STADT
GESCHICHTE
BASEL

A Long-Term Archival Pipeline for the Forschungsdatenplattform

omeka2dsp

Moritz Mähr
Moritz Twente

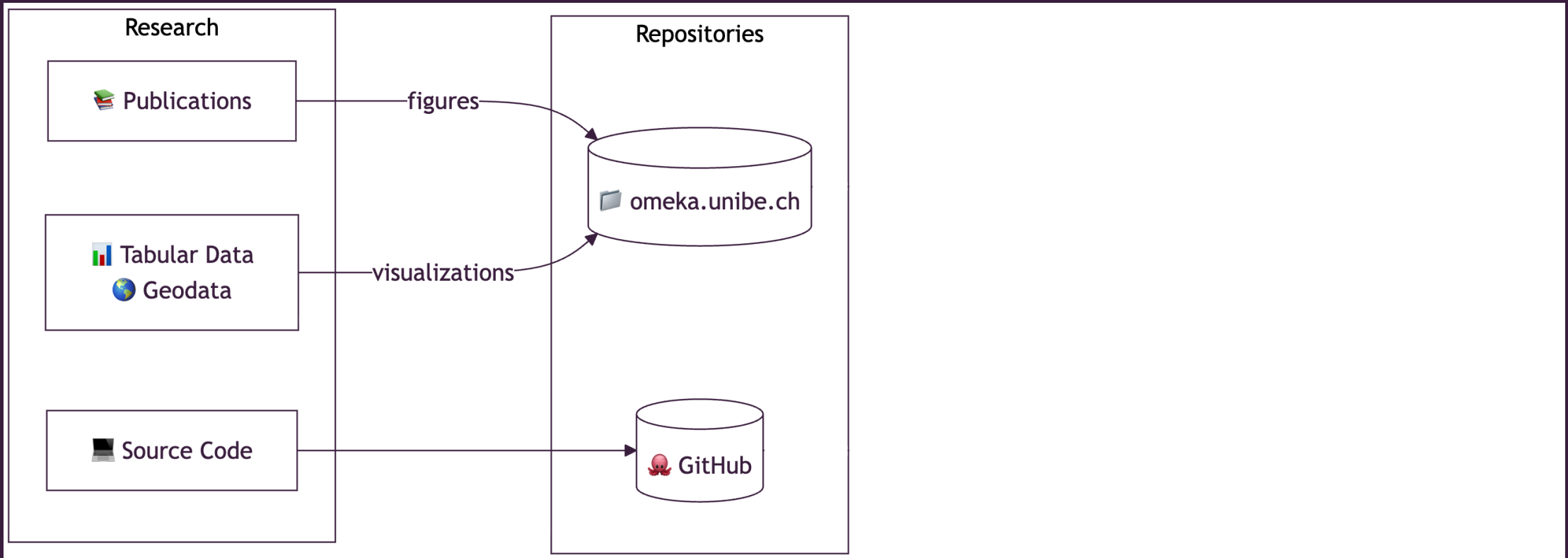
2025-10-15

Stadt.Geschichte.Basel

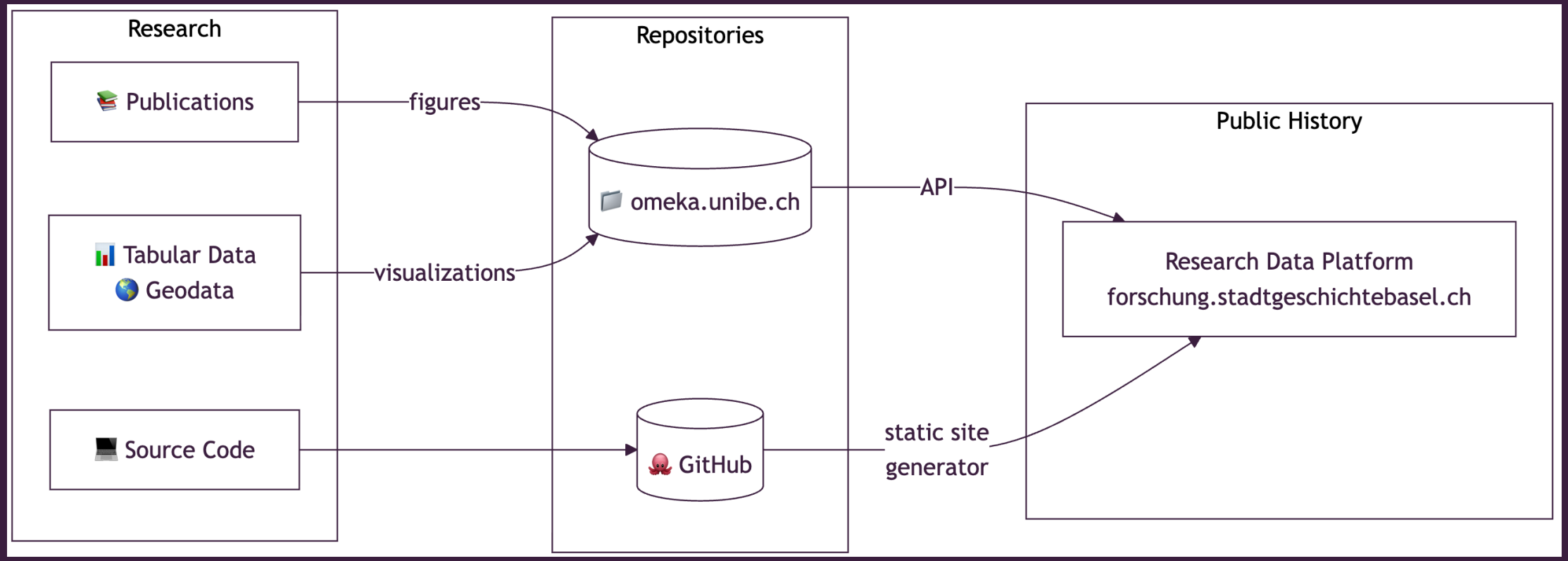
- Large-scale historical research project, initiated in 2011 by the Association for Basel History and carried out 2017–2026 at the University of Basel
- More than 70 researchers studying the history of Basel from the earliest settlements to the present day
- Funded with more than 9 million Swiss francs by the Canton of Basel-Stadt, the Lottery Fund, and private sponsors
- Specialized Team for Research Data Management and Public History
- Various research outputs (incl. books, papers, data stories, figures, and source code)

