

Digital Scholarship & Data Science Essentials for Library Professionals

Peter Verhaar and Neha Moopen



**Universiteit
Leiden**
The Netherlands

Discover the world at Leiden University

This workshop is organised by

- The Digital Scholarship and Digital Cultural Heritage Collections Working Group
- The Data Science in Libraries Working Group



L I B E R
L I B E R

Workshop agenda:

<https://edu.nl/pue8y>

Applications of data science

- Survey conducted by WG Data Science in Libraries
- Ryan Cordell, *Machine Learning + Libraries: A Report on the State of the Field*
- Joan Lippincott, *Directions in Digital Scholarship: Support for Digital, Data-Intensive, and Computational Research in Academic Libraries*
<<https://doi.org/10.56561/ULHJ1168>>
- Lisa Federer, *Defining data librarianship: a survey of competencies, skills, and training*
<0.5195/jmla.2018.306>

Library Analytics

- Evidence based acquisition
- Analyses of circulation data
- Usage data and statistics in the form of COUNTER statistics
- Number of questions
- Number of participants in workshops
- Assessment of the impact of cancellations of journals via unSUB
- Data about space occupancy in study spaces
- Data on wifi activity



Research Intelligence

- Analyses of CRIS metadata to study the journals researchers publish in
- Bibliometric analyses based on research output and citation analysis
- Analysis based on altmetrics
- Use of APIs from funders system
- Use of the ORCID API
- Research on the uptake of open access (e.g. using DOAJ or OPenAlex)



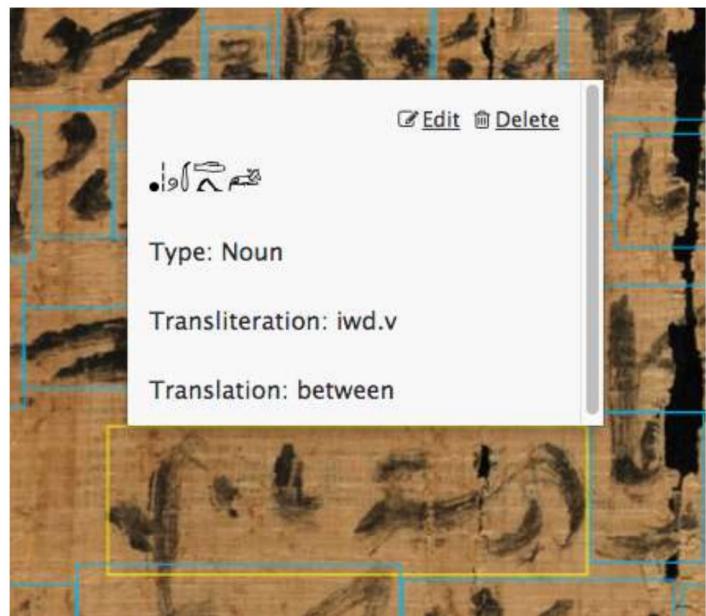
DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS



Digitisation and metadata

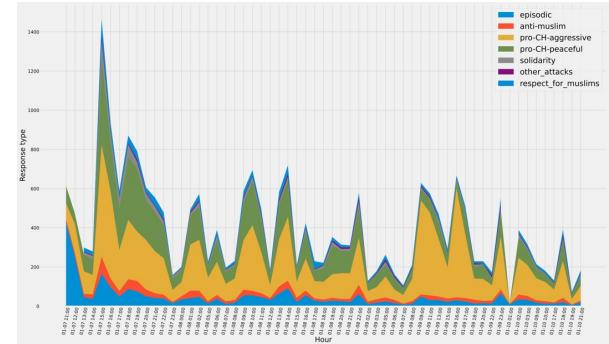
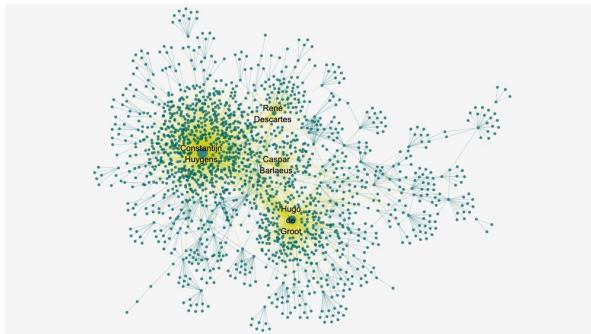
- Metadata for images generated by computer vision
- LLMs used to propose subjects headings from controlled thesaurus
- Linking named entities such as authors to authority files and identifies
- Handwritten Text Recognition
- Crowdourcing platforms to georeference historical maps





