National Initiatives for Open Science in Europe

Speaking FAIR implementation

moving from recommendations to supporting practical implementation by service providers

FAIRSFAIR
Fostering Fair Data Practices in Europe

or: Why implementation matters

- From FAIR data to FAIR metadata
- Allows the cooperation of human and machines

- Framework for FAIR semantics → FAIRsFAIR
- Implementation of FAIR framework → NI4OS-Europe

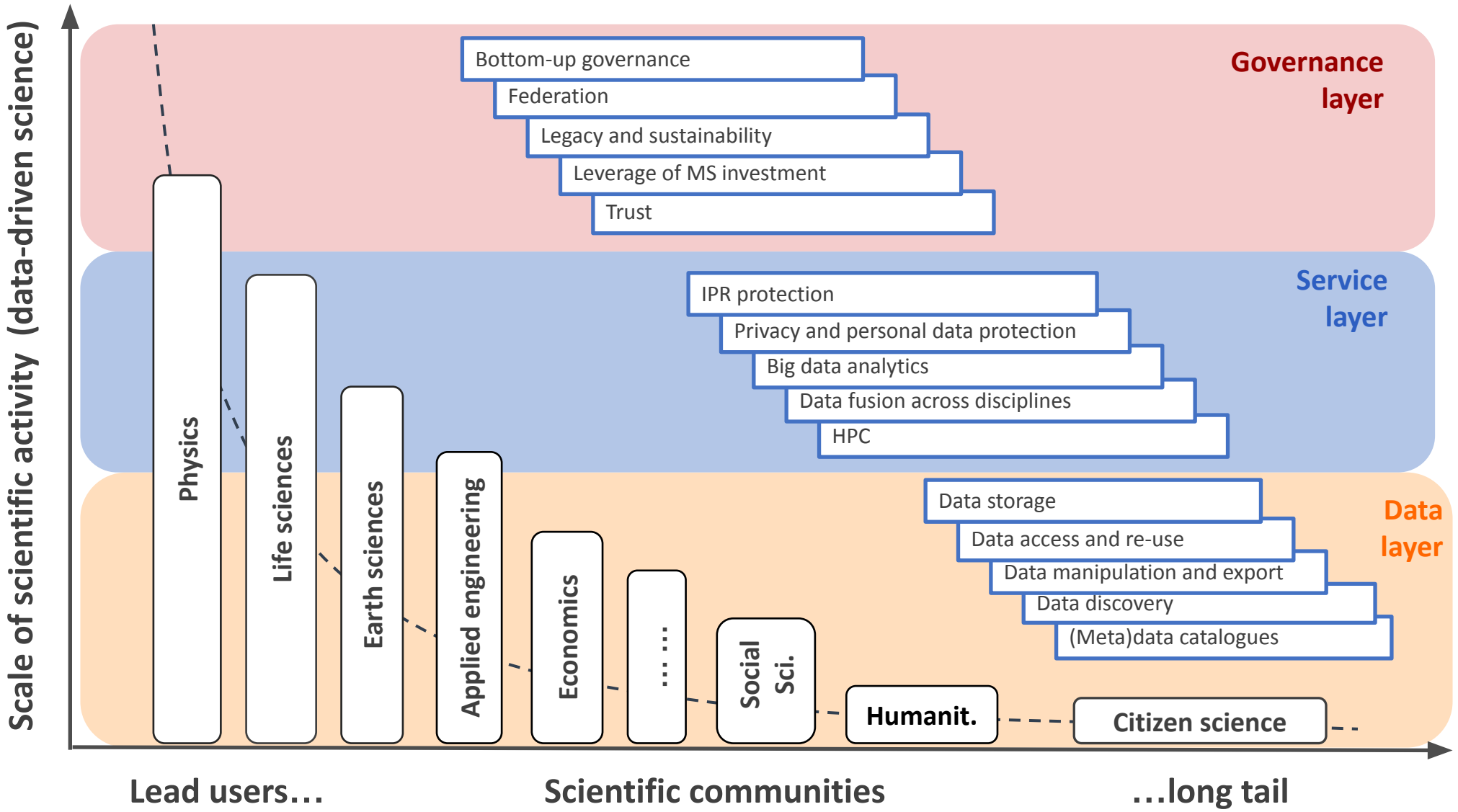
- Best practices for FAIR implementation
- Integration pathways
- Onboarding of services to EOSC

- NI4OS-Europe & FAIRsFAIR collaboration

Agenda

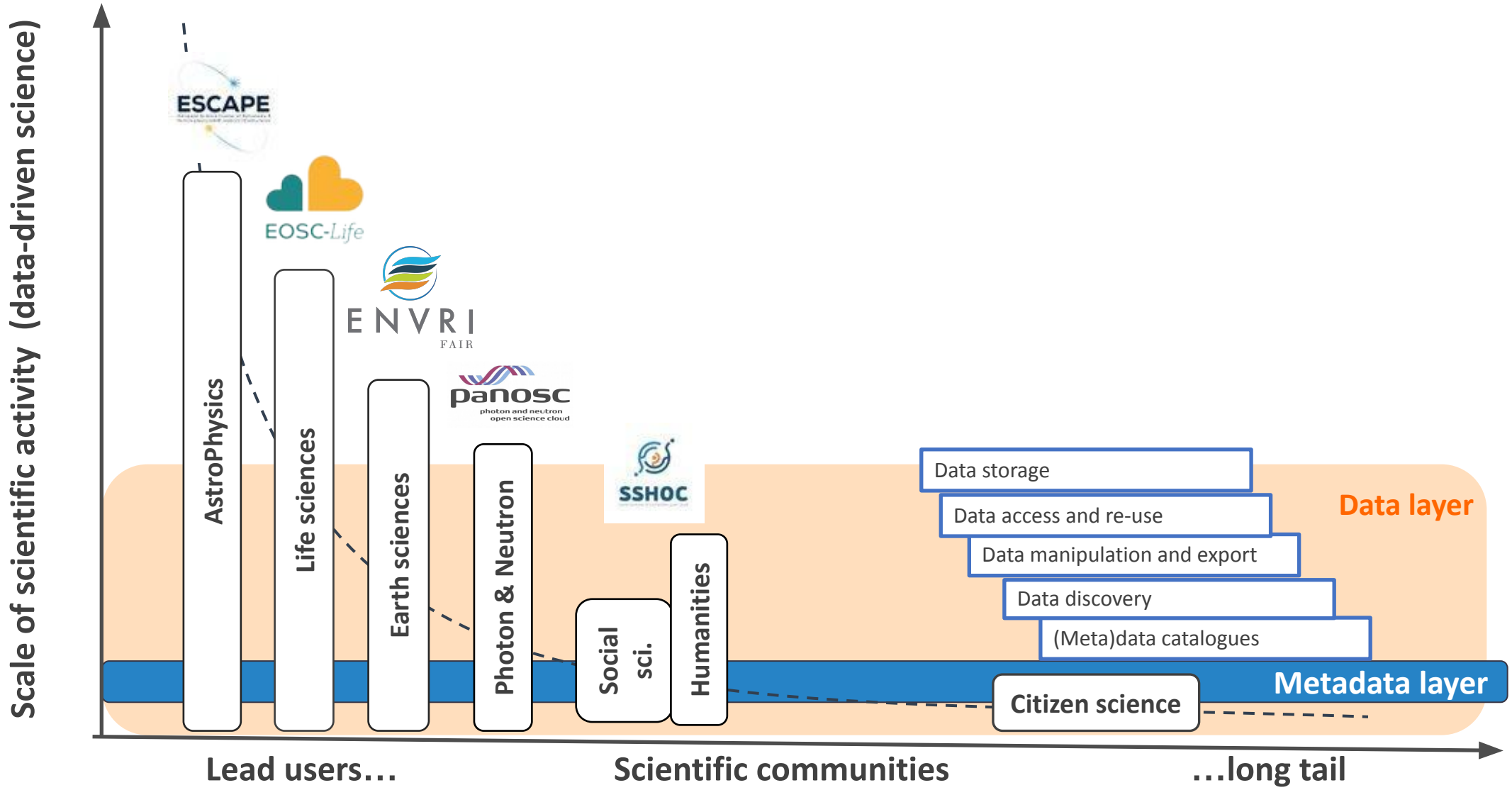
Welcome from the Director of NI4OS Europe	Eleni Toli, University of Athens / ATHENA
Agenda, & brief intro to the topic	Gerard Coen, DANS
Let's speak! Mentimeter session	Elli Papadopoulou, OpenAIRE / ATHENA
FAIRsFAIR "Recommendations for FAIR Semantics"	Gerard Coen, DANS
"NI4OS-Europe: servicing the service providers"	Andreas Athenodorou, CY
Presentations from the NI4OS Semantics Expert Group: ● 'A semantic knowledge integration framework for interdisciplinary research communities' ● 'FAIR semantics a case study' at the University of Debrecen	Valentina Vassalo, CY Ádám, Száldobágyi, Uni Debrecen
Q&A with Mentimeter	Elli Papadopoulou, OpenAIRE / ATHENA
Wrap-Up	ALL

Scientific landscape for EOSC



Based on Burgelman

Semantics, metadata, and ontologies are a core component for interoperability



Based on Burgelman

Many challenges

Metadata are like toothbrushes...

…everyone thinks that it is a good idea to have it, but nobody wants to use someone else's.

Problems
Lack of (or overabundance of)

P1: explicit definitions
P2: common semantics (general ontologies)
P3: reference repository
P4: common metadata scheme across communities
P5: metadata models

Needs

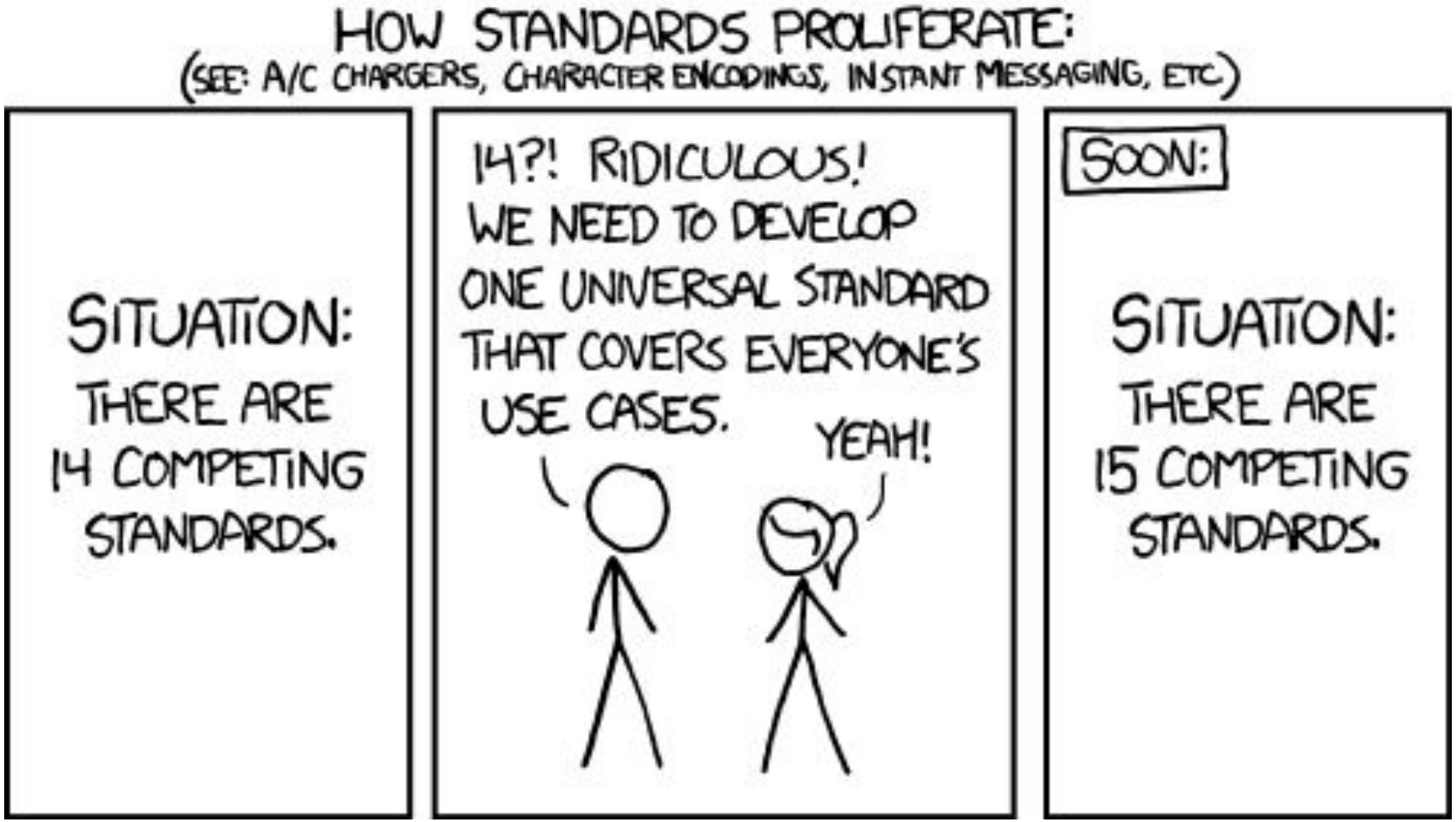
N1: principle approaches/tools for ontology and metadata schemes
N2: harmonisation across disciplines
N3: harmonisation of data of the same type
N4: federated access to existing research data repositories