Research Data

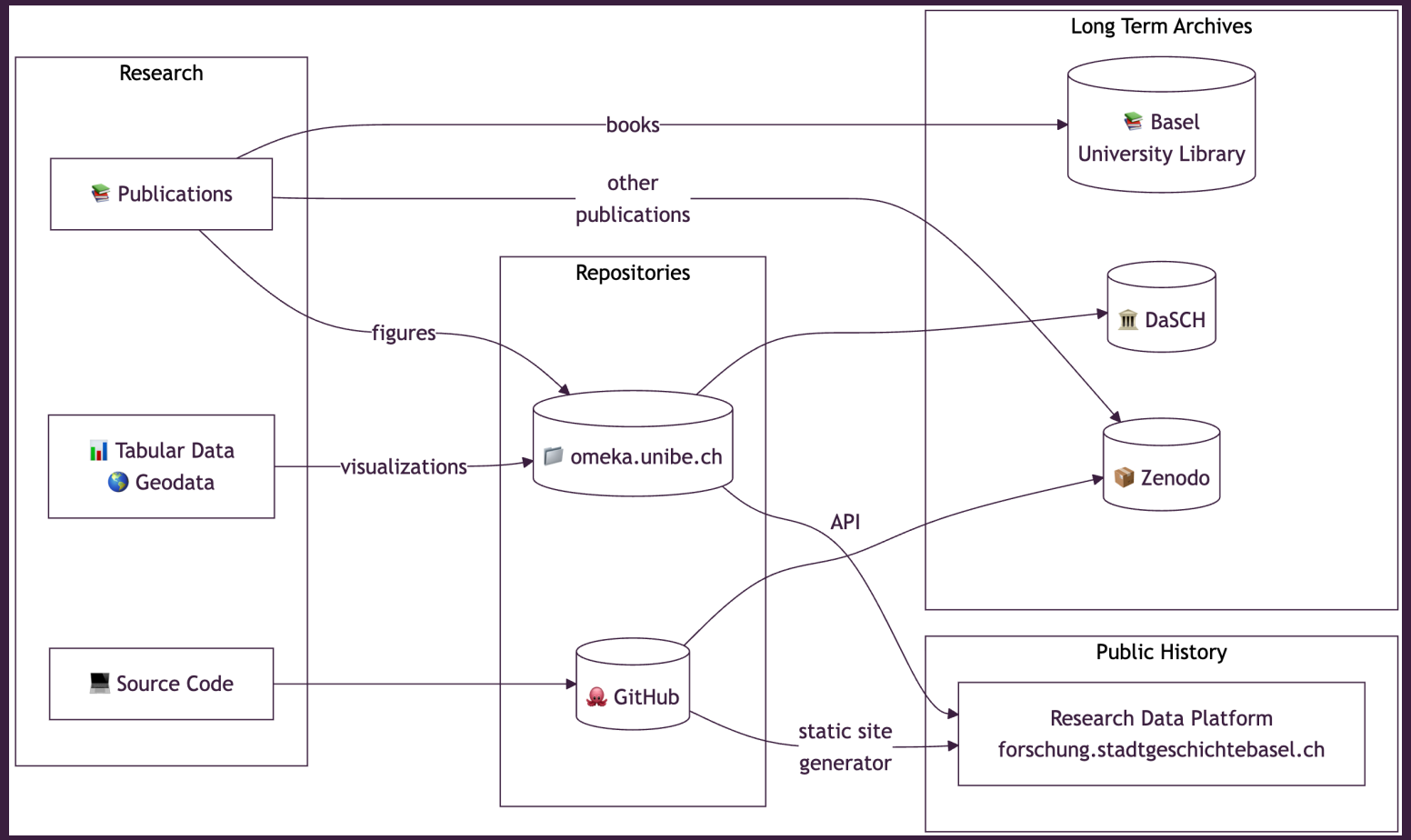
Collecting and Managing Research Data



Public History with Research Data



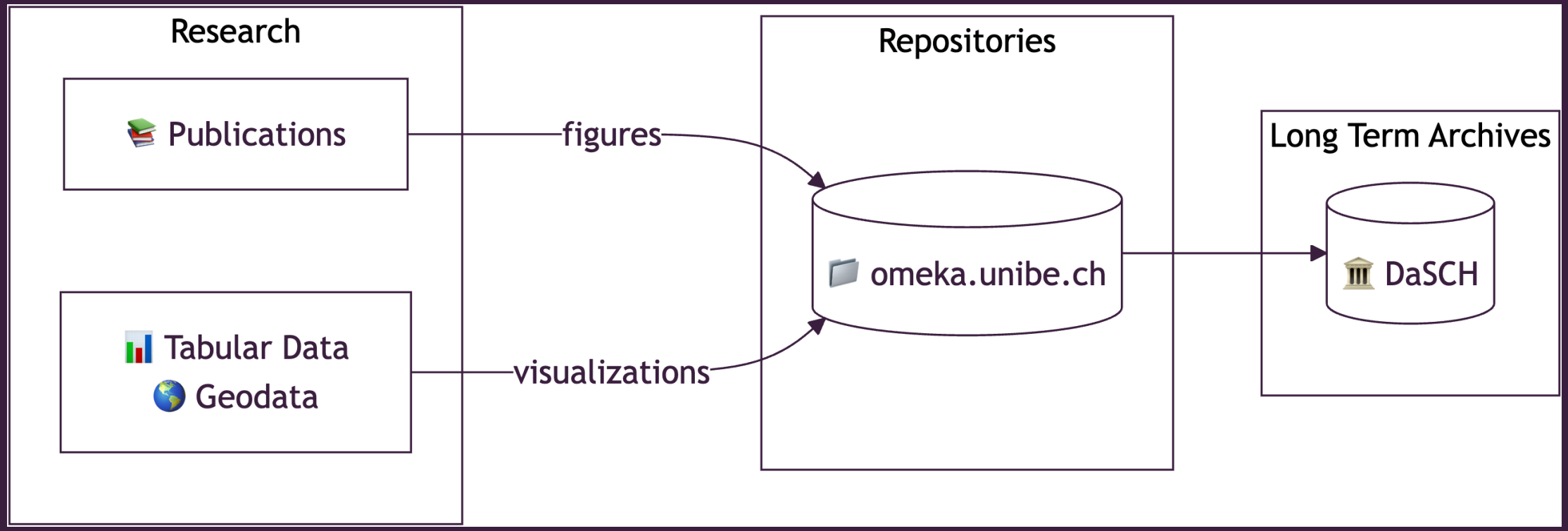
Archiving Research Data for the Long Run



Long-Term Preservation

- **Institutional sustainability:** External dependencies (Omeka at UniBern) may not be permanently funded
- **FAIR principles:** Research data must be Findable, Accessible, Interoperable, and Reusable with proper metadata
- **Citability:** Researchers need persistent identifiers (PIDs) (stable URLs or DOIs) to reference historical materials
- **Scalability:** Current minimal computing approach (GitHub Pages) cannot handle large files

Where DaSCH Fits in: omeka2dsp



Challenges

- Data model differences (Omeka vs DaSCH)
- Metadata transformation and crosswalks
- Automation of deposit and version control

STADT
GESCHICHTE
BASEL

Data Model Differences

STADT
GESCHICHTE
BASEL

Omeka

DH@UNIBE Omeka

SIGNED IN AS
Moritz Maehr

Logout

Search items

Find resources...