Involvement in research

- Advice and consultation
- Direct participation in research projects
- Research Software Engineers



Education



- Courses on programming
- Workshops on Tools and methods





Universiteit
Leiden

mouse&manuscript

HOME

LESSONS

GLOSSARY

ABOUT



Lessons in codicology and palaeography

based on manuscripts from the Middle East,
Islamic Africa and beyond



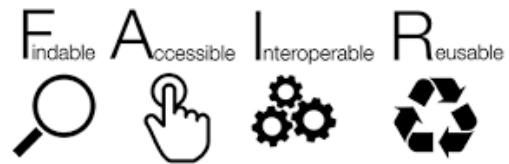
<https://mouse.digitalscholarship.nl/>

Research Data Management

- Advice on data formats
- Advice on data modelling
- Disambiguation of named entities by linking these to identifiers
- Cleaning up data using OpenRefine
- Converting data to linking open data



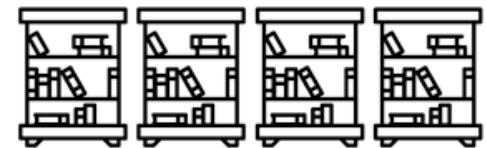
OpenRefine



Collections as Data

- Thomas Padilla et al. “Always Already Computational: Collections as Data, Final Report”
[<https://doi.org/10.5281/zenodo.3152935>](https://doi.org/10.5281/zenodo.3152935)
- Toma Tasovac, “Cultural Heritage Data from a Humanities Research Perspective: A Dariah Position Paper”
[<https://hal.science/hal-02961317/document>](https://hal.science/hal-02961317/document)

Always Already Computational



KB LAB

Datasets about our collections



GLAM Workbenches

API | GET: <https://api.data.netwerkdigitaalergoed.nl/queries/rkd/opvragen-kunstenaars/run>