Recommendations

R1: definitions of concepts, metadata and data schemes
R2: creating semantic artefacts with open licenses
R3: associated documentation for semantic artifacts
R4: repositories of semantic artefacts
R5: minimum metadata model and cross walks discovery
R6: extensible options for disciplinary metadata
R7: apply a broad definition of data (datasets, workflows, lab protocols, software, methods, hardware design, etc.)
R8: clear protocols and building blocks for catalogues

Summary of the Problems, Needs and Recommendations for Semantic Interoperability (EOSC Interoperability Framework)

Toothbrush effect of metadata



Source: https://xkcd.com/927/ *(CC BY-NC 2.5)*


FAIRSFAIR
Fostering Fair Data Practices in Europe


NI4OS
Europe

OPEN SCIENCE FAIR

Speaking FAIR implementation for service providers session

Go to

www.menti.com

Enter the code

58 93 72 7

Or use QR code

FAIRSFAIR

Fostering Fair Data Practices in Europe

Recommendations for FAIR Semantics:
Approaches for enabling I2 of the FAIR Principles

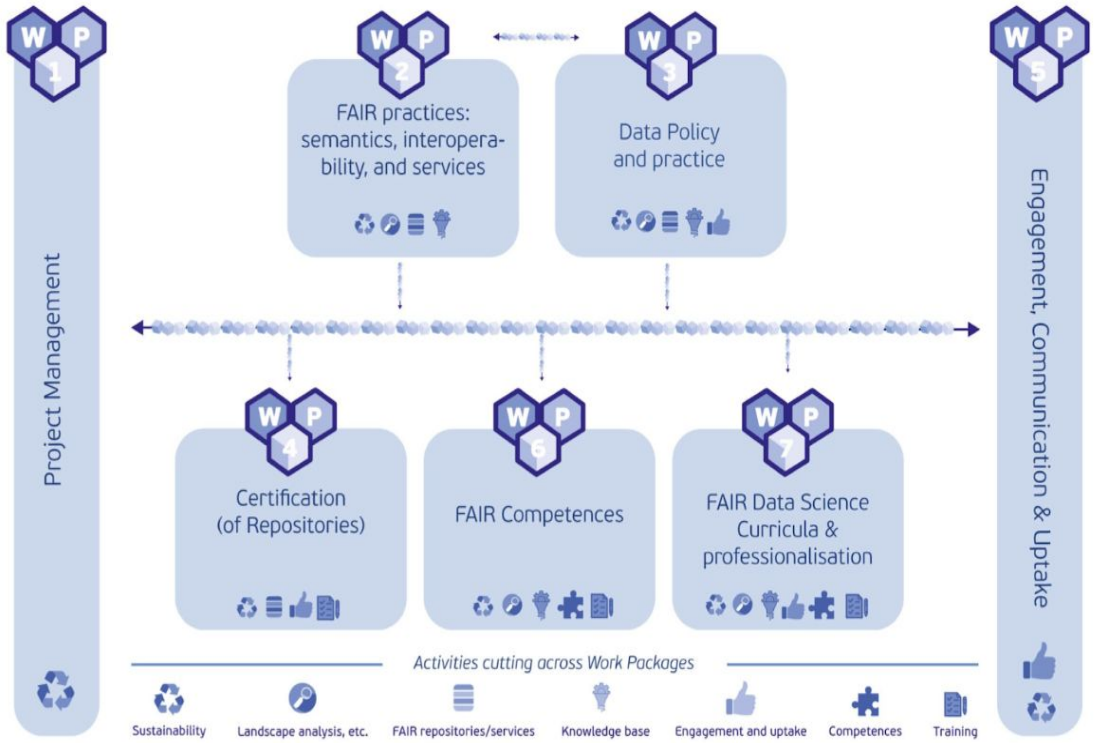
Gerard Coen (DANS), Open Science FAIR, 21st September 2021, Virtual

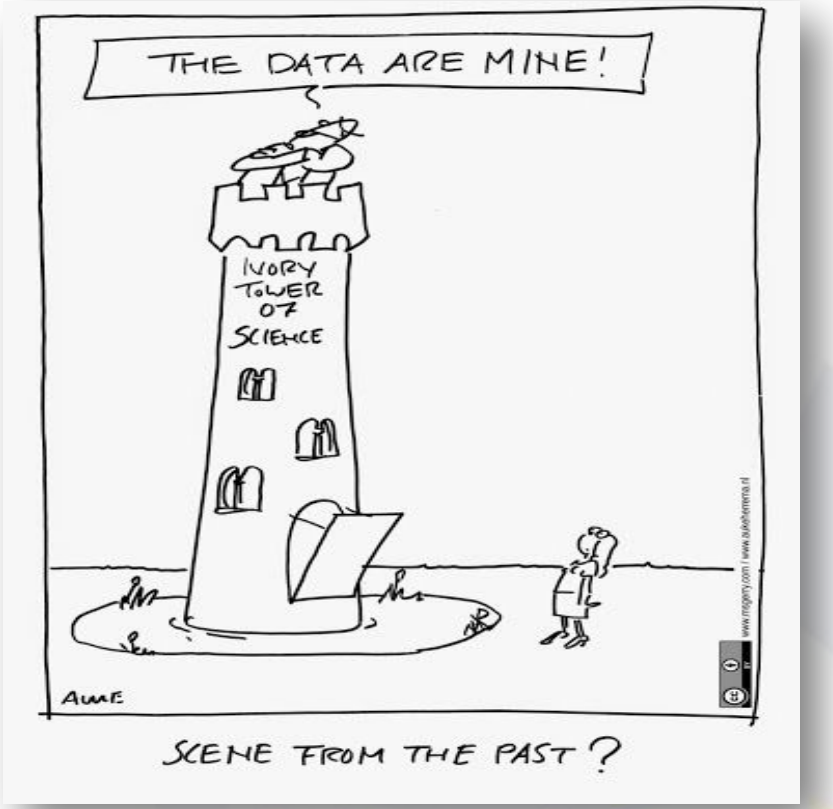
@gerardcoen
ORCID: 0000-0001-9915-9721

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

FAIRsFAIR overarching objective

To supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. Emphasis is on fostering FAIR data culture and the uptake of good practices in making data FAIR.





What is meant by FAIR Semantics?

"Semantic artefact" is a broad term including research resources such as ontologies, terminologies, taxonomies, thesauri, vocabularies, metadata schemas, and standards.

They are key components to enabling FAIR but these semantic artefacts have themselves have to be FAIR.

"FAIR Principle I2: (Meta)data use vocabularies that follow FAIR principles"

"FAIR Semantics" are semantic artefacts which adhere to the FAIR principles.

The end goal of the FAIR Semantics team...

...is to co-create both recommendations for making semantic artefacts FAIR, and a set of agreed best practices to follow together with the semantics community at large

- Based on community input & feedback processes

- ***First version*** released in March 2020
  - DOI: 10.5281/zenodo.3707984.
    - ***Second version*** released in January 2021
      - DOI:10.5281/zenodo.4314320
        - ***Third version*** due to be released in February 2022

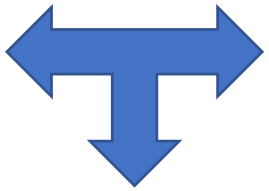
- 17 Priority Recommendations & 14 Best Practices Recommendations

One Example - Priority Recommendation #3

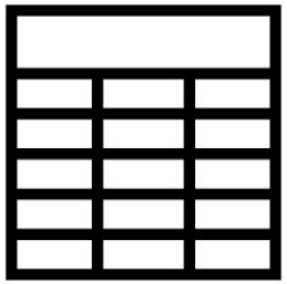
P-Rec 3: « A common minimum metadata schema **must** be used to describe semantic artefacts and their content »

