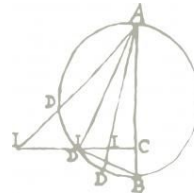


# Digital Research Infrastructure for Digital Humanities

## Building a Central Knowledge Graph for Research Projects

Robert Casties, Steffen Hennicke, Kim Pham  
2022 ELAG

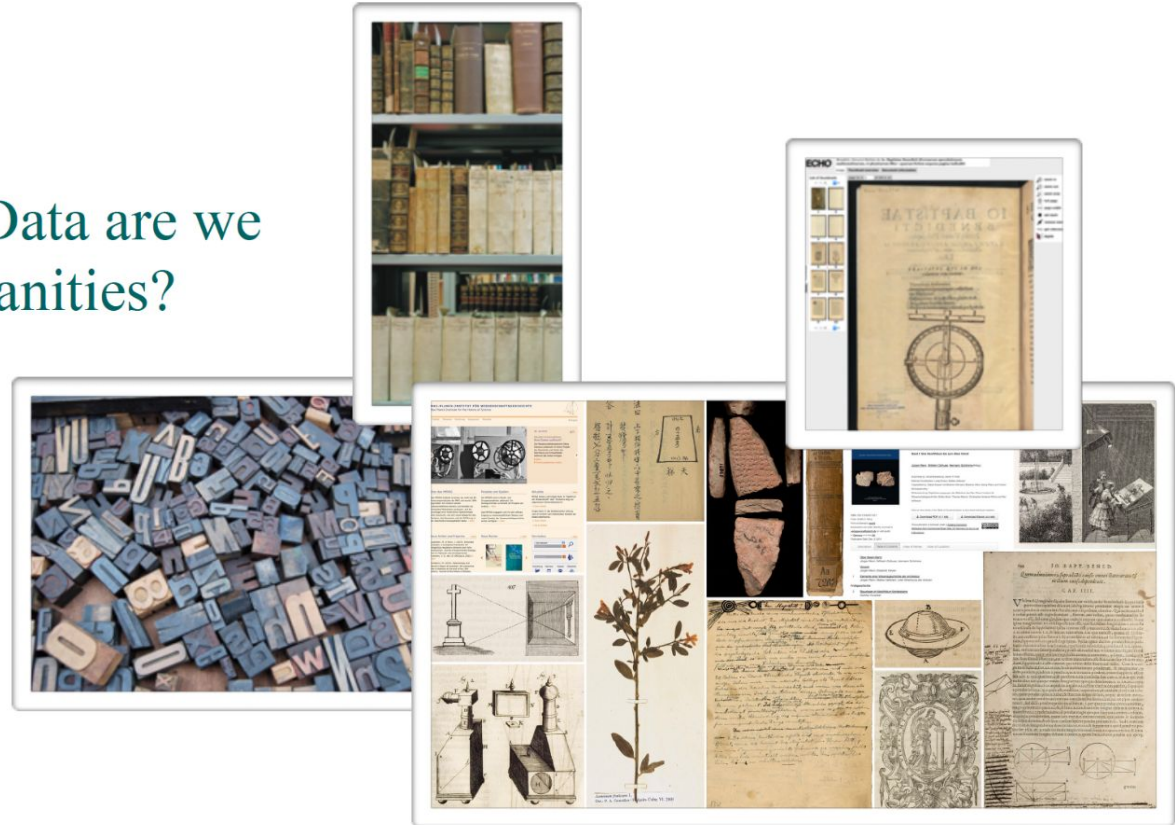


MPIWG  
MAX PLANCK INSTITUTE  
FOR THE HISTORY OF SCIENCE

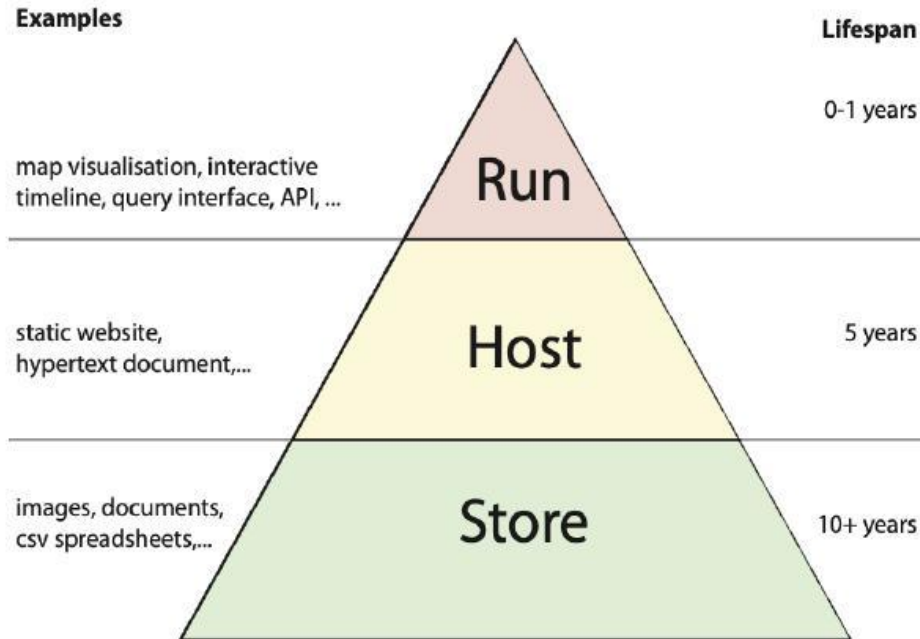
# Max Planck Institute for the History of Science



# What kind of Research Data are we dealing with in the Humanities?



# Digital Preservation



The screenshot shows a web browser window with the URL `db.sphaera.mpgw-berlin.mpg.de/resource/Start`. The page has a dark green header with the text "Sphaera CorpusTracer" on the left and "Login" on the right. Below the header is a navigation menu with links for "Books", "Persons", "Places", "Terms", and "Search". A search bar is located below the menu with the placeholder text "type to search, minimum 3 symbols ...".

