley, at the Princes Armes in Pauls Church-Yard	1638	A01446	STC 1157	ESTC S100504
Anonymous	Londini :	In officina Iohannis Haviland	1626	A01639	STC 1177	ESTC S115271
Godwin, Francis, 1562-1633	In Vtopia [i.e. London?]	[J.Bill]	1629	A01809	STC 11944	ESTC S118694
Godwin, Francis, 1562-1633	[Oxford] :	J. Barnes	1603]	A01812	STC 11948	ESTC S118380

Table 2 A sample of the project data showing the information categories captured in the .CSV file. The punctuation marks serve to capture uncertainty about the date, the “Publisher” column cells contain several data points each.

Rather than provide the learners with the entire, rather complex TEI P5 XML¹⁴ files, a simplified .CSV version of the data has been used for the hands-on activities at LD4DH. These tabular datasets were initially generated as part of the workflow for the ElePHäT project using a set of Python scripts to pull out the data for author, publication place, publisher, date, six distinct ID numbers, and three separate titles. The various .CSV files, the Python script and the custom-build project ontology EEBOO are all available from the project Github page.¹⁵ It is worth noting that the data wrangling at the stage of generating the .CSV files did not involve any semantics. This is significant as the match between the modern metadata category and the data contained within it is not always exact. For example, the “Imprint” category (displayed in *Table 2* as “Publisher”) contains a large degree of additional information about the historical printing process, its funding model, and even geographical location, as these details were recorded in the original historical text. An example of this is from the record A00648/STC 10783/ESTC S114801. The imprint data contains information about the individual carrying out the printing (“G.Eld”); the individual commissioning the print (“Roger Barnes”); the location of the shop that sold it (“S. Dunstans Church-yard”); and the name of the street of said shop (“Fleet Street”). Indeed, so rich is this information that in 2016, we carried out an investigation into the extraction of specific details from this data category using natural language processing (Khan, Nurmikko-Fuller, & Page, 2016).

Due to the dataset’s internal richness and diversity, many learners opt to engage in some degree of data wrangling themselves, although it is possible to complete the workflow process without an additional step of data tidying (beyond the minting of URIs). Many participants

¹⁴ That is to say, information that was captured as XML, in adherence to the Text Encoding Initiative’s (or TEI’s) P5 guidelines. For more information about the TEI’s P5 guidelines, see <https://tei-c.org/guidelines/p5/>.

¹⁵ <https://github.com/oerc-elephant/preprocessed-elephant>.

also opt to spend some of their free time engaged in additional research around the subject, merging their aim of technological up-skilling (during the workshop) and their existing sense of needing or desiring to have an understanding of the data prior to working with it.

Many workshop participants opt to separate out surname and first name, or at the very least, the date and the name (see column “Author” in [Table 2](#)). Although the punctuation (such as the colon and square brackets) are meaningful as indicators of ambiguity for the library professionals who created and painstakingly curated the original EEBO-TCP data, these additional characters can be problematic in later stages of the RDF production workflow. For this reason, many learners choose to edit these characters out, essentially applying a reductionist approach to simplify the messy historical data into the categories of modern information representation systems.

4 THEORY

Each LD4DH workshop combines theory and praxis. The former includes aspects and considerations that are relevant to the entire DH community, far beyond the scope of the niche group researchers who choose to use the LOD methodology in their research. A major component of this theoretical part of the workshop is to equip the learners with enough information regarding the pragmatics, the challenges, and the opportunities presented by LOD so that they can critically evaluate the method for its strengths and weaknesses. This in turn enables the participants to make an informed decision as to whether or not to engage with it in their research beyond DHOxSS. It is not always the right tool. There are no silver bullets.