A guiding use-case: searching for ontology

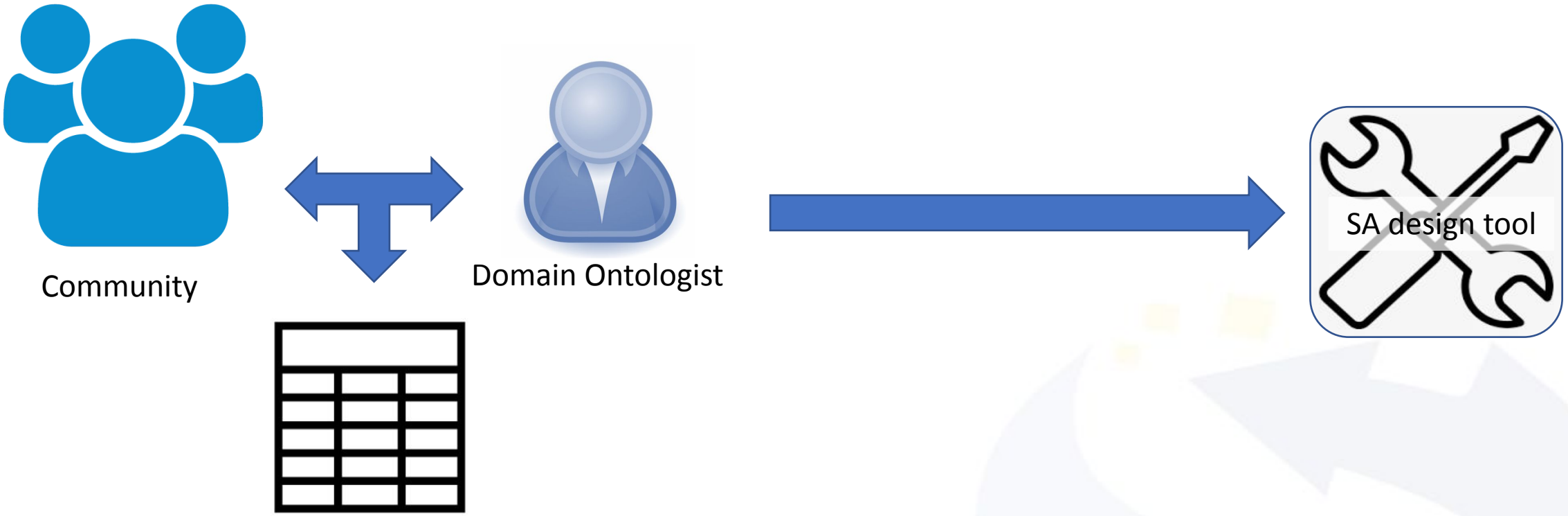
Community



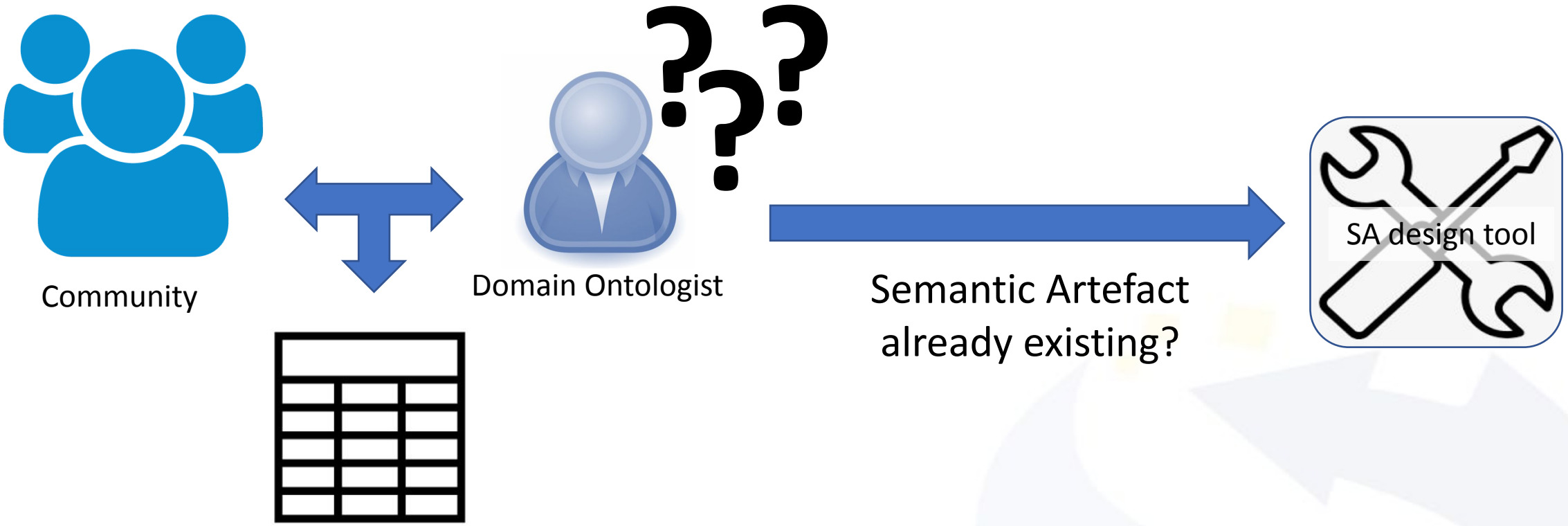
Domain Ontologist



A guiding use-case: searching for ontology



A guiding use-case: searching for ontology



A search engine for Semantic Artefacts

Domain Ontologist

Semantic artefact ??

A search engine for Semantic Artefacts

Domain Ontologist

Semantic artefact ??

Natural Environment Research Council

National Oceanography Centre
British Oceanographic Data Centre BODC

The NERC Vocabulary Server (NVS)

Service Status

BioPortal

AgroPortal

EcoPortal

Research Data Australia

Examples of the changes 1st >> 2nd version (D2.5)

D2.2 >	*P-Rec. 3: Use a common minimum metadata schema to describe semantic artefacts and their content*	F2, R1.1, R1.2 and R1.3

D2.5 >	Mandatory	P-Rec. 3: A common minimum metadata schema must be used to describe semantic artefacts and their content	F2, R1.1, R1.2 and R1.3	Metadata

Alignment with RFC 2119 (9 MUST, 7 SHOULD, 1 MAY Recommendations)

Examples of the changes 1st >> 2nd version (D2.5)

D2.2 >	*P-Rec. 4: Publish the Semantic Artefact and its content in a semantic repository*	F4

D2.5 >	Optional	*P-Rec. 4: Semantic Artefact and its content should be published in a trustworthy semantic repository*	F4	Repository

Integrating the idea of trustworthiness for repositories. Proposed as an 'Optional' requirement.

*CoreTrustSeal plus FAIR overview: https://doi.org/10.5281/zenodo.4003630

*See: Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

** Relate to trustworthiness in data repositories but the notion of adherence to quality standards, certification and compliance is still relevant.*

Ways to get involved

Workshop to discuss the future of the recommendations
(in planning Winter 2021)

Suggestions, contributions, and feedback are always welcome...

GitHub

GitHub is being used by the team to collect feedback:
https://github.com/FAIRsFAIR/FAIRSemantics

Use of labels:

Please use "Clarification Needed" where you feel like a recommendation lacks clarity

Please use "Relevance" to comment on the relevance (or lack thereof) for the stakeholder you represent.

Please use "Implementation Example" to suggest practical implementations or initiatives that are missing for this recommendation.

It is also possible to submit problems encountered, suggestions, questions, recommendation proposals etc. as issues.

Working with the RDA VSSIG

TG - Minimum metadata for FAIR Semantic Artefacts and DCAT profiles

Lead: Clement Jonquet (INRAE)

- Defining a minimum metadata schema for minimally FAIR semantic artefacts
- Defining a DCAT profile for semantic artefacts

TG - FAIR Semantic Repositories

Lead: Alexandra Kokkinaki (BODC) & Gerard Coen (DANS)

- Evaluate the recommendations from the perspective of repositories & service providers
- Establish a list of technical implementations for the recommendations

Email: alexk@bodc.ac.uk & gerard.coen@dans.knaw.nl

Acknowledgements

FAIR Semantics team:

Yann le Franc (e-Science Data Factory) - Task Lead T2.2

Jessica Parland-von Essen, Hanna Koivula (CSC)

Gerard Coen (DANS)

Luiz Bonino (GOFAIR, UTwente)

Thank you!

www.fairsfair.eu

@FAIRsFAIR_EU

In/fairsfair

info@fairsfair.eu

NI4OS-Europe: Servicing the Service Providers

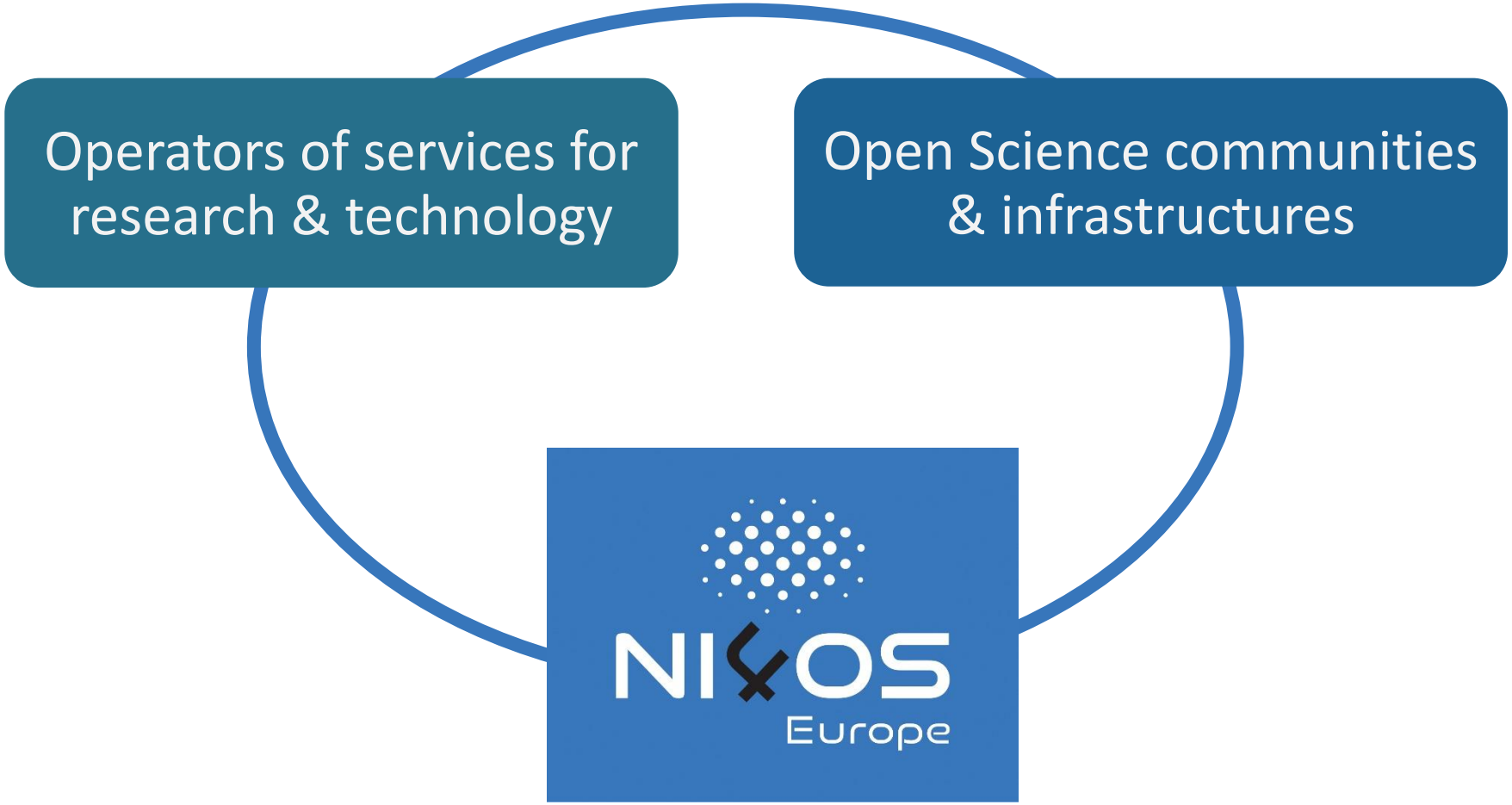
Open Science Fair 2021

Andreas Athenodorou
NI4OS-Europe WP6 leader
The Cyprus Institute

NI4OS-Europe

15 Member States and Associated Countries | 22 Partners

Partnership building blocks



Mission

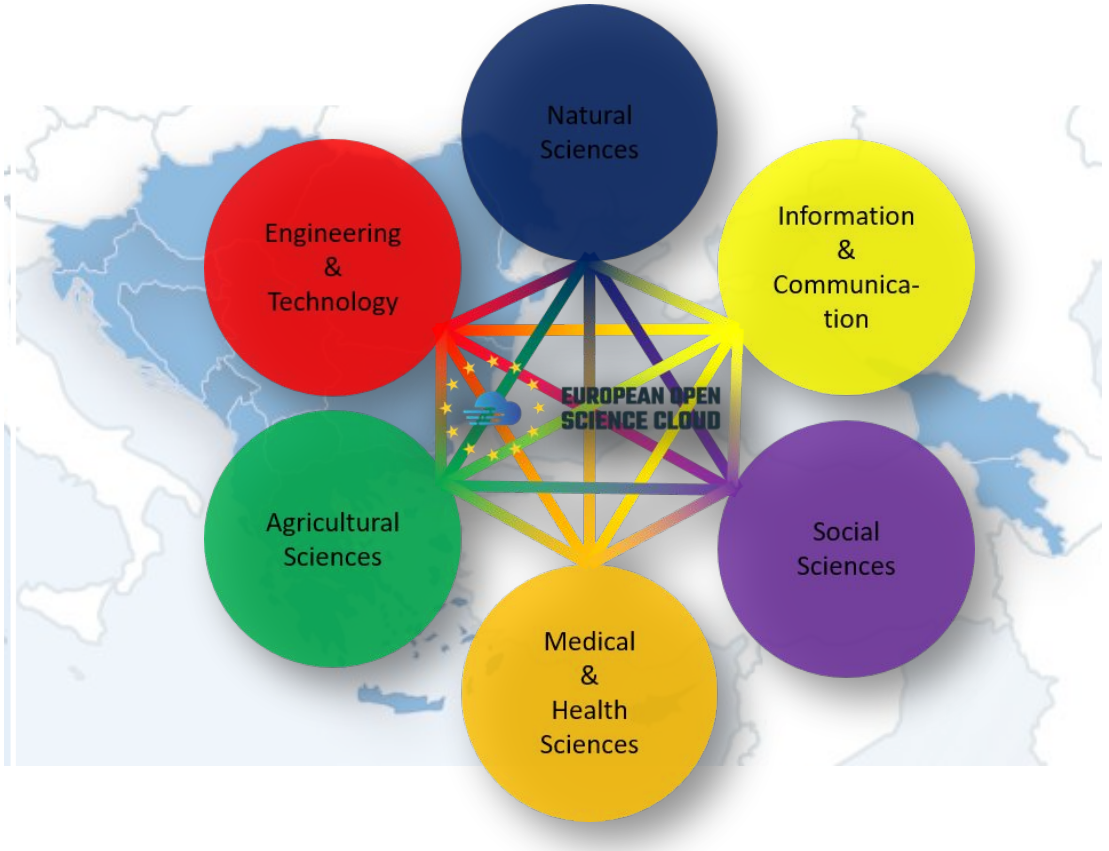
Support the development and inclusion of the national Open Science Cloud (OSC) initiatives in 15 Member States and Associated Countries in the overall scheme of EOSC governance

Spread the EOSC and FAIR principles in the community and train it

Provide technical and policy support in on-boarding of the existing and future service providers into EOSC

NI4OS-Europe supports OPEN SCIENCE!