The main content area is divided into two columns: "Books" and "Persons".

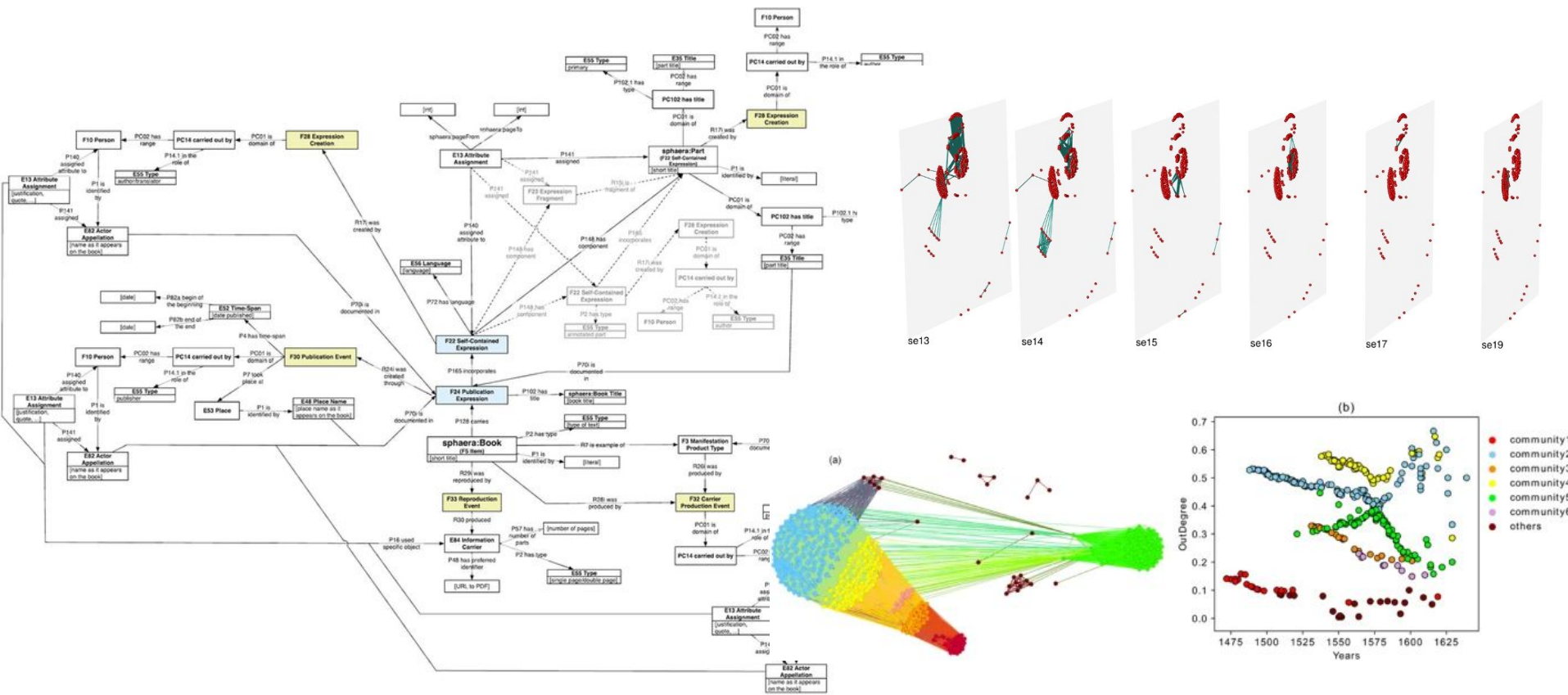
**Books section:**

- Text: "There are currently 320 books in the database."
- Button: "List all books"
- Section: "Recently edited"
- Book entry:
  - Title: "El filosofo. Ocupacion de nobles, y discretos contra la cortesana ociosidad. Sobre los libros de cielo, y Mundo, Meteoros, Parvos Naturales, Ethica, Economica, Politica de Aristoteles, y Esfera de Sacro Bosco. Epitome claro, y curioso. Tratanse estas materias con rigor escolastico: y dividense en dos libros, Filosofo Natural, y Filosofo Moral. Por el licenciado Cosme Gomez Texada de los Reyes. Capellan mayor de las Bernardas Descalças, y Patronazgo en S. Iñefonso de Talavera. Al licenciado Frey Martin Rodriguez de Corrales, del Abito de San Iuan, Prior de San Christoval de Salamanca, y Vicario General de dicha Sacra Religion, etc."
  - Author: "Johannes de Sacrobosco, Aristotle, Cosme Gómez Tejada de los Reyes"
  - Publisher: "Santiago Martin Redondo"
  - Date of publication: "1650"
  - Button: "View"
- Partial entry below:
  - Title: "The Art of Navigation. First written in the Spanish Tongue by"

**Persons section:**


- Text: "There are currently 342 persons in the database."
- Button: "List all persons"
- Section: "Recently edited"
- Person entry:
  - Name: "Philipp Melanchthon"
  - Dates: "1497 - 1560"
  - Image: A portrait of Philipp Melanchthon, a German humanist, theologian, and reformer, wearing a black cap and a dark robe.