Another aspect of the theoretical component is jargon-busting. These sessions are crucial in establishing a shared vocabulary, to facilitate communication, and to support engagement with the material. These sessions also enable participants to engage in meaningful conversations with other members of community of LOD practitioners. They also help to boost confidence in terms of using the appropriate terminology when discussing their project, and their technical needs, with their colleagues and the IT service provision at their home institutions. We introduce core concepts such as the Five Star Linked Open Data Standard;¹⁶ the idea of knowledge graphs;¹⁷ and the RDF triple.¹⁸ All the examples of data as RDF that the participants encounter in the workshop are expressed in one specific syntax, .TTL (pronounced “turtle”, and one of several possible options, the most common alternatives currently being JSON-LD and RDF-XML)¹⁹ to provide learners with a sense of consistency between examples but specific activities during the week also enable them to learn about the possibility of using different syntaxes for representing RDF.

Among the participants at each iteration of LD4DH, there has been a small minority of those who attend for reasons other than wanting to learn how to use LOD in their research. These include industry representatives, and the occasional “scout” – those who had been sent by their superiors to find out what LOD is all about. To address their needs (as well as help those participants interested primarily in the the research potential of this method), the LD4DH lesson materials contain information to help them engage with the IT support services at their own institutions. These cover the practical considerations of setting up a LOD project that go beyond issues like researcher aims and institutional policies (both of which I discuss at length in Nurmikko-Fuller, in press), such as the need for a server (and a person to manage that server!); and the process by which to decide which triplestore is best for the project, and so on. We also compare LOD to markup languages (such as XML) and standard relational databases. These discussions can help those who have prior experience of either of the two alternatives to quickly visualise the differences between them and RDF.

The theoretical component of the workshop discusses vexed ethical issues associated with the use of this digital methodology. One of these (and arguably the one that is easiest for all of to relate to at a personal level), is the enormous potential it has to invade individual privacy. At the core of the Linked Data paradigm is a potential privacy crisis. It is the promise that this method can

16 https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data.

17 <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>.

18 <https://www.w3.org/TR/rdf11-primer>.

19 <https://www.w3.org/TR/turtle/>.

unlock knowledge by bringing information from disparate but complementary datasets together. When dealing with data such as historical records, this is of an immense benefit: scholars are able to create much more comprehensive pictures of the past by bringing together information from several different sources. But as I argue in my forthcoming book (Nurmikko-Fuller, in press), what if this technology is uncritically applied to us? Many of us categorise information in specific places, and recoil from the idea that a third party would have access to all that information simultaneously. Imagine finding out someone else was accessing our financial information, health records, employment history, and social media habits. Even if information is held in separate databases with different details removed to anonymise the data, Linked Data has potential for bringing all these fragments together, thus effectively removing any and all anonymisation.

The discussion of the theoretical foundations is supported, enriched, and diversified with daily hands-on activities, exposing the participants to both theory and praxis. The activities constitute the structural backbone of the workshop, and are arranged to follow the order of a RDF production workflow.

5 ACTIVITIES

There are two types of workflow that structure LD4DH. At the macro-level, there is the RDF production process, which gives the workshop its cohesion. At the micro-level, each hands-on session has its own specific workflow, with task-appropriate software, and session-specific learning objectives.

Throughout the week, participants focus on the same dataset, but move from familiarising themselves with the data (as illustrated in [Table 2](#) above) to modelling the content (converting .CSV to .TTL). Towards the end of the week, they progress from RDF production to writing SPARQL queries. This workflow has been reported on in the context of specific projects (Nurmikko-Fuller, Bangert, & Abdul-Rahman, 2017; Nurmikko-Fuller, Bangert, Dix, Weigl, & Page, 2018), so I will limit the discussion to an outline of activities to illustrate the learning objectives that bring the workshop together.

There are four discrete tasks as summarised in [Table 3](#). The first task requires participants to engage with data represented as RDF through a non-SPARQL endpoint (the so-called Follow-Your-Nose approach to information discovery). In the past, this activity has focused on the use of the Pubby UI,²⁰ however, from 2022 onwards the plan is to use the four different UIs available for DBpedia:²¹ DBpedia’s own resource page,²² OpenLink Faceted Browser²³, OpenLink Structured Data Editor,²⁴ and the LodLive Browser.²⁵ This activity is most useful to those participants who have prior knowledge of database design and management, as it helps them to start thinking about and understanding the notion of information captured in RDF as an interconnected graph. As part of this session, participants also practice converting between different syntaxes of RDF such as Turtle, RDF/XML, and JSON-LD using EasyRDF.²⁶