- We support OPEN SCIENCE
- By providing all the necessary tools to the **LONG TAIL OF SCIENCE** throughout the **EOSC**
- Servicing **ALL** possible disciplines
- Namely we support the on-boarding of
  - Thematic Services
  - Generic Services
  - Repositories
- We provide access to **ORDM tools**
- We provide training on **FAIR**

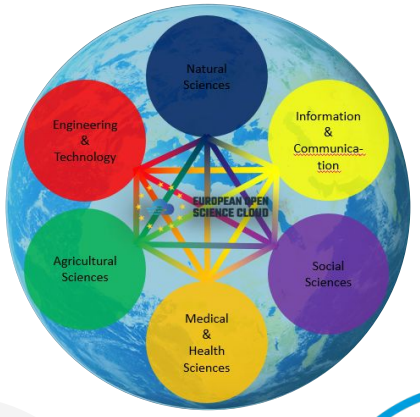


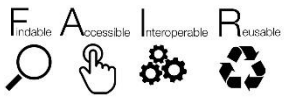
User engagement, training and demonstrators

Involving and supporting scientific communities

Ensuring take-up of core EOSC services in the community

- WP6





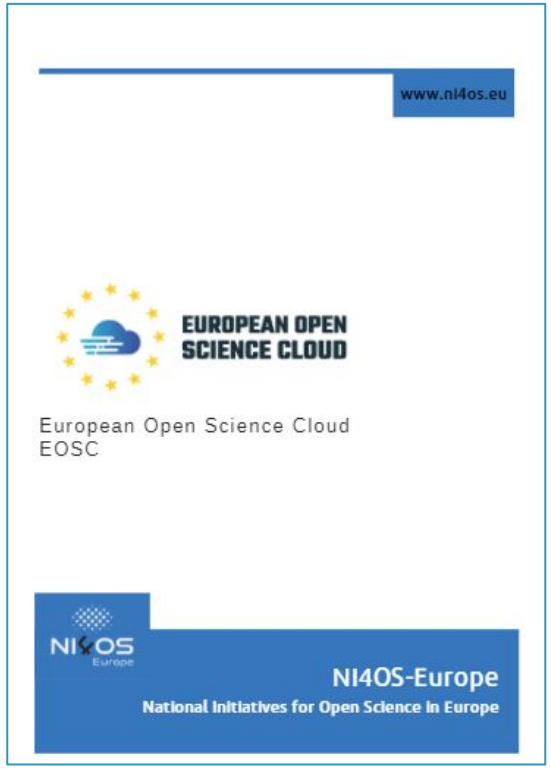
Promoting uptake of FAIR among research communities

Promoting and ensuring EOSC research outcomes through concrete support to users

Training for federated services, interoperability, ORDM principles, repository certification, FAIR

Support to EOSC service & FAIR uptake in communities

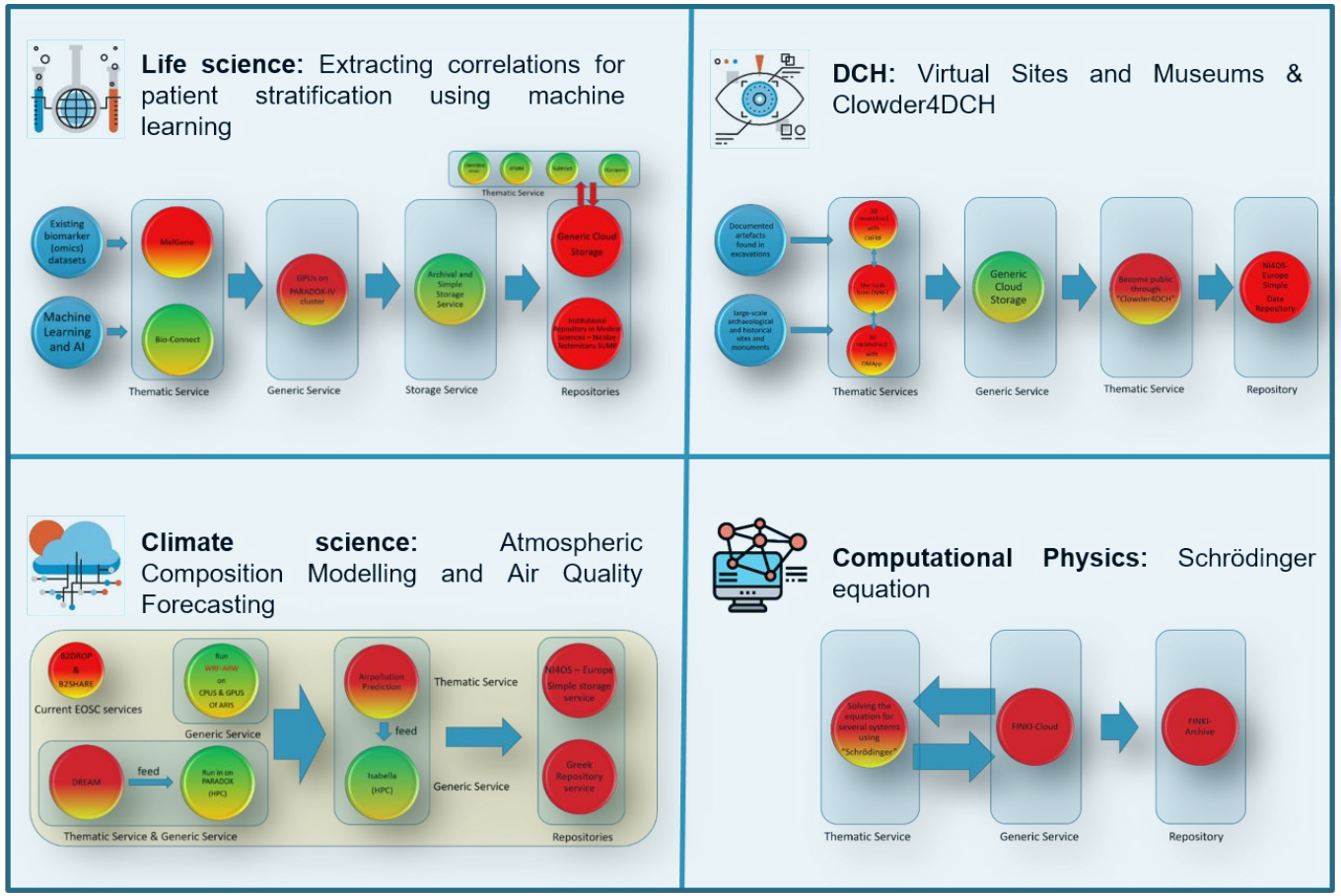
- Policy/Strategy: **Ambassadors** from each country assigned as EOSC promoters
- Training and dissemination: **material for FAIR and EOSC service uptake** is available in all different mother languages of the NI4OS-Europe area. Webinars for disseminating EOSC and FAIR principles in each country
- Infrastructures and tools: Provide **ORDM tools** and enhancing current practices



Flagship scientific communities

We have identify four highly cross-disciplinary, real-user communities:

- Life Science
- Digital Cultural Heritage
- Climate Science
- Computational Physics



EOSC candidate generic services: examples

- HPC Resources
  - CPU
  - GPU
  - Xeon Phi

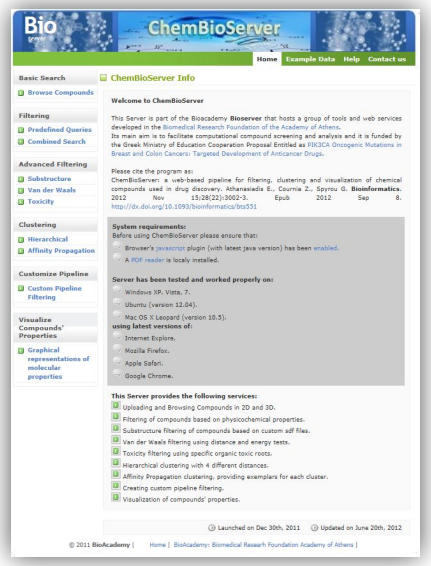
- Cloud Virtual Machines

- Generic Storage

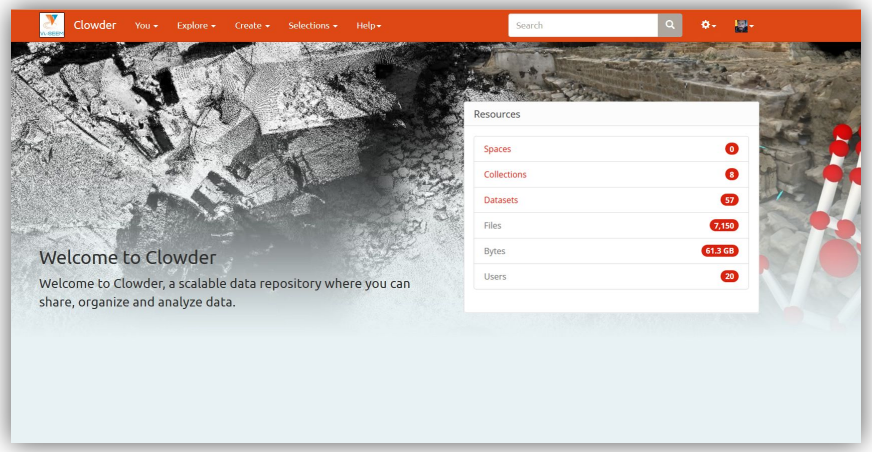
- Data management services (Archival, Repository, Data discovery, Hadoop on-demand, Data analysis service, Simple storage)