Sites

RESOURCES
Items
Media
Item sets
Vocabularies
Resource templates

ADMIN
Users
Modules
Jobs
Settings
Assets
Logs

MODULES
Search manager
CSV Import
Custom Vocab
Bulk Check
Easy Install
Export

Items

Add new item

Item set | Stadt.Geschichte.Basel objekt medium model

1 of 31 | 1–25 of 772

Advanced search

Created | Descending | Sort

Batch actions | Go

	Title	Class	Owner	Created
	Europäische Handelswege im Hoch- und Spätmittelalter		Moritz Twente	Oct 6, 2025
	Verbreitung von Basler Papier in Europa im 15. Jahrhundert		Moritz Twente	Oct 6, 2025
	Tuchreviere in Flandern und am Mittelrhein im späten 14. Jahrhundert		Moritz Twente	Oct 6, 2025
	Grenzposten in den Langen Erlen. Aus dem Album von Emil Seiler-La Roche		Noelle Schnegg	Oct 5, 2025
	Badende Kinder in der Wiese		Noelle Schnegg	Oct 5, 2025
	Die Walderholungsstätte in den Langen Erlen kurz nach der Eröffnung		Noelle Schnegg	Oct 5, 2025
	Postkarte mit den verschiedenen Attraktionen des Erlenparks		Noelle Schnegg	Oct 5, 2025
	Mercerie- und Kurzwarengeschäft der Familie Bornstein-Littner an der Feldbergstrasse 96		Noelle Schnegg	Oct 5, 2025
	Das Krematorium auf dem Horburg-Gottesacker		Noelle Schnegg	Oct 5, 2025
	Rückseite der Mehrfamilienhäuser an der Ackerstrasse im äussersten Klybeck		Noelle Schnegg	Oct 5, 2025
	Plan für neue Strassenanlagen, Kirchplatz und Gottesacker		Noelle Schnegg	Oct 5, 2025
	Wettbewerbsvorschlag für die Neugestaltung des Marktplatzes, Zeichnung von Karl Moser		Noelle Schnegg	Oct 5, 2025

DH@UNIBE Omeka

SIGNED IN AS
Moritz Maehr
Logout

Search items

Find resources...

Sites

RESOURCES
Items
Media
Item sets
Vocabularies
Resource templates

ADMIN
Users
Modules
Jobs
Settings
Assets
Logs

MODULES
Search manager
CSV Import
Custom Vocab
Bulk Check
Easy Install
Export

ITEMS **Europäische Handelswege im Hoch- und Sp...**

Public view | Edit item | < | ^ | >

Metadata | Linked resources

Identifier	abb38506
Title	Europäische Handelswege im Hoch- und Spätmittelalter
Description	Basel lag im europäischen Handel vor 1500 günstig im Schnittpunkt einer wichtigen Nord-Süd-Achse und der West-Ost-Achse (Kartengrundlage: Irsigler 2003, S. 233).
Era	de Mittelalter
ISO639	de
Is Part Of	Burkart, Lucas (Hg.): Stadt in Verhandlung. 1250–1530. Basel 2024 (Stadt.Geschichte.Basel).

ID
13183

Visibility
Public

Item sets
Stadt.Geschichte.Basel objekt medium model

Created
Oct 6, 2025

Owner
Moritz Twente

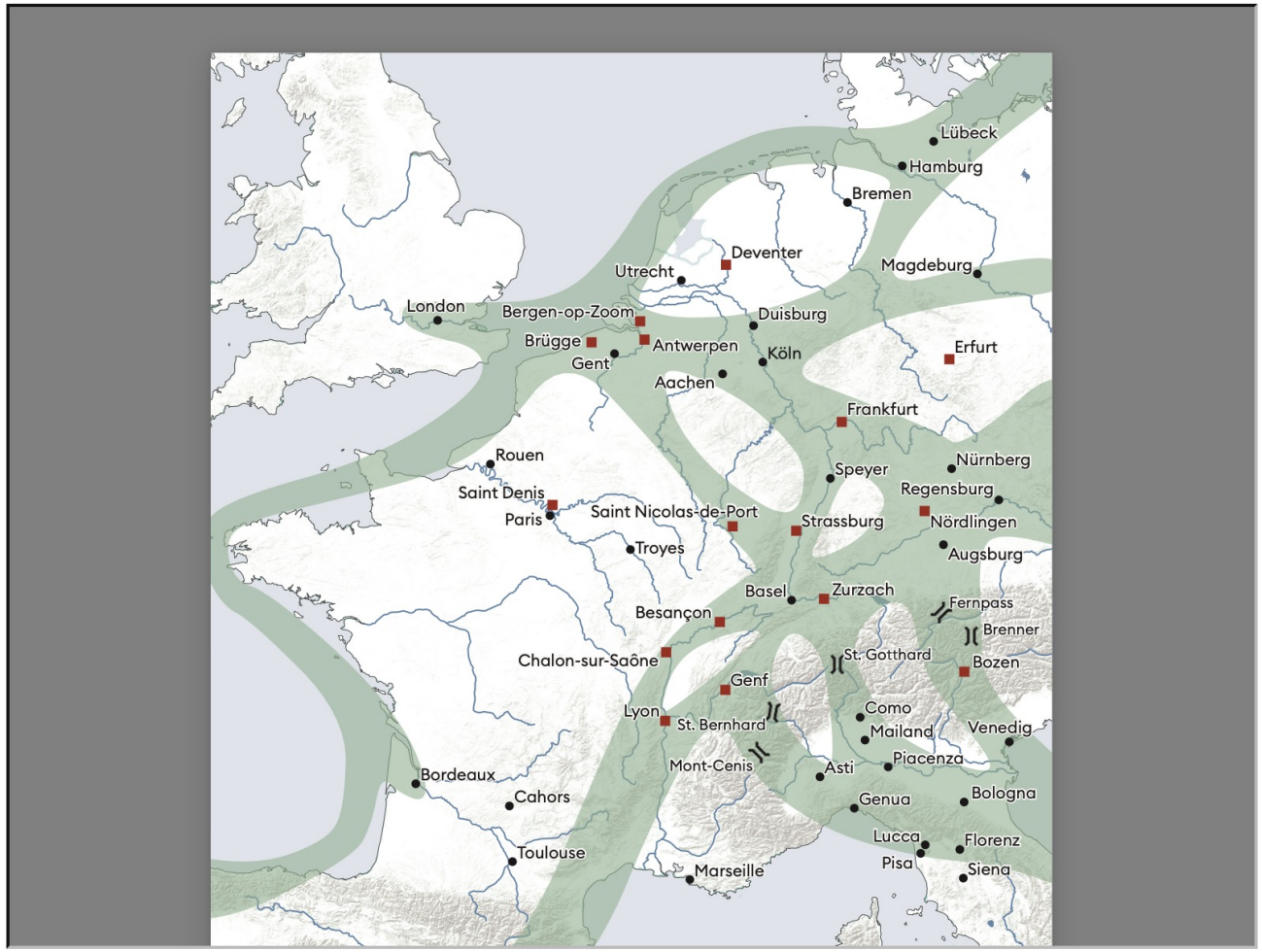
Media (4)