DAY	TASK	SOFTWARE
Monday	Follow-Your-Nose approach	DBpedia interfaces, EasyRDF
Tuesday	Design and implement ontologies	pen and paper, Protégé
Wednesday	Producing instance-level RDF	Web-Karma, Blazegraph
Thursday	Using triplestores and SPARQL	SPARQL Playground, Blazegraph
Friday	Exploring the British Museum’s collections	ResearchSpace

Table 3 Table of the assigned tasks for each day of the five days of LD4DH. In the middle are the tasks for each day; the right-hand column lists the specific software used for each task.

²⁰ <https://github.com/cygri/pubby>.

²¹ The homepage for DBpedia is at <https://dbpedia.org>, but the dropdown list for the three UIs is best accessed through a page for a resource. An example of such a page might be <https://dbpedia.org/page/Oxford>.

²² As above, for Oxford.

²³ For example <https://dbpedia.org/describe?uri=http%3A%2F%2Fdbpedia.org%2Fresource%2FOxford>.

²⁴ <https://osde.demo.openlinksw.com/#/editor?uri=http%3A%2F%2Fdbpedia.org%2Fdata%2FOxford.ttl&view=statements>.

²⁵ <http://en.lodlive.it/?http%3A%2F%2Fdbpedia.org%2Fresource%2FOxford>.

²⁶ <https://www.easyrdf.org/converter>.

The second task is to develop an ontological structure. As part of this process, participants must first familiarise themselves with the data (as illustrated in [Table 2](#)). They are also asked to formulate the research questions their ontological structure would be able to answer, and spend some time reading over the scope notes or specifications of other existing ontologies, so as to establish an overview as to how this type of data has been represented by others. In the case of the ElePHāT project data, learners focus almost exclusively on bibliographic metadata ontologies such as Bibframe,²⁷ FaBio,²⁸ FRBROO,²⁹ MODS/RDF,³⁰ MADS/RDF,³¹ and [Schema.org](#).³² They can also opt to incorporate aspects of these ontologies into their own ontological model, and to examine other vocabularies, schemas, and ontologies such as FOAF.³³ The activity includes a “show and tell” session where learners take turns to present their ontological models for peer-review by the others in the class.

Most participants are eager to engage with the praxis part of each day. Engaging with software and producing results or finding answers creates a sense of *doing* DH. For the ontology development stage however, the “software” of choice is pen and paper. Drawing and redrawing concepts is a lengthy, iterative process, and throughout the session the learners discuss and change aspects such as type of information category. Some examples might include: Is “Person” sufficiently detailed? Do we need “Author” and “Publisher” as different types of People? But what about when either is an institution? Should the data be modelled as authors and publishers as types of “Agent”? The aim is to create a schema-level representation of the data – to define the possible information categories and the relationships between those categories that are present in the dataset.

It is only once a consensus of sorts has been reached – normally under increasing time pressure as the end of the workshop approaches – that the participants progress to the implementation stage. This is done using Protégé.³⁴ It was chosen for two reasons: first, it is a popular tool, used across various different disciplines beyond Humanities, Arts, and Social Sciences; and, the point-and-click UI means that users do not need to acquire additional (potentially distracting) programming skills to complete this stage of the workflow.

Once complete, the ontological model is exported from Protégé as a .TTL file. This syntax of RDF is selected because it is most suitable for use in the next stage of the workflow.

The third step is to combine the original dataset (available as a .CSV and illustrated in [Table 2](#)) and the ontology (exported as a .TTL) to produce instance-level RDF. The tool of choice at DHOxSS for this part of the process has been Web-Karma,³⁵ a free and Open software from the University of Southern California. Like Protégé, this tool also has a point-and-click UI, which makes the mapping between the .CSV and the .TTL file possible without the need for programming. It provides a visualisation of the resulting graph, which is a clear and convenient way to check the progress and accuracy of the data mapping, and fix any possible errors.