EOSC candidate thematic services: examples

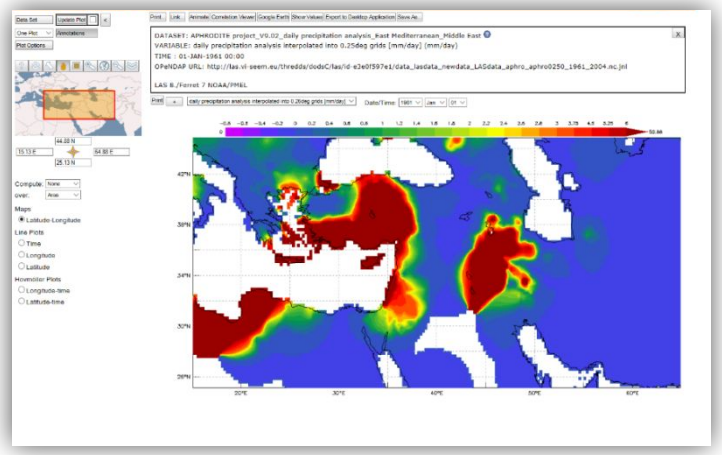
ChemBioServer



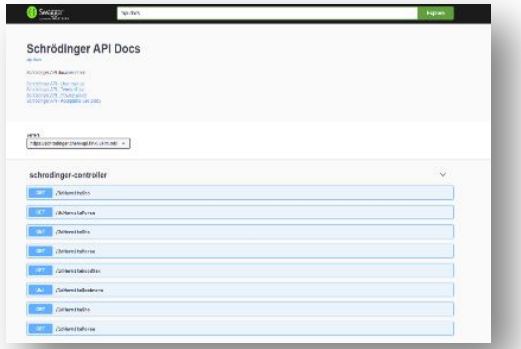
DCH Clowder



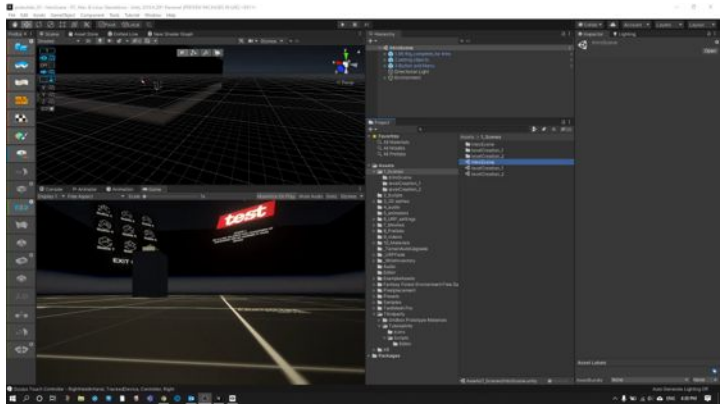
Live Access Server



Schrödinger API



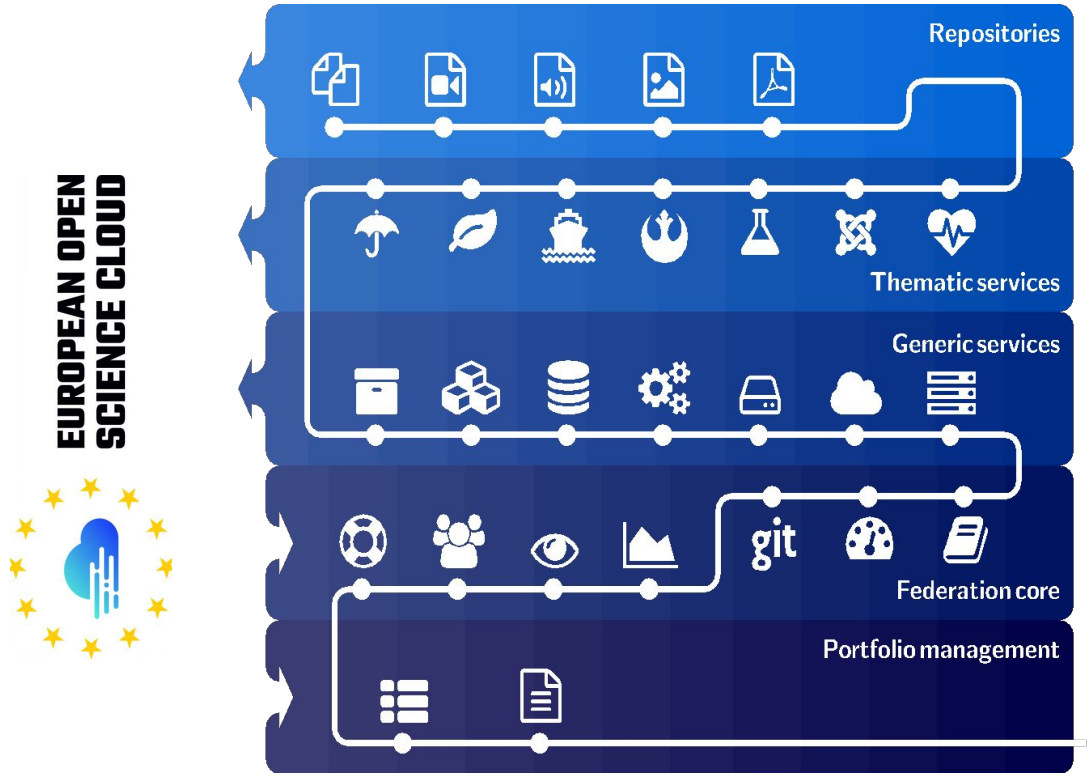
OVRET



Airquality

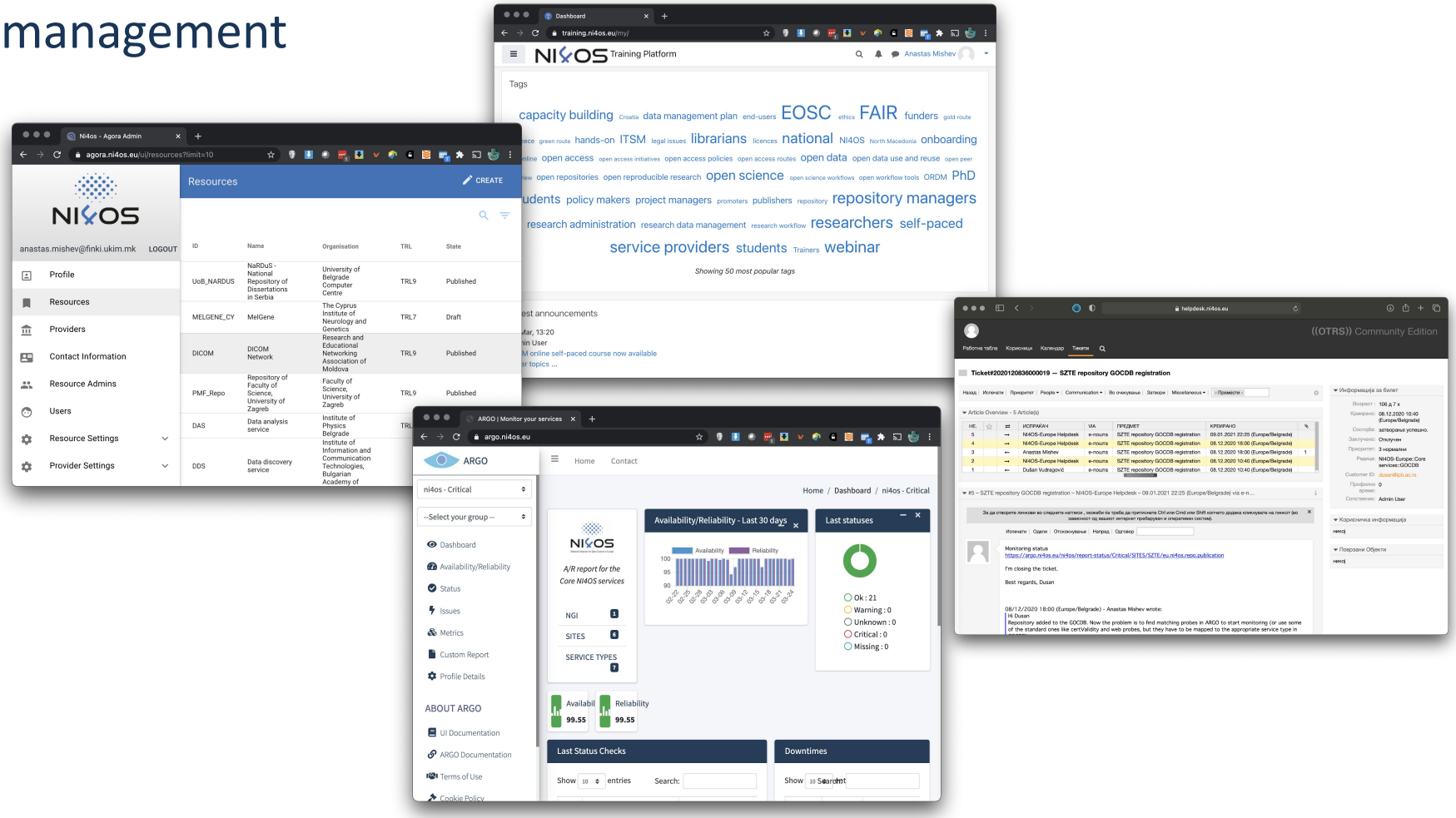
Service integration and onboarding

- Pre-production environment – validate readiness and maturity level for EOSC onboarding
- Service portfolio management system based on the EOSC provider and service profile
- Integration with federation core services
- Service categorization
- Onboarding of
  - generic services
  - thematic services
  - repositories



NI4OS-Europe pre-production environment

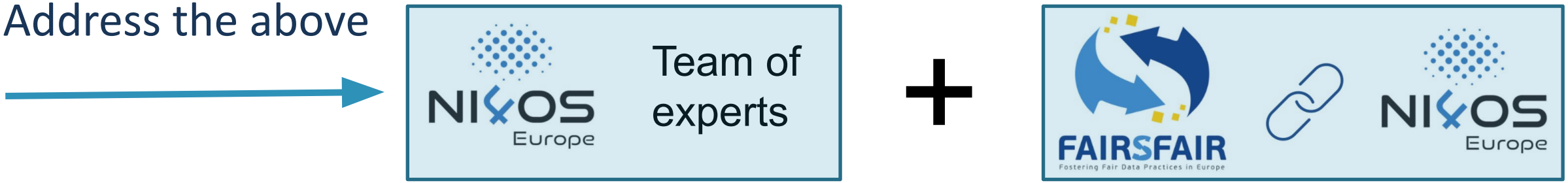
- Federating core
  - Service catalogue management system (AGORA)
  - AAI
  - Helpdesk
  - Monitoring
  - Accounting



How FAIR-enabling are the services?

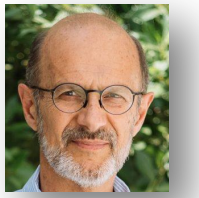
- Does a service support metadata?
- Does a service support semantics?
- Are there metadata standards available for a specific domain?
- How much FAIR-educated are the scientists developing a service?
- How about semantics interoperability?
- ….
- ….
- How do we facilitate the inclusion of semantics, metadata schemas, ontological solutions on a service at practical level?

- Address the above



NI4OS-Europe Task-Force on Semantics/Metadata/Ontologies

- NI4OS-Europe Team of experts:
  - Consists of academics with experience on metadata and controlled vocabularies



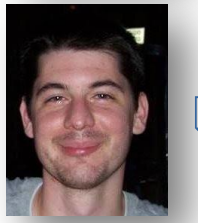
- Panos Constantopoulos (ATHENA)

- Agiatis Benardou (ATHENA)

- Georgios Artopoulos (CYI)

- Zoe Cournia (BRFAA)

- Valentina Vassallo (CYI)



- Adam Szaldobagyi (UD)

- Vicky Liakopoulou (ATHENA)

NI4OS-Europe Task-Force on Semantics/Metadata/Ontologies

- NI4OS-Europe Team of experts:
- Additional action to DoA
- Goals:
  - Provide more context on the foggy subject of FAIR-enabling services
    - Answer questions and increase awareness on the technicalities of FAIR implementation
    - Everyone to understand the basics of semantics and the role of metadata and controlled vocabularies (ontologies, taxonomies, etc)
  - Work with service providers to analyse the different types of metadata
    - Provide tailored advice for appropriate use in a research data management lifecycle
  - Contribute to some parts of the implementation of domain data protocols (to later be implemented in the Argos DMP service)
  - Contribute to the EOSC FAIR TF and EOSC FAIR Metrics and Data Quality TF conversations

NI4OS-Europe Task-Force on Semantics/Metadata/Ontologies

- Activities:
  - **Understanding the FAIR needs and readiness of services included in on-boarding by running a survey (Running now)**
  - Organisation of webinars for on-boarded service providers and repository managers
  - Providing a pathway on integrations and/or alterations that are necessary for enhancing existing services or new services
  - Feedback on ORDM tools (RePol, LCT, RoLECT)
    - LCT – License Clearance Tool: https://lct.ni4os.eu
    - RePol – Repository Policy Generator https://repol.ni4os.eu/
    - RoLECT – EOSC RoP Legal & Ethics Compliance https://rolect.ni4os.eu/

Thanks for your attention!

