RESEARCH DATA MANAGEMENT: USEFUL TOOLS

ORGANIZING • CLEANING • STORING

DH Lab

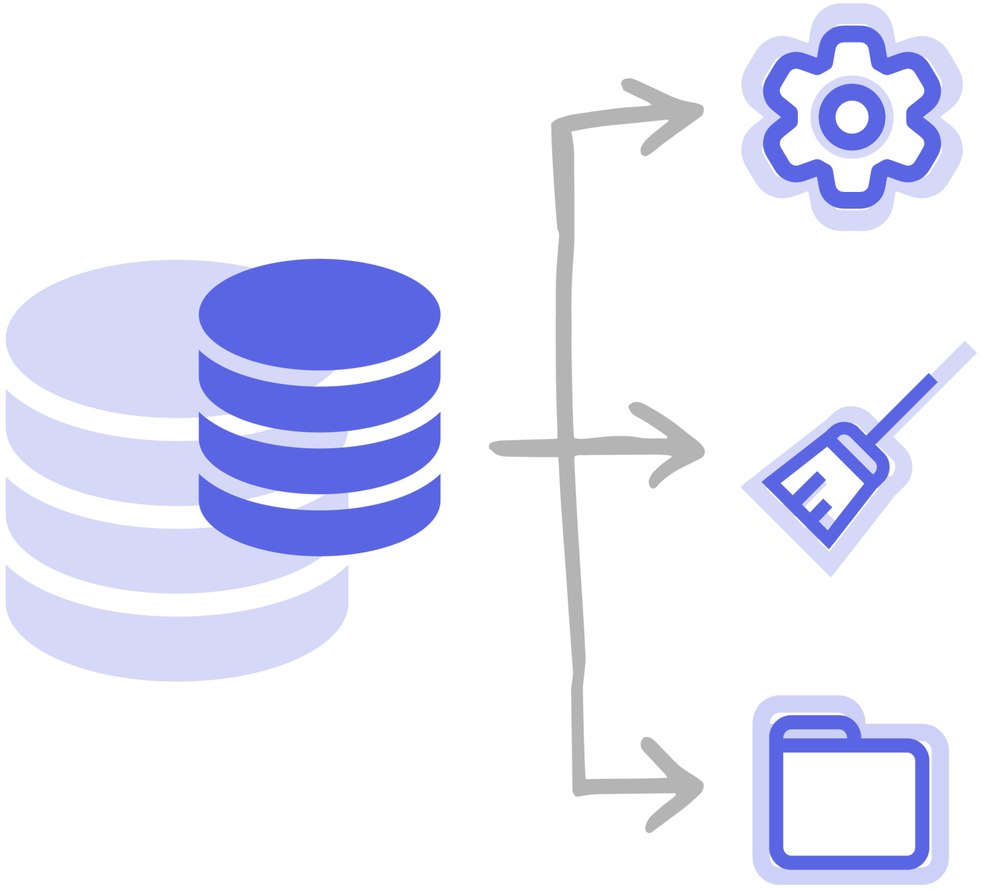
KUNSTHISTORISCHES INSTITUT IN FLORENZ
MAX-PLANCK-INSTITUT

Nov 25th 2025

DH Day

DH_Lab

OUR FOCUS



Organizing

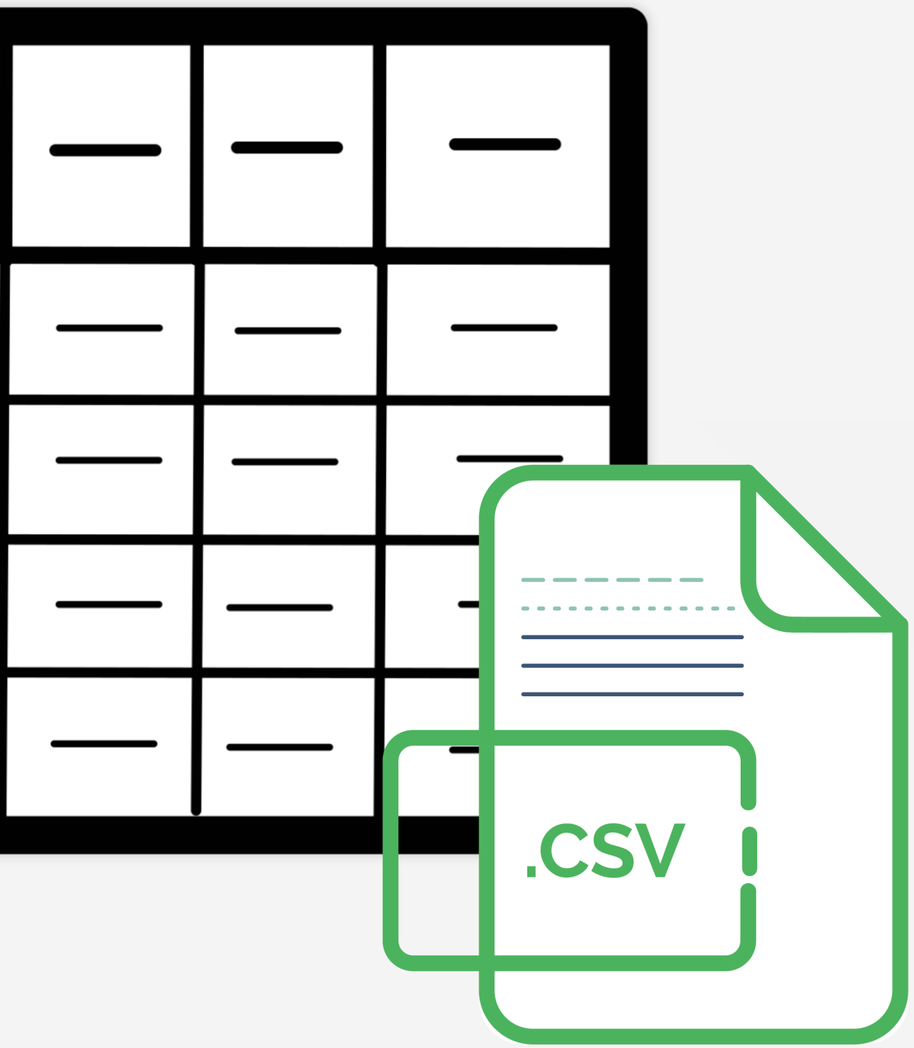
Structure at the beginning

Cleaning & transforming

Making data usable