The fourth and final stage of the process is two-fold. First, participants upload their RDF into a triplestore, and second, they learn to explore the new knowledge graph using the SPARQL Protocol and RDF Query Language (SPARQL).³⁶ A conscious decision has been made not to explore the protocol aspect, focusing exclusively on the query language. Our decision reflects existing advice (DuCharme, 2013) that describes the protocol as “rules for how a client program and a SPARQL processing server exchange SPARQL queries and results. These rules are ... mostly

27 <https://www.loc.gov/bibframe/>.

28 <https://sparontologies.github.io/fabio/current/fabio.html>.

29 <http://www.cidoc-crm.org/frbroo/home-0>.

30 <https://www.loc.gov/standards/mods/modsrdfl/>.

31 <https://www.loc.gov/standards/mads/rdf/>.

32 <https://schema.org/>.

33 <http://xmlns.com/foaf/spec/>.

34 <https://protege.stanford.edu/>.

35 <https://usc-isi-i2.github.io/karma/>.

36 SPARQL is a recursive acronym, meaning that the ‘S’ in SPARQL stands for “SPARQL”! It combines a Protocol and a Query Language for RDF, but for the purposes of LD4DH, we have chosen to focus on the query language aspect exclusively.

an issue for SPARQL processor developers". Given that the LD4DH workshops are not for SPARQL processor developers, the protocol is not covered.

Anecdotally, most participants have appeared to exhibit the most uncertainty and lack of confidence when asked to engage with SPARQL. They seemed to regard it as the most technical of the tasks. Undoubtedly this was due in part to a lack of readily available WYSIWYG (or What You See Is What You Get) UIs or graphical user-interfaces (GUIs). These would make the task less daunting by hiding the code behind a more familiar search box. The solution for this was the introduction of the SPARQL Playground³⁷ into the curriculum – this provided participants with simple and easy-to-read examples that allowed them to build up their familiarity with SPARQL in a step-by-step process.

For a number of years, the triplestore of choice was Virtuoso³⁸ based on two factors: first, it, like the other tools encountered in the context of LD4DH, was at one time a free tool (in the sense of both *gratis* and *libre*³⁹); second, it was the triplestore of choice for the original ElePHäT project. This benefited the workshop as the teaching staff were familiar with the triplestore and its idiosyncrasies. In 2018, the decision was made to switch to Blazegraph.⁴⁰ It emerged as the triplestore of choice because it remains free and open, as well as being relatively intuitive to manage.

By the end of Thursday afternoon, the learners have completed an entire workflow for converting tabular data into a knowledge graph. They have familiarised themselves with the data; produced an ontological model to enable them to represent the information within that dataset in a meaningful way, and to answer their desired research questions; they have produced instance-level RDF; and, successfully uploaded those triples into a triplestore, and completed SPARQL queries over them. In many ways, Thursday represents the pinnacle of the DHOxSS experience: it is the day where the challenges are the greatest, the frustrations the deepest, the euphoria of success the highest.

The week culminates in a day-long exploration of British Museum data. Friday's activities are primarily focused on applying the skills that have been acquired throughout the week, as opposed to up-skilling in a new area or technical ability. In 2018 and 2019, the LD4DH workshop concluded with a hands-on practical session (a mini-workshop, of types) exploring the British Museum's collection using the ResearchSpace tool.⁴¹ The day is largely run by Dominic Oldman and Diana Tanase of the British Museum, and presents the learners with an opportunity to apply their knowledge to a genuine, real-world LOD project, and to see how much they have learnt in the course of the week. They are also able to assess which skills they find most useful, relevant, and worth developing further. We also provide links and suggestions for additional tools (such as Open Refine⁴² for tidying data) and publications (DuCharme, 2013; Wood, Zaidman, Ruth, & Hausenblas, 2013; Van Hooland & Verborgh, 2014), as well as solving idiosyncratic problems (usually connected to projects the learners are working on outside of the LD4DH workshop). The afternoon trip to the Royal Oak⁴³ is, of course, purely optional, but well-attended.