NI4OS_eu NI4OS ni .eu

A semantic knowledge integration framework for interdisciplinary research communities

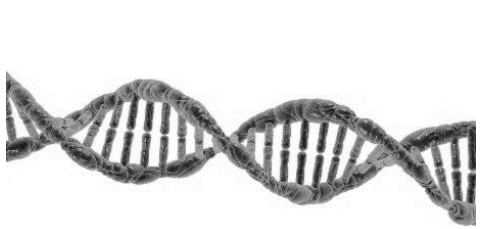
Data management and FAIR principles in the VI-SEEM case study

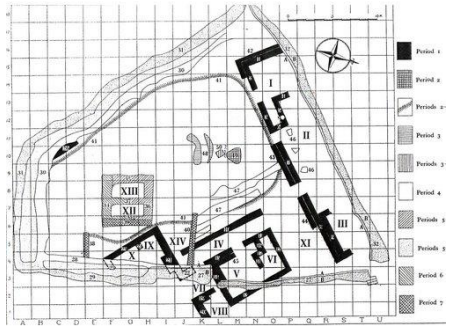
Open Science Fair 2021

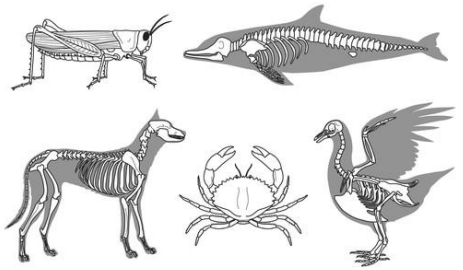
Valentina Vassallo
The Cyprus Institute

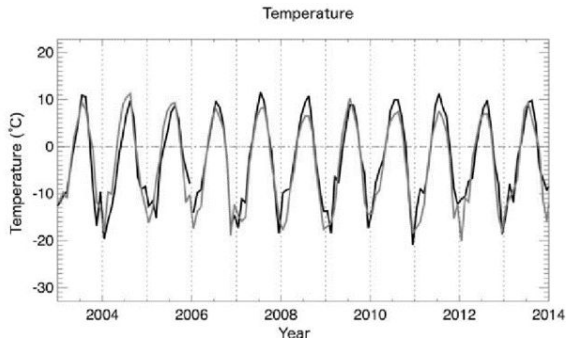
Multidisciplinary data and scientific communities

Various types of content coming from different scientific communities



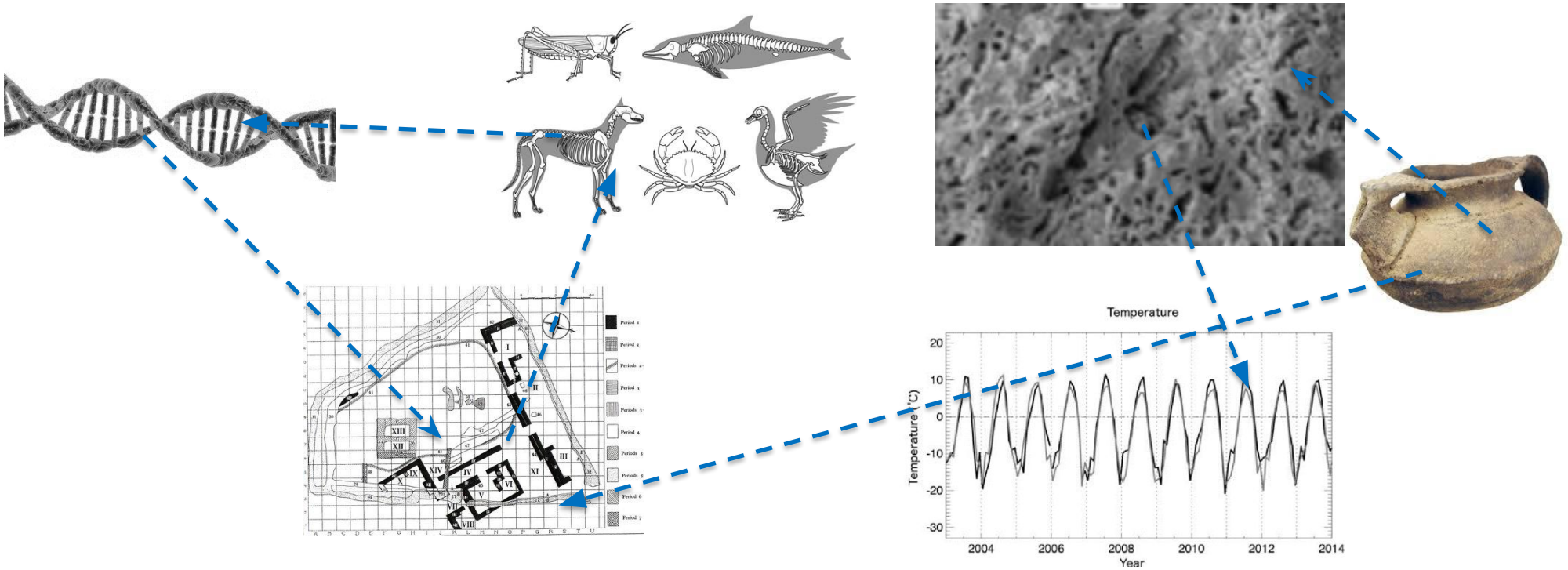






Every scientific community produces datasets that are documented following different approaches and according to different data descriptions

Knowledge communication framework



Establishing a knowledge communication framework that guarantees: a comprehensive description and documentation of cross-disciplinary digital resources, their long term preservation, publication, access, use and reuse

VI-SEEM case study

- NI4OS is the evolvement of VI-SEEM and builds on the research infrastructure previously established

That's FAIR!

Why are these aspects so important?

- **F** Findability
- **A** Accessibility
- **I** Interoperability
- **R** Reusability

Findability and Accessibility are guaranteed by the common access at the research infrastructure level, giving direct access to providers' data and enabling search for relevant information through their metadata, by linking them. The possibility to access data at the source enables to investigate data reliability and data transparency.

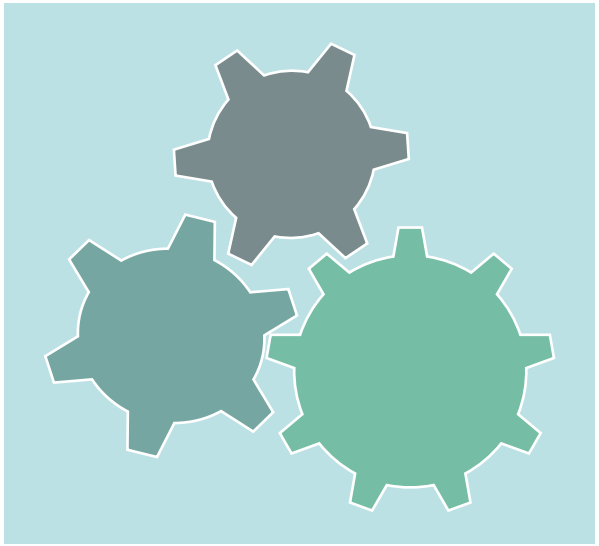
Interoperability is guaranteed by the use of an ontological solution at the research infrastructure level, aligning the multidisciplinary datasets homogeneously: it facilitates the management, integration, and access to research data by describing their semantic relationships.

Such choice brings to data Reusability, giving users the possibility to both find data within a specific research field and reuse them within other research communities: e.g., climate data to be reused for cultural heritage conservation purposes; biological data to trace the evolution of species, earth and agricultural data for the identification of archaeological remains, and so forth. The aim is to integrate data from heterogeneous sources and to efficiently discover further scientific data of interest, enabling the answering of complex queries that could not be answered from individual sources. Essentially, datasets can be reused for many purposes, not necessarily connected to the research domain of the user or related to the aims data was created for.

That’s FAIR!

How?

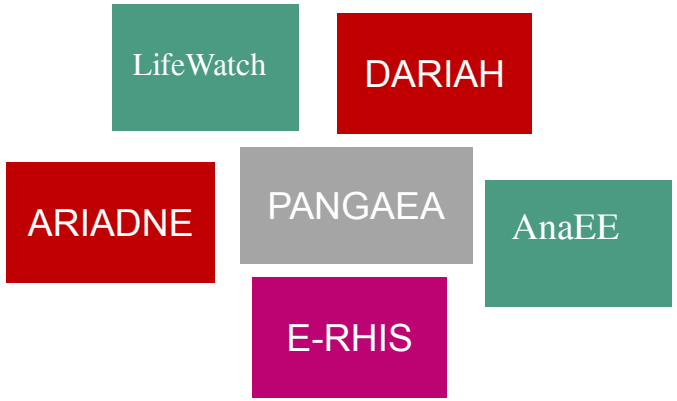
- Tools, standards and guidelines
- Harmonization and interoperability within and across communities and with core initiatives



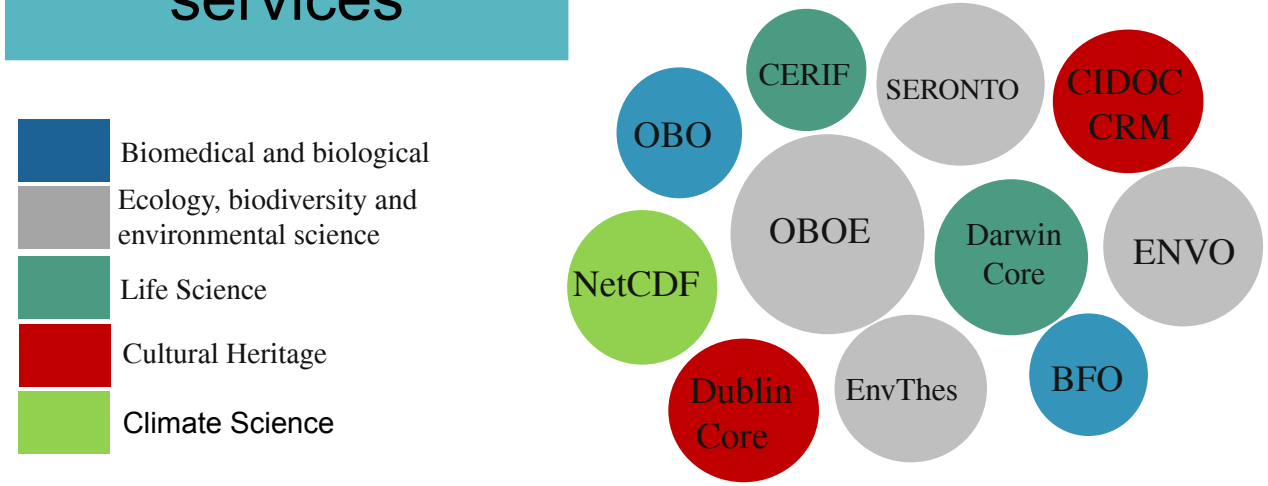
Background

Research Infrastructures (RI)

- State of the art about integration and interoperability tools and services



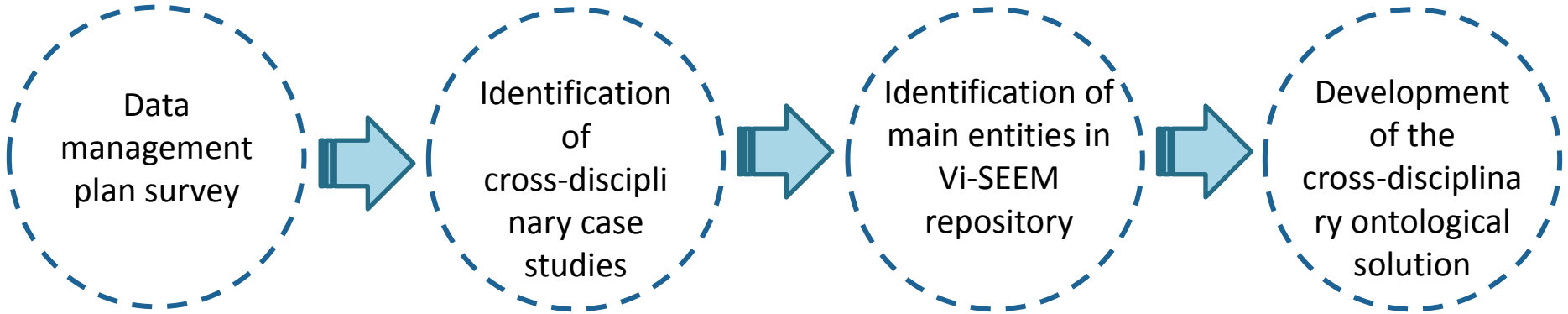
Tools and services



- Different semantic resources (data models, ontologies and vocabularies) used for describing, integrating, and normalizing datasets within related domain research infrastructures

Methodology

Establishment of a methodology towards the development of a cross-disciplinary ontological solution.