# Data modelling with linked data and CIDOC CRM



1

Digitization and Ingest System  
METSMODS xml + images



2

Library frontend



4



Digital repository  
custom DRI app







5

Central knowledge graph

6

Knowledge Graph frontend



3

Storage and backup



CDSTAR  
CWDC common data storage architecture

Glue services - components connected by additional services

- Rabbit MQ "message broker" event communication
- Docker - environment setup
- Gitlab - Configuration and deployment
- Keycloak - security and authorization

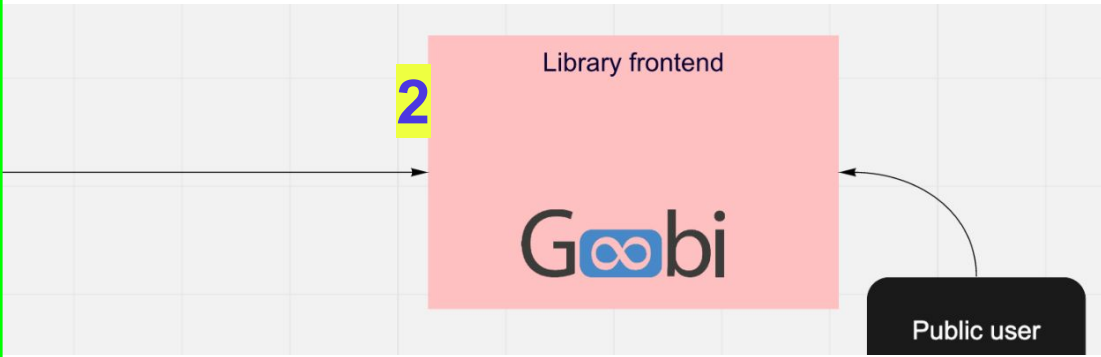





7

Public user

Architecture of  
DRIH  
(Digital  
Research  
Infrasatrucre  
for the  
Humanities)



Ingest Platform  
Goobi UI



1

Digitization and Ingest System  
METSMODS xml + images

Goobi

2

Library frontend

Goobi

Public user

DIGITAL LIBRARIES CONNECTED  
 MAX-PLANCK-INSTITUT FÜR WISSENSCHAFTSGESCHICHTE  
 English Search   
 Login

DLC | COLLECTIONS MPIWG | BROWSE | SEARCH

Image view  
 Contents  
 Thumbnail gallery  
 Bibliographic data  
 Full text

Search in:

Search results  
 1 / 71

Is part of  
 Collection:  
 Rara-Sammlung des MPIWG

Cite and reuse

Entire work

Downloads  
 Here you can download the full text of the work as a TEI file

Volumen secundum. Opera geometrica. Opera astronomica. Varia de optica. (2)

CHRISTIANI HUGENII  
 A ZULICHEM,  
*Dum viveret Zelæmi Toparchæ,*  
 OPERA VARIA.  
 VOLUMEN SECUNDUM.

Rare books knowledge graph Digital Libraries  
Connected <https://dlc.mpg.de/browse/> UI

1

Digitization and Ingest System  
METSMODS xml + images

Goobi

4

Digital repository

custom DRI app

Solr

X3ML Engine

iiif

ePIC  
Persistent Identifiers for eResearch

Fedora™

3

Storage and backup

cdstar  
GWDG common  
data storage architecture

1

Digitization and Ingest System  
METSMODS xml + images

Goobi

4

Digital repository

custom DRI app

Solr

X3ML Engine

ePIC  
Persistent Identifiers for eResearch

iiif

Fedora™

3

Storage and backup

cdstar  
GWDG common  
data storage architecture

## Storage Service CDSTAR UI

cdstar About Discover Repositories Documentation Server: dri ▾

### Archive 557b71bfd46 in dri

dri / vault: dri / archive: 557b71bfd46

ID	dri / 557b71bfd46
Owner	mpiwg-dri-epic@GWDG
Created	6 months ago
Modified	6 months ago
Size	1200 files
Profile	default

[Download as Zip](#)

**Permissions**  
Sowner: OWNER

**Attributes**  
No attributes yet. [+ Add first attribute](#)

**Files**

File Name	Type	Size
2430/derivate/0001.jpg	image/jpeg	2M <a href="#">↓</a>
2430/derivate/0002.jpg	image/jpeg	1M <a href="#">↓</a>
2430/derivate/0003.jpg	image/jpeg	2M <a href="#">↓</a>