6 EVALUATIONS

There are two things that most learners at the LD4DH have in common: they have identified LOD as a methodology they are interested in or that might have value for them. At the same time, it is very rare to have participants who have prior experience of the methodology (see

37 <https://sparql-playground.sib.swiss/>.

38 <https://virtuoso.openlinksw.com/>.

39 For those interested in the topics of Open Access and Open Source, the Wikipedia article on *gratis* and *libre* provides an succinct and easy-to-read summary of the differences between the two, and their application to intellectual property, computer code, and other relevant outputs" https://en.wikipedia.org/wiki/Gratis_versus_libre.

40 <https://blazegraph.com/>.

41 Note that this refers specifically to the ResearchSpace tool (<https://researchspace.org/>), and not to the British Museum's defunct original SPARQL endpoint (which at one time was available from <http://collection.britishmuseum.org/>). At the time of writing, the latter has been inaccessible for at least half a decade, and has never been used for the exercises at LD4DH.

42 <https://openrefine.org/>.

43 The pub on Woodstock Road, which, according to signage within the building was in the 1770s "a desolate spot".

for example, the independent reports from 2017⁴⁴ and 2018⁴⁵). The workshop presents an opportunity for up-skilling, but it is also an intense experience.

Participant feedback was available for four years, 2018–2021 inclusive. This provides a balanced set of two years when the workshop was taught in person in Oxford, and two of which were online. The number of respondents was small, but increased annually: only three of the 20 participants filled in the feedback form in 2018. This increased to six in 2019, eleven in 2020, and twelve in 2021, resulting in just 32 responses across four years (the size of the group is largely capped at 20 due to the limitations of available teaching space, but some additional students attend each year). The questions also changed with the move to the online medium: in 2018 and 2019 there is no data regarding the country of origin of the participants, but the corresponding information from 2020–2021 (“the COVID-years”) shows a spread of eleven different countries (the Netherlands, UK, Italy, India, Canada, Portugal, Germany, China, Switzerland, Mexico, and Spain). Participants from earlier years (although this is not captured in the survey data explicitly) are known to include those from at least France, Austria, Norway, and Sweden, as well as the UK.

None of the feedback forms collected demographic details of participants such as age or gender, focusing instead on levels of professional development and domain. With just one exception (a software researcher), all participants across all years were either academics (including students, early career, mid-career, and late career) or GLAM-sector professionals. Of the 32 respondents, nine described themselves as falling into more than one category, most frequently as being both researchers and practitioners – this is not surprising given the nature of DH more generally and the interest and uptake of LOD specifically.

With such a low response rate (32 attendees) it is difficult to draw conclusions of any statistical significance. Having said that, the respondents were a heterogeneous group and, albeit self-selected, in some respects they could be seen as a maximum diversity sample. At the very least, they offer impressions worth noting.

The feedback was overwhelmingly positive, with most critical feedback reflecting the challenges of having to move to the online delivery method with short notice in 2020. Throughout the four years, almost all aspects of the Linked Data workshop were categorised as either “good” or “excellent” (and only two as “satisfactory” and none as “poor”). Qualitatively, the comments from participants include phrases such as “inspiring” (2018), “brilliant”, “excellent”, “great job” (2019), “successful event” (2020), and “extremely well moderated and extraordinarily well organised”, “...just great. It is very helpful...”, “Very interesting talks and very good overall experience”, and “Excellent Organisation. Wonderful you get to have the presentations” (2021). In 2018, the benefit of having several tutors in particular were highlighted: “The workshop was inspiring and challenging (in the best way). Terhi, John, and Graham were so generous with their time and knowledge. I enjoyed having their different points of view. I have already recommended the summer school to several colleagues”.