Europäische Handelswege im Hoch- und Spätmittelalter

Legende zur Karte 'Europäische Handelswege im Hoch- und Spätmittelalter'

MEDIA **Europäische Handelswege im Hoch- und Sp...**

Metadata



Class
Image

Identifier	m38506_1
Title	Europäische Handelswege im Hoch- und Spätmittelalter

ID
13184

Visibility
Public

Part of item
Europäische Handelswege im Hoch- und Spätmittelalter

Created
Oct 6, 2025

MIME type
application/pdf

Size
5,915,682 bytes

Ingester
Upload

Source
abb38506_hauptachsen_handel_AI.pdf

File Derivatives
original
large

MEDIA **Europäische Handelswege im Hoch- und Sp…**

Public view | Edit media | < | ^ | >

Class
Image

Identifier	m38506_1
Title	Europäische Handelswege im Hoch- und Spätmittelalter
Caption	Basel lag im europäischen Handel vor 1500 günstig im Schnittpunkt einer wichtigen Nord-Süd-Achse und der West-Ost-Achse (Kartengrundlage: Irsigler 2003, S. 233).
Creator	Stadt.Geschichte.Basel
Publisher	Stadt.Geschichte.Basel
EDTF	..14XX
Era	de Mittelalter
DCMI Type	Image
MIME Type	en application/pdf
Extent	116.06x140.41 mm
Source	Irsigler, Franz: Jahrmärkte und Messen im oberrheinischen Raum vom 14. bis 16. Jahrhundert, in: Krimm, Konrad; Brüning, Rainer (Hg.): Zwischen Habsburg und Burgund. Der Oberrhein als europäische Landschaft im 15. Jahrhundert, Ostfildern 2003, S. 233; Terrain: Esri Inc., USA; Bearbeitung: Nico Görlich / Moritz Twente
ISO639	de
Relation	Burkart, Lucas (Hg.): Stadt in Verhandlung. 1250–1530. Basel 2024 (Stadt.Geschichte.Basel).
Rights	Public Domain Mark
License	en https://creativecommons.org/publicdomain/mark/1.0/

