NFDI4Objects
Research Data Infrastructure for the Material Remains of Human History

de RSE CONFERENCE WÜRZBURG 2024

Research Software Engineering in NFDI4Objects: Community building, implementation of FAIRification Tools and scripting in Computational Archaeology

deRSE24 - Conference for Research Software Engineering in Germany
Julius-Maximilians-Universität Würzburg | 05. - 07. March 2024
Session: RSE in Digital Humanities

F. Thiery & L. K. Schubert & F. Fricke /w A. Schneider & J. Landauer

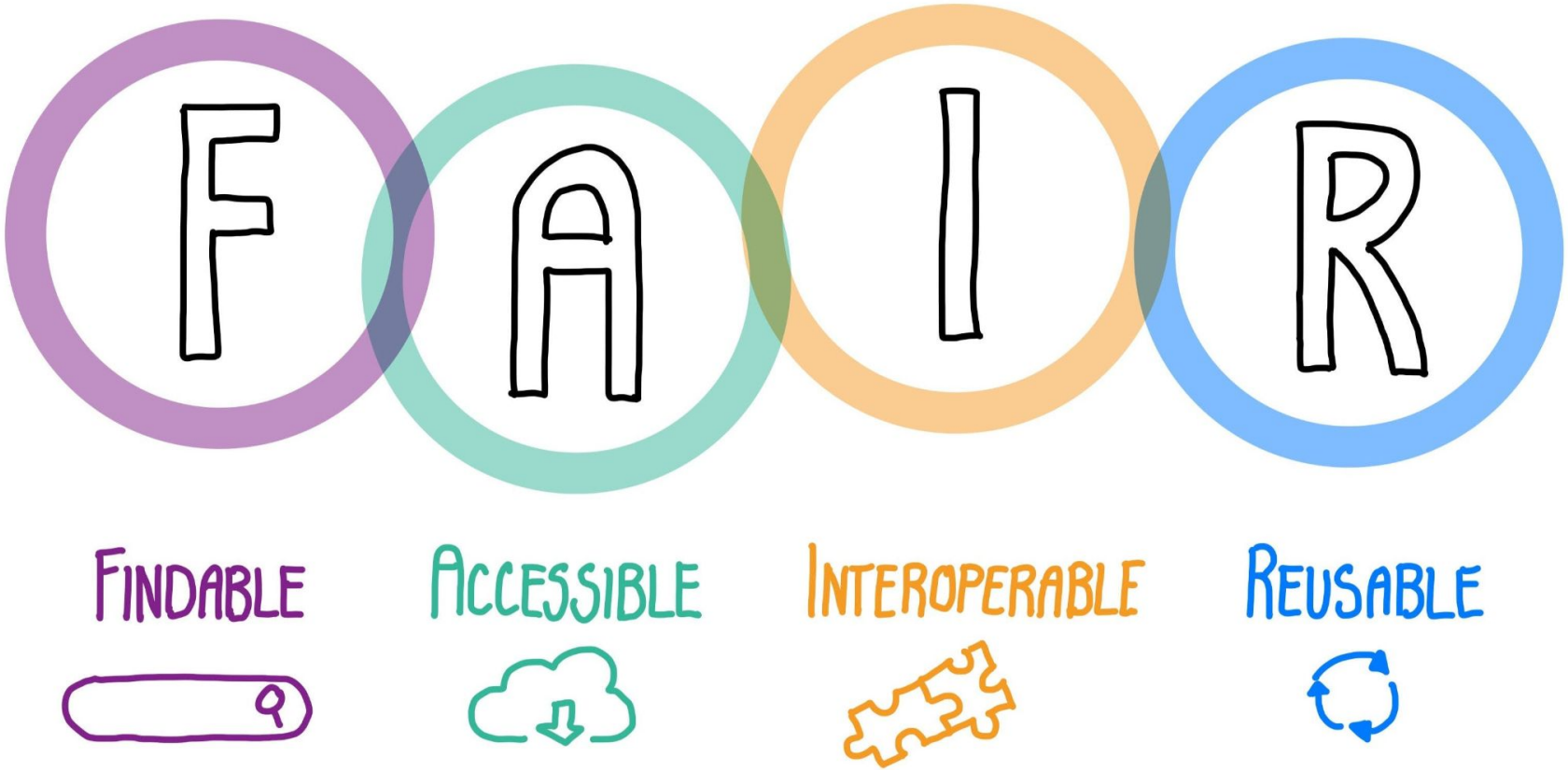
DOI 10.5281/zenodo.10774878

Research Software plays an increasing role in the context of Humanities to support the analysis of the vast and ever-growing data.

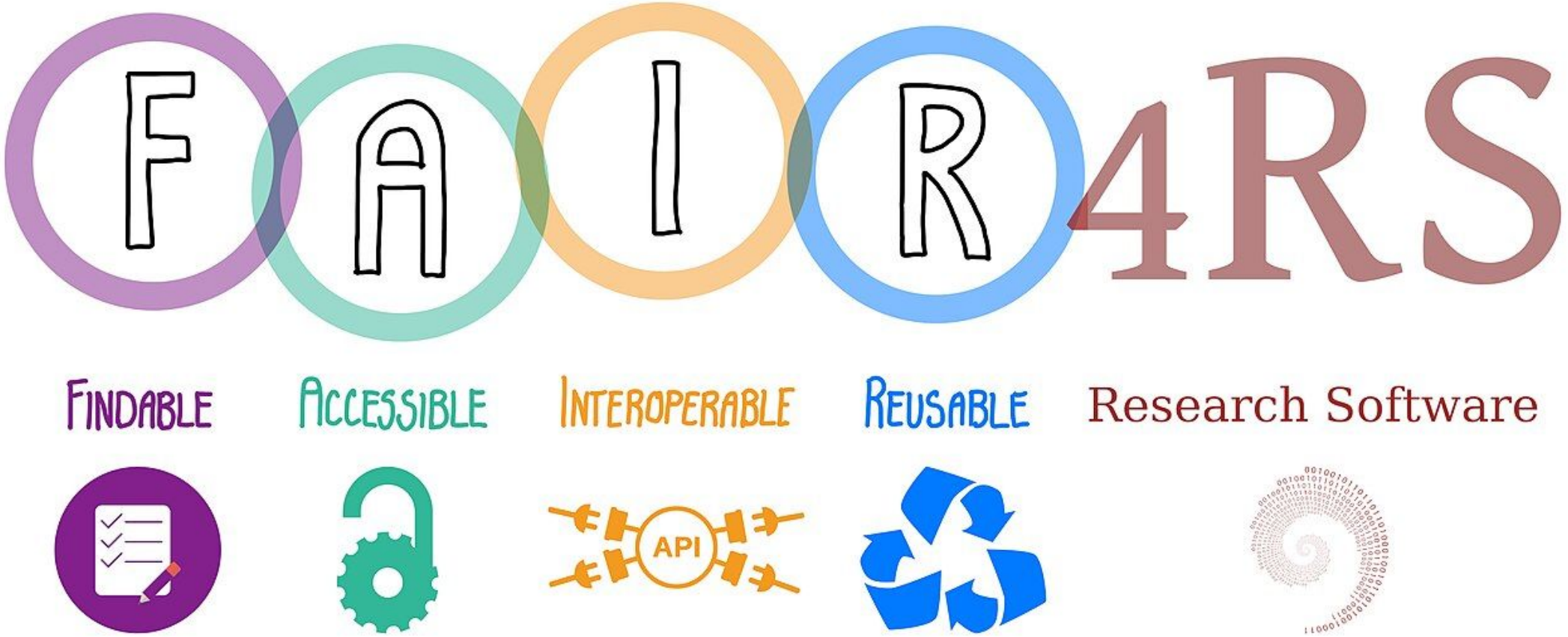
More and more disciplines come together performing advanced analyses, the demand for reproducible and testable results becomes more serious.

So far, most tools have been created ad-hoc to test a hypothesis, but this does not comply with modern objective research practices.

Instead, well-designed, and proven tools and methods are needed that allow reproducible and well-structured results.



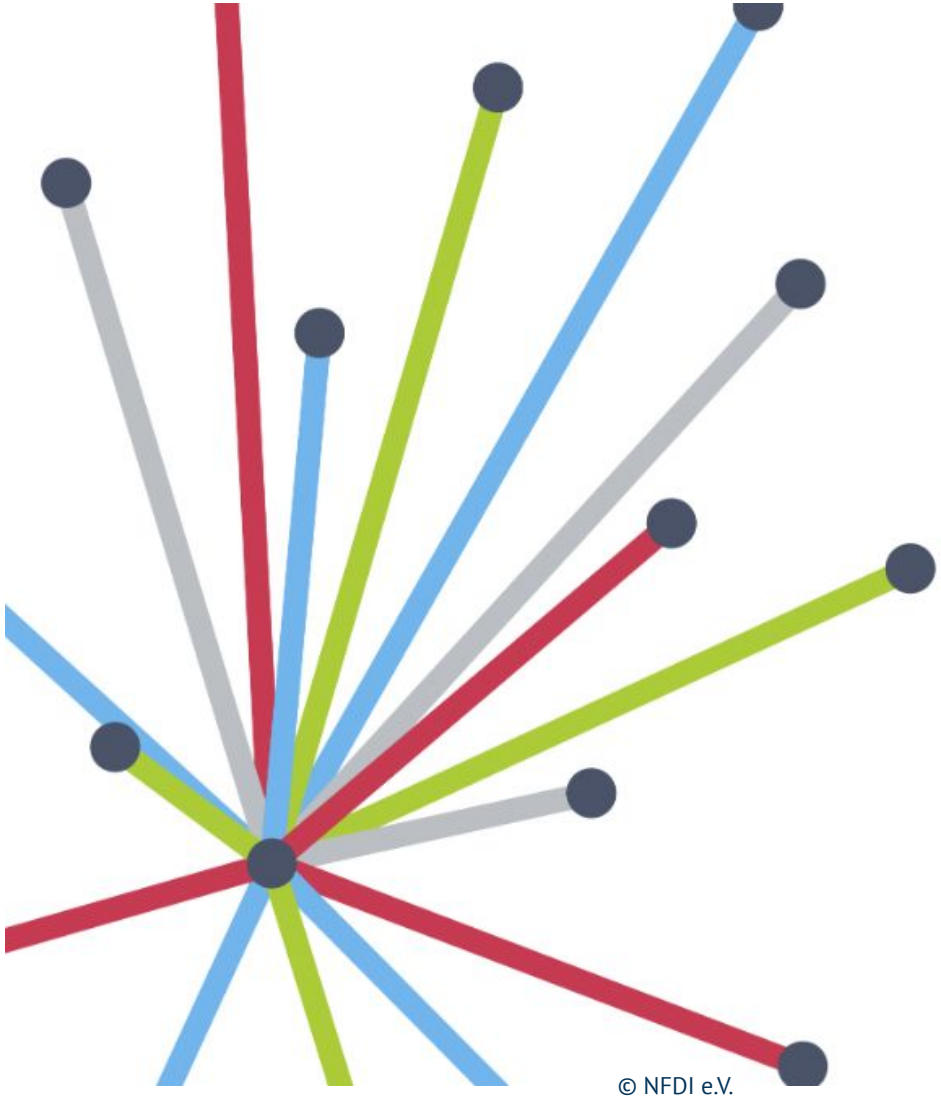
Tools must thereby be equally accessible and FAIR as the data itself, in compliance with the standard right of access to and participation in culture.



Florian Thiery, Heidi Seibold, CC BY 4.0, via Wikimedia Commons

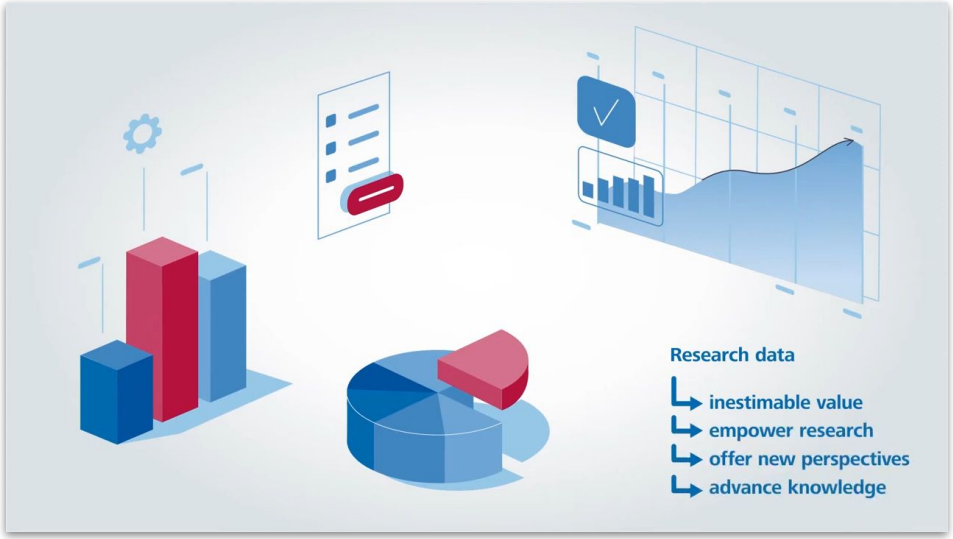
Solution: RSE combined with FAIR is the key!

(German) National Research Data Infrastructure (NFDI)



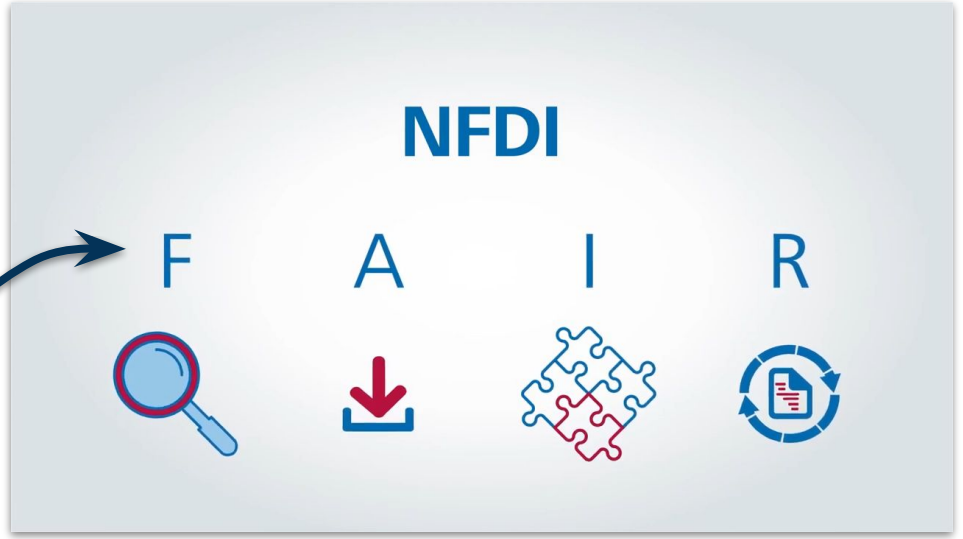
© NFDI e.V.