Two aspects of the workshop received criticisms in subsequent years. The first highlights the importance of expectation management. Unfortunately, the survey data does not cover those iterations of the workshop which took place in the facilities at the University of Oxford: in those years, the software was pre-installed on desktop machines in a small computer lab. In 2018 and 2019, participants were asked to arrive at the Summer School with the software pre-installed on their personal devices: not all participants complied with this requirement in either year. This represented a major challenge for the organisers: at least one tutor and often more than one had to shift their focus from teaching content or explaining tasks to the whole group to focus on problems arising from an individual machine. In some cases participants with institutional laptops had limited access rights, preventing software installation and/or the downloading of prerequisite libraries. Other participants attended with a tablet rather than a laptop, others refused to power up their machines with an alternative operating system from the USB-stick they had been given. Participant expectation can also present some challenges: there is an underlying assumption that the tutors of the workshop are also experts at installing

⁴⁴ <https://dhh.uni.lu/2017/07/12/dhoxss-2017-linked-open-data/>.

⁴⁵ <https://www.hirmeos.eu/2018/08/07/discovering-linked-open-data-at-the-digital-humanities-at-oxford-summer-school/>.

the necessary software on any and all machines, regardless of operating system, prerequisites, or administrative restraints. Feedback from 2018 illustrates this point: “The software tools did not work and I felt this could have been sorted out in advance”.⁴⁶

The main difference between 2018–2019 and the COVID-years was the move to online delivery. In an era of Zoom-fatigue and relentless online delivery, we may be quick to categorise the latter as less preferable. The participant feedback paints a more complex picture, however. In 2018–2019, there was negative feedback about the physical room and conditions in which the workshop took place, an aspect of in-person teaching, which we may be prone to forget: “Room was physically uncomfortable and layout did not suit the style of workshop, hence the low score for learning environment” (2018); “The teaching was excellent but the size of the room for the number of attendants and facilitators was not appropriate. It was very difficult to move around the room, see presentations from various angles of the room, for facilitators to communicate to attendees in small groups/individually and for us to break out to do group work. The unexpected heat wave made this even more unbearable” (2019); and “Our room was way too small for the amount of people, at times it was very loud” (2019).

The sudden move to online delivery for LD4DH in 2020 elicited negative feedback on some aspects of the workshop, in particular the hands-on element. It was inevitably more difficult to provide a seamless experience without the necessary time to develop the appropriate mode of delivery: “The Linked Data workshop felt like another lecture and was not really hands-on” and “The interactive workshop I attended seemed completely unprepared for teaching in an online environment” (2020).

Having learnt our lesson, and perhaps reflecting evolving attitudes as to the benefits of online learning, the feedback in 2021 was very positive: “the theoretical part of the morning sessions connected perfectly to better understand the practical part of the afternoon workshops. Congratulations!” (2021). A very welcome result of the move to the online medium was that it opened the workshop up to an international audience (as illustrated by the inclusion of participants, for the first time, from China and India) as well as for at least one neurodivergent attendee: “...Format worked well. As someone who is autistic, aspects of this worked better than in person. It would be great if there was a way to make your next in person event more accessible to neurodivergent participants by including some hybrid elements from the online event. You might see if a few neurodivergent people in DH could make specific suggestions to help” (2020). Future iterations of LD4DH will seek to find ways to replicate some of these successes and affordances, and to continue to cater for the needs of diverse cohorts.

7 CONCLUSION

Thinking back to my experience as a participant of the LOD workshop in 2012 has provided an opportunity to stand back and evaluate how it has evolved during my time as a lecturer and how it meets the needs of those who participate today. None of the lecturers wore sparkly trousers, for one. Software develops and changes, new projects emerge, some of the conceptual and philosophical debates remain the same. What else is different? How has our pedagogy evolved? Has the market shifted, or has the typical participant changed?

I believe that the feedback from the participants, with all its caveats, shows that the approaches we have applied to LD4DH have been successful in meeting and even exceeding the expectations of our diverse cohort of students. But the workshop is only one part of the much greater experience of DHOxSS itself. The unique and undoubtedly strongest asset of the Summer School is that it brings together some of the very best of the DH community. It creates an international, open, and dynamic learning environment for participants, providing ample opportunities for up-skilling, knowledge transfer, and networking. All these aspects have contributed to success. And so, as is so often the case with examples of successes in DH, at the core here too is the most important thing that makes the Summer School what it is: the people.

⁴⁶ Please note that in 2018 the participants were asked to arrive with the workshop with the prerequisite software already installed.