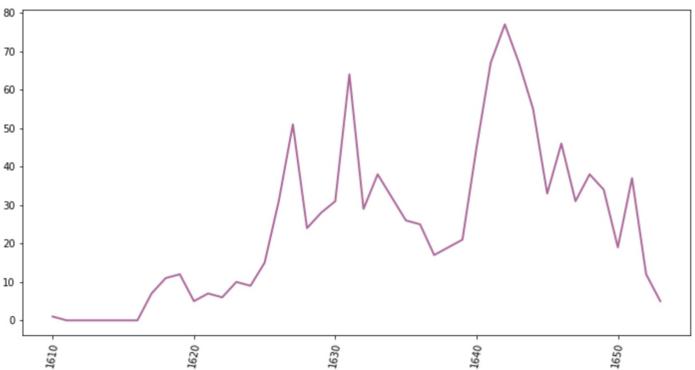
Variables [+](#)

```
3 PREFIX rkdo: <http://data.rkd.nl/def#>
4 PREFIX schema: <http://schema.org/>
5 prefix wd: <http://www.wikidata.org/entity/>
6 prefix wdt: <http://www.wikidata.org/prop/direct/>
7
8 SELECT ?name ?gender ?nationality ?id ?add ?alternate ?pob ?birth_date
9 ?pod ?death_date ?edu_name ?title ?image ?event ?event_desc ?event_location ?event_lat ?event_long ?event_date
10
11 WHERE {
12
13 ?artist schema:identifier '56854' .
14 ?artist schema:name ?name .
15 OPTIONAL { ?artist schema:gender ?gender . } .
16 OPTIONAL { ?artist schema:nationality ?nationality . } .
```



```
import matplotlib.pyplot as plt
import seaborn as sns

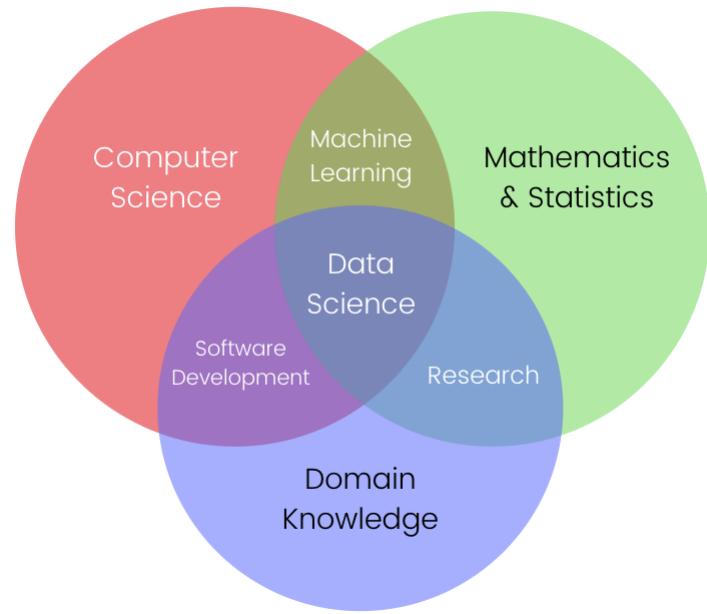
fig = plt.figure( figsize = ( 12, 6 ))
ax = sns.lineplot( x= x_axis , y= y_axis ,
color= '#AA6799' , linestyle='solid', linewidth=2 )
plt.xticks(rotation= 80)
plt.show()
```



- Do you have comments, additions or questions?
- To what extent is your own institution involved in these types of activities?
- Can you think of other applications that should be added to this overview?

Challenges

- Lack of qualified staff
- Lifelong learning
- Scalability
- Data science not always articulated in job description
- Weak participation by library staff in other units
- Working with research and not only for researcher



Data Science Terms
Explained

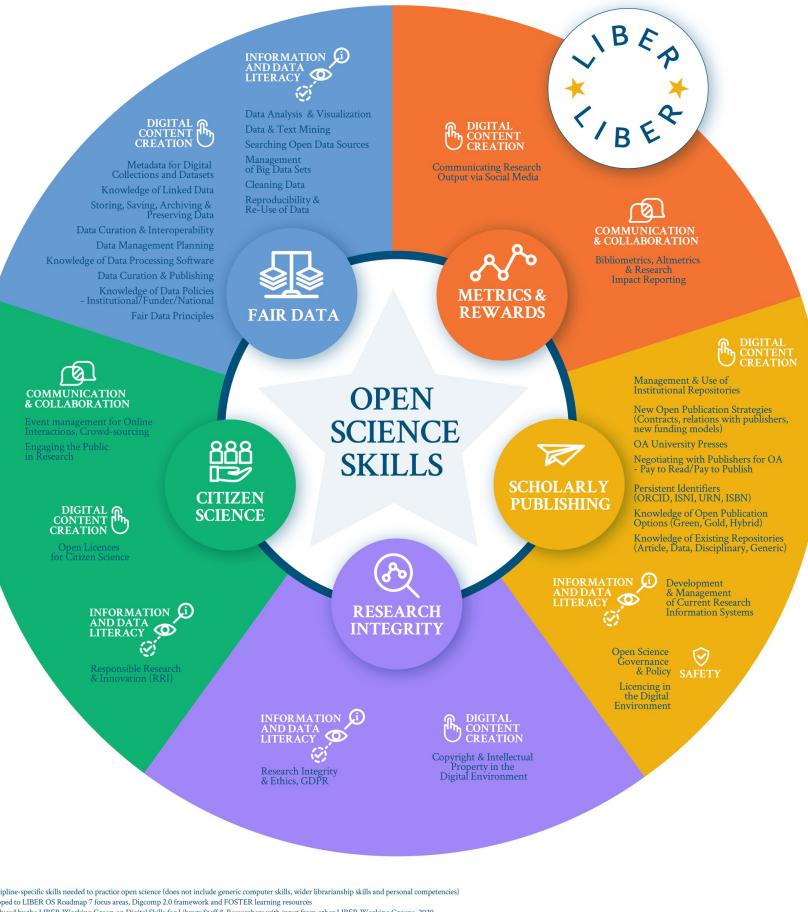
Challenges (continued)