Storing & preserving

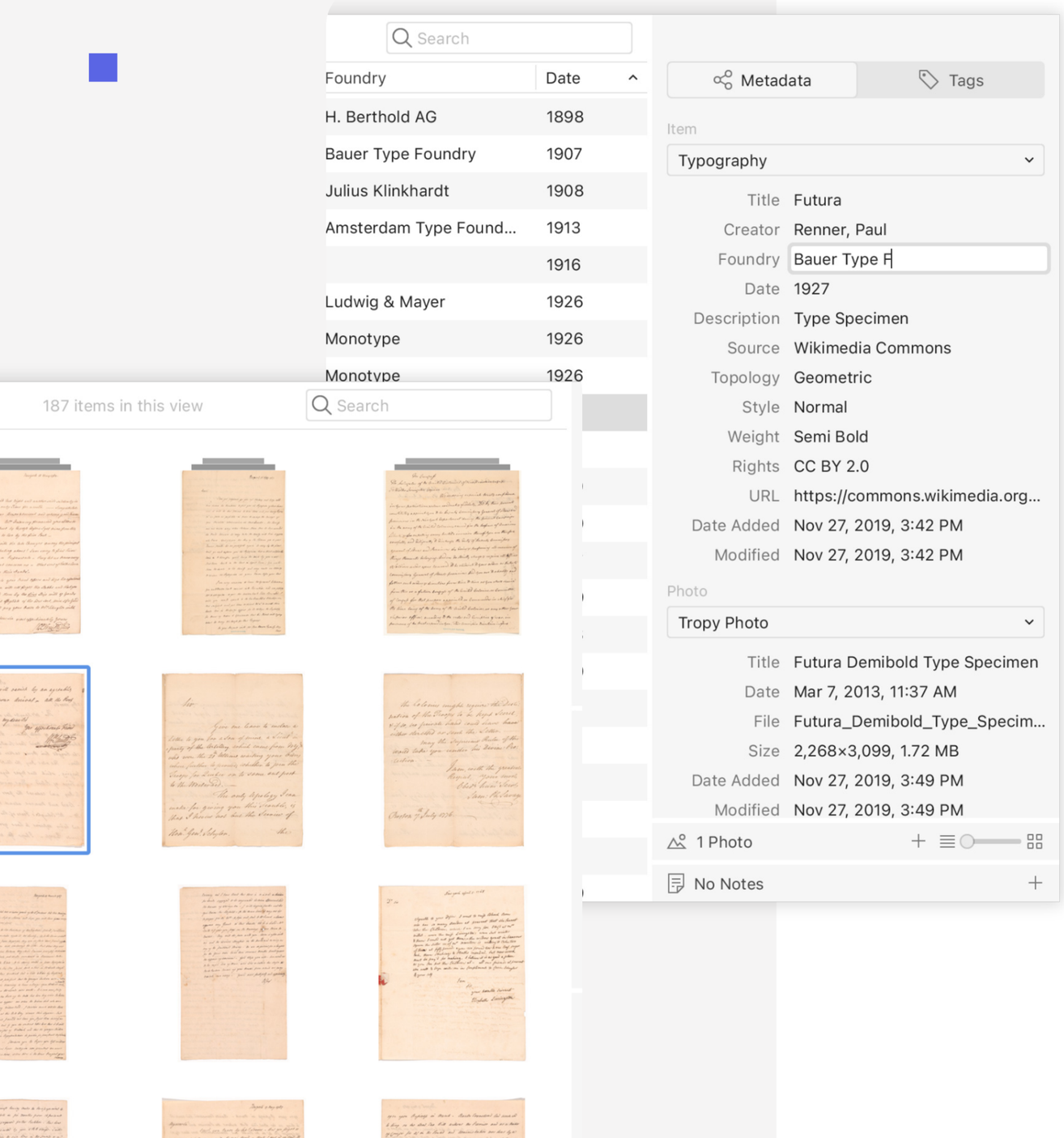
Ensuring long-term access



SPREADSHEET STRUCTURE BASICS

- One row = one item
- One column = one attribute
- No merged cells
- Avoid color-coding or formatting as “data”
- Use controlled vocabularies consistently (AAT, ULAN, TGN)
- Save as CSV for interoperability

```
Artist,Title,Date
Artist 1,Artwork A,1500