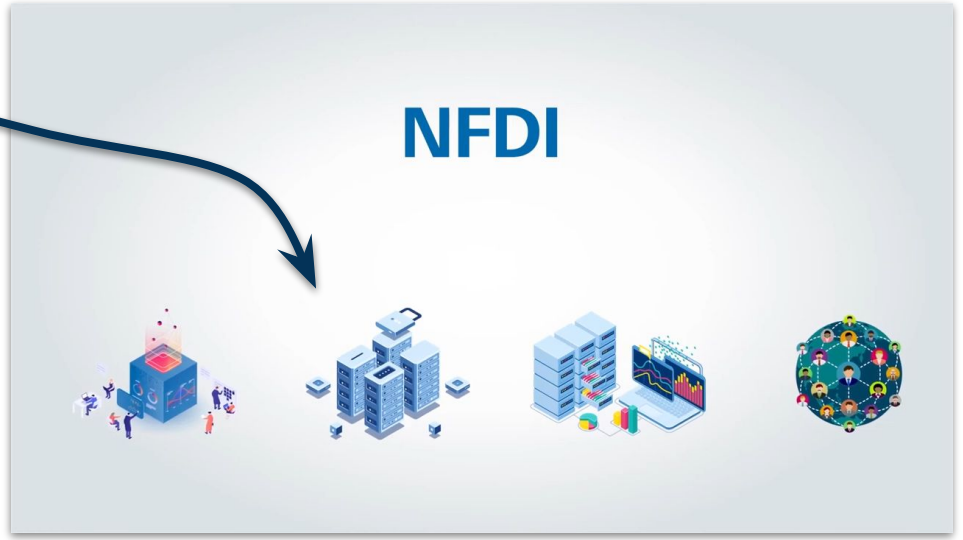
Images via https://youtu.be/uJ01g9m8uE4?si=_w-8HXCVMZtcle6G @DFGbewegt



The NFDI helps to FAIRify research data …

Images via https://youtu.be/uJ01g9m8uE4?si=_w-8HXCVMZtcle6G @DFGbewegt



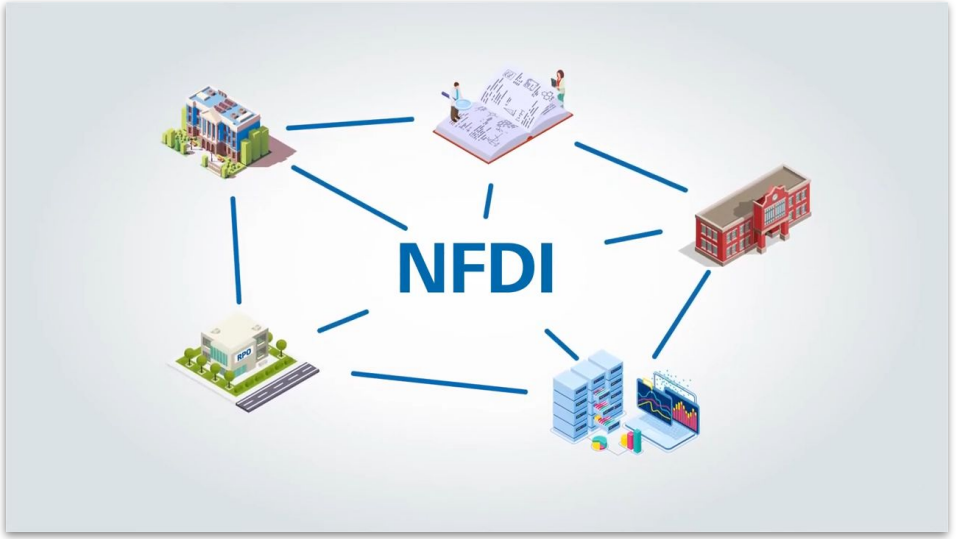


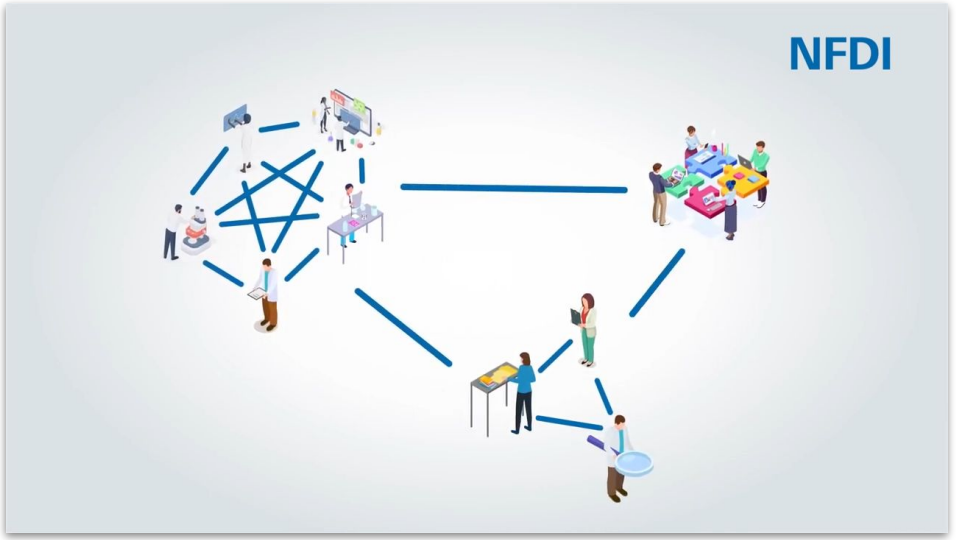
… creating data quality with the help of RSEng.

Images via https://youtu.be/uJ01g9m8uE4?si=_w-8HXCVMZtcle6G @DFGbewegt

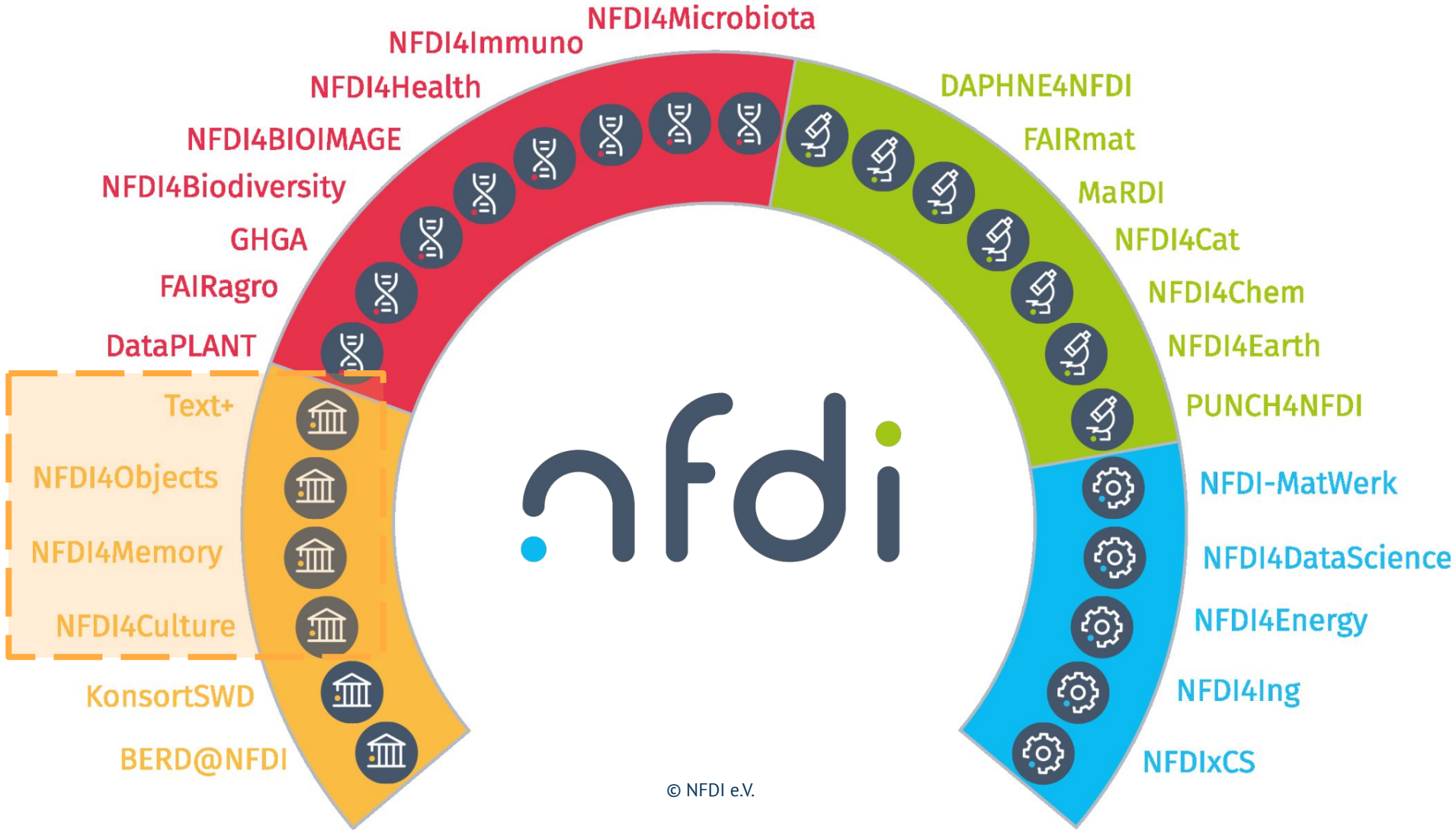
In NFDI, both, unis and RPOs, work together …

Images via https://youtu.be/uJ01g9m8uE4?si=_w-8HXCVMZtcle6G @DFGbewegt

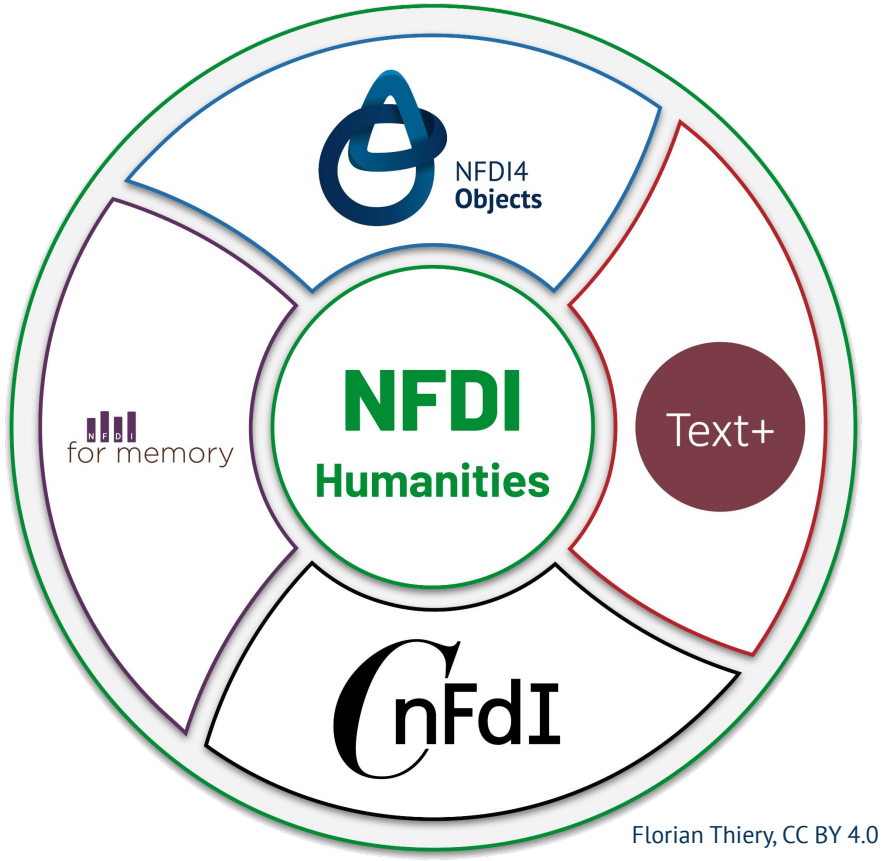




… to create an interlinked research infrastructure!