- Issues with Data Quality
 - Low orthographic quality of texts created via OCR
 - Poor data management
 - Lack of interoperability
- Infrastructural Issues
 - No HPC facilities
 - Lack of storage or back-up facilities
- Legal issues and copyright issues
 - Inconsistencies between licences
 - Laws may hinder accessibility
- No institutional policy or mandate
 - Lack of support and no budget

- Do you recognise these challenges?
- Are there specific challenges that should be added to this list?

Competencies

- British Library and AHRC, *Scoping Skills and Developing Training Programme for Managing Repository Services in Cultural Heritage Organisations*
- IFLA, *Core Competencies of Library Professionals*
- CARL, *Competencies for Librarians in Canadian Research Libraries*
- ALA, *Library Competencies*
- ALCTS, *Core Competencies for Cataloging and Metadata Professional Librarians*



LIBER, Open Science Skills Visualisation,
<https://zenodo.org/records/3702401>

Digital Humanities at Oxford Summer School

Digital Humanities @ Oxford Summer School 2023





Donate to *Programming Historian* today!

Programming Historian

ABOUT ▾ CONTRIBUTE ▾ LESSONS EVENTS SUPPORT US ▾ BLOG EN ES FR PT



Lesson Index

Our lessons are organized by typical phases of the research process, as well as general topics. Use the buttons to filter lessons by category. If you can't find a skill, technology, or tool you're looking for, please [let us know!](#)

ACQUIRE (12) TRANSFORM (35) ANALYZE (30) PRESENT (24) SUSTAIN (2)

APIS (8) PYTHON (30) DATA MANAGEMENT (9) DATA MANIPULATION (30) DISTANT READING (14)

SET UP (7) LINKED OPEN DATA (1) MAPPING (14) NETWORK ANALYSIS (6) WEB SCRAPING (5)

DIGITAL PUBLISHING (12) R (8) MACHINE LEARNING (4) CREATIVE CODING (1)



The Digital Humanities Literacy Guidebook



FINDING THE KLAN WITH NETWORK ANALYSIS

Elaine Frantz (Kent State University)

Historical network analysis.

 HISTORY, NETWORK SCIENCE



GIS MAPPING

Susan Grunewald (University of Pittsburgh)

GIS mapping with an emphasis on history projects.

 HISTORY



HISTORICAL GAZETTEERS

Ruth Mostern (University of Pittsburgh)

Building historical gazetteers.

 HISTORY



HOW TO GROW DATA FORESTS WITH XML TREES

Elisa Beshero-Bondar (University of Pittsburgh)

eXtensible Markup Language (XML).

 LITERATURE, ENGLISH



IMPROVING ACCESS TO VIDEO ORAL HISTORIES

Michael Christel (Carnegie Mellon University)

Video oral history projects.

 HISTORY, PUBLIC HISTORY



LOGISTIC REGRESSION

Matthew J. Lavin (University of Pittsburgh)

Machine learning for literary analysis.