Artist 2,Artwork B,1620
```



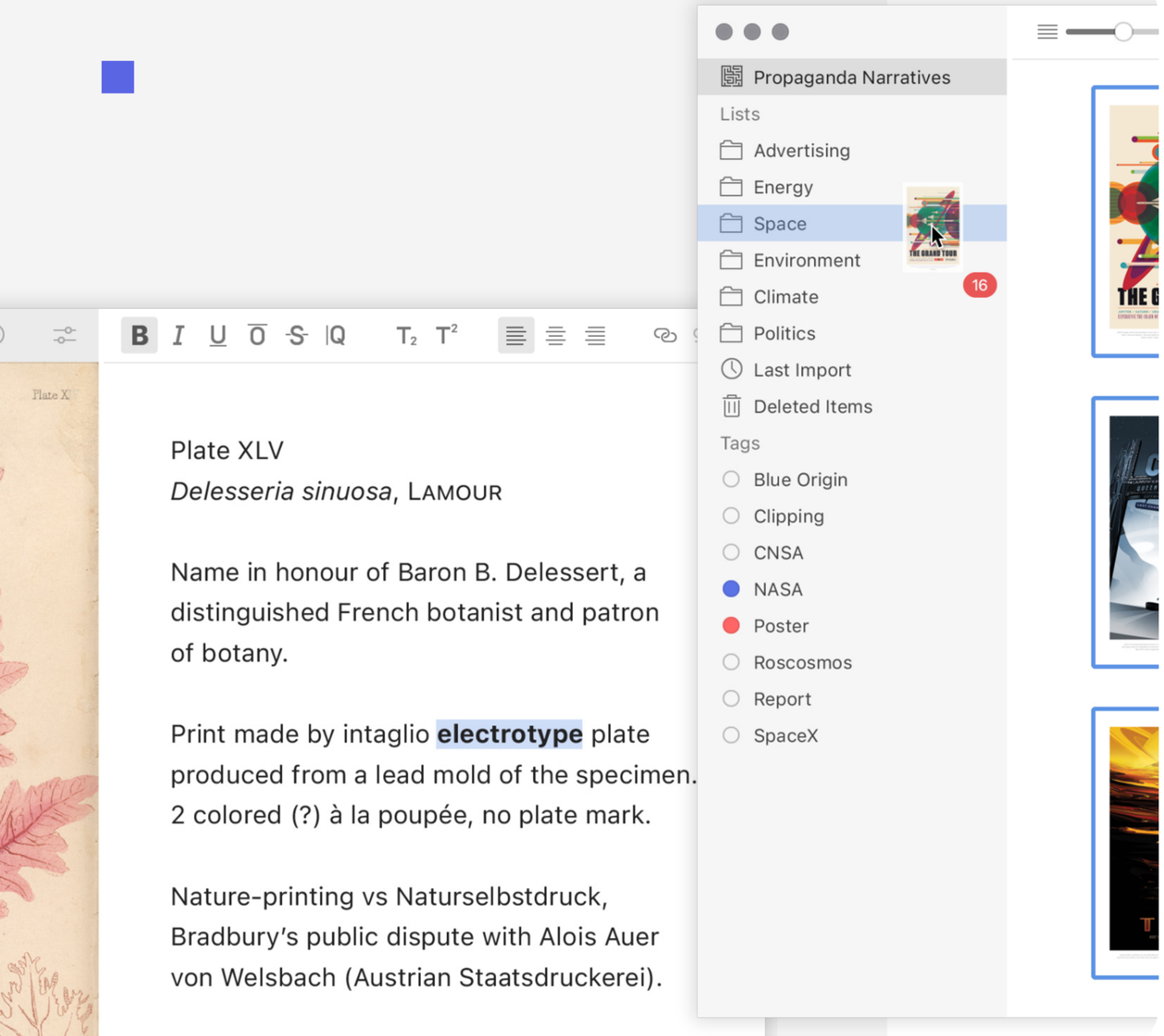
TROPY

HTTPS://TROPY.ORG

WHAT TROPY DOES:

- Organizes research images (artworks, archival materials)
- Adds structured metadata using customizable templates
- Supports tags, notes, and lists for better navigation
- Keeps images meaningful—no more folders full of unlabeled photos