NFDI contains of several consortia

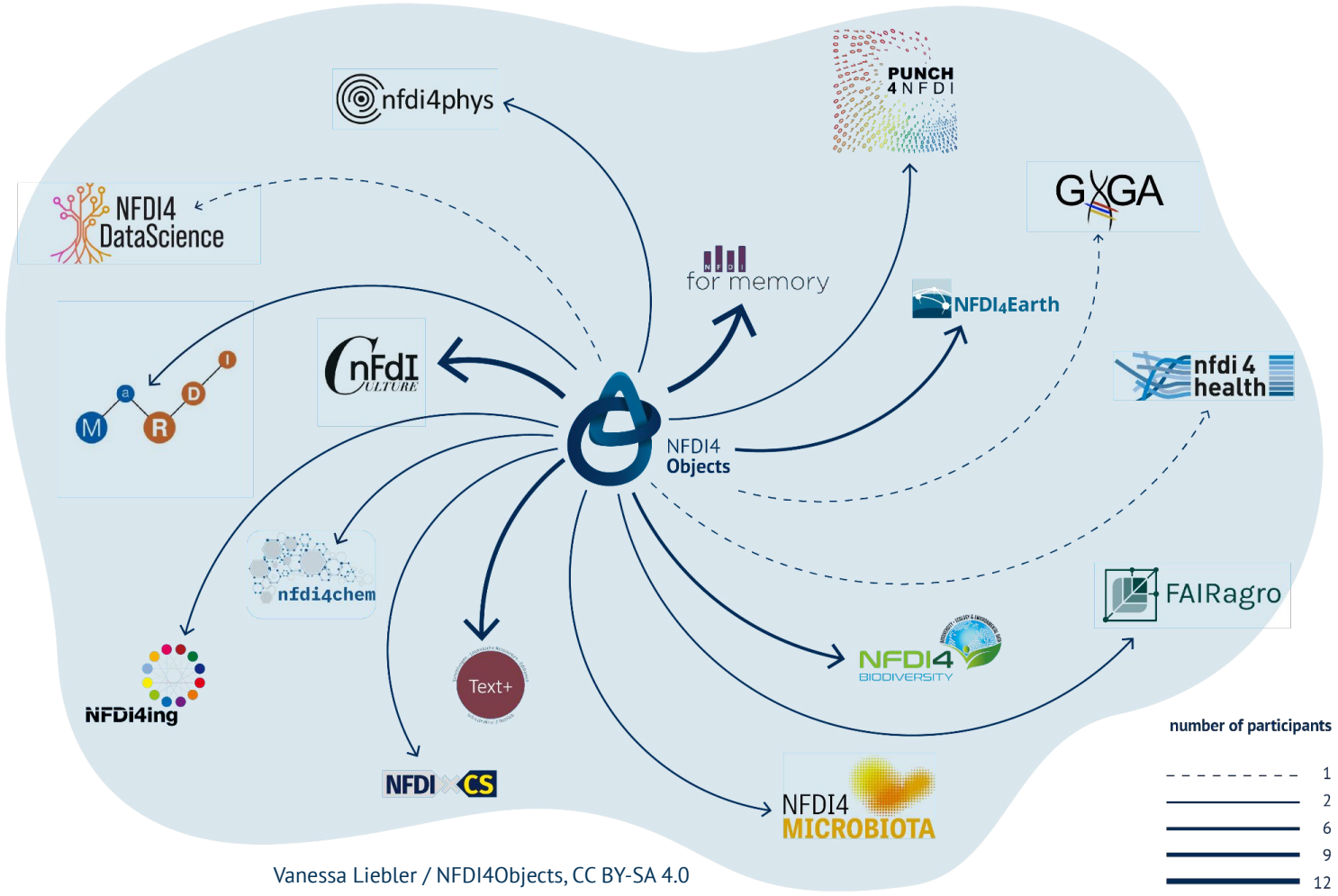


The consortia from the humanities cooperate closely within the MoU-Group

NFDI4Objects

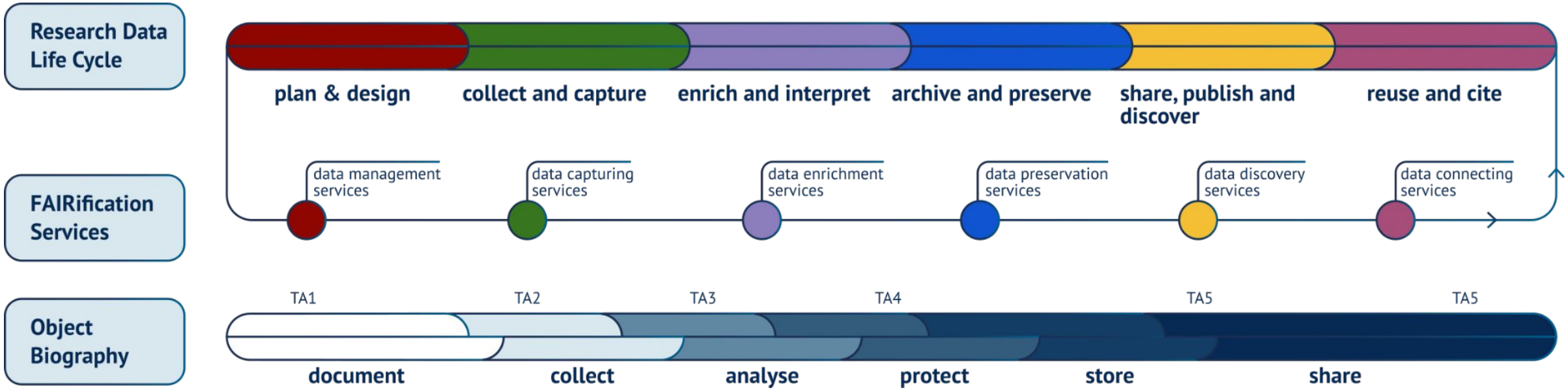
Research Data Infrastructure for the Material Remains of Human History

NFDI4Objects (N4O) is a broad community dealing with material remains of human history, the FAIR and CARE principles as well as FAIR4RS.

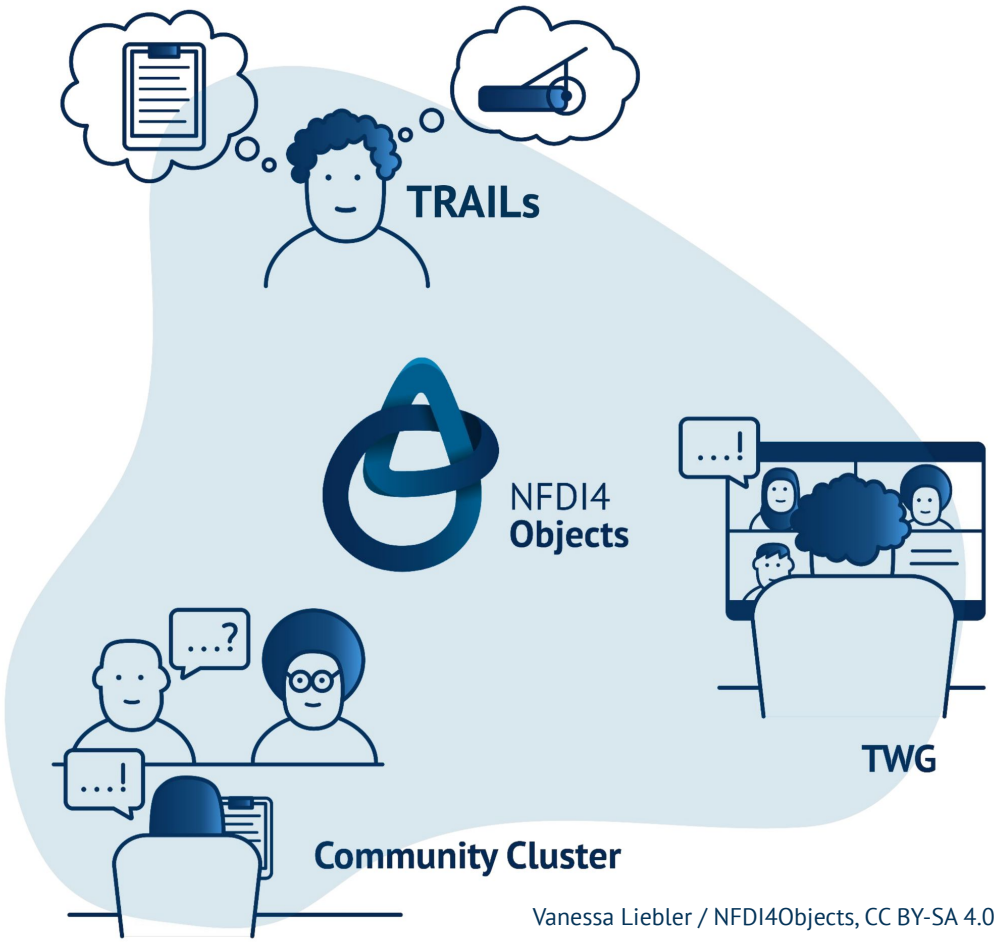


Vanessa Liebler / NFDI4Objects, CC BY-SA 4.0

NFDI4Objects is interdisciplinarily interconnected

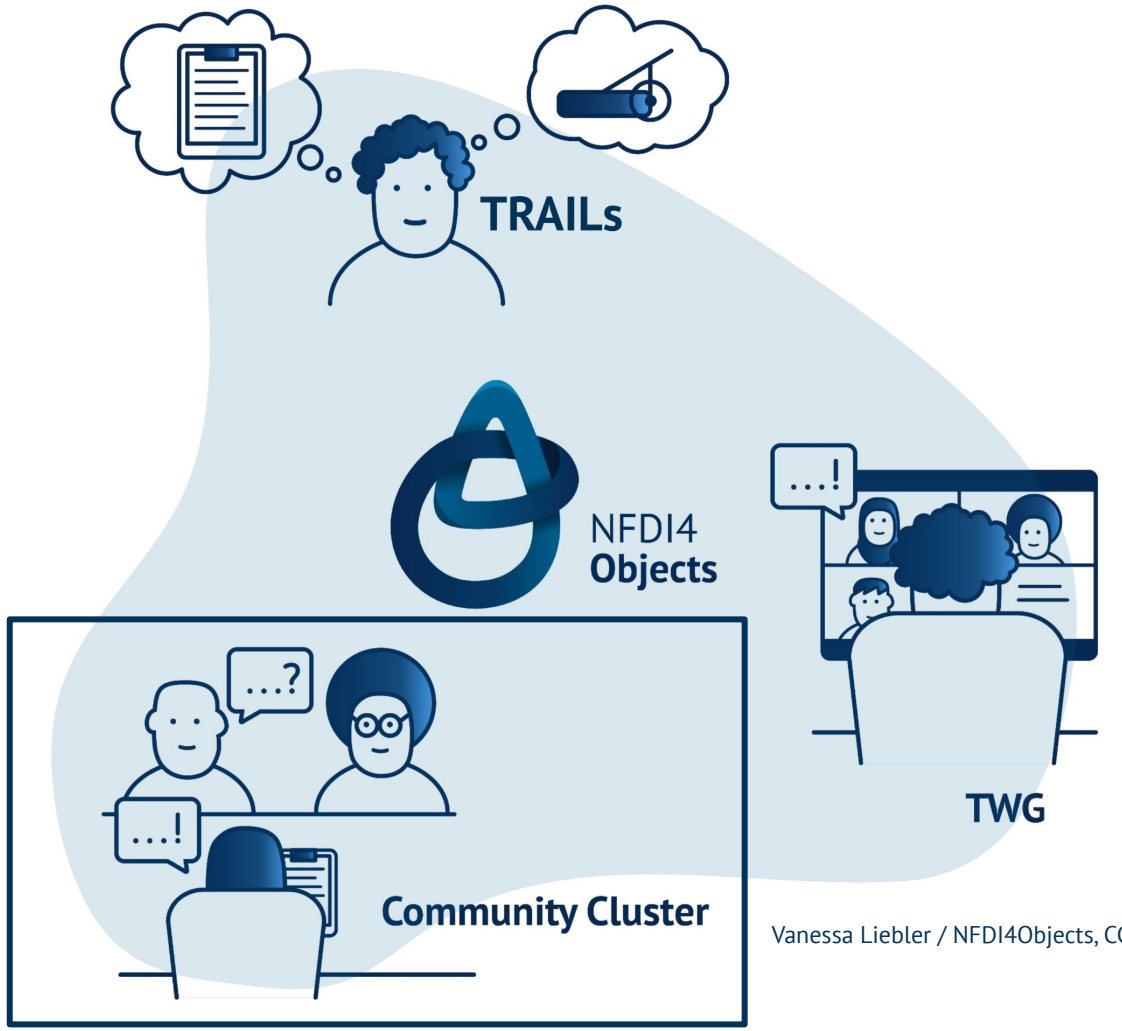


The NFDI4Objects Research Data Lifecycle, goes along with the Object Biography, and needs RSE!



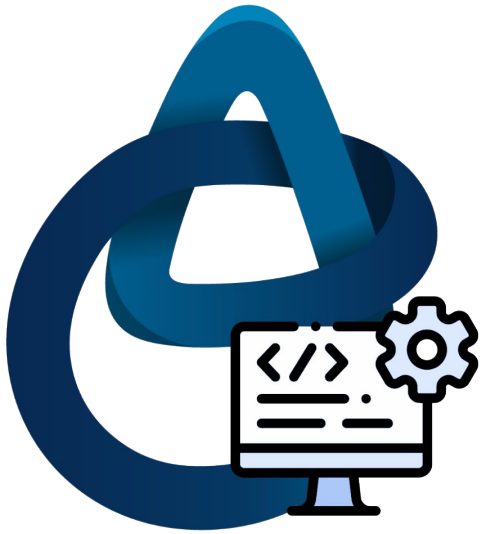
Community Participation in NFDI4Objects is done in Community Cluster, TWGs and TRAILs.

Community Participation Possibilities



Discuss with us in the N4O Community Clusters

- **Strengthening Research Software Engineering** and the community in Computational Archaeology and the Digital Humanities
- **Implementation** of the **FAIR4RS principles** in the scientific process
- Development and further development of **sustainable FAIRification tools**
- **Teaching Research Software Engineering** principles in Computational Archaeology and the Digital Humanities

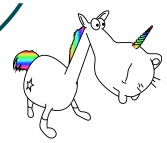


NFDI4**Objects**

Community Cluster
Research Software Engineering (RSE)

Main Goals of the N4O RSE Community Cluster

AG Research Software Engineering in den Digital Humanities



SPARQL UNICORN

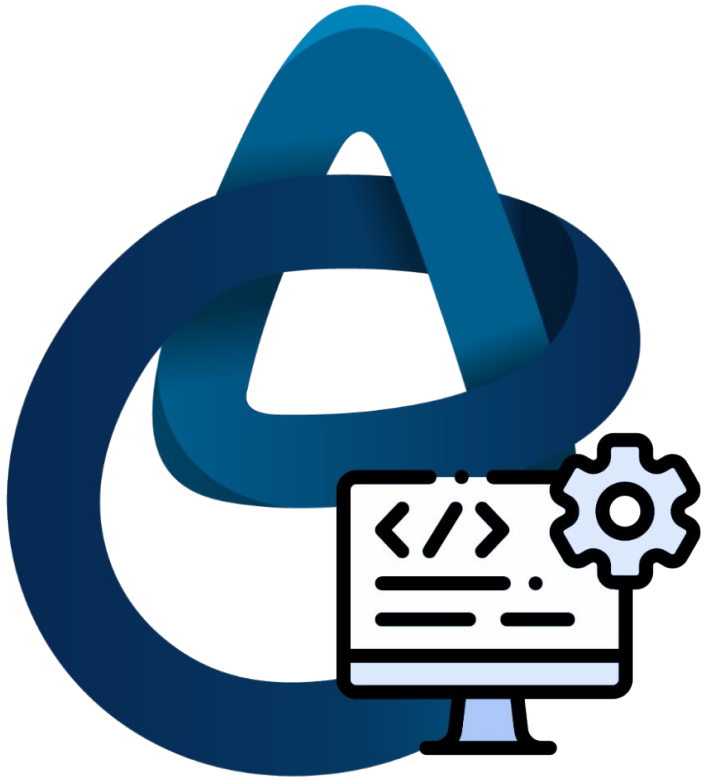
CAA Special Interest Group (SIG) on Special Interest Group on Scientific Scripting Languages in Archaeology (SIG-SSLA)

GESELLSCHAFT FÜR FORSCHUNGSSOFTWARE

The Community Cluster is networking with …

https://www.listserv.dfn.de/sympa/info/n4o_cc_rse

https://zenodo.org/doi/10.5281/zenodo.10512603

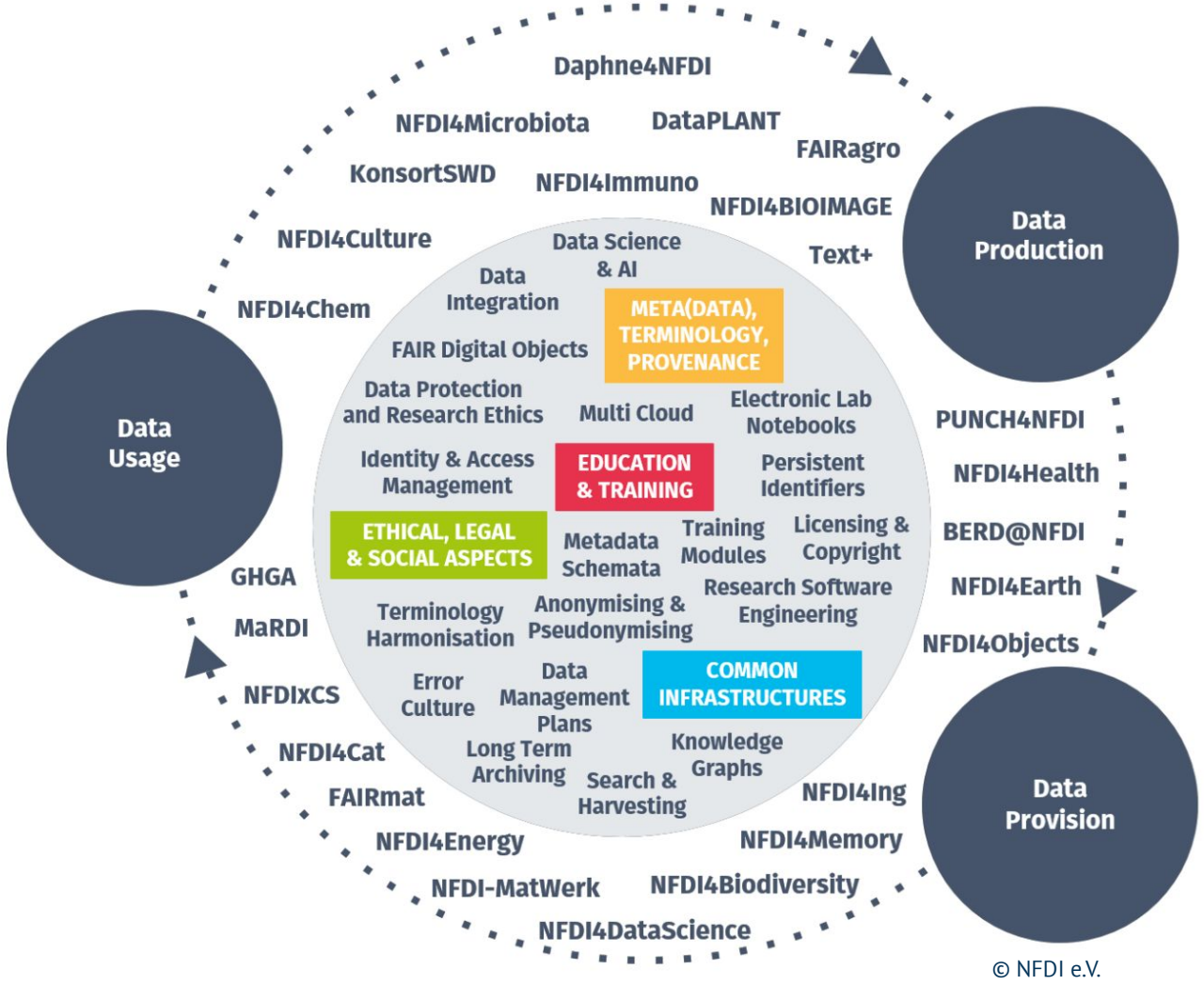


NFDI4Objects

Community Cluster

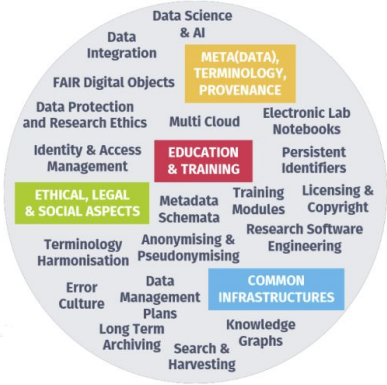
Research Software Engineering (RSE)

Join us!