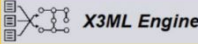



1

Digitization and Ingest System  
METSMODS xml + images



4

Digital repository  
custom DRI app

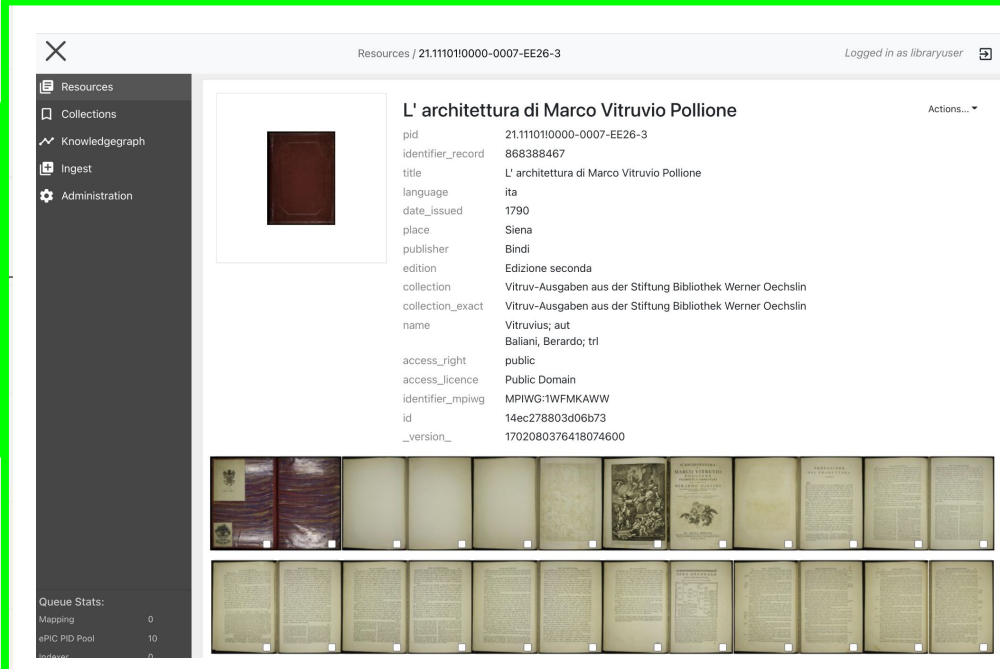






3

Storage and backup



GWDC common data storage architecture



Resources / 21.1110110000-0007-EE26-3 Logged in as libraryuser

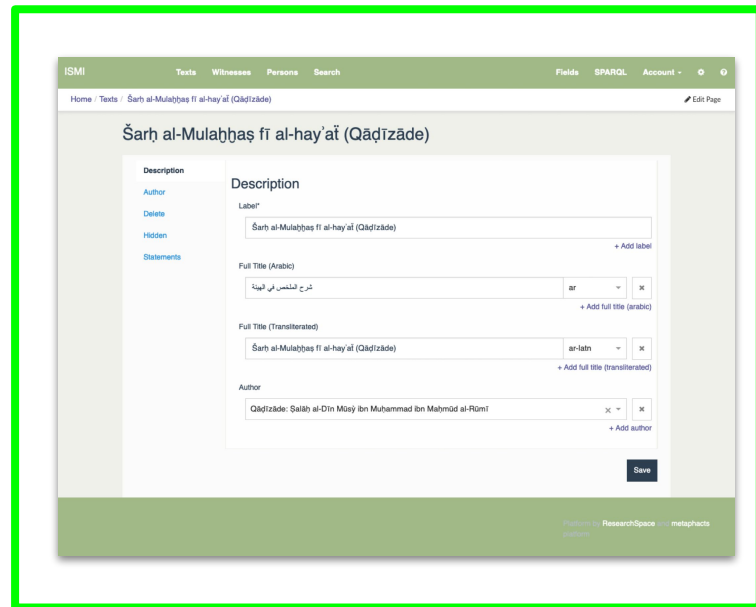
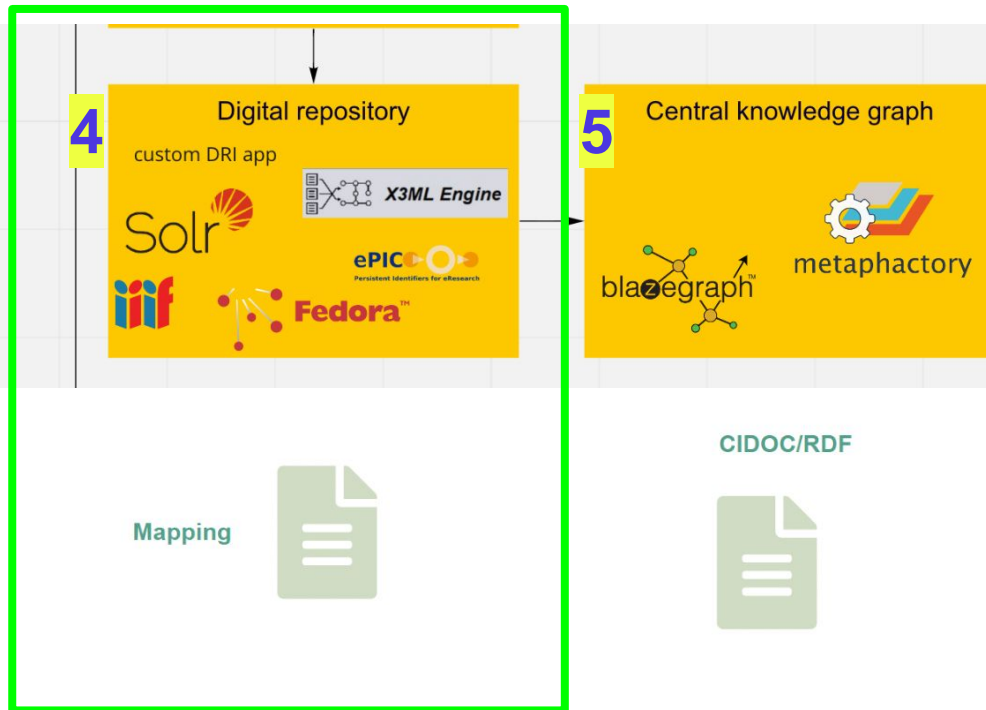
**L' architettura di Marco Vitruvio Pollione** Actions...

pid	21.1110110000-0007-EE26-3
identifier_record	868388467
title	L' architettura di Marco Vitruvio Pollione
language	ita
date_issued	1790
place	Siena
publisher	Bindi
edition	Edizione seconda
collection	Vitruv-Ausgaben aus der Stiftung Bibliothek Werner Oechslin
collection_exact	Vitruv-Ausgaben aus der Stiftung Bibliothek Werner Oechslin
name	Vitruvius; aut Baliani, Berardo; trl
access_right	public
access_licence	Public Domain
identifier_mpiwg	MPIWG:1WFMKAWW
id	14ec278803d06b73
_version_	1702080376418074600