ORGANIZING DATA



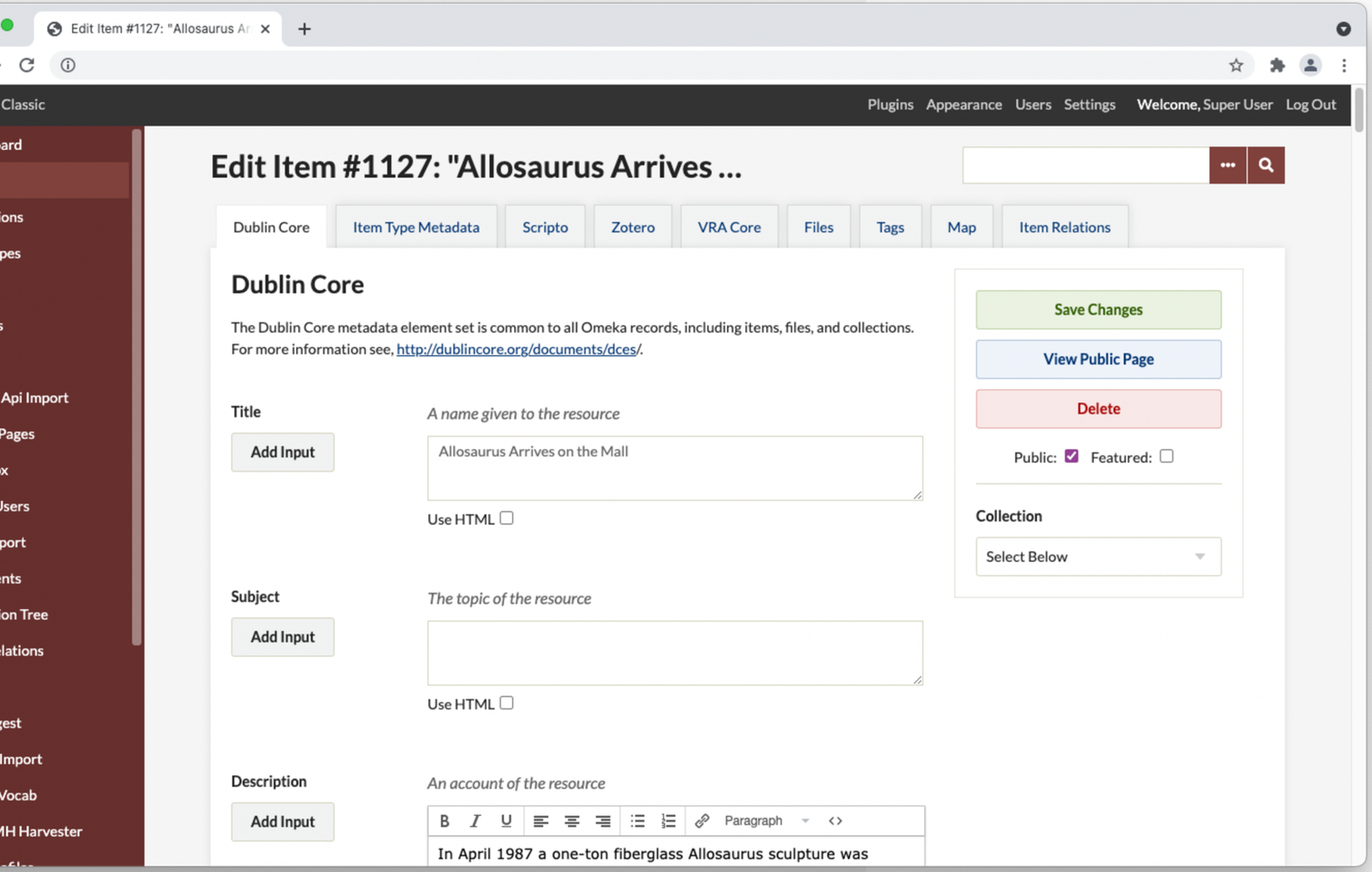
TROPY

HTTPS://TROPY.ORG

WHY IT'S USEFUL:

- Keeps images + metadata together
- Enables quick retrieval of content
- Exports structured metadata for use in spreadsheets or Omeka
- Integrates with Zotero for bibliographic links

