Getting Started in DS

Eleonora Gandolfi   Nora McGregor

2024-06-04

This guide aims to define Digital Scholarship & Data Science in a research library context and provide an overview of the key competency frameworks, reports, networks and communities of practice for library professionals to explore and learn more.

Introduction

In this Topic Guide we'll look at broad definitions of digital scholarship & data science in a research library context, and provide recommended resources for library professionals to learn more about how new technologies and approaches are changing our traditional work of curation, creation, collecting, sharing at cultural heritage institutions and the benefits of gaining new skills in this area!

What is "digital scholarship"?

Though there are many definitions of what constitutes "digital scholarship" out there, we tend to favour the broadest of interpretations, that is, roughly, any type of innovative research that combines the methodologies from traditional humanities & social science disciplines with computational tools and digital methods provided by computing disciplines, such as data science.

Though closely aligned to, and generously informed by, the academic discipline Digital Humanities, "digital scholarship" allows us to consider more broadly the full range of innovative scholarly activities our users seek to undertake with our digital collections and data, across a diverse range of disciplines.

> **"Digital scholarship" allows us to define a space for us heritage professionals, where research is undertaken, in our own right, during the course of our daily work, utilizing computational methods in the curation, creation, collecting and sharing of our digital collections and data, but is not confined to formal academic pursuits or a particular discipline.**

Is this all just "data science"?

Not quite. Data science, also an academic field, can be defined as a set of computational methods for the identification of novel and actionable insights from data.

Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data.

So when we talk about data science in libraries, we're talking about the specific skills and use of particular computational methods in undertaking some types of digital scholarship activities.

Relevance to the Library Sector (Case Studies/Use Cases)

So what might "digital scholarship and data science in libraries" actually look like?

- a subject librarian using a digital tool to clean up a set of catalogue records in order to understand gaps in the metadata and collection scope
- a collaborative project to automatically transcribe handwritten texts from old manuscripts
- a library director reading up on the latest in AI and seeking expert perspectives in order to write a strategy document
- a metadata specialist packaging up digital collections into datasets that can be used by researchers
- a digitisation project looking to improve the searchability of printed books in under-resourced languages
- a reference librarian pointing a researcher to datasets that might help them with their research enquiry
- an imagining technician creating a 3D model of a collection item
- a licensing manager keeping up to date on the latest uses of Text and Data Mining (TDM) and AI in research so that digital collections meets the needs of library users and staff who want to work with them at scale
- a major interdisciplinary research project using the latest technologies to ask research questions of digital heritage collections https://livingwithmachines.ac.uk/
- a curator creating an online exhibit with annotations https://www.exhibit.so/
- a research software engineer contributing to international knowledge exchange networks (e.g. IIIF, AI4LAM)
- an assistant librarian attending a summer school https://www.cdh.cam.ac.uk/dataschools/

Why is it important for library staff to learn these skills?

In the recommended reading section of this Topic Guide we have linked to a number of key competency and skills reports and frameworks that define the need for such skills in library work. But in the most high-level terms we know that:

- Libraries need to continually keep apace this digital turn in order to understand the change in service requirements and support colleagues and each other keen to make the most of it.
- Digital scholarship work is collaborative, requires input across disciplines and domain expertise, our curatorial experts have an essential role to play in that.
- We've so much to gain from understanding digital methods and having closer collaborations with digital scholars—there's a synergy in solving shared issues (e.g. correcting OCR, enriching collections metadata, conquering back-cataloguing).
- Digital scholars are, today, using technology in innovative ways, expectations have already changed, they're seeking access at scale to our collections for computational analysis, they're using Generative AI to ask their research questions, we need to understand these technologies to understand how the nature of archival enquiry is changing
- Cultural heritage digital collections are only going to grow and we need the digital skills to work confidently with them at scale

Hands-on activity and other self-guided tutorial(s)

Each Topic Guide on this site includes a range of specific hands-on activities and other self-guided tutorials colleagues across Europe personally recommend. When you're ready to go further and have a better idea of the specific skills you need for a particular task, we can recommend having a good search through these excellent platforms which host or link to a great many in-depth training materials:

DARIAH-Campus https://campus.dariah.eu/

DARIAH is a pan-European infrastructure for arts and humanities scholars working with computational methods. It supports digital research as well as the teaching of digital research methods. Though not specific to the library professional context, tutorials here are useful for applying techniques to digital collections.

The Glam Workbench https://glam-workbench.net/

The GLAM Workbench is the brainchild of Tim Sherratt, a historian, and is a collection of Jupyter notebooks to help you explore and use data from GLAM institutions (galleries, libraries, archives, and museums). It includes tools, tutorials, examples, hacks, and even some

pre-harvested datasets. It's aimed at researchers in the humanities but has useful tutorials for anyone interested in working with GLAM data.