Workflow for the development of the cross-disciplinary ontological solution

State of the art and assessment

- Data Management Plan
- Solutions adopted (e.g., metadata standards and ontologies, controlled vocabularies, persistent IDs)

Survey

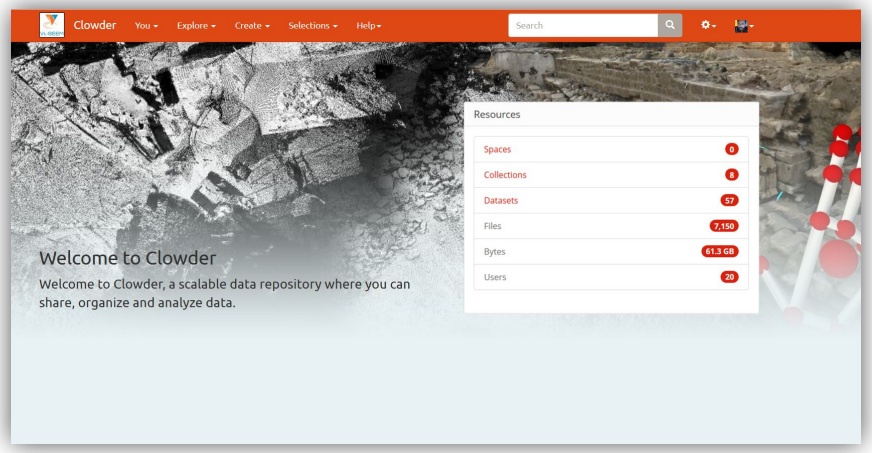
(1) INSTITUTION NAME	(2) DATA STORAGE SERVICE	(2) DATA STORAGE SERVICE Web address	(3) DESCRIPTION OF DATA	(4) METADATA MODEL/SCHEMA / FORMAT	(5) DISCIPLINE / AREA OF RESEARCH	(6) AMOUNT OF METADATA	(7) AMOUNT OF DIGITAL OBJECTS
Acronym of the Institution, as in DoW	*The name and the web address of your collection of data*		*the description of data as in DoW*	*metadata model, schema or format currently used to describe your collection. Please provide separately, as an attached file, an example of your metadata model with the definition of each field/tag*	*Discipline/area of research that the metadata belong to*		

(8) PERSISTENT IDs	(9) CONTROLLED VOCABULARY	(10) METADATA EXPORT PROTOCOL	(11) LANGUAGE(S))	(12) OBJECT TYPES	(13) RIGHTS	(14) COMMENTS	(15) PRIMARY CONTACT
specify if you use or require to use any persistent Identifiers for your collections, and detail it	*which are (if any) the controlled vocabularies used in each of your collections*	*the protocol(s) (if known) currently used in order to export metadata: e.g. OAI-PMH, FTP protocols*	*the language(s) of your metadata*	*the object types: image, text, sound, video (e.g. jpg, pdf, 3D, etc.)*	*details on the Right status of the collection. E.g. Creative Commons, Public Domain, Rights Reserved, etc.*	*further comments*	*the person responsible for the collection at institutional level*

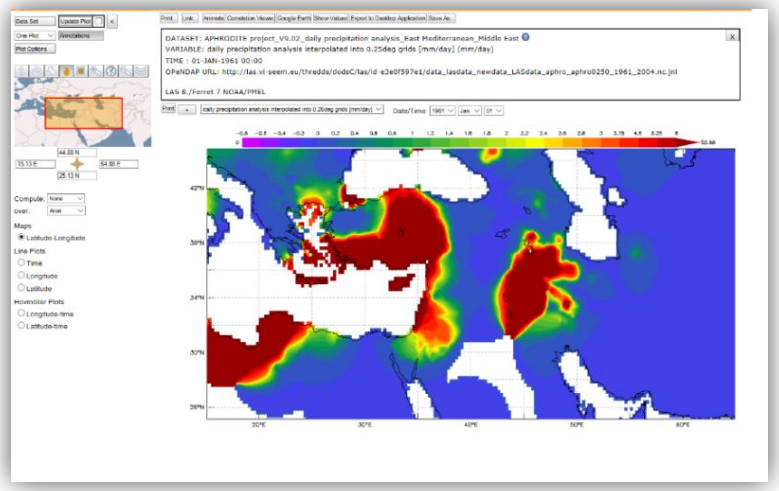
Use cases

Life Science (LS)

Digital Cultural Heritage (DCH)



Climate Sciences (CS)



- An extended Dublic Core was created and used to integrate the data

dc.contributor
dc.contributor.advisor
dc.contributor.author
dc.contributor.editor
dc.contributor.illustrator
dc.contributor.other
dc.coverage.spatial
dc.coverage.temporal
dc.creator
dc.date
dc.date.accessioned
dc.date.available
dc.date.copyright
dc.date.created
dc.date.issued
dc.date.submitted
dc.date.updated
dc.description
dc.description.abstract
dc.description.provenance
dc.description.sponsorship
dc.description.statementofresponsibility
dc.description.tableofcontents
dc.description.uri
dc.description.version
dc.format
dc.format.extent
dc.format.mimetype
dc.identifier
dc.identifier.citation
dc.identifier.govdoc
dc.identifier.isbn
dc.identifier.ismn
dc.identifier.issn
dc.identifier.other
dc.identifier.sici
dc.identifier.slug
dc.identifier.uri
dc.language
dc.language.iso
dc.language.rfc3066
dc.provenance
dc.publisher
dc.relation
dc.relation.haspart
dc.relation.hasversion
dc.relation.isbasedon
dc.relation.isformatof
dc.relation.ispartof
dc.relation.ispartofseries
dc.relation.isreferencedby
dc.relation.isreplacedby
dc.relation.isversionof

Example of the extended Dublin Core fields

Information management approaches

Aims

- Make accessible and relate the resources coming from the three research communities
- Retrieve information that can be used by different disciplines (Interoperability and Reusability)

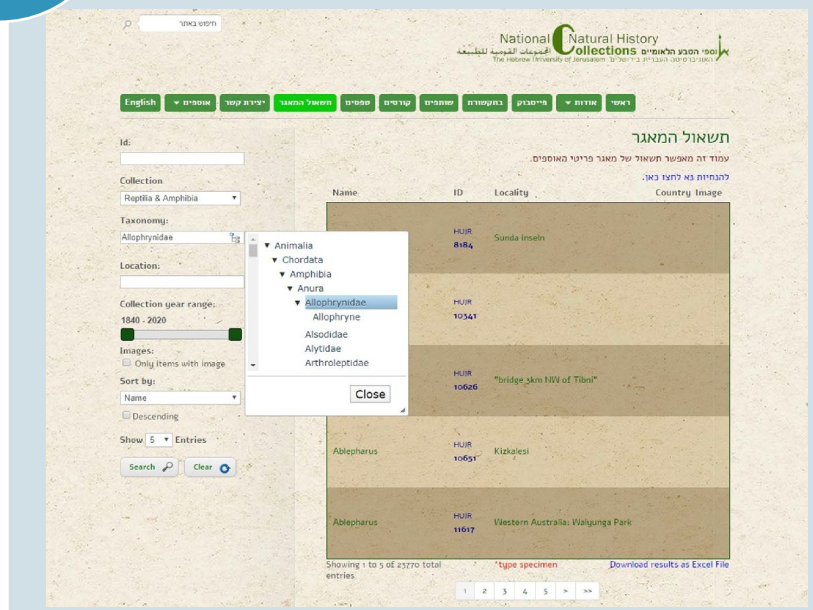
Data-oriented

Ontology developed on the base of the data

Users' needs-based

Ontology developed on specific users' needs

Use cases



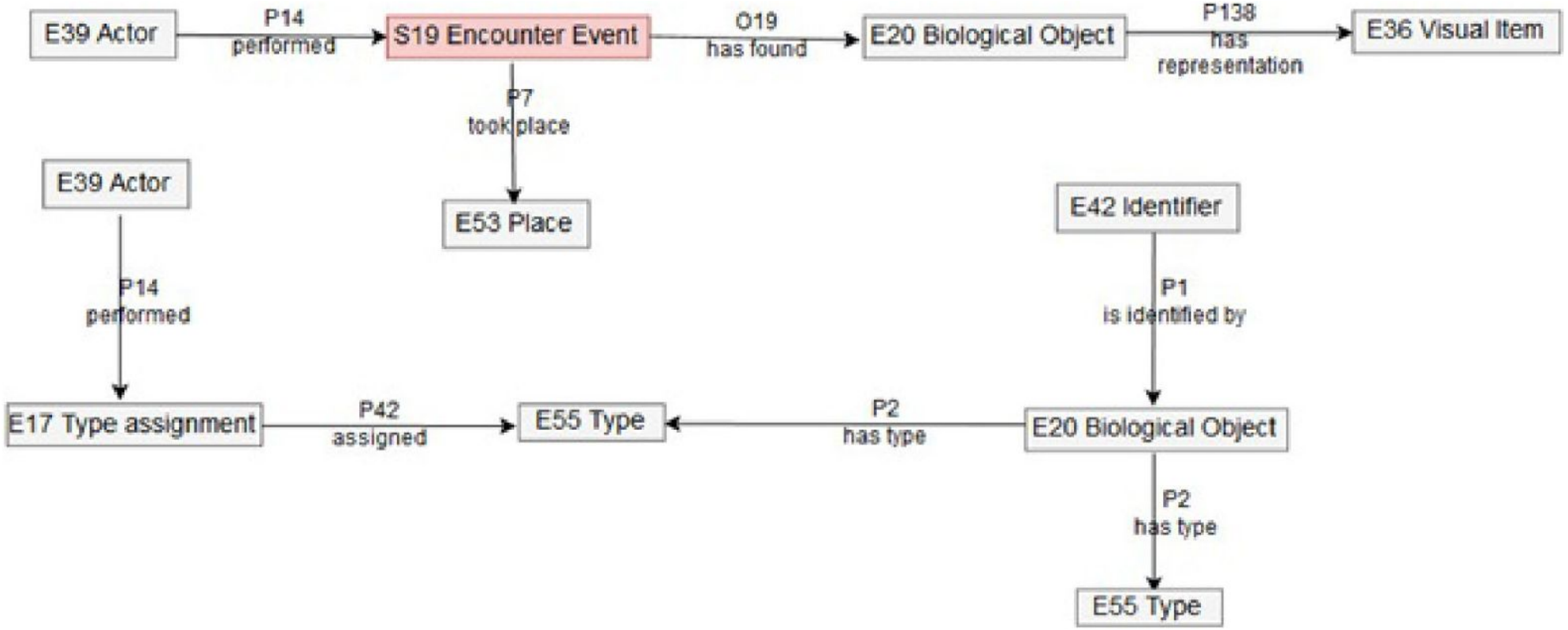
Aharoni Online Digitised collection datasets with their extended taxonomy

ARCHAEOLOGY	ETHNOLOGY	HISTORY	NATURE HISTORY	HISTORY OF ART
NAME AND LOCATION OF THE MUSEUM	Inventory number	Inventory number	NAME AND LOCATION OF THE MUSEUM	NAME AND LOCATION OF THE MUSEUM
Number of main documentation	Number of main documentation	Number of documentation	International museum code	Inventory number
Collection	Number of pieces	Object	sex, age, development stage: number of specimens or fragments	Number of main documentation
Object	Name of the object	Collection	Object name, code	Name of the artist
Placement	Folk term	Negative number and mark	Collected by: name, date	Negative number and label
Negative mark and number	Collection	Photograph	Identified by: name, date	Title