The Cluster is embedded into the NFDI Sections

Nationale Forschungsdaten Infrastruktur



Research Software Engineering (RSE)

Die Arbeitsgruppe vernetzt die NFDI-Fachkonsortien in Software-bezogenen Aspekten. Sie berücksichtigt drei Schwerpunkte: Forschungssoftware, Software-Communities und Software-Infrastruktur bei NFDI. In beratender und unterstützender Funktion betreibt die Arbeitsgruppe ein zentrales Forum und etabliert das erforderliche Software-Ökosystem innerhalb NFDI für die professionelle Entwicklung von Software-Infrastrukturkomponenten, die in ihrer Gesamtheit einen integralen Bestandteil von NFDI darstellen. Zusätzlich dient die Arbeitsgruppe als Schnittstelle für NFDI zu vergleichbaren europäischen und internationalen Initiativen, um die Anschlussfähigkeit von NFDI mit weiteren Infrastrukturen zu fördern.

Section “Common Infrastructures” WG-RSE

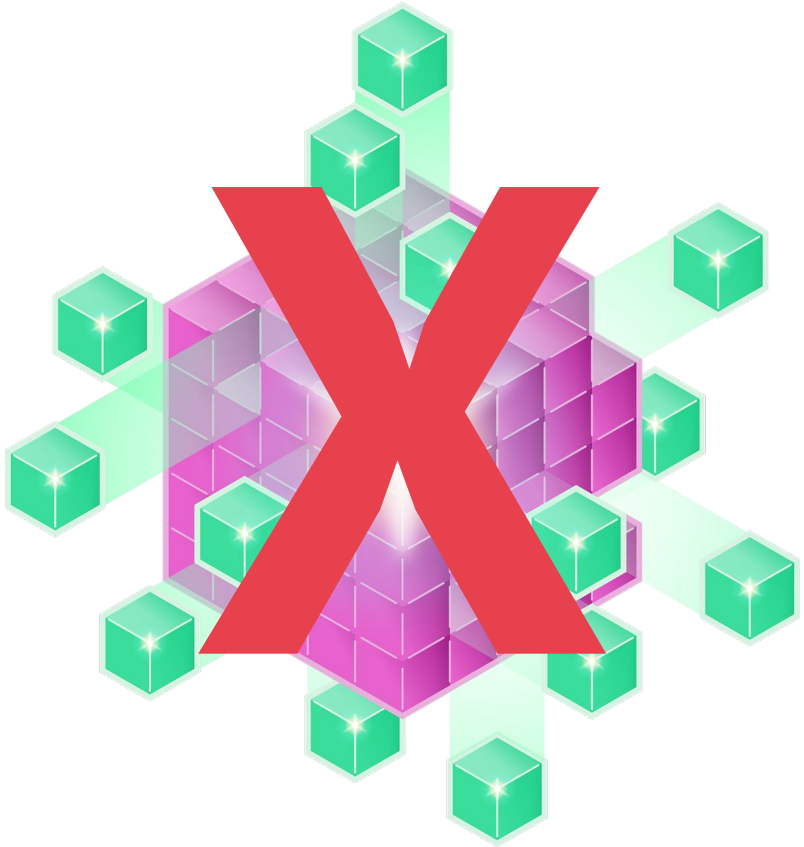
Research Software

… in NFDI4Objects
… the context of CAA

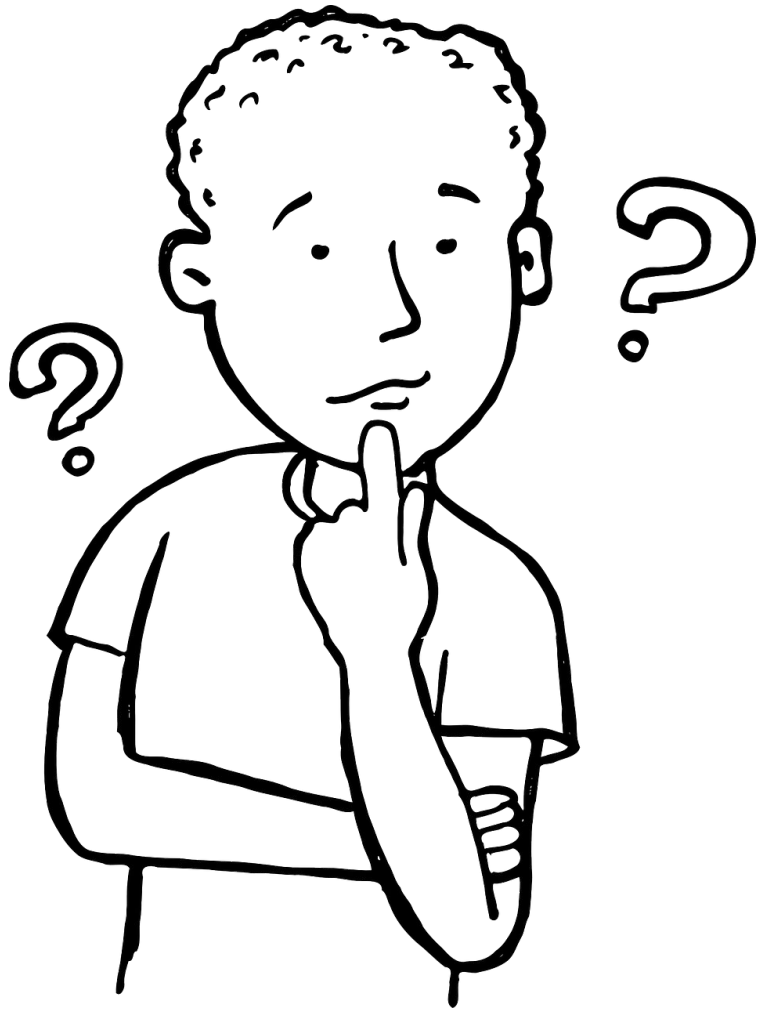
NFDI4**Objects**
Research Data Infrastructure
for the Material Remains of
Human History

Research Software in Archaeology are implemented using different tools, programming methods and methodologies.

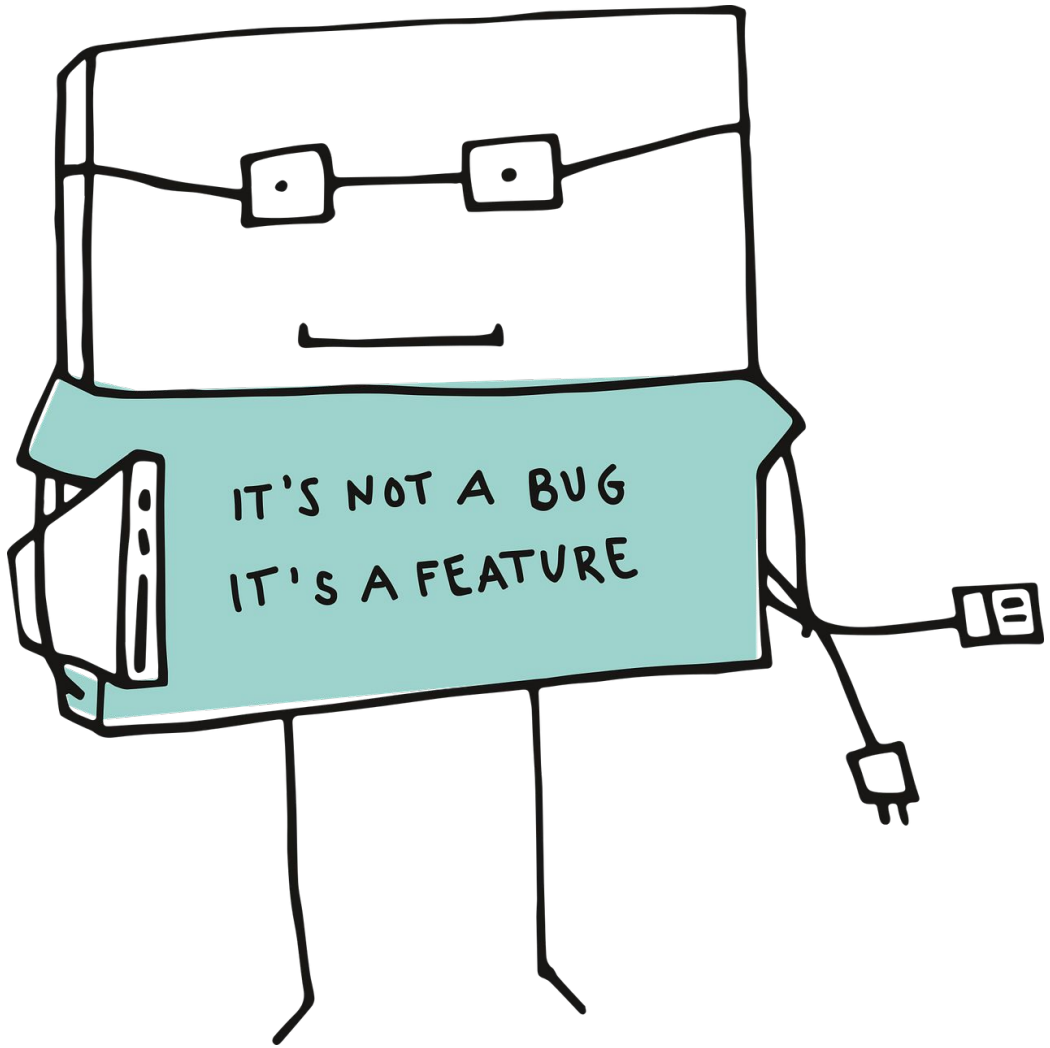
Research Software often do not follow (Research) Software Engineering principles, making it difficult for any uptaker to understand or re-use the code.



What is worse, few were published or made accessible, as the results (aka the data) were deemed more important than the means for generating them.



This impacts reproducibility and therefore, reduces the value and credibility of the results.



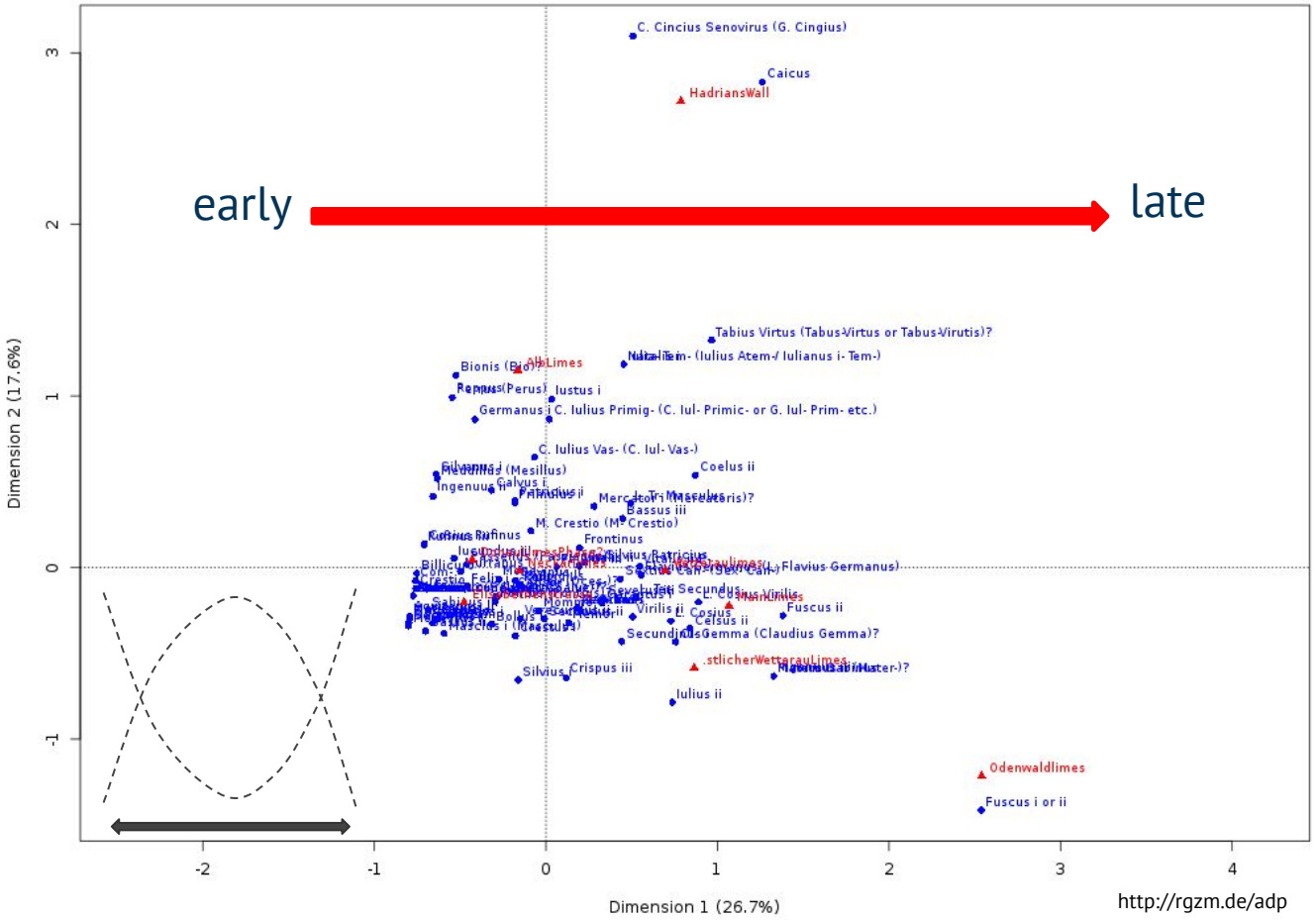
However, good examples are available which will be presented in the following slides.