ID
13184

Visibility
Public

Part of item
Europäische Handelswege im Hoch- und Spätmittelalter

Created
Oct 6, 2025

MIME type
application/pdf

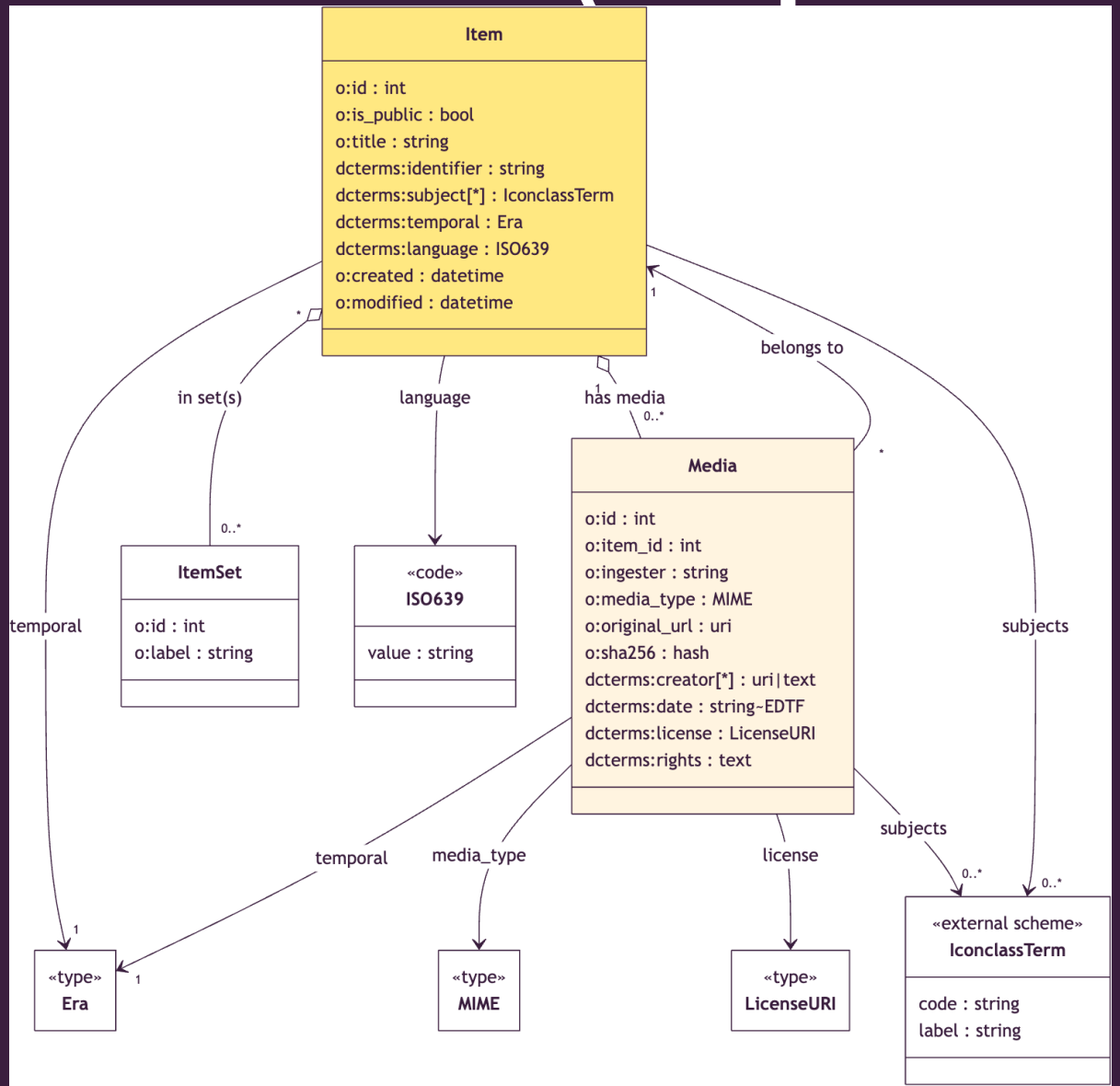
Size
5,915,682 bytes

Ingester
Upload

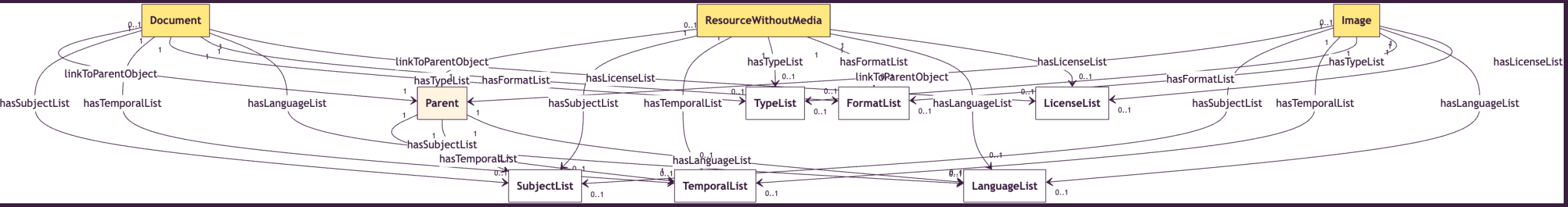
Source
abb38506_hauptachsen_handel_AI.pdf

File Derivatives
original
large

Data Model Omeka (simplified 😇)



Data Model DaSCH (simplified 😅)



Key Differences

- **Modeling approach:**
  - *DaSCH*: Class hierarchy (`Resource` → `Document` / `Image` …), explicit value classes (`TextValue`, `ListValue`).
  - *Omeka S*: Flat JSON-LD model (`Item`, `Media`, `ItemSet`), Dublin Core–centric.
- **Normalization & constraints:**
  - *DaSCH*: Strict cardinalities and mandatory fields (`hasTitle [1]`).
  - *Omeka S*: More flexible, “validation” through templates.
- **Hierarchy representation:**
  - *DaSCH*: Explicit `Parent` class and `linkToParentObject`.
  - *Omeka S*: Uses field `ItemSet` or `Media` to model relations.