The author has no competing interests to declare.

AUTHOR CONTRIBUTIONS

Conceptualization, Formal Analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing.

AUTHOR AFFILIATION

Terhi Nurmikko-Fuller  orcid.org/0000-0002-0688-3006

Centre for Digital Humanities Research, Australian National University, Canberra, Australia

REFERENCES

- Brier, S.** (2012). Where's the pedagogy? In M. K. Gold (Ed.), *The role of teaching and learning in the Digital Humanities*, 1, 390–401. Minneapolis, MN: University of Minnesota Press. DOI: <https://doi.org/10.5749/minnesota/9780816677948.003.0038>
- DuCharme, B.** (2013). *Learning SPARQL: querying and updating with SPARQL 1.1*. Sebastopol, CA: O'Reilly Media.
- Imlawi, J., & Gregg, D.** (2014). Engagement in online social networks: The impact of self-disclosure and humor. *International Journal of Human-Computer Interaction*, 30(2), 106–125. DOI: <https://doi.org/10.1080/10447318.2013.839901>
- Jett, J., Nurmikko-Fuller, T., Cole, T. W., Page, K., & Downie, J. S.** (2016). Enhancing scholarly use of digital libraries: A comparative survey and review of bibliographic metadata ontologies. In *Proceedings of the 16th ACM/IEEE Joint Conference on Digital Libraries* (pp. 35–44). Newark, NJ: ACM. DOI: <https://doi.org/10.1145/2910896.2910903>
- Khan, N. J., Nurmikko-Fuller, T., & Page, K.** (2016). BABY ELEPHANT: Building an analytical bibliography for a prosopography in Early English imprint data. In *IConference 2016 Proceedings*. Urbana, IL: iSchools. DOI: <https://doi.org/10.9776/16588>
- Needham, J., & Haas, J. C.** (2019). Collaboration adventures with primary sources: Exploring creative and digital outputs. *The Journal of Interactive Technology and Pedagogy*, 14(9). Retrieved from <https://jitp.commons.gc.cuny.edu/collaboration-adventures-with-primary-sources-exploring-creative-and-digital-outputs/>
- Nurmikko-Fuller, T.** (in press). *Linked Data for Digital Humanities*. Oxford, UK: Routledge.
- Nurmikko-Fuller, T., Bangert, D., & Abdul-Rahman, A.** (2017). All the things you are: Accessing an enriched musicological prosopography through JazzCats. In *Proceedings of the international conference of Digital Humanities* (pp. 554–556). Montreal, Canada: Alliance of Digital Humanities Organizations. Retrieved from <https://dh2017.adho.org/abstracts/305/305.pdf>
- Nurmikko-Fuller, T., Bangert, D., Dix, A., Weigl, D., & Page, K.** (2018). Building prototypes aggregating musicological datasets on the Semantic Web. *Bibliothek Forschung und Praxis*, 42(2), 206–221. DOI: <https://doi.org/10.1515/bfp-2018-0025>
- Page, K., Nurmikko-Fuller, T., Cole, T. W., & Downie, J. S.** (2017). Building worksets for scholarship by linking complementary corpora. In *Proceedings of the international conference of Digital Humanities* (pp. 319–321). Montreal, Canada: Alliance of Digital Humanities Organizations. Retrieved from <https://dh2017.adho.org/abstracts/606/606.pdf>
- Van Hooland, S., & Verborgh, R.** (2014). *Linked Data for libraries, archives and museums: How to clean, link and publish your metadata*. London, UK: Facet.
- Wood, D., Zaidman, M., Ruth, L., & Hausenblas, M.** (2013). *Linked Data: Structured data on the Web*. New York, NY: Manning Publications.

TO CITE THIS ARTICLE:
Nurmikko-Fuller, T. (2022). Teaching Linked Open Data using Bibliographic Metadata. *Journal of Open Humanities Data*, 8: 6, pp. 1–11. DOI: <https://doi.org/10.5334/johd.60>

Published: 10 March 2022

COPYRIGHT:
© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.