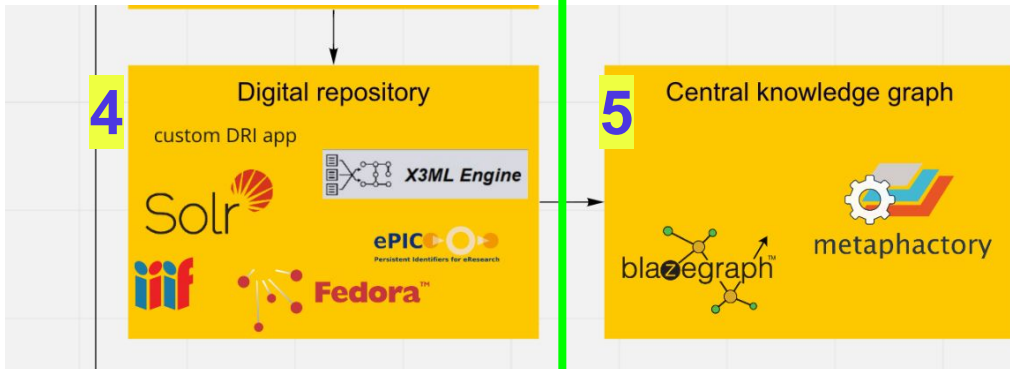
Queue Stats:

Mapping	0
ePIC PID Pool	10
idriver	0

**Custom Digital  
Repository Admin UI**



X3ML Mapping UI



Mapping



CIDOC/RDF



blazegraph workbench  
Ultra-scalable, high-performance database from Blazegraph

WELCOME QUERY UPDATE EXPLORE NAMESPACES STATUS PERFORMANCE

Current namespace: kb

Wik - SPARQL Query Namespace shortcuts: |Bigdata\* |W3C\* |Dublin Core\* |SocialIDhar\* |Custom\* |Edit

```

1 PREFIX wp: <http://vocabularies.wikispathways.org/wp#>
2 PREFIX dc: <http://purl.org/dc/element/1.1/>
3 PREFIX dcterms: <http://purl.org/dc/terms/>
4
5 SELECT DISTINCT
6   ?pathwayRes [str(?resid) as ?metway]
7   (str(?title) as ?pathwayTitle)
8   (substr(str(?checbid), 0) as ?CHEBI)
9   (substr(str(?checbid), 50) as ?CHEBI_Short)
10 WHERE {
11   ?chem a wp:Metabolite ;
12         dcterms:isPartOf ?id ;
13         dcterms:isPartOf ?pathwayRes ;
14         wp:isInCHEBT ?checbid ;
15         ?pathwayRes a wp:Pathway ;
16         dcterms:isPartOf ?psid ;
17         dc:title ?title .
18 }

```

Advanced features

Execute | Clear

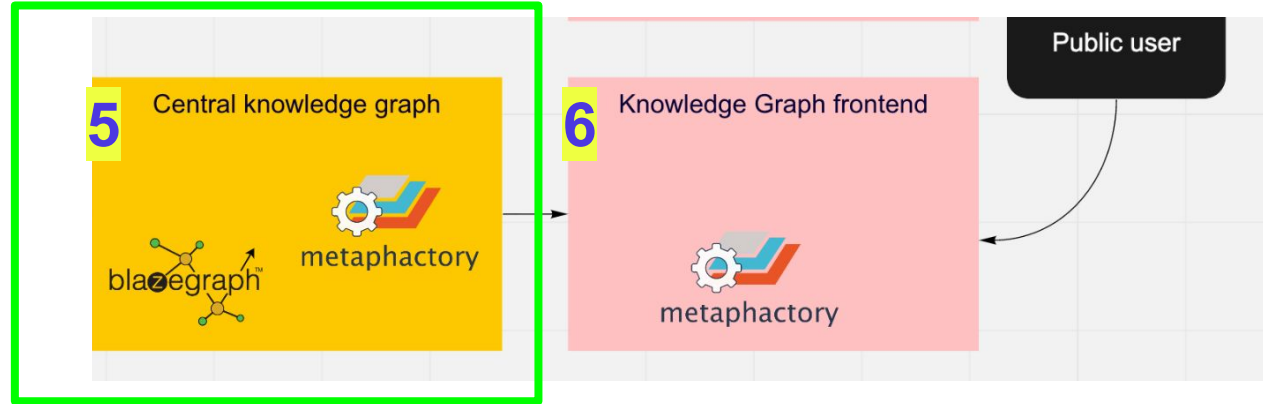
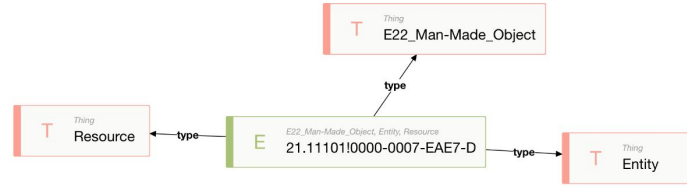
pathwayRes | pathway | pathwayTitle | CHEBI | CHEBI\_Short

Total results: 0, displaying 50 per page

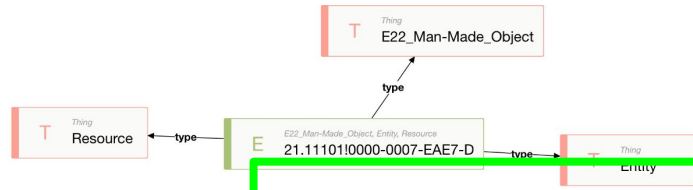
Time	Query	Results	Execution Time	Delete
2016-07-10T18:12:09.962Z	PREFIX wp: <http://vocabularies.wikispathways.org/wp#> PREFIX dc: <http://purl.org/dc/element/1.1/> PREFIX dcterms: <http://purl.org/dc/terms/>  SELECT DISTINCT ?pathwayRes [str(?resid) as ?metway] (str(?title) as ?pathwayTitle) (substr(str(?checbid), 0) as ?CHEBI) as ?CHEBI_Short	0	585ms	Δ

**Blazegraph  
Triplestore UI**

Diagram



Diagram



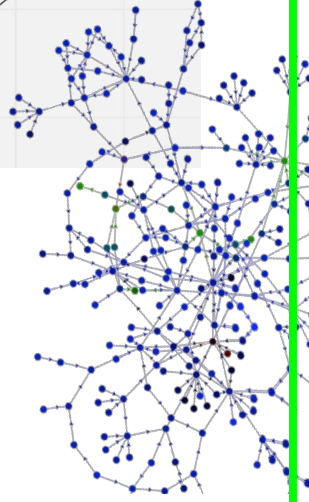
# Knowledge Graph Metaphacts Platform UI