STADT
GESCHICHTE
BASEL

Metadata Transformation

Data Validation

- Custom Python scripts using `pydantic` for schema validation as Omeka does not enforce strict validation
- Validation also helps identify data quality issues
- Lots of manual cleaning required, all done in Omeka

Crosswalk Examples

Omeka Property	DaSCH Property	Notes
dcterms:title	hasTitle	Required in DaSCH (cardinality 1)
dcterms:identifier	hasIdentifier	Not by DSP used for stable references
dcterms:subject	hasSubjectList	Iconclass codes preserved
dcterms:temporal	hasTemporalList	Era mapping required
dcterms:creator	hasCreatorList	Multiple creators supported
dcterms:license	hasLicenseList	License URIs validated
Media → Item	linkToParentObject	Hierarchy explicitly modeled

Version Control and Updates

- **Challenge:** DaSCH supports versioning, but requires careful planning
- **Strategy:**
  - Initial deposit: Create new resources via REST API
  - Updates: Use PUT requests with resource IDs to create new versions
  - Identifier stability: ARK IDs remain constant across versions

API vs DSP-Tools

REST API Approach

- Direct HTTP requests
- Fine-grained control
- Supports versioning
- Complex error handling
- Used for updates

DSP-Tools Approach

- XML-based bulk import
- Good for initial deposits
- Less flexible for updates
- Comprehensive validation
- Used for large-scale ingestion

STADT
GESCHICHTE
BASEL

omeka2dsp

omeka2dsp

Long-Term Archival Pipeline for Stadt.Geschichte.Basel

MODIFIED
September 22, 2025

This repository contains the pipeline and data model for the long-term preservation of the research data of Stadt.Geschichte.Basel (SGB) on the DaSCH Service Platform (DSP).
It enables the transfer of metadata and media files from the SGB Omeka S instance to the DSP. The pipeline detects changes, updates existing records, and ensures reproducible and open research.

📚 Documentation

Comprehensive documentation is available in the `docs/` directory:

- 📖 Complete Documentation – Full system documentation
- 🏗️ Architecture Overview – System design and components
- 🔄 Workflows – Data migration workflows with Mermaid diagrams
- 🔧 API Reference – Python function documentation
- 🧩 Data Model – Data model documentation

📒 Quick Start Guides

- ⚡ Installation & Setup
- ⚙️ Configuration
- 📋 Usage
- 🛠️ Development
- 🔍 Troubleshooting