Computational Archaeology
aka Archaeoinformatics

NFDI4Objects FAIRification Tools

ALLIGATOR

Alligator | Allen Transformator



http://rgzm.de/adp

	A	B	C
1	potter	limes	count
2	Abitus (Habitus)	Elisabethenstrasse	2
3	Acutus i	Elisabethenstrasse	1
4	Aemilius i	Elisabethenstrasse	1
5	Aemilius i	Wetteraulimes	1
6	Albanus ii	DonauLimesPhase2	1
7	Albanus ii	Elisabethenstrasse	1
8	Amandus ii	AlbLimes	1
9	Amandus ii	Elisabethenstrasse	2
10	Amandus iii (Amandinus)?	Elisabethenstrasse	1
11	Aper i	DonauLimesPhase2	2
12	Aper i	Elisabethenstrasse	2
13	Apro (Apro-)?	DonauLimesPhase2	1
14	Apro (Apro-)?	Elisabethenstrasse	1
15	Aquitanus	DonauLimesPhase2	7
16	Aquitanus	Elisabethenstrasse	19
17	Ardacus ii	Elisabethenstrasse	3
18	Ardanus	Elisabethenstrasse	2
19	Astaurus (Asaurus or Tastaurus?)	Wetteraulimes	1
20	Atticus i	AlbLimes	2
21	Auro (Aro)?	DonauLimesPhase2	1
22	Ave (i) or Avetu or Ave tu or Ave Vale etc.	Elisabethenstrasse	2
23	Aveus ii/Avevus?	Elisabethenstrasse	1
24	Avitus ii	Elisabethenstrasse	2
25	Balbus i	Elisabethenstrasse	1
26	Bamsinus or Bamasinus?	Elisabethenstrasse	1
27	Bassinus i	AlbLimes	2
28	Bassus ii	DonauLimesPhase2	4
29	Bassus ii	Elisabethenstrasse	18
30	Bassus ii	ÖstlicherWetterauLimes	1
31	Bassus ii-Coelus	Elisabethenstrasse	7
32	Bassus iii	AlbLimes	2
33	Bassus iii	Elisabethenstrasse	2
34	Bassus iii	Wetteraulimes	4
35	Bassus iii	ÖstlicherWetterauLimes	1
36	Bellicus i	Elisabethenstrasse	1
37	Bilicatus (Bilicatos)	Elisabethenstrasse	3
38	Billicuro	DonauLimesPhase2	3
39	Billicuro	Elisabethenstrasse	1
40	Bionis (Bio)?	AlbLimes	2

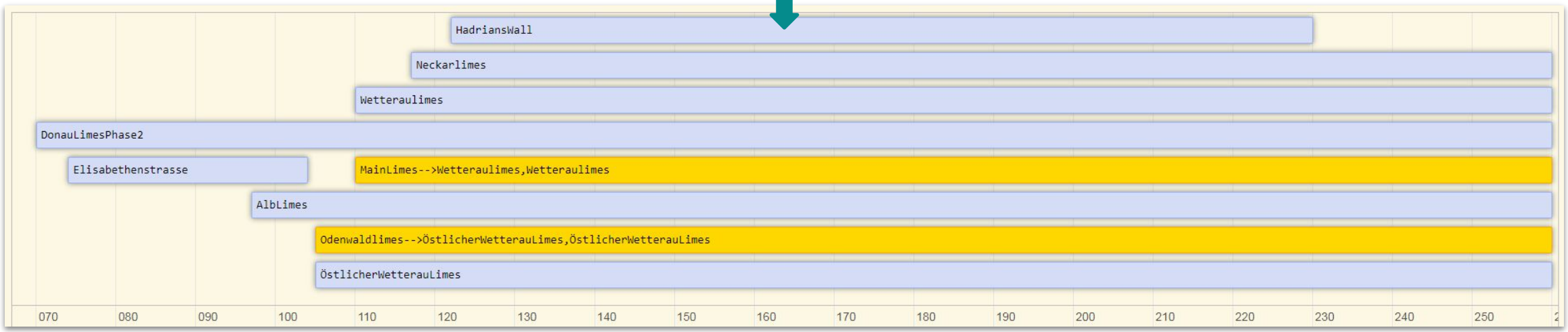
Following the horseshoe paradigm, a CA result may provide a measure of chronological overlap.

	A	B	C	D	E	F	G
1	**name**	**x**	**y**	**z**	**start**	**end**	**fixed**
2	AlbLimes	-0.162	1.149	-0.519	97	260	fixed
3	DonauLimesPhase2	-0.43	0.046	-0.372	70	260	fixed
4	Elisabethenstrasse	-0.479	-0.204	0.270	74	104	fixed
5	HadriansWall	0.787	2.717	2.279	122	230	fixed
6	MainLimes	1.067	-0.223	0.273	0	0	floating
7	Neckarlimes	-0.155	-0.021	-0.973	117	260	fixed
8	Odenwaldlimes	2.540	-1.215	1.228	0	0	floating
9	Wetteraulimes	0.695	-0.019	-0.092	110	260	fixed
10	ÖstlicherWetterauLimes	0.864	-0.585	-0.103	105	260	fixed

	A	B	C	D	E	F	G
1	**name**	**x**	**y**	**z**	**start**	**end**	**fixed**
2	AlbLimes	-0.162	1.149	-0.519	97	260	fixed
3	DonauLimesPhase2	-0.43	0.046	-0.372	70	260	fixed
4	Elisabethenstrasse	-0.479	-0.204	0.270	74	104	fixed
5	HadriansWall	0.787	2.717	2.279	122	230	fixed
6	MainLimes	1.067	-0.223	0.273	**110**	**260**	floating
7	Neckarlimes	-0.155	-0.021	-0.973	117	260	fixed
8	Odenwaldlimes	2.540	-1.215	1.228	**105**	**260**	floating
9	Wetteraulimes	0.695	-0.019	-0.092	110	260	fixed
10	ÖstlicherWetterauLimes	0.864	-0.585	-0.103	105	260	fixed

The Alligator calculates the missing dates using the CA information …

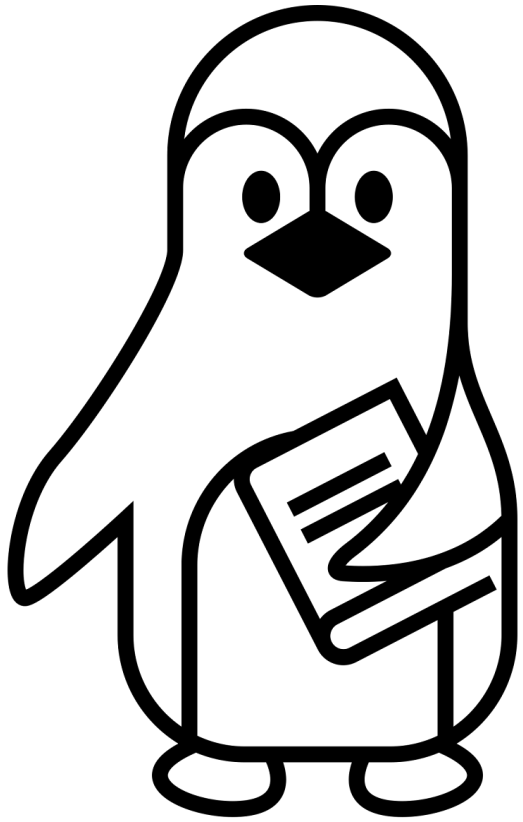
	A	B	C	D	E	F	G
1	**name**	**x**	**y**	**z**	**start**	**end**	**fixed**
2	AlbLimes	-0.162	1.149	-0.519	97	260	fixed
3	DonauLimesPhase2	-0.43	0.046	-0.372	70	260	fixed
4	Elisabethenstrasse	-0.479	-0.204	0.270	74	104	fixed
5	HadriansWall	0.787	2.717	2.279	122	230	fixed
6	MainLimes	1.067	-0.223	0.273	**110**	**260**	floating
7	Neckarlimes	-0.155	-0.021	-0.973	117	260	fixed
8	Odenwaldlimes	2.540	-1.215	1.228	**105**	**260**	floating
9	Wetteraulimes	0.695	-0.019	-0.092	110	260	fixed
10	ÖstlicherWetterauLimes	0.864	-0.585	-0.103	105	260	fixed



… which can be visualised in a timeline …

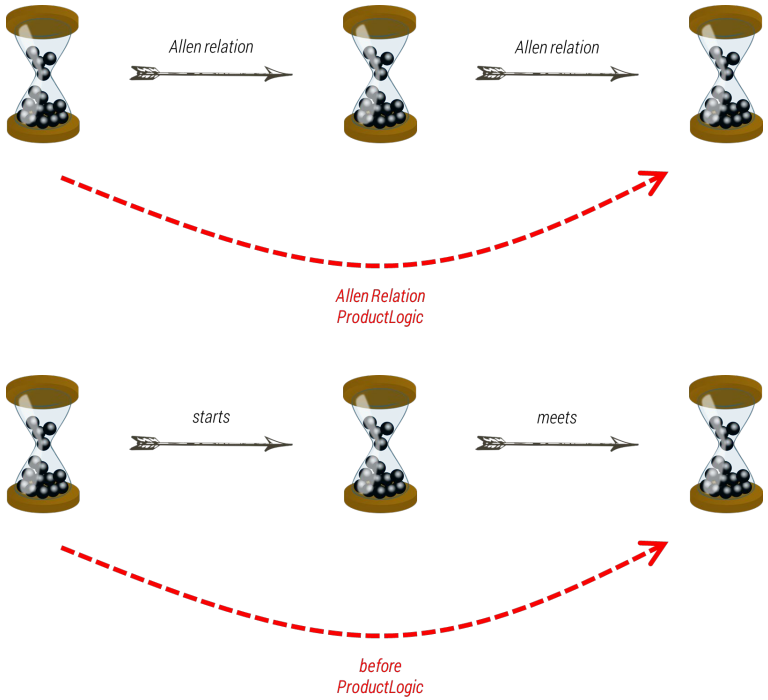
	AL	DL2	ES	HW	ML	NL	OL	WL	ÖWL	
AL	e	f	oi	di	fi	fi	fi	fi	fi	AlbLimes
DL2	fi	e	di	di	fi	fi	fi	fi	fi	DonauLimesPhase2
ES	o	d	e	b	b	b	b	b	b	Elisabethenstrasse
HW	d	d	a	e	d	d	d	d	d	HadriansWall
ML	f	f	a	di	e	fi	f	e	f	MainLimes
NL	f	f	a	di	f	e	f	f	f	Neckarlimes
OL	f	f	a	di	fi	fi	e	fi	e	Odenwaldlimes
WL	f	f	a	di	e	fi	f	e	f	Wetteraulimes
ÖWL	f	f	a	di	fi	fi	e	fi	e	ÖstlicherWetterauLimes

… and transformed in a relative chronology.



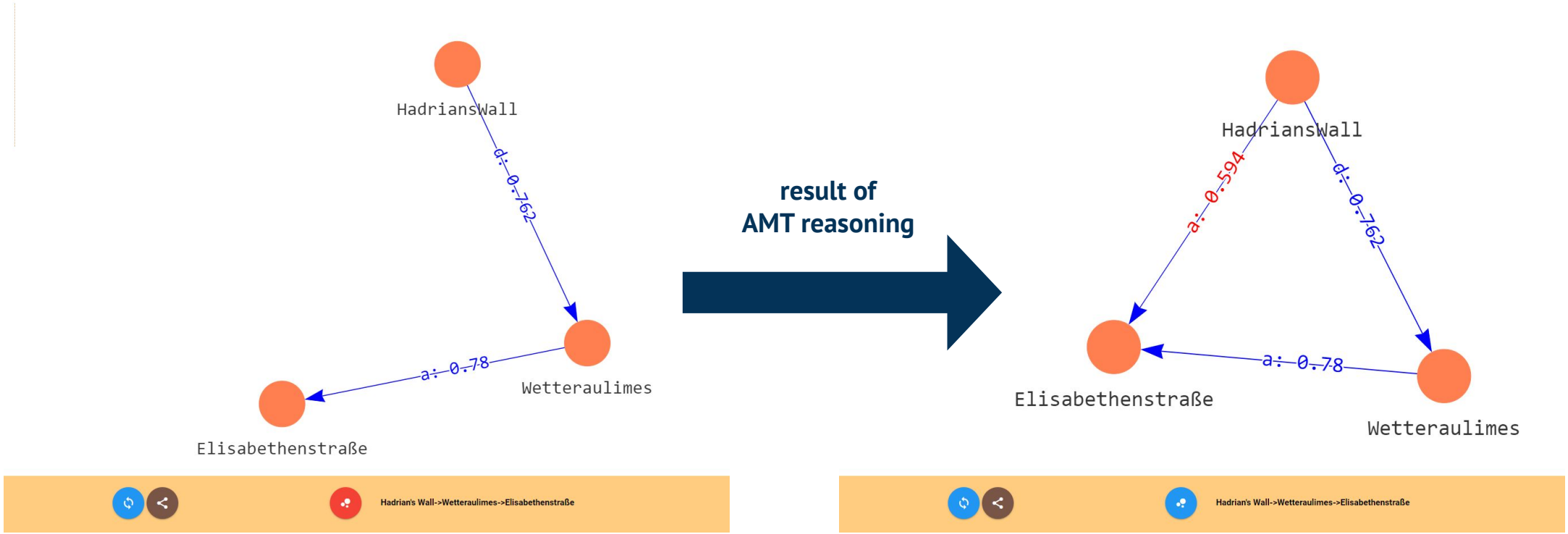
ACADEMIC **META TOOL**

AMT | Academic Meta Tool



	b	m	o	fi	di	si	e	s	d	f	oi	mi	a	q
b	b	b	b	b	b	b	b	b	sb	sb	sb	sb	q	q
m	b	b	b	b	b	m	m	m	bc	bc	bc	tt	sv	q
o	b	b	ob	ob	ol	oc	o	o	bc	bc	ct	sc	sv	q
fi	b	m	ob	fi	di	di	fi	o	bc	tt	sc	sc	sv	q
di	ol	oc	oc	di	di	di	di	oc	ct	sc	sc	sc	sv	q
si	ol	oc	oc	di	di	si	si	hh	yc	oi	oi	mi	a	q
e	b	m	o	fi	di	si	e	s	d	f	oi	mi	a	q
s	b	b	ob	ob	ol	hh	si	s	d	d	yc	mi	a	q
d	b	b	sb	sb	q	yo	di	d	d	d	yo	a	a	q
f	b	m	bc	tt	sv	ys	fi	d	d	f	ys	a	a	q
oi	ol	oc	ct	sc	sv	ys	oi	yc	yc	oi	ys	a	a	q
mi	ol	hh	yc	mi	a	a	mi	yc	yc	mi	a	a	a	q
a	q	yo	yo	a	a	a	a	yo	yo	a	a	a	a	q
q	q	q	q	q	q	q	q	q	q	q	q	q	q	q

Axioms of Allen's interval algebra can be used in reasoners to generate new knowledge.