ORGANIZING DATA



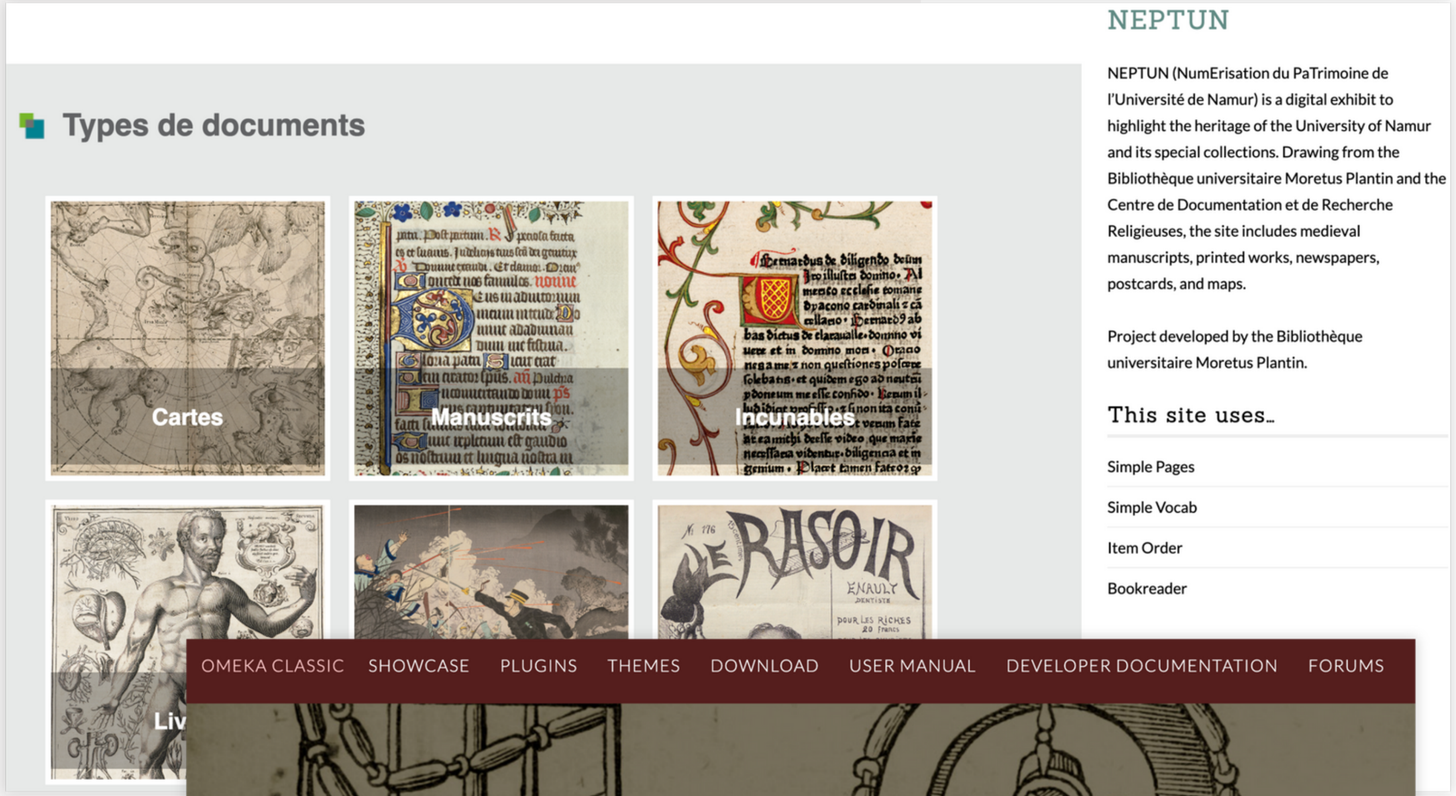
OMEKA

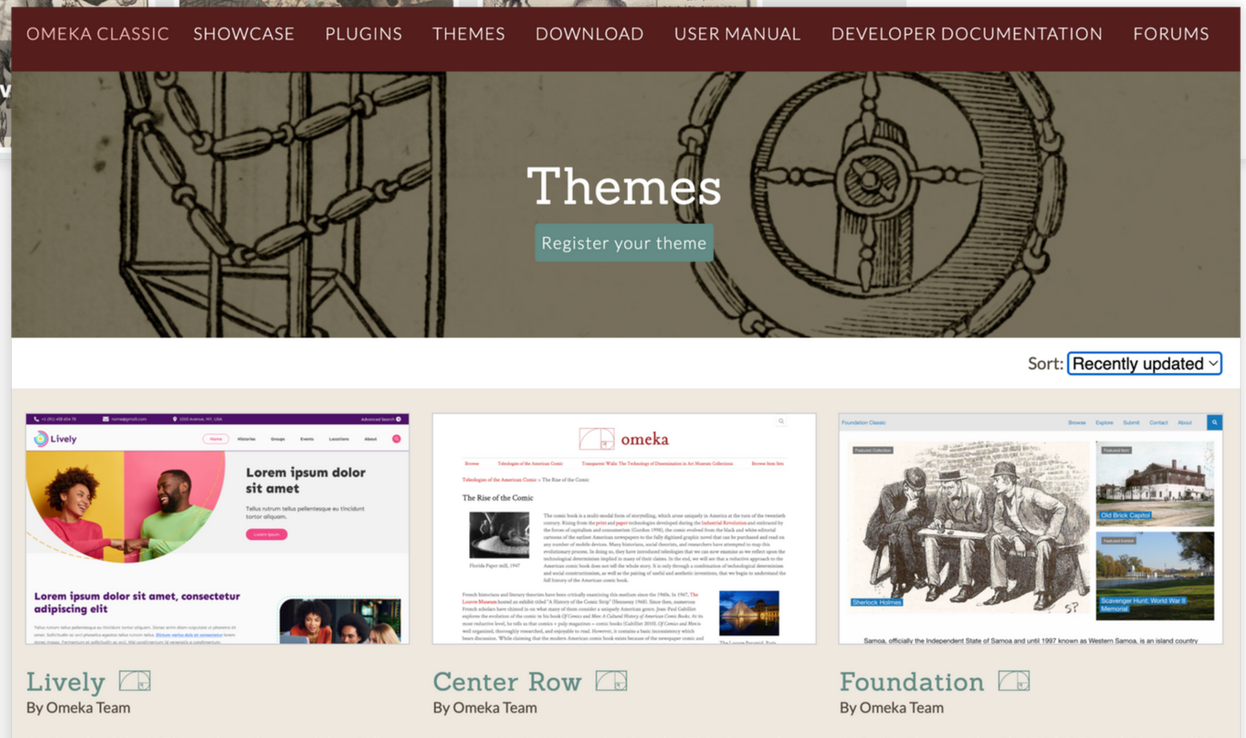
HTTPS://OMEKA.ORG/

WHAT OMEKA DOES:

- Creates structured collections
- Uses metadata standards (Dublin Core, custom vocabularies)