Edit this page
Report an issue

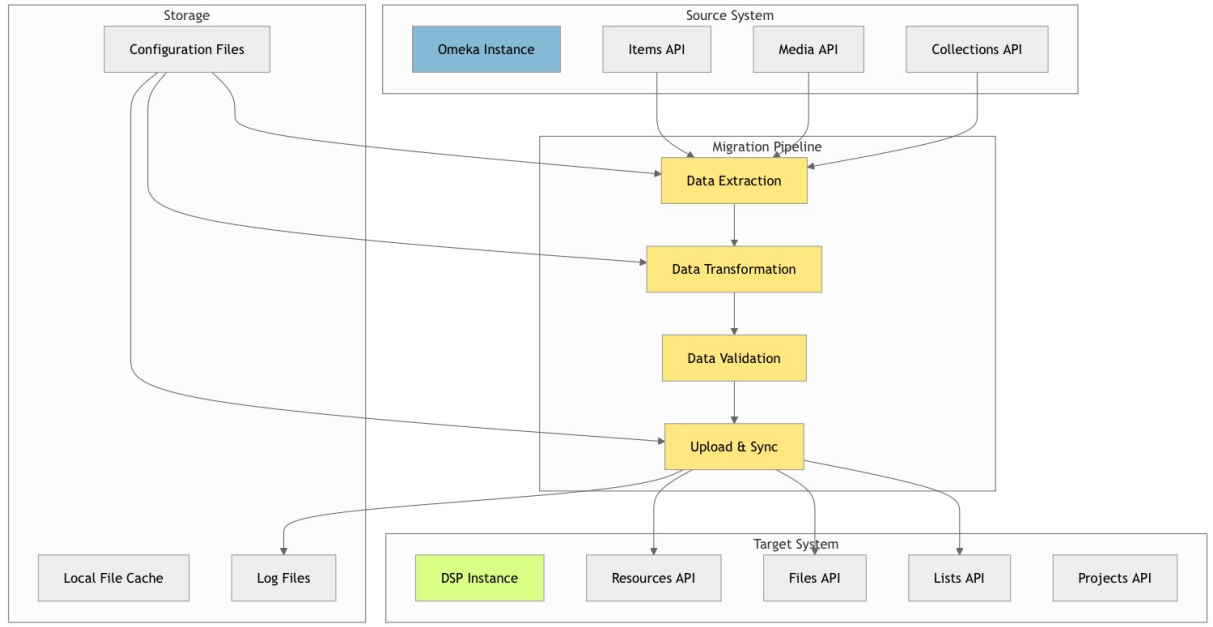
System Architecture

MODIFIED
September 22, 2025

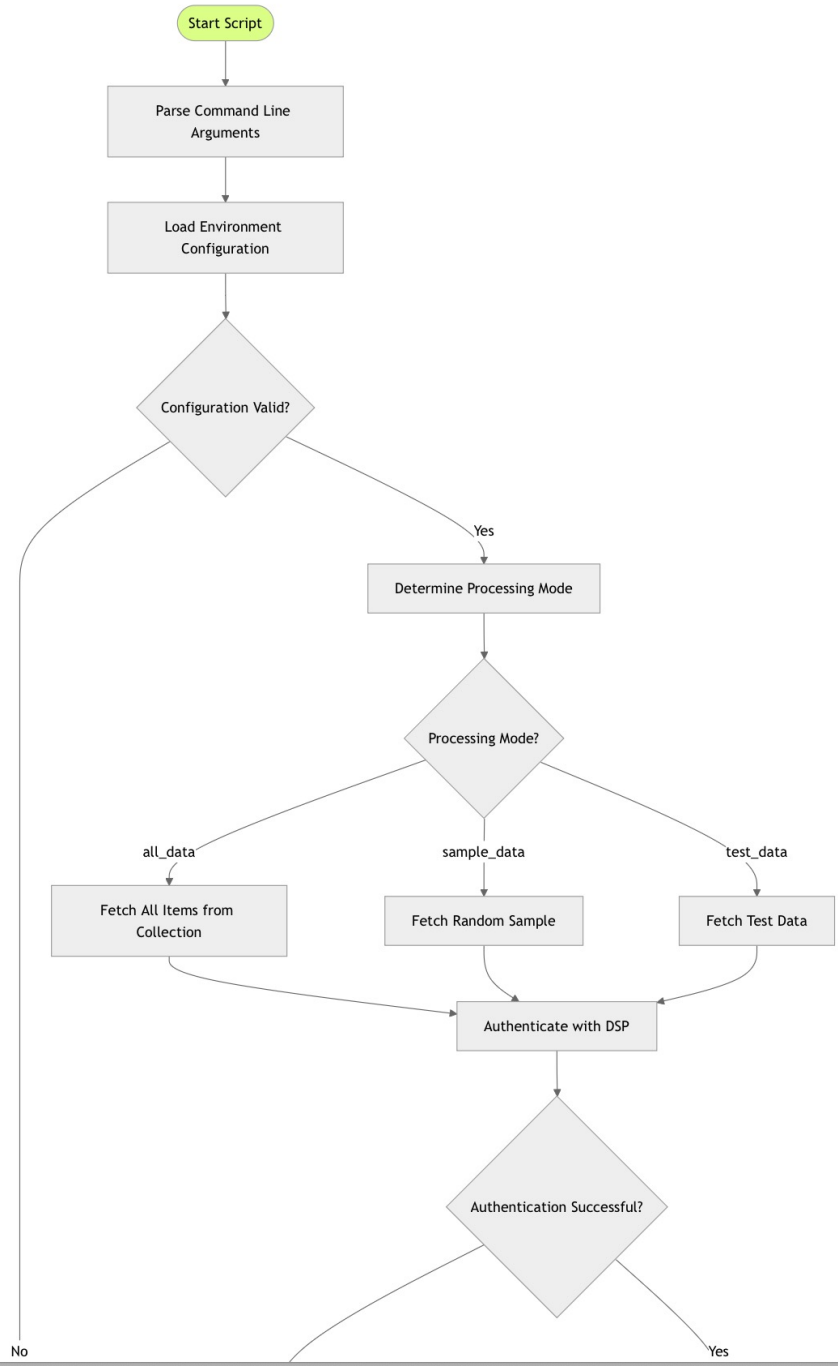
Overview

The omeka2dsp system is designed as a data migration and synchronization pipeline that transfers cultural heritage data from Omeka (a digital collections platform) to the DaSCH Service Platform (DSP) for long-term preservation.

High-Level Architecture



Edit this page
Report an issue



Edit this page
Report an issue

Lessons Learned

- **Metadata crosswalks and data quality**: Mapping Dublin Core to the DaSCH ontology required custom logic and validation. Omeka's flexible schema led to inconsistent metadata, demanding extensive cleaning.
- **Identifiers and file handling**: Synchronizing identifiers between Omeka and DaSCH proved complex. Large files (>100 MB) necessitated chunked uploads and tailored handling.
- **Workflow timing and validation**: Determining the right moment to shift from active curation to archival mode was key. Early validation with `pydantic` schemas and subset testing prevented costly errors.
- **Documentation and reproducibility**: Precise mapping documentation ensured consistency across transformations and supported reproducible workflows.
- **Architecture and infrastructure**: Decoupling archival (DaSCH) from presentation (CollectionBuilder) enhanced flexibility. Lightweight public interfaces can coexist with robust preservation systems when APIs enable automation.

Key Takeaways

For the Community

- Lightweight publishing can work with robust infrastructure
- FAIR principles are achievable in practice
- Decoupled architectures provide flexibility
- Open-source tools enable customization

For DaSCH Users

- Plan metadata transformation early
- Use validation extensively
- Consider hybrid API/DSP-Tools approach
- Document mapping decisions thoroughly

Resources

- Research Data Platform
- omeka2dsp Documentation
- Documentation Platform