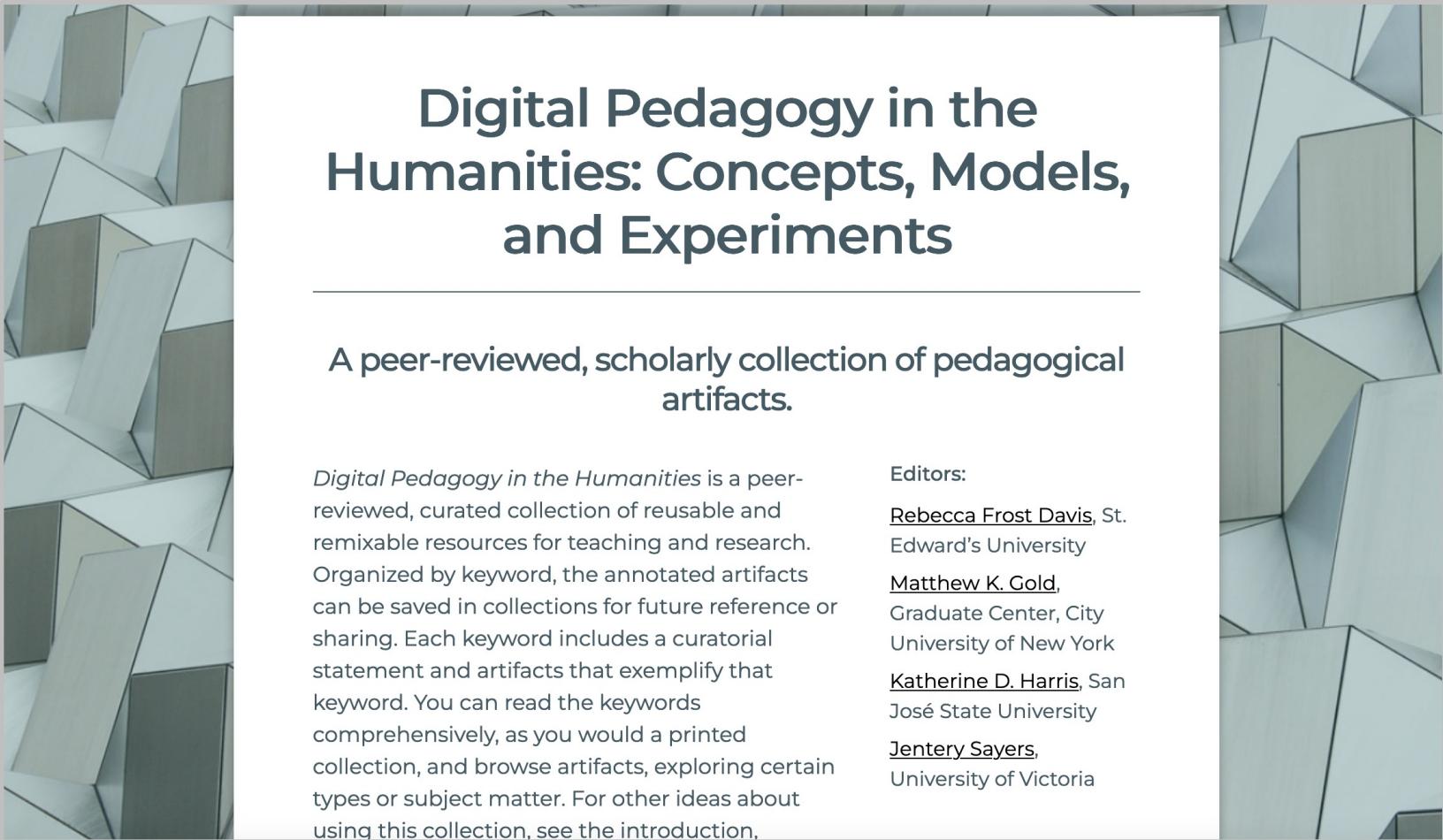
 LITERATURE, ENGLISH

The screenshot shows the homepage of the Open Educational Resources For The Digital Arts & Humanities website. The header features a logo with a stylized sun-like icon and the text '#dariahTeach'. Below the header, the main title 'Open Educational Resources For The Digital Arts & Humanities' is displayed. To the right is a login form with fields for 'Username' and 'Password', and buttons for 'Shibboleth Login' and 'Login as Guest'. A link for 'Forgotten your username or password?' is also present. The main content area has a teal background. On the left, there's a large image of a vintage typewriter with the word 'IGNITE' overlaid. To the right, there are several cards representing different resources:

- Unit III: Museums and Digital Storytelling**: Includes a video thumbnail of Chimamanda Ngozi Adichie speaking at TED.
- Unit IV: Physical Heritage Design, Interaction, and Evaluation**: Includes a thumbnail showing a 3D model of a building complex.
- Unit V: Talking Stories**: Includes a thumbnail of a book titled 'The Talking Book'.
- Unit VI: Design Objectives & Key Qualities of the Digital Medium**: Includes a thumbnail of a person working in a workshop.
- Unit VII: Design Objectives & Key Qualities of the Digital Medium**: Includes a thumbnail of a painting.
- Unit VIII: Physical Heritage Design, Interaction, and Evaluation**: Includes a thumbnail of a person working in a workshop.

Navigation links like 'Home', 'Contact', 'Testimonials', and 'About' are visible at the top, along with a search bar.

<https://teach.dariah.eu/>



Digital Pedagogy in the Humanities: Concepts, Models, and Experiments

A peer-reviewed, scholarly collection of pedagogical artifacts.

Digital Pedagogy in the Humanities is a peer-reviewed, curated collection of reusable and remixable resources for teaching and research. Organized by keyword, the annotated artifacts can be saved in collections for future reference or sharing. Each keyword includes a curatorial statement and artifacts that exemplify that keyword. You can read the keywords comprehensively, as you would a printed collection, and browse artifacts, exploring certain types or subject matter. For other ideas about using this collection, see the introduction,

Editors:

Rebecca Frost Davis, St.
Edward's University

Matthew K. Gold,
Graduate Center, City
University of New York

Katherine D. Harris, San
José State University

Jentery Sayers,
University of Victoria

<https://digitalpedagogy.hcommons.org/>

Digital Scholarship



edusources

Collections (3)



Collection

[Text Encoding](#)

1 item



Collection

[Databases](#)

2 items



Collection

[Linked Open Data](#)

1 item

<https://search.edusources.nl/en/communities/digital-scholarship>

Challenges



- Where to get started?
- Difficult to evaluate the relevance of tutorials beforehand
- Focus is occasionally on specific disciplines
- Sites are sometimes difficult to navigate
- Prior knowledge may be needed
- Focus generally on research

Learning hub

- “Digital Scholarship & Data Science Essentials for Library Professionals”
- Curated guidance for librarians wanting to get introductory grounding in data science
- No duplication of efforts
- Connect learners to the right type of materials
- Pointers for beginners and for more advanced learners
- Discussion of applications within libraries



LIBER Learning Hub

Prototype:

<https://nehamoopen.github.io/liber-learning-hub/>

Break

Discussion 1

- Central questions:
 - Which skills would you like to develop?
 - About which topics would you like to have more knowledge?
 - Which topics should be covered on the learning hub?
 - What are the gaps in our knowledge?
- Discussion in small groups: 10 minutes
 - Write topics on the flashcards
- Next, we will group and prioritise the topics. You can express approval with the coloured dots

Guides

- Topics guides will be developed on the basis of a template:

<https://edu.nl/ub6we>

- Template prescribes
 - Metadata
 - (links to) tutorial for beginners and for more advanced learners
 - Exercises and recommendations for practical activities
 - References to Cookbooks and GLAM Workbenches
 - Discussion of application within libraries

Discussion 2

- Central questions:
 - Do you think that these types of guides would indeed help you to develop knowledge about these topics?
 - Do you miss specific aspects?
 - Are there components which can be removed from the template?
 - Do you think that the first topic guide on LLMs could be helpful? Is there anything you would to add or change?
- Assume that you have been given time to study these topics
- Keep notes in the Google Doc provided. Appoint a notetaker!

Other questions

- Who should write the guides? Can we commission authors?
- How can we collect user feedback?
- Do we need to appoint an editorial board?
- How can we promote the use of the learning hub?
- Should we monitor the usage?
- How can we make the resource findable?
- Should LIBER host the learning hub?
- How to organise the maintenance? Who can we make sure that guides are updated, if necessary

Thank you!

- Peter Verhaar
p.a.f.verhaar@library.leidenuniv.nl

- Neha Moopen
n.moopen@uu.nl