Here is an example of the relative chronology of Hadrian's Wall and the inferred relative chronological results using AMT.

SPARQLing Unicorn QGIS Plugin

There is a lack of
FLOS GIS tools
for **Linked Open Data**.

The **SPARQLing Unicorn QGIS Plugin** addresses the problem of the lack of availability of tools for Semantic Web geodata.

What is the SPARQL Unicorn?

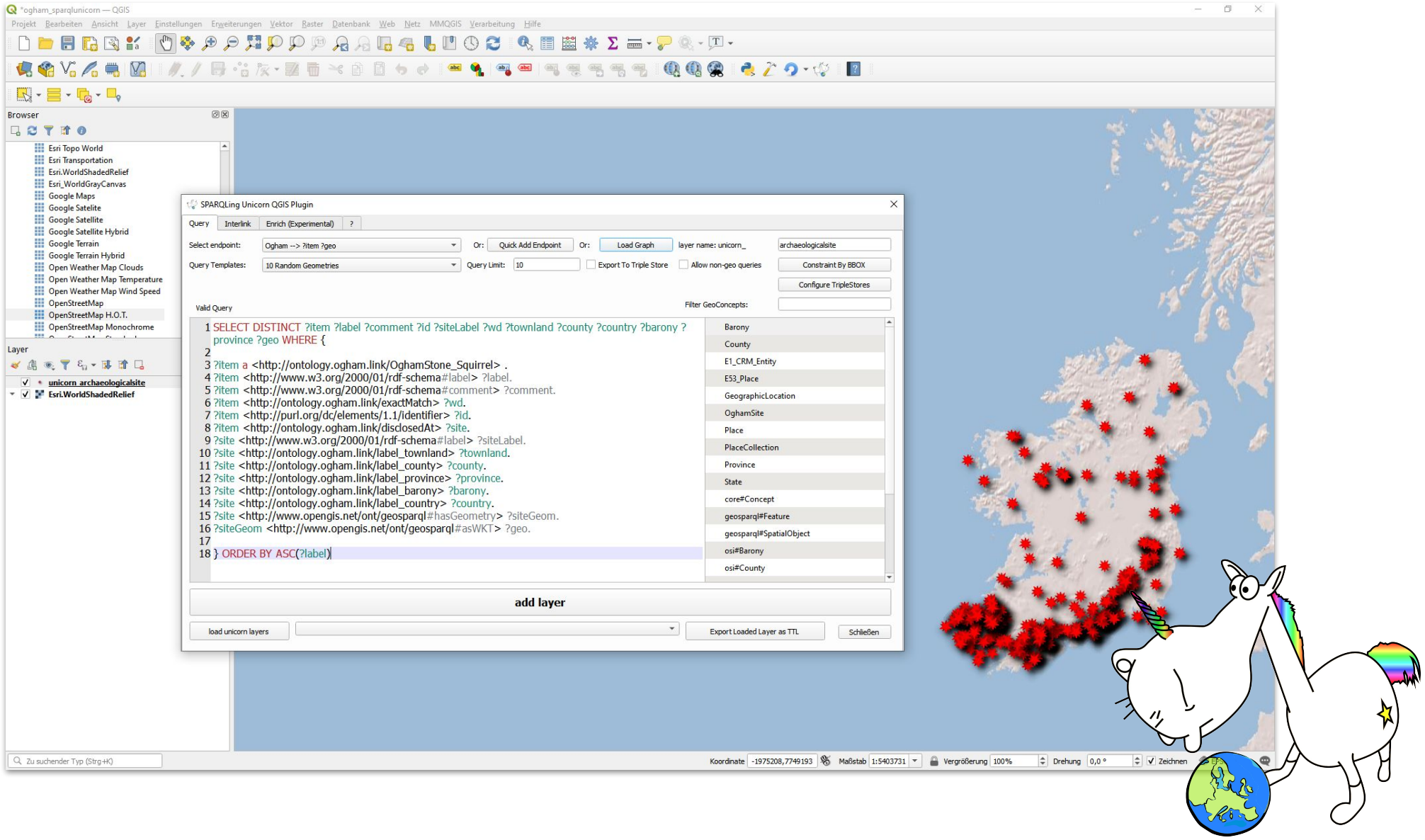
⇒ **new release v0.17**: https://github.com/sparqlunicorn/sparqlunicornGoesGIS/releases/tag/v0.17

The **SPARQL Unicorn** allows the execution of Linked Data queries in (Geo)SPARQL to selected triplestores and geo-enabled SPARQL endpoints and thus prepares the results of the queries in QGIS for the geocommunity.

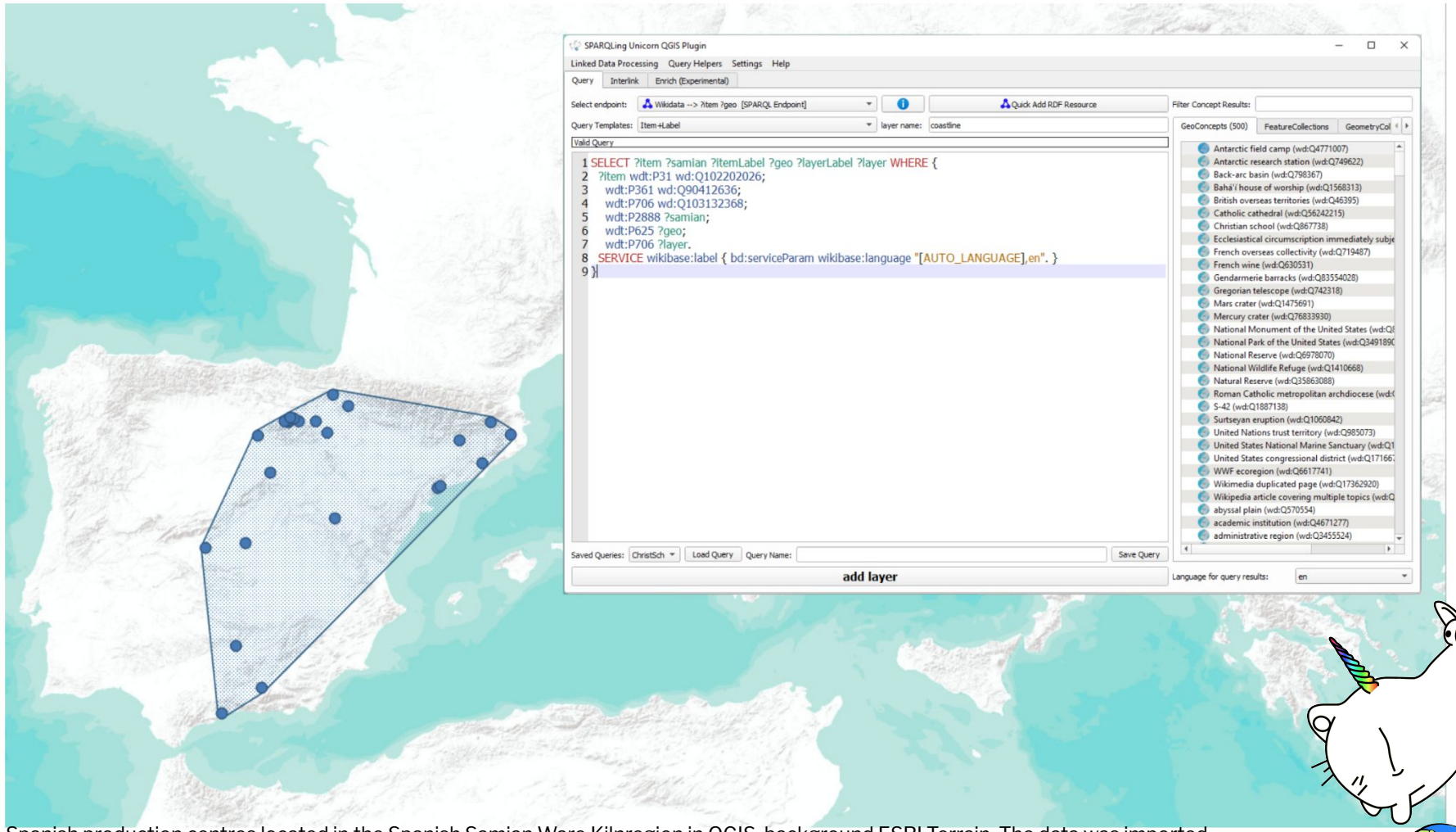
QGIS

https://github.com/sparqlunicorn/sparqlunicornGoesGIS via 10.5281/zenodo.3786814

What does the SPARQL Unicorn?



Example: Wikidata Ogham Sites



Spanish production centres located in the Spanish Samian Ware Kilnregion in QGIS, background ESRI Terrain. The data was imported with the 'SPARQLing Unicorn QGIS Plugin' using the queries https://w.wiki/4pVW and https://w.wiki/4pVY, Florian Thiery, CC BY 4.0.

Example: Samian Research Kilnregions

⇒ **new release v0.17**: https://github.com/sparqlunicorn/sparqlunicornGoesGIS-ontdoc/releases/tag/0.17

SPARQL Unicorn Ontology Documentation

DOI 10.5281/zenodo.8190763

This repository hosts a standalone version of the HTML documentation feature included in the SPARQLing Unicorn QGIS Plugin.

Rather than initiating the documentation generation within the SPARQLing Unicorn QGIS Plugin, this python script allows the generation of the documentation standalone or as a Github Action.

The standalone script does not rely on QGIS classes and does not provide the full functionality available in the SPARQLUnicorn QGIS Plugin.

Deviations from the SPARQLing Unicorn Plugin are listed as follows:

- Support for less geometry literals: Only WKT and GeoJSON literals are supported for rendering

Usage Example as Github Action

For a usage example please refer to this repository: https://github.com/sparqlunicorn/sparqlunicornGoesGIS_testdata

https://github.com/sparqlunicorn/sparqlunicornGoesGIS-ontdoc via 10.5281/zenodo.8190763

Use the a Open Source Tool to create LOD which is working as, e.g., GitHub Action

Computational Archaeology
aka Archaeoinformatics

R & Python Scripts

Preliminary hierarchical cluster analysis of ceramic vessels (Stroke Band Pottery between Central Germany and Poland) using Gower's distance measure and the Ward method.

Image by Sophie C. Schmidt, CC BY-SA 4.0

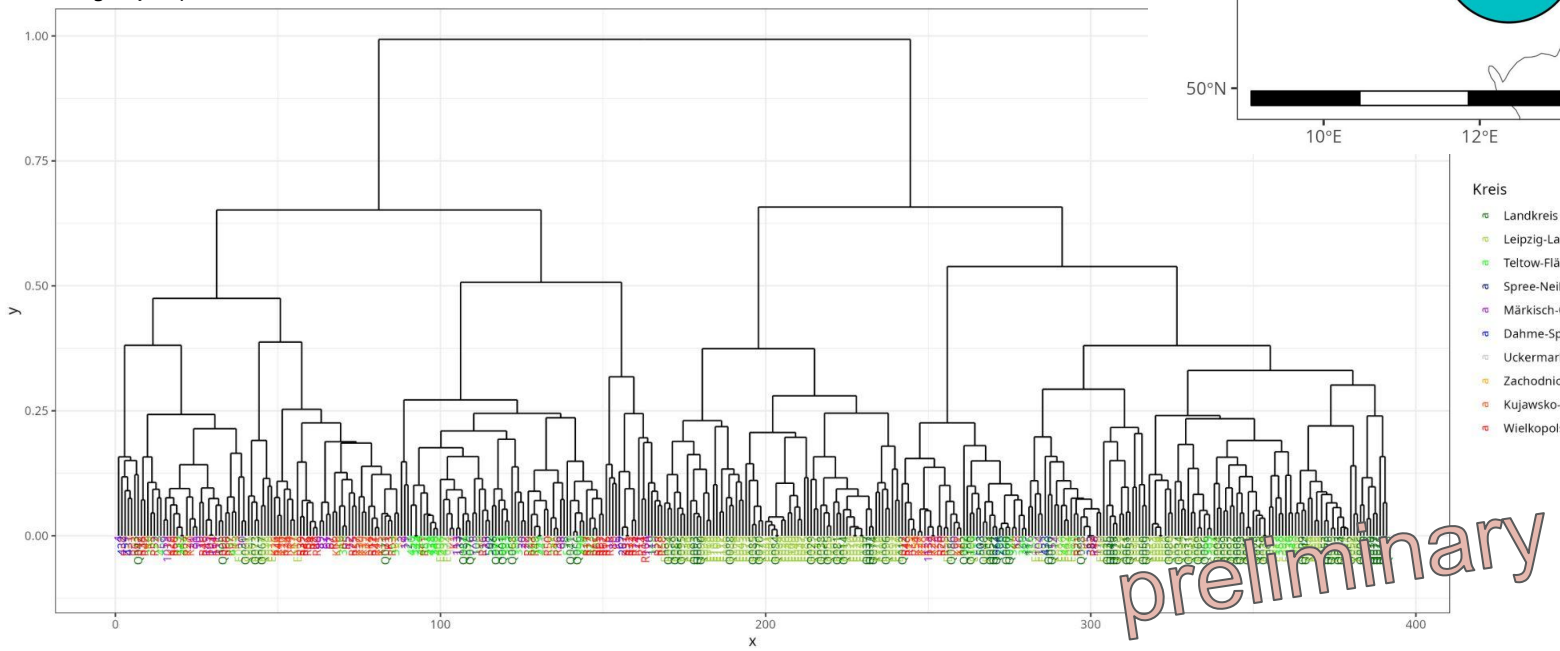
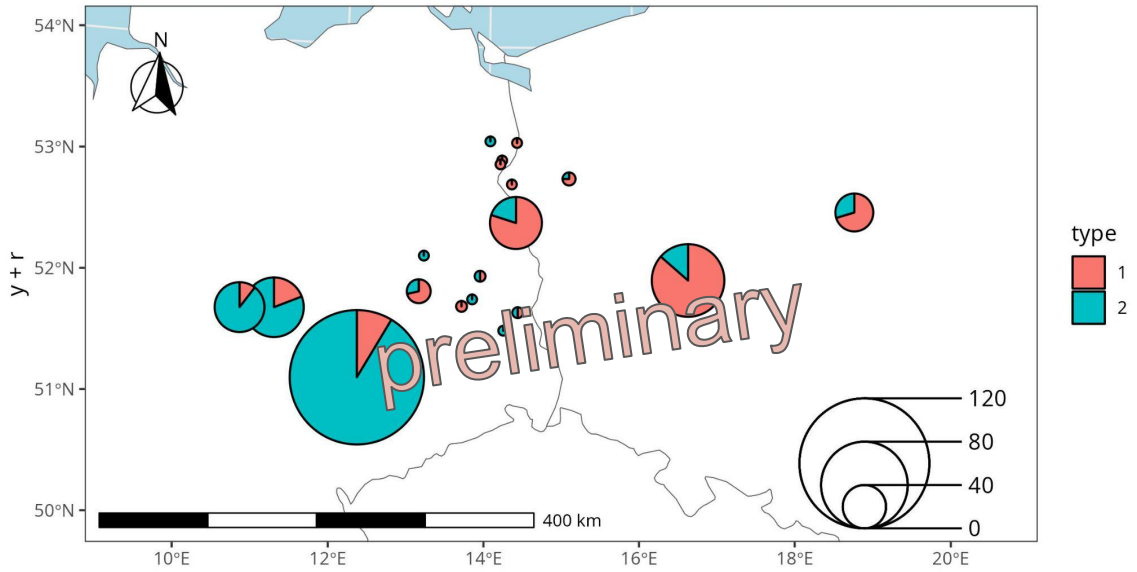
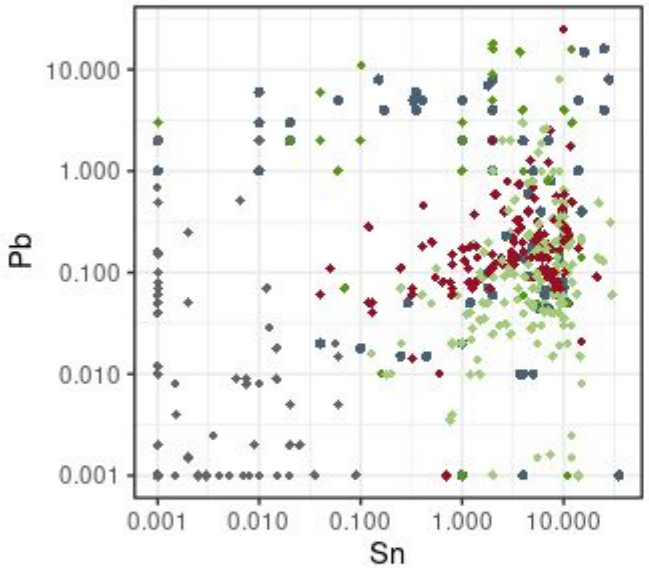