- Ideal for cataloguing and small digital collections
- Can serve as a project-level organizing platform

ORGANIZING DATA





OMEKA

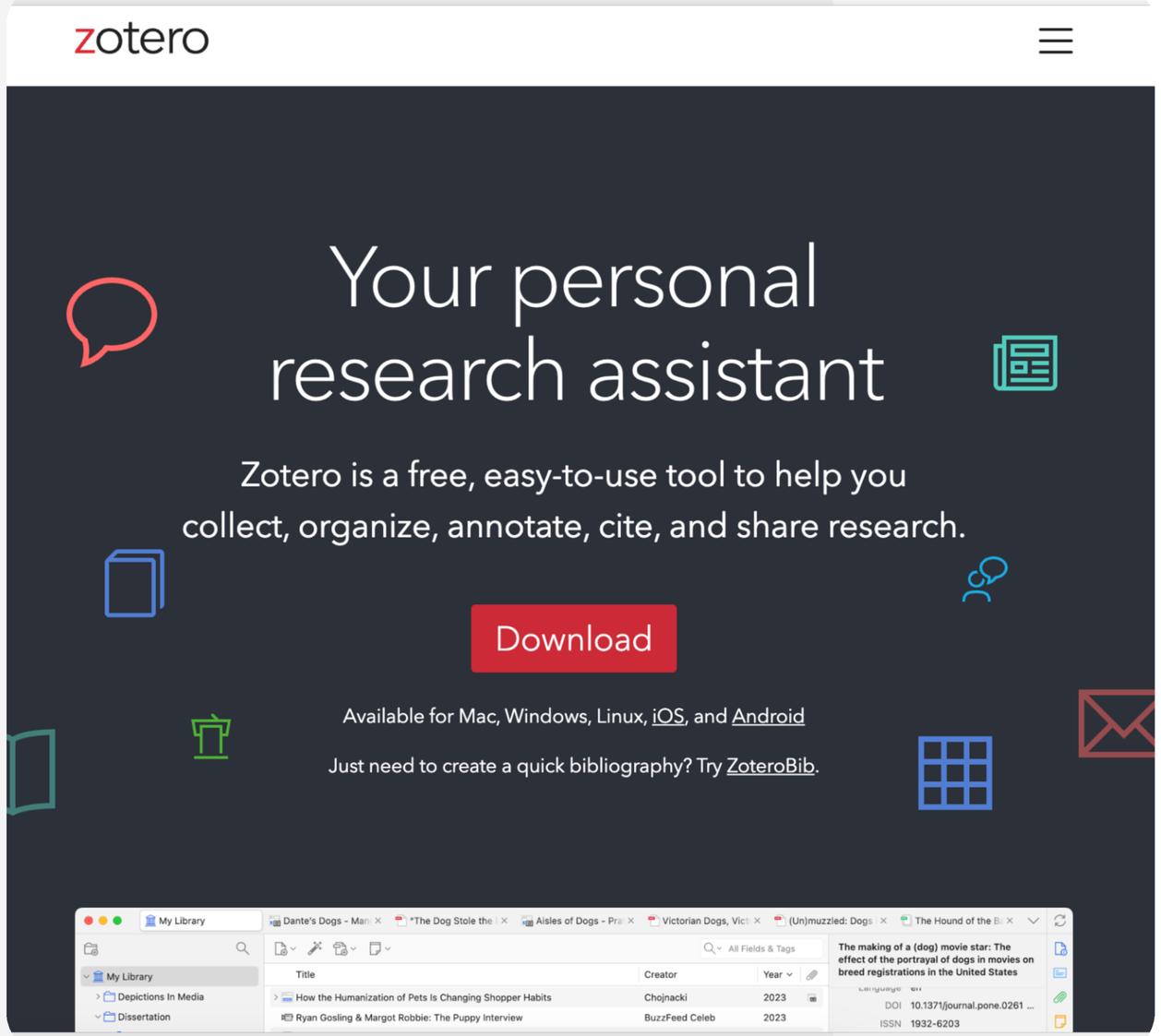
HTTPS://OMEKA.ORG/

WHY IT'S USEFUL:

- Presents data cleanly internally or publicly
- Supports multimedia objects
- Facilitates metadata consistency across a project
- Integrates with IIIF (high-resolution image viewers)

→ **https://omeka.org/classic/directory/**

ORGANIZING DATA



TOOLS FOR BIBLIOGRAPHIC DATA: ZOTERO

HTTPS://WWW.ZOTERO.ORG/

KEY FEATURES:

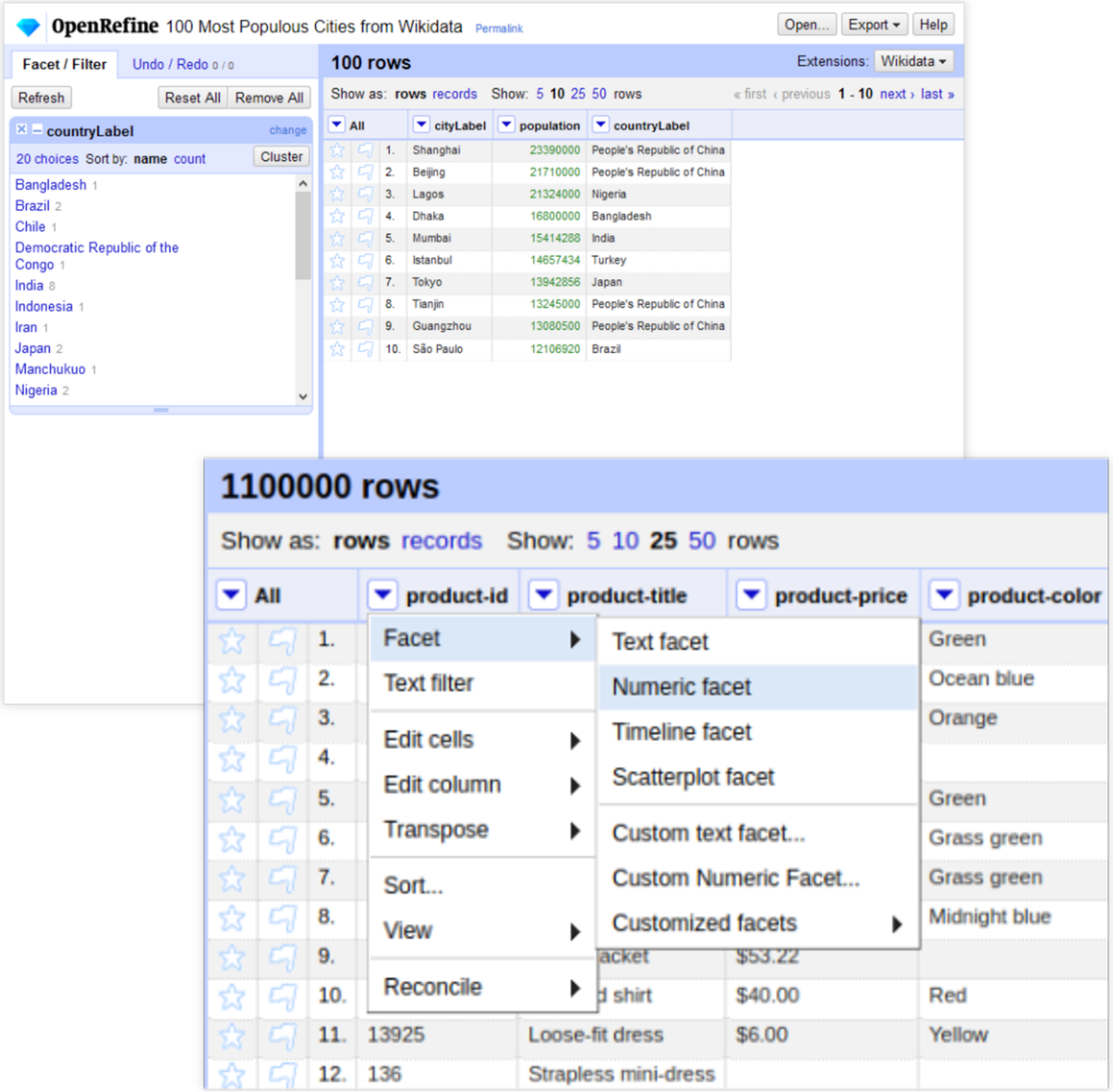
- Automatic metadata extraction from PDFs and catalogues
- Tags, collections, notes
- Group libraries for collaboration
- Better metadata quality