**5** Central knowledge graph

**6** Knowledge Graph frontend

Public user

Linked Open Data



ResearchSpace Platform

Home / Glaser 103

### Glaser 103

URI: [http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/DMDLOG\\_0000/berlin-Z/](http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/DMDLOG_0000/berlin-Z/)

Type: E35\_Title

Knowledge Graph Source

#### Outgoing Statements

Property	Object	Named Graph
P2_has_type	mainFile	<a href="http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context">http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context</a>
type	E35_Title	<a href="http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context">http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context</a>
label	Glaser 103 (archiving)	<a href="http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context">http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context</a>

Execute as SPARQL query

#### Incoming Statements

Property	Subject	Named Graph
P1_is_identified_by	text 21.111010000-0007-EFF2-B/DMDLOG_0000	<a href="http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context">http://content.metaphactory.de/berlin/resource/21.111010000-0007-EFF2-B/context</a>

Graph

E22\_Man-Made\_Object



## 7 GLUE - components connected by additional services

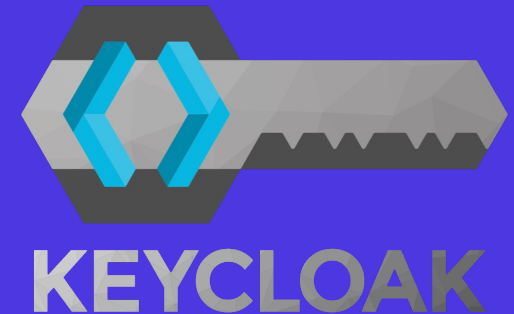
“message broker”  
event  
communication



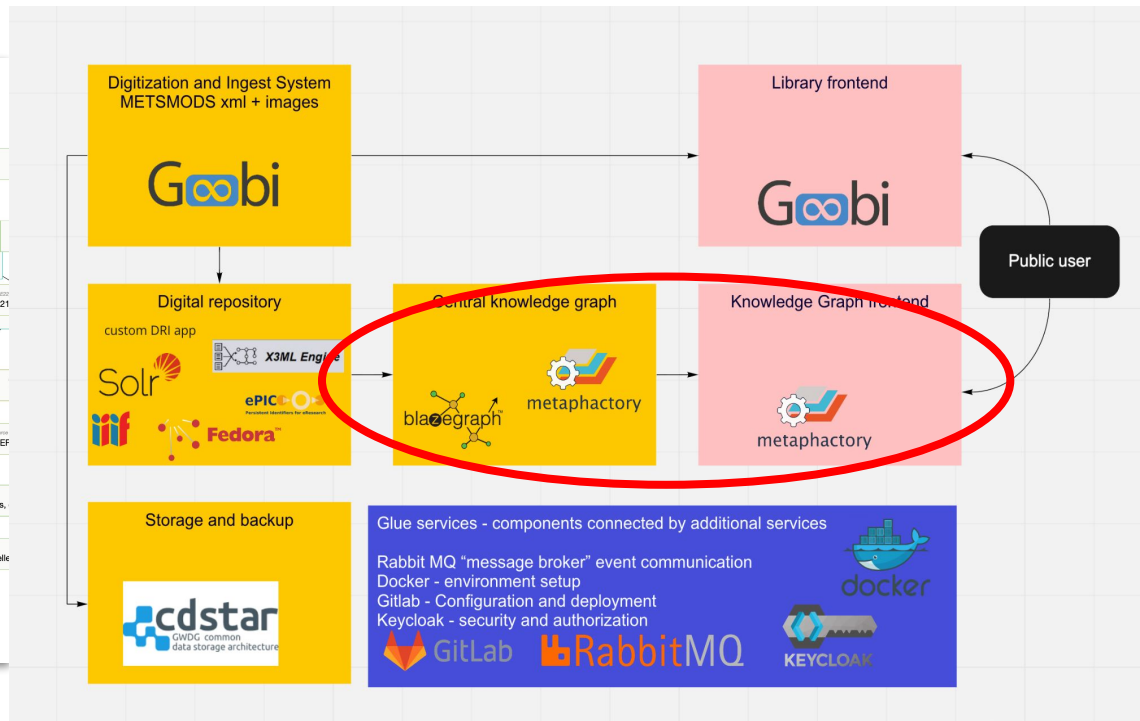
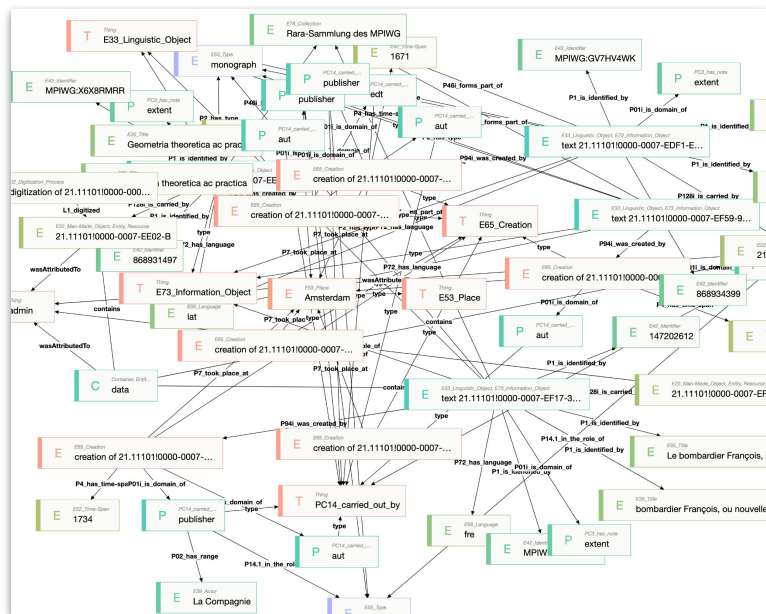
environment setup  
Configuration and deployment