Ineo https://www.ineo.tools/

Ineo is a project developed and maintained by CLARIAH that lets you search, browse, find and select digital resources for research in humanities and social sciences. At the end of 2024 it will offer access to thousands of tools, datasets, workflows, standards and learning material. It is a work in progress so do keep that in mind when browsing.

Library Carpentry https://librarycarpentry.org/

Library Carpentry is an international volunteer community, under the Carpentries, focussed building software and data skills within library and information-related communities. The lessons here are meant to be taught as workshops led by a Carpentries certified instructor (for a fee) but you may find it useful to have a read through the content which is open and available to all.

The Programming Historian https://programminghistorian.org/en/

The Programming Historian has been publishing peer-reviewed tutorials on digital tools and techniques for humanists since 2008 and though they're generally aimed at academic researchers, staff at British Library have found them highly useful over the years in their own work!

Social Sciences & Humanities Open Marketplace
https://marketplace.sshopencloud.eu/search?order=score&categories=training-material

Built as part of the Social Sciences and Humanities Open Cloud project (SSHOC), the Social Sciences and Humanities Open Marketplace is a discovery portal which pools and contextualises resources for Social Sciences and Humanities research communities: tools, services, training materials, datasets, publications and workflows. The Marketplace highlights and showcases solutions and research practices for every step of the SSH research data life cycle.

Recommended Reading/Viewing

There is no shortage of recommended reading lists out there, and again, each Topic Guide on this site will have their own recommended reading on a particular topic. Here we present a number of key reports and publications which articulate the broad sectoral view of why such skills are necessary for library professionals.

Digital Scholarship & Data Science Skills Competency Frameworks

LIBER Publications

- LIBER Digital Skills for Library Staff & Researchers Working Group - LIBER Europe have lots of resources here, including a very useful diagram Identifying Open Science Skills for Library Staff & Researchers

- LIBER Job Description Repository Contains job description examples for Digital Curator and other digital roles which reference the types of skills required for such work.

- Europe's Digital Humanities Landscape: A Study From LIBER's Digital Humanities & Digital Cultural Heritage Working Group is a report based on a Europe-wide survey run by LIBER's Digital Humanities & Digital Cultural Heritage Working Group. The survey focused on digital collections and the activities libraries undertake around them. It covered the following topics and themes including staffing/skills

Key Publications specific to digital scholarship and data science skills for research library staff

- The British Library and the Arts and Humanities Research Council published a report on skills: Scoping Skills and Developing Training Programme for Managing Repository Services in Cultural Heritage Organisations. There is a very useful section (Section 3.) that references several other digital skills frameworks for research library staff across Europe.

- Lippincott, Joan K. Directions in Digital Scholarship: Support for Digital, Data-Intensive, and Computational Research in Academic Libraries. Coalition for Networked Information, June 2023. https://doi.org/10.56561/ULHJ1168

- Padilla, Thomas. 'Responsible Operations: Data Science, Machine Learning, and AI in Libraries'. OCLC, 26 August 2020.

- Cordell, R. C. (2020). Machine Learning + Libraries: A Report on the State of the Field. LC Labs, Library of Congress.

- Federer L. Defining data librarianship: a survey of competencies, skills, and training. J Med Libr Assoc. 2018 Jul;106(3):294-303. doi: 10.5195/jmla.2018.306. Epub 2018 Jul 1. PMID: 29962907; PMCID: PMC6013124.

General Competencies for Librarians which include reference to digital

- American Library Association (ALA) Library Competencies (Various roles): Library Competencies | Tools, Publications & Resources (ala.org) (USA)

- Canadian Association of Research Libraries Competencies for Librarians in Canadian Research Libraries Publications and Documents (including specifically Competencies-Final-EN-1-2.pdf (Canada)

- CILIP: the library and information association Professional Knowledge & Skills Base - (UK)

Finding Communities of Practice

As you embark on learning more about digital scholarship and data science in a library context, you might want to explore and join existing communities of practice.

LIBER Working Groups

Working groups are open to staff at participating LIBER Member institutions:

- LIBER Data Science in Libraries
- LIBER Digital Scholarship & Digital Cultural Heritage
- Or have a look at the other LIBER Working Groups - LIBER Europe

International Networks

- AI4LAM
- Code4Lib
- Electronic Literature Organisation
- Flickr Commons
- Glam Labs International
- IIIF/UV Open Collective
- IMPACT Centre of Competence
- Museums Computer Group
- Transkribus
- Wikidata Community)

National Networks (European)

Ireland/UK - RLUK Digital Scholarship Network

💡 Think we're missing something?

We'd love to hear it! Suggest edits by opening a new Issue or adding to the discussion on existing Issues on the project Github. If you're new to GitHub don't worry, we have a Topic Guide for that: GitHub: How to navigate and contribute to Git-based projects! Or just drop us a line!