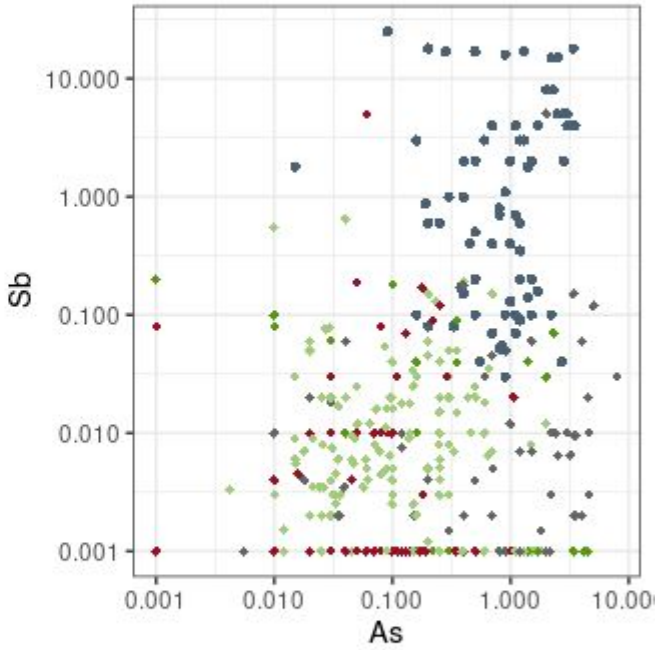
Image by Sophie C. Schmidt, CC BY-SA 4.0

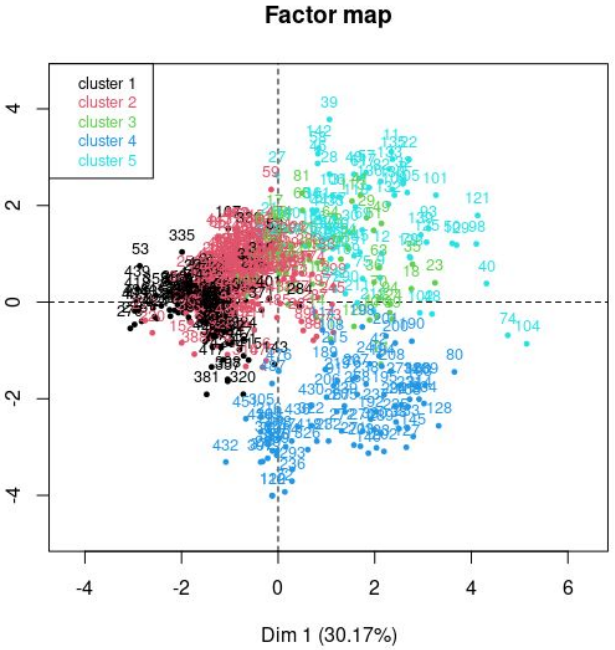


The tree was cut to create two groups and their occurrences mapped to the sites.

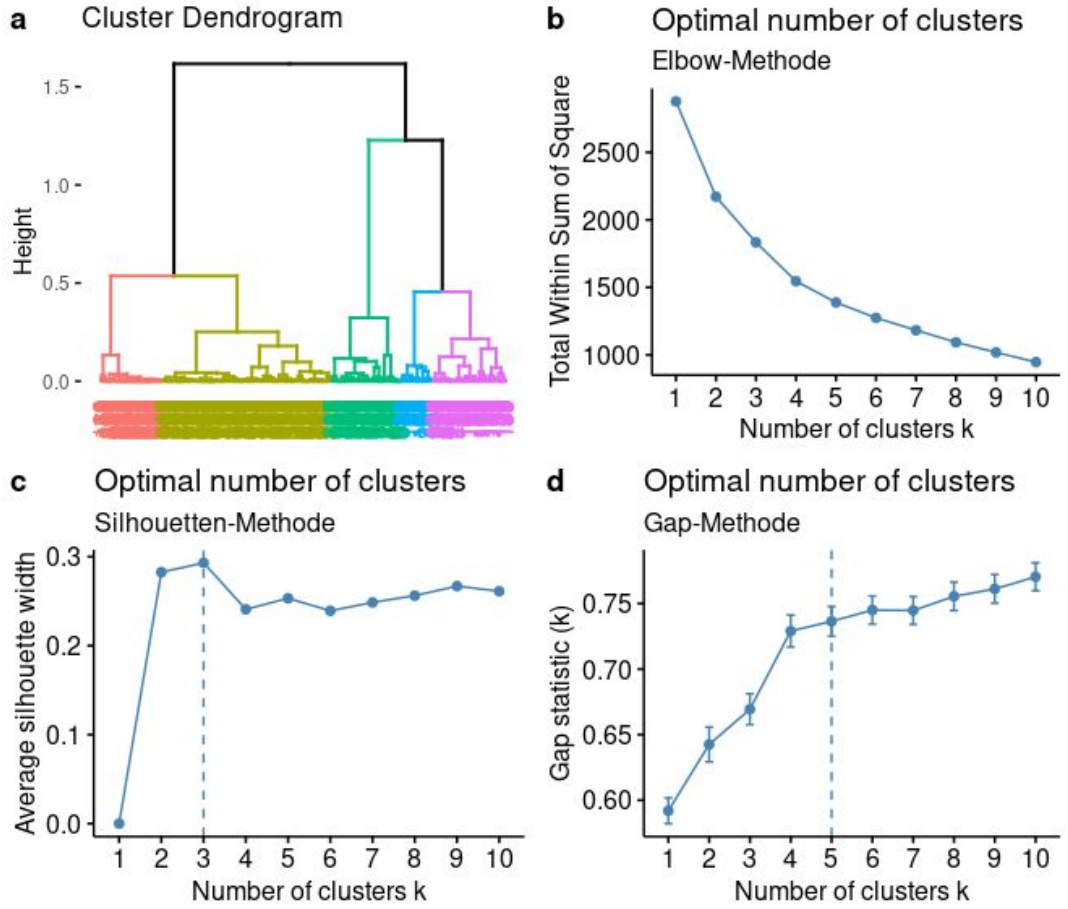
R-Analysis by Sophie C. Schmidt (curr. PhD stud.)





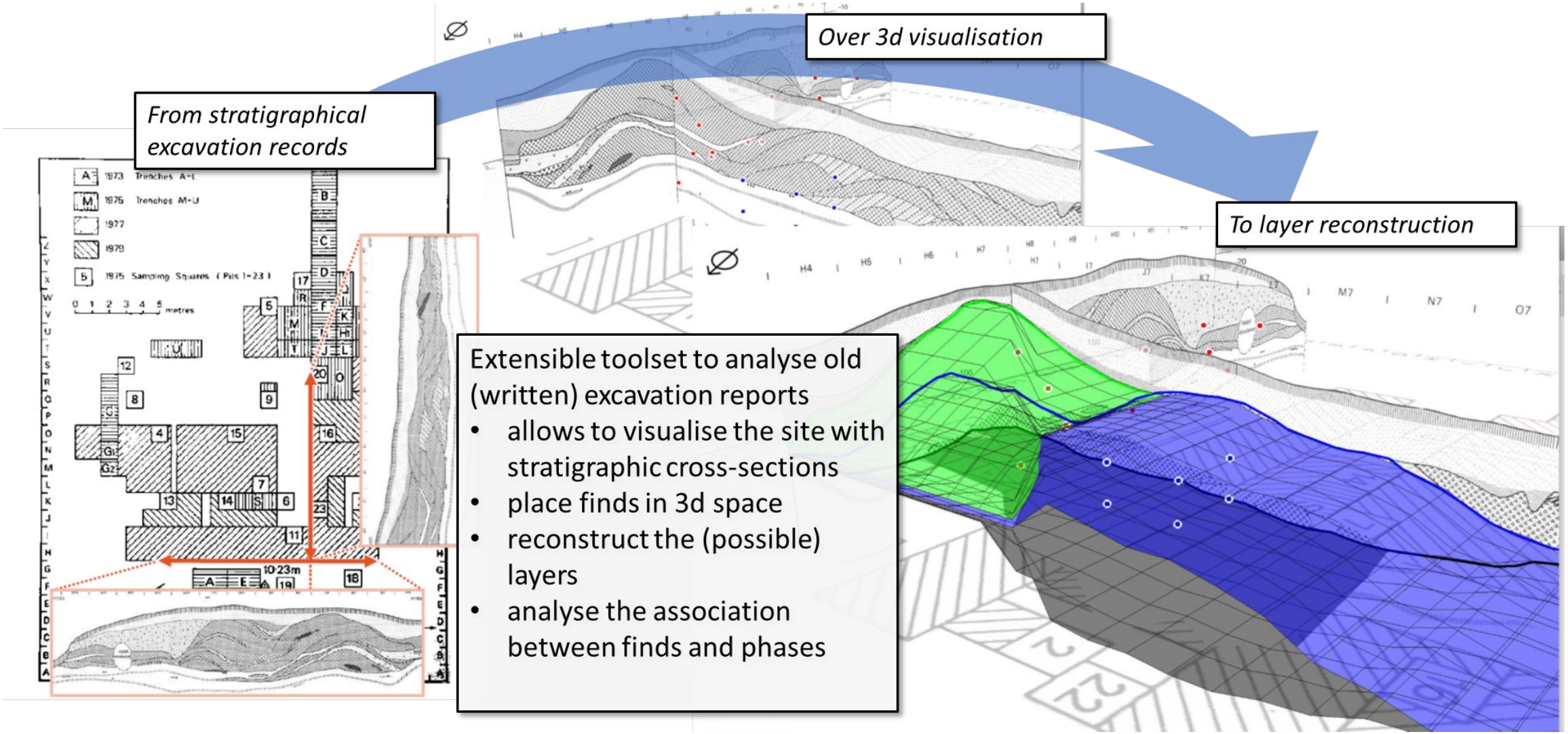


Usage of the HCPC-function of the R-package FactoMineR (a combination of PCA, hierarchical clustering and k-means clustering) on archaeometallurgical data



Determination of number of clusters by metrics

R-Analysis by Fabian Fricke

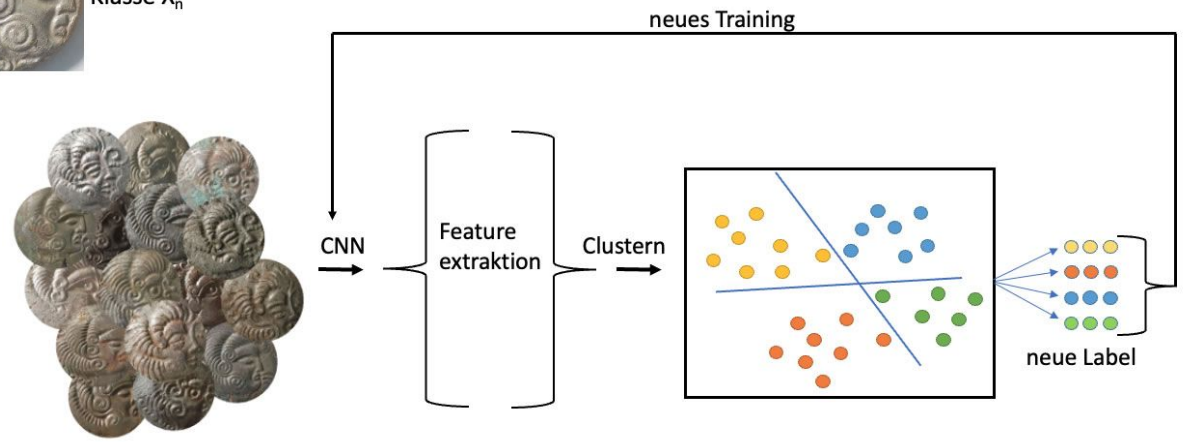


Stratigraphical Reconstruction using Python
by T. Noack, L. Schubert, A. Maier

Computational Archaeology
aka Archaeoinformatics

AI Technologies

Graphic: C. Deligio, Big Data Lab, ClaReNet. Coin images: N.N., Jersey Heritage, via https://clarenet.hypotheses.org/it-teil1



Graphic: C. Deligio, Big Data Lab, ClaReNet. Coin images: N.N., Jersey Heritage, via https://clarenet.hypotheses.org/it-teil2

Supervised / Unsupervised Learning & Neural Networks

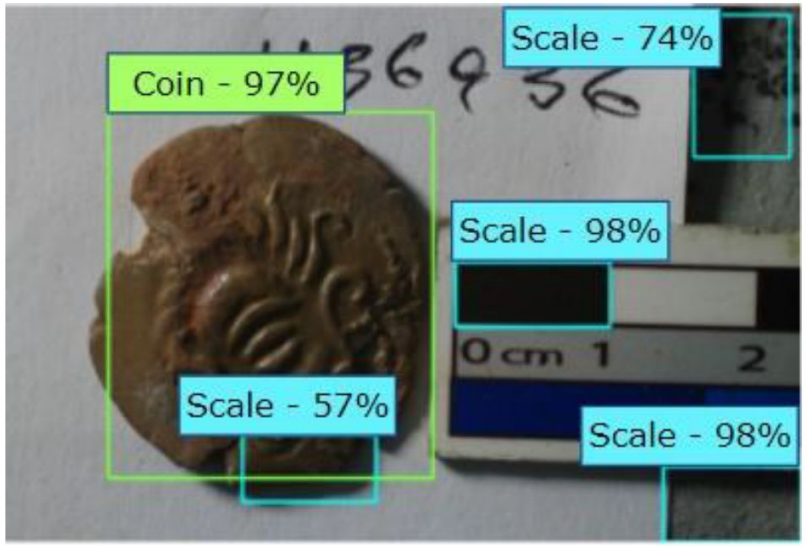


Figure 4 - Shadowy areas can lead to a wrong prediction by the model, resulting in an incorrect size calculation. (Photo: Jersey Heritage. Graphic: C. Deligio, Big Data Lab)

Images taken from https://zenodo.org/doi/10.5281/zenodo.8301464 . Chrisowalandis Deligio, Karsten Tolle, & David Wigg-Wolf. (2023, August 30). Supporting the analysis of a large coin hoard with AI-based methods.