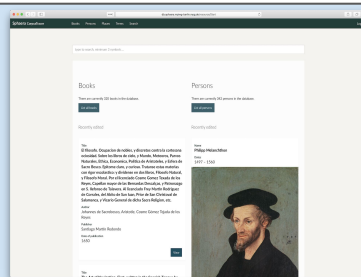
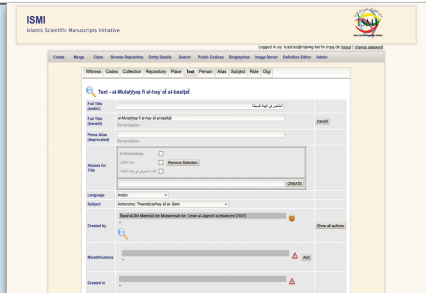
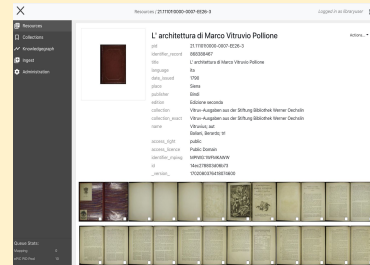
security and authorization



# A Central Knowledge Graph (CKG) - Integrating project research data

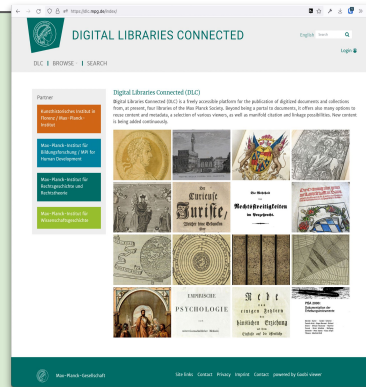


## Project metadata Parthenos

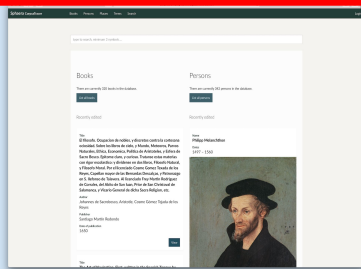
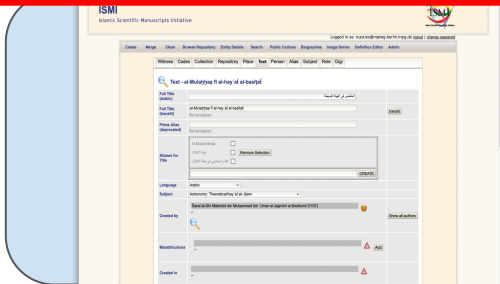
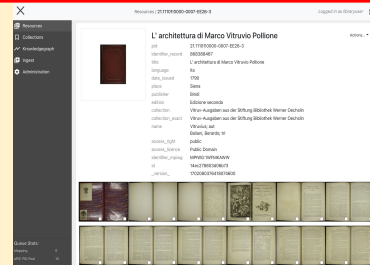


## Project data CRM-Dig

## Rare books data FRBRoo

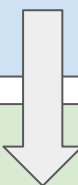
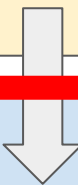
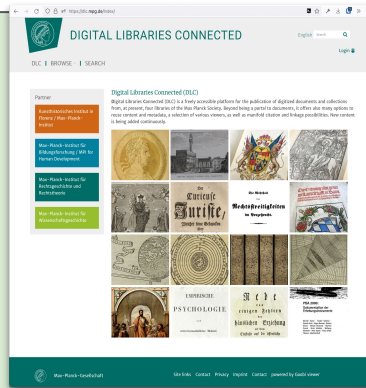