TOOLS FOR BIBLIOGRAPHIC DATA: MENDELEY

HTTPS://WWW.MENDELEY.COM/

KEY FEATURES:

- Good PDF management
- Annotation tools
- Cloud syncing
- Citation styles for publications

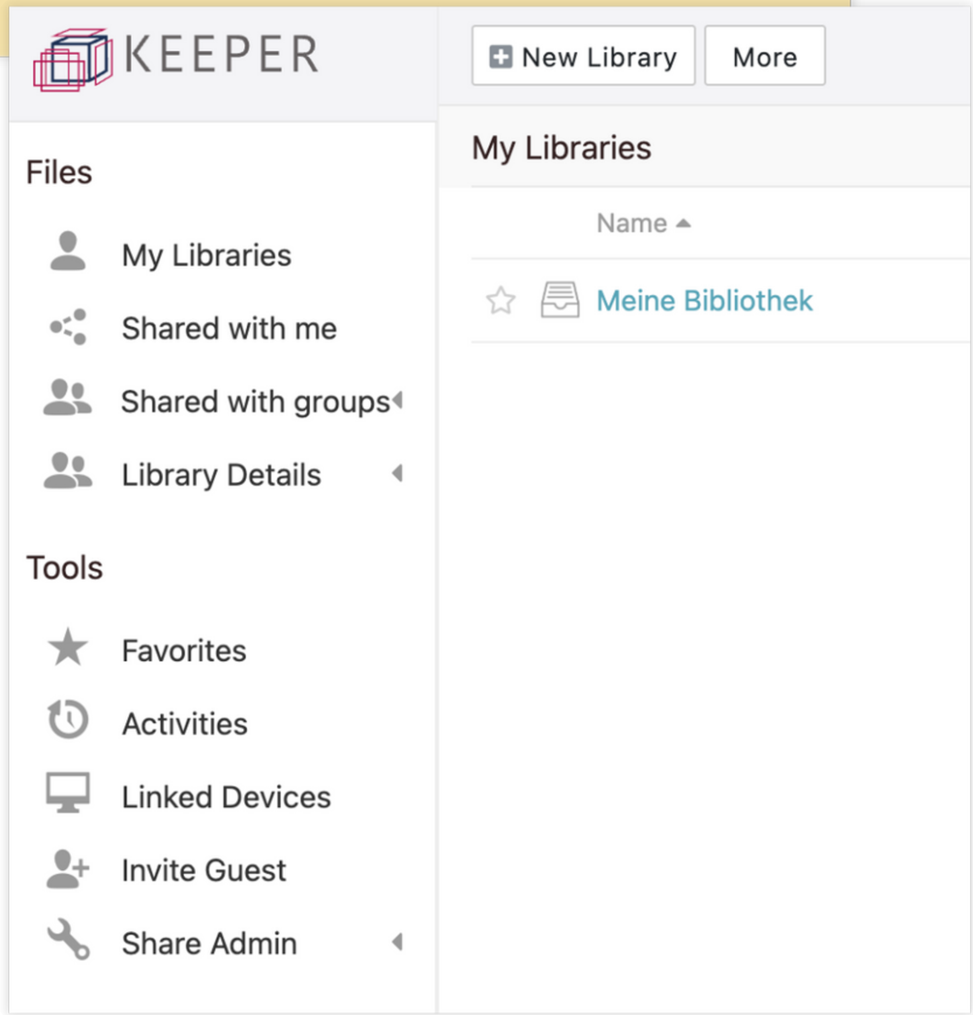


OPENREFINE

HTTPS://OPENREFINE.ORG/

KEY FEATURES

- Detects inconsistencies in names, dates, places
- Clusters similar values
- Splits or reshapes complex fields
- Tracks every change
- Exports clean CSV, JSON, TTL, and more



KEEPER

HTTPS://KEEPER.MPDL.MPG.DE/

KEY FEATURES

- Cloud storage with long-term archiving (10+ years)
- Organizes data into Libraries with optional metadata and DOIs
- Easy sharing with MPG researchers and external collaborators
- Supports large files and cloud synchronization
- Enables collaborative document authoring

PREFERRED FORMATS & DOCUMENTATION

USE OPEN, DURABLE FORMATS

- Tabular data: CSV, TSV
- Text: TXT, PDF/A
- Images: TIFF (master) + JPEG (derivatives)
- Structured data: XML, JSON

BEST PRACTICES FOR DOCUMENTATION

- Include a README describing the dataset
- Provide a data dictionary explaining each field
- Track version numbers (v1, v2…)
- Keep raw and cleaned copies separately

COMPACT WORKFLOW

Photograph materials → Tropy

Build dataset → Spreadsheet/CSV

Clean data → OpenRefine

(Optional) Organize for display → Omeka

Store and preserve → Keeper

Manage bibliography → Zotero / Mendeley

THANK YOU

Feel free to share any suggestions, questions, or additional tools.

Nov 25th 2025

DH Day

DH_Lab