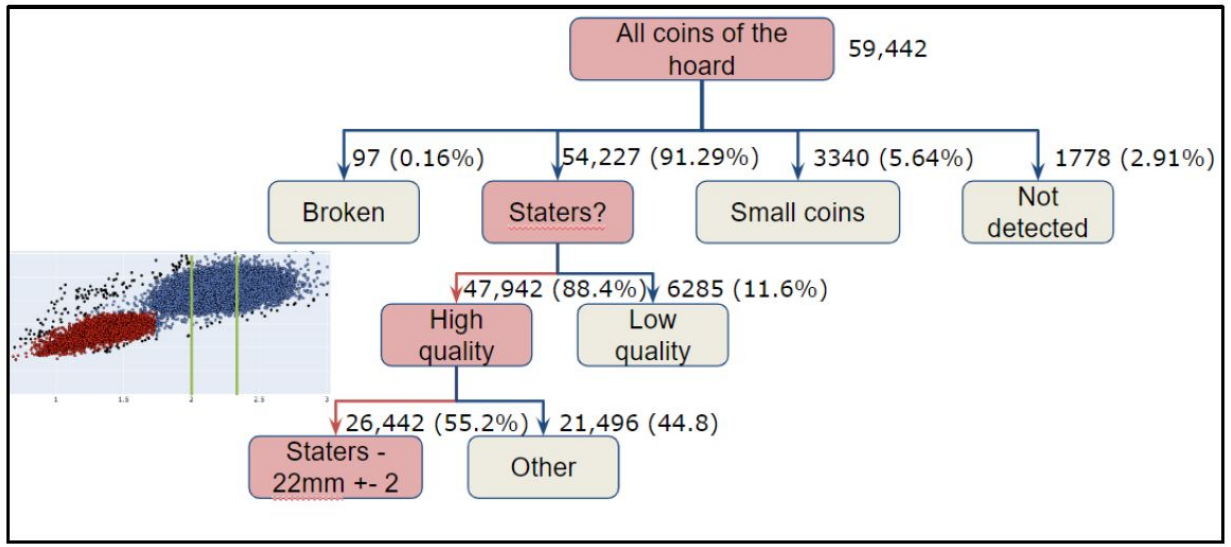


Figure 8 - Using the divide and conquer methodology, the data set could be divided step by step into more easily analysable parts. (Graphic: C. Deligio, Big Data Lab)

Object Detection & Unsupervised Learning

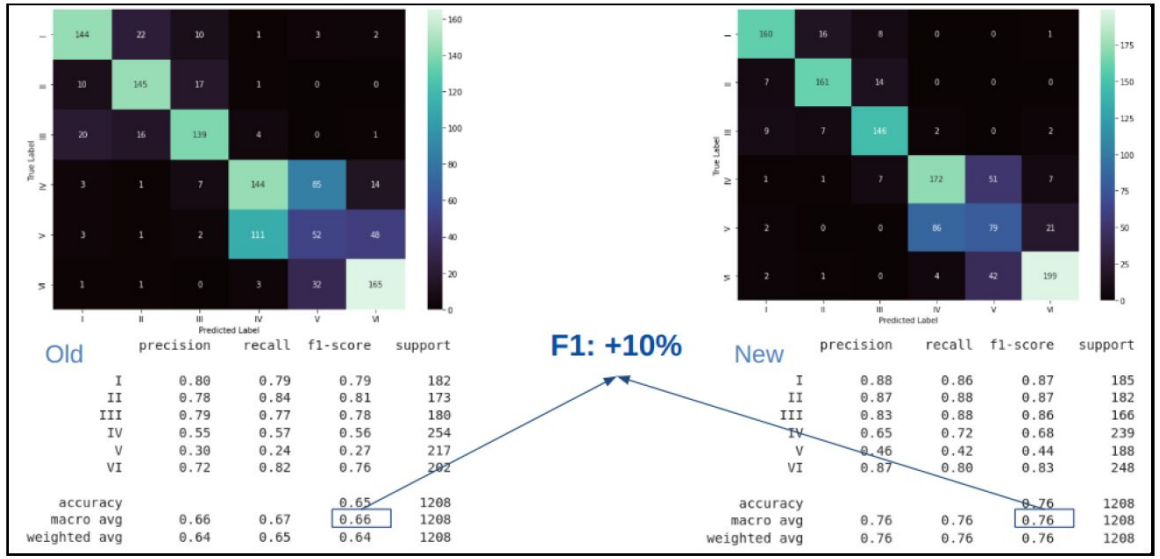


Figure 10 - Same predictions, different results. Comparison between two classifications (old vs revised). (Graphic: C. Deligio, Big Data Lab)

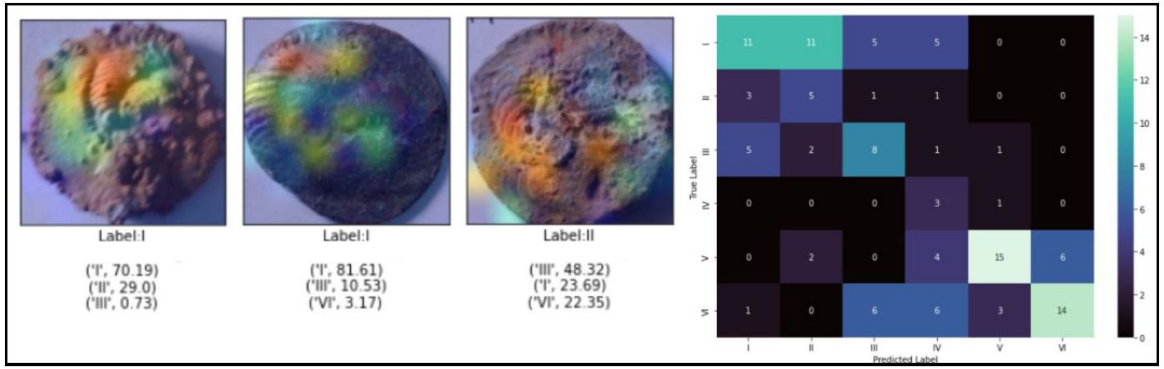


Figure 11 - On the left is the visualisation of the prediction with the top 3, on the right is the matrix of the 120 predictions. An F1 value of 44% and an accuracy of 47% were achieved. (Photos: Jersey Heritage. Graphic: C. Deligio, Big Data Lab)

Images taken from https://zenodo.org/doi/10.5281/zenodo.8301464 . Chrisowalandis Deligio, Karsten Tolle, & David Wigg-Wolf. (2023, August 30). Supporting the analysis of a large coin hoard with AI-based methods.

From unsupervised to supervised learning

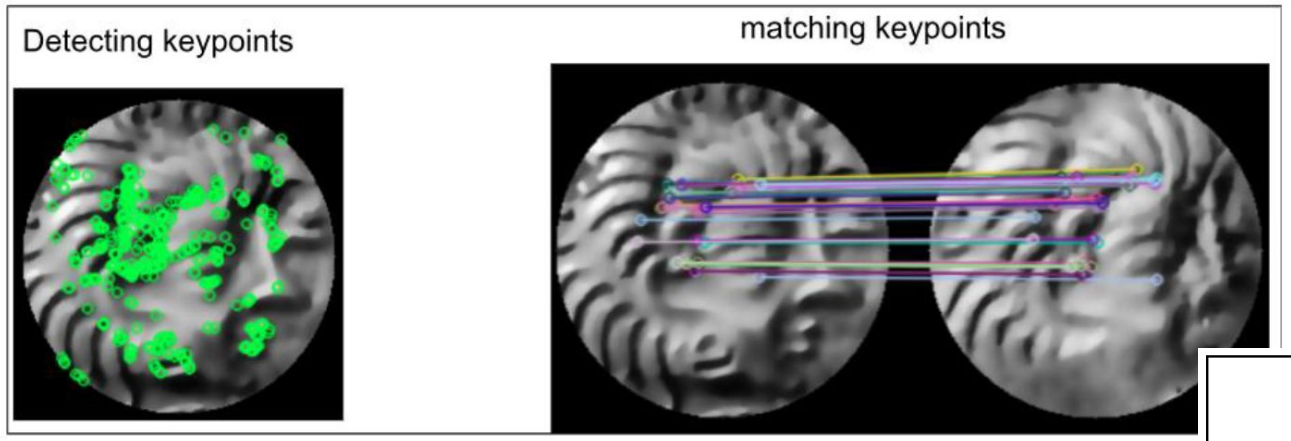


Figure 14 - On the left, an image with detected keypoints. On the right, an example of two supposedly identical die pieces (according to the GT) and their matches. (Photos: Jersey Heritage. Graphic: C. Deligio, Big Data Lab)

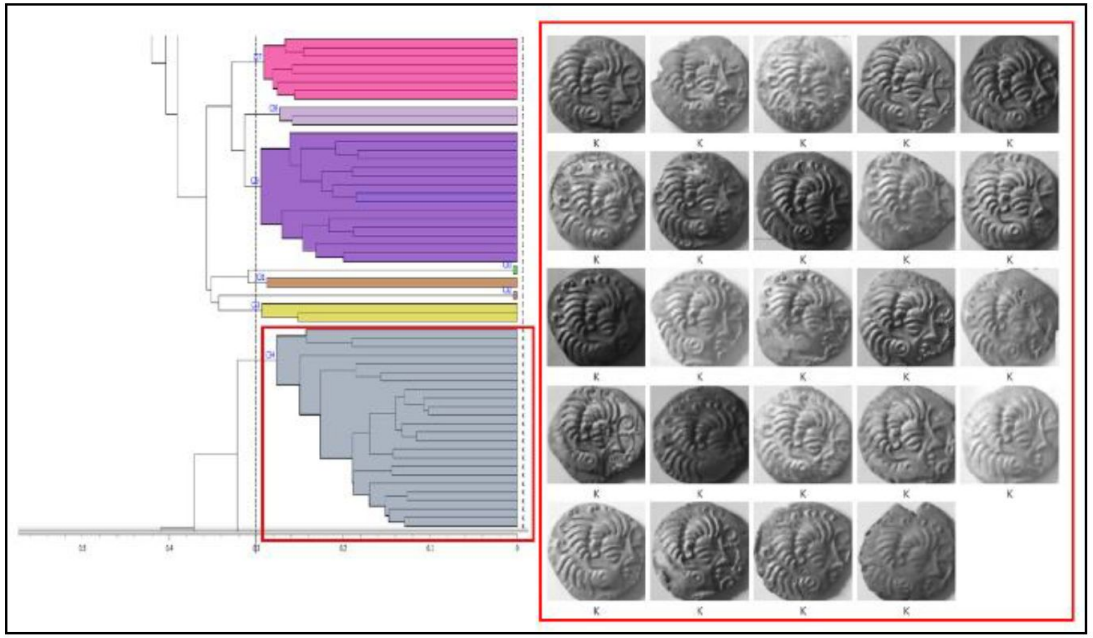
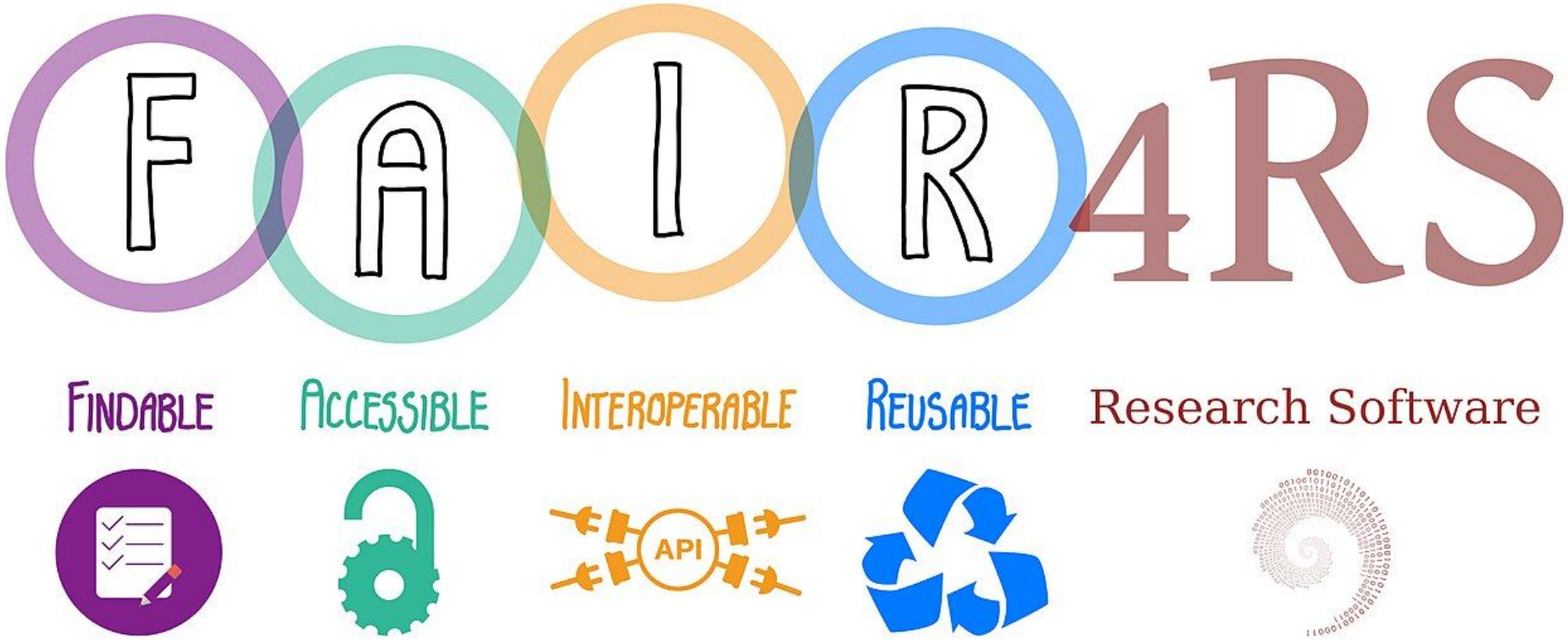


Figure 15 - On the left, part of the dendrogram. On the right, an overview of a cluster successfully containing the coins of one die. (Photos: Jersey Heritage. Graphic: C. Deligio, Big Data Lab)

Images taken from https://zenodo.org/doi/10.5281/zenodo.8301464 . Chrisowalandis Deligio, Karsten Tolle, & David Wigg-Wolf. (2023, August 30). Supporting the analysis of a large coin hoard with AI-based methods.

Implementing a die study

Summary



Florian Thiery, Heidi Seibold, CC BY 4.0, via Wikimedia Commons

Join us and help us to strengthen RSE in the Digital Humanities and Computational Archaeology also within the NFDI.

Thank you for your attention.

Questions?

NFDI4Objects

Research Data Infrastructure for the Material Remains of Human History

DOI 10.5281/zenodo.10774878