## Project metadata Parthenos

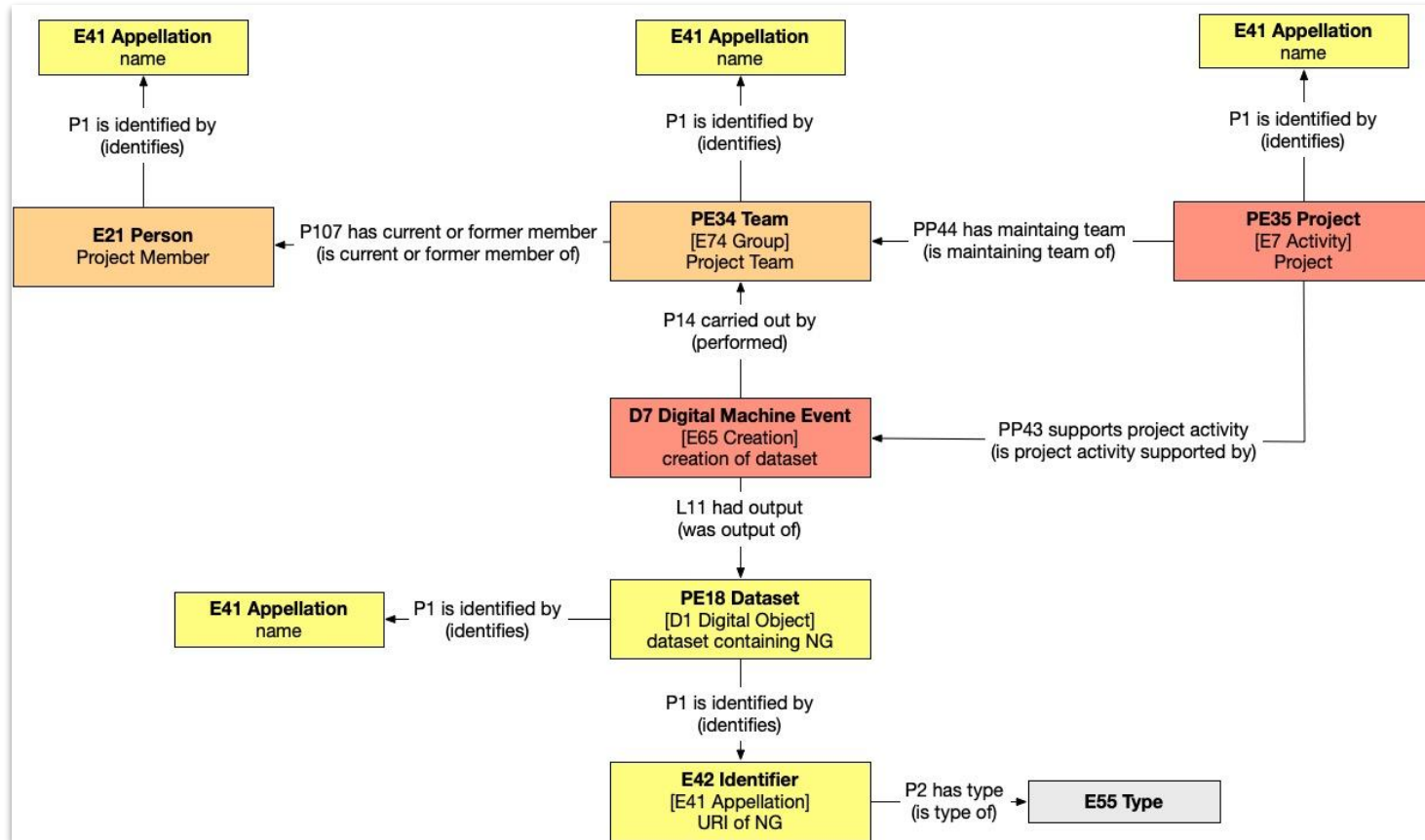


## Project data CRM-Dig

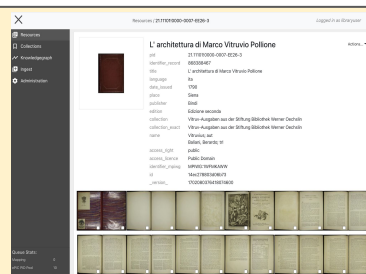
## Rare books data FRBRoo



# Managing Research Data Management



## Project metadata Parthenos

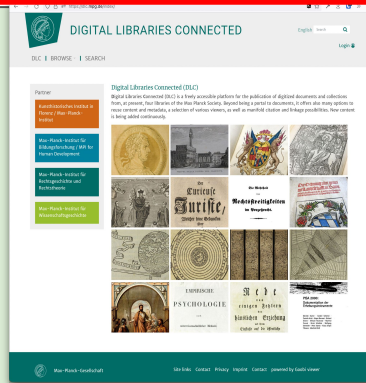


A screenshot of the ISMI website. The page shows a metadata entry for a text titled 'Text: al-Buhārī's Kitāb al-Asfāl'. The interface includes fields for 'Title', 'Author', 'Date', and 'Language', along with a 'Description' field. The ISMI logo is visible in the top right corner.

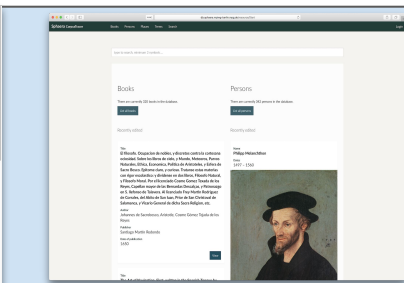
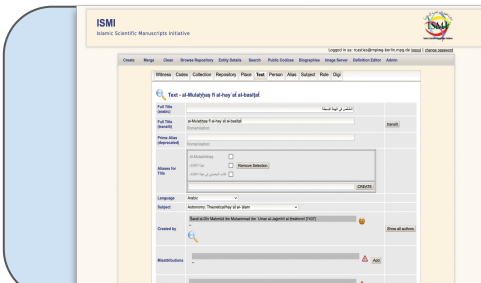
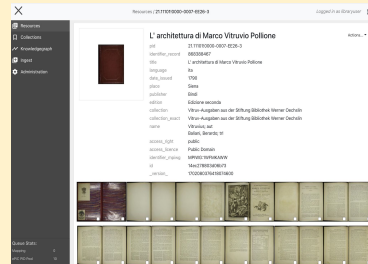
A screenshot of the CRM-Dig interface. It displays a 'Books' section with a list of entries and a 'Persons' section with a portrait of a man in a black cap. The interface is clean and modern, with a white background and blue accents.

## Project data CRM-Dig

## Rare books data FRBRoo

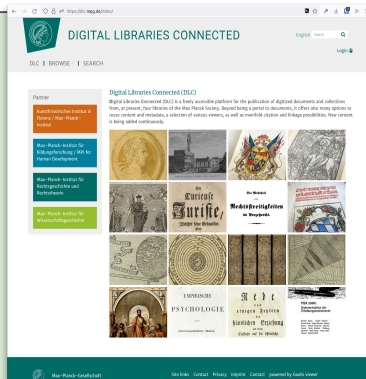


# Project metadata Parthenos



# Project data CRM-Dig

# Rare books data FRBRoo



# Challenges

**Mapping data** - Heterogenous, reconciliation choices, amount of modification, updates

**Modelling decisions** - granularity to capture small/mid/big projects, what should we actually store?

**Complex system** - downstream effects, troubleshooting integrations, interdependence, multiple partners + maintenance

<https://gitlab.gwdg.de/MPIWG/infrastructure>