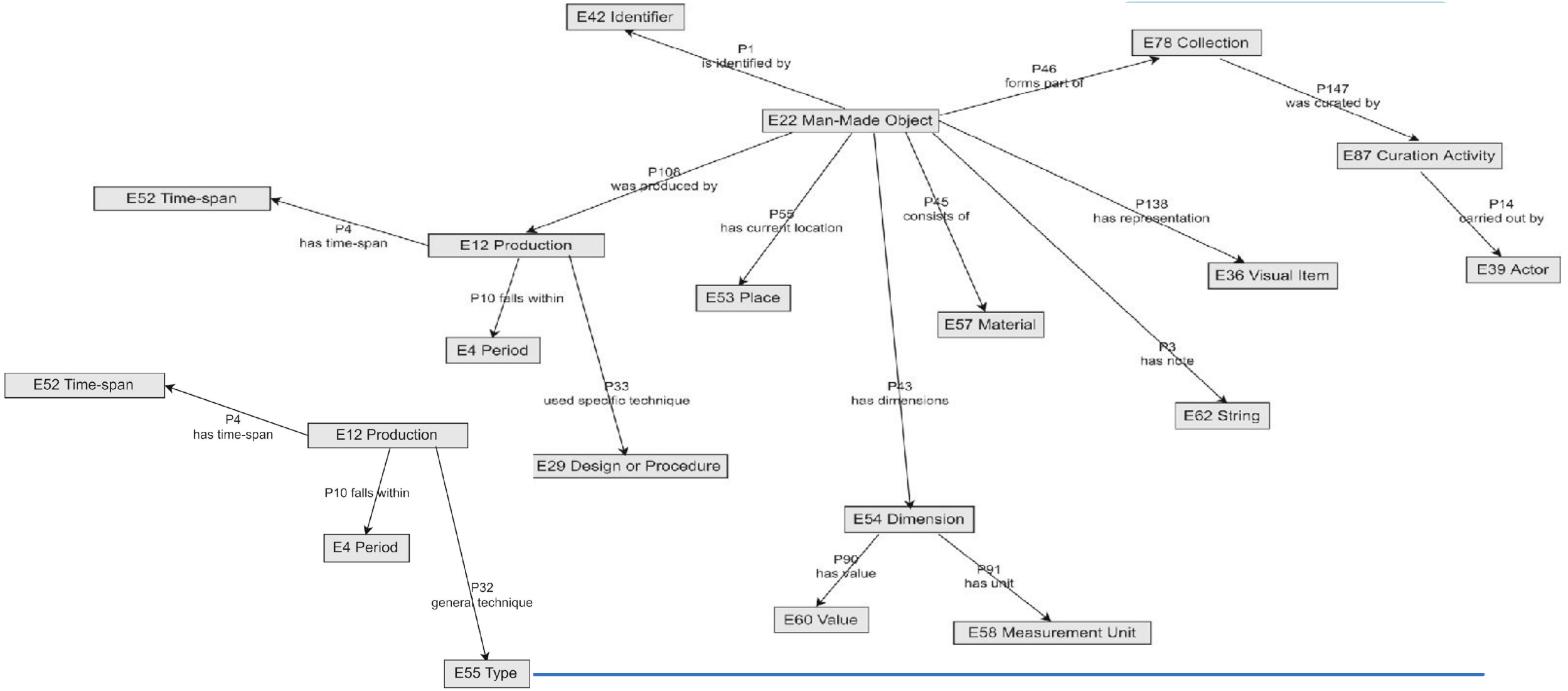
Museum of Republic of Sprska datasets description

The semantic solution

Semantic alignment

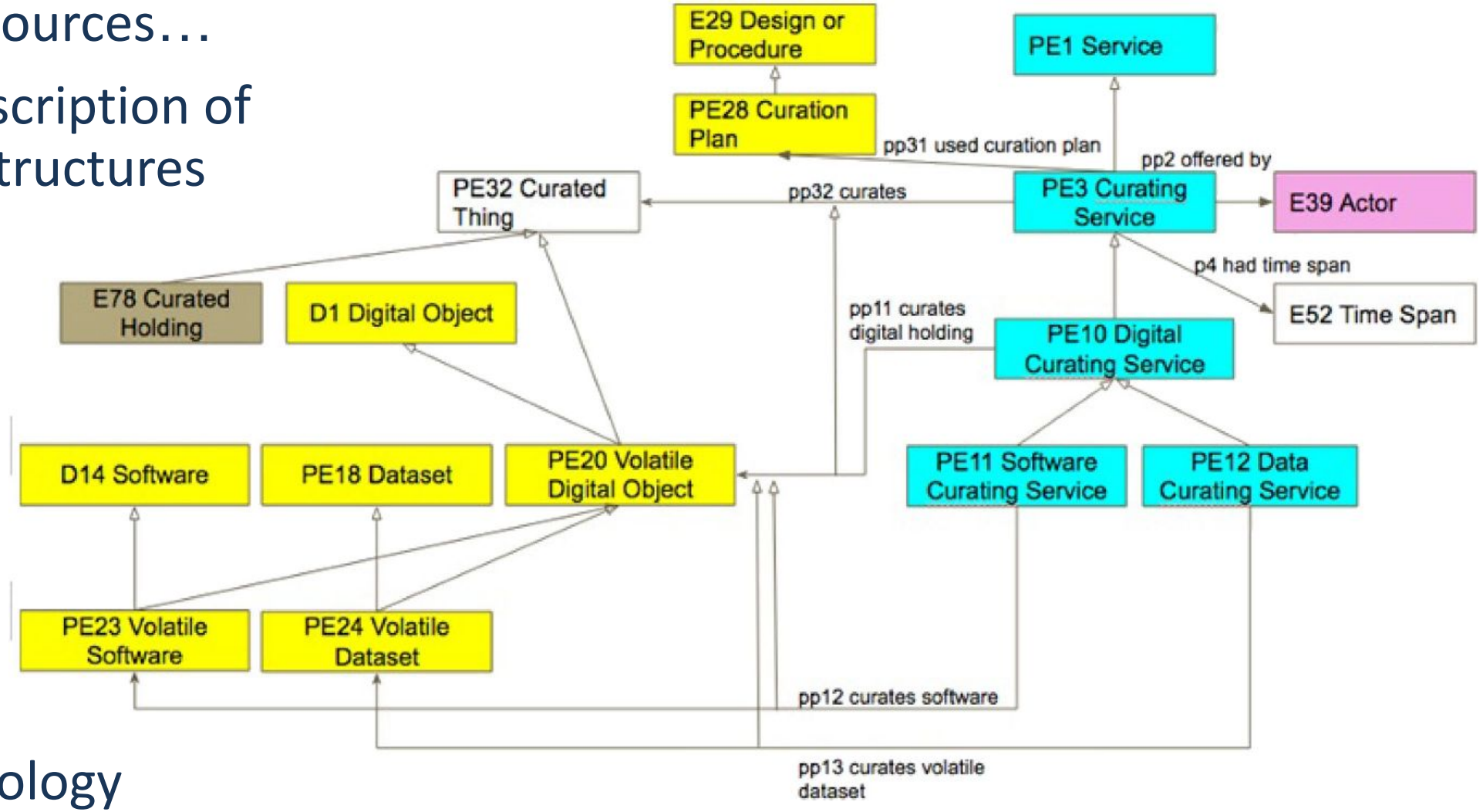


The semantic solution



The semantic solution

- Beyond the resources…
- Ontological description of research infrastructures



Parthenos ontology

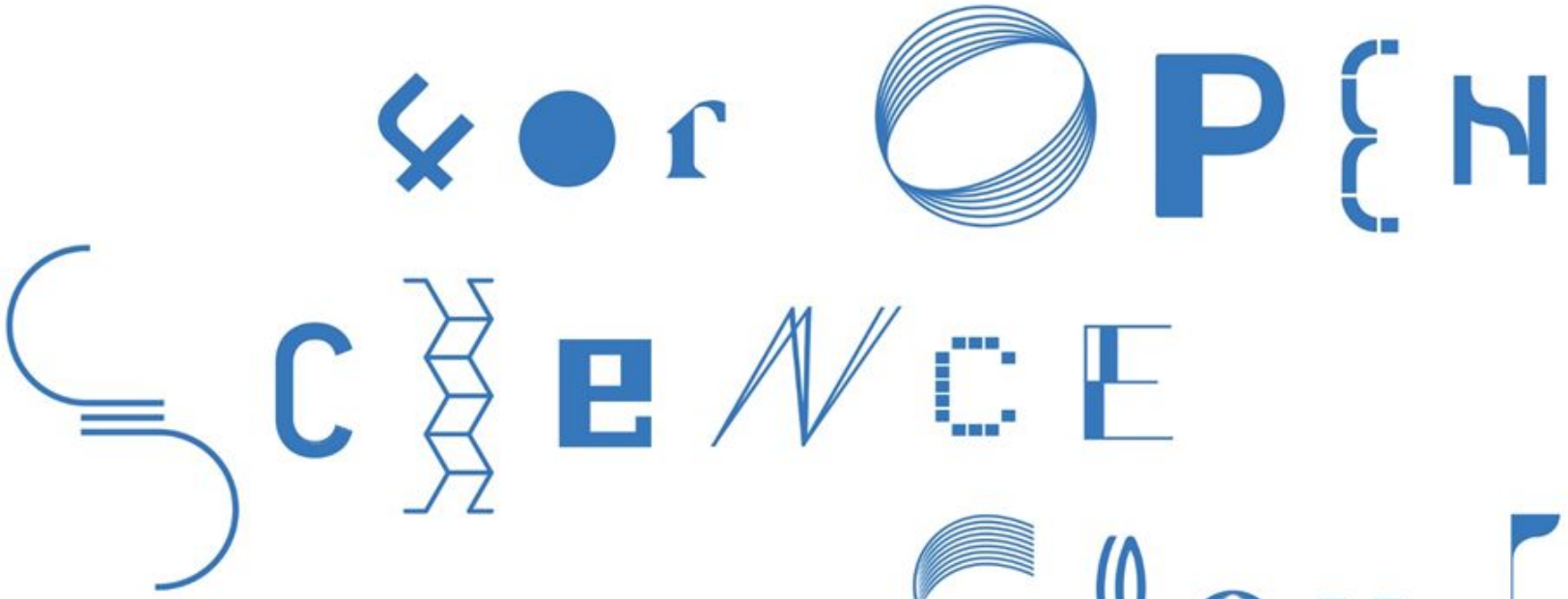
Conclusions

Recommendations

- Case study addressing the issue of the development of a semantic knowledge framework to support the data management of multidisciplinary research infrastructures and guarantee its FAIRness

- Semantic solutions (e.g., ontologies, controlled vocabularies, thesauri and other knowledge organization systems) are fundamental for the implementation of the FAIR principles, especially, for instance, concerning the Interoperability principle within multi-disciplinary research infrastructures

Thanks!



NI4OS_eu NI4OS ni .eu

FAIR semantics a case study at the University of Debrecen

Adam Szaldobagyi

„All stakeholders should recognise that research data repositories are an essential part of the infrastructure for open science.”[1]

1, Business models for sustainable research data repositories: https://doi.org/10.1787/302b12bb-en.

FAIR

Monitoring the open access policy of Horizon 2020[2]

When can we say that a data set meets FAIR requirements?

- 1. A dataset is **findable** if its metadata includes a PID of the dataset and a valid URI to the data file.
- 2. A dataset is **accessible** if the data file can be accessed (fetched) via a valid URL in its metadata.
- 3. A dataset is **interoperable** if the data file is in a machine-readable format.
- 4. A dataset is **reusable** if it has a Creative Commons (CC) licence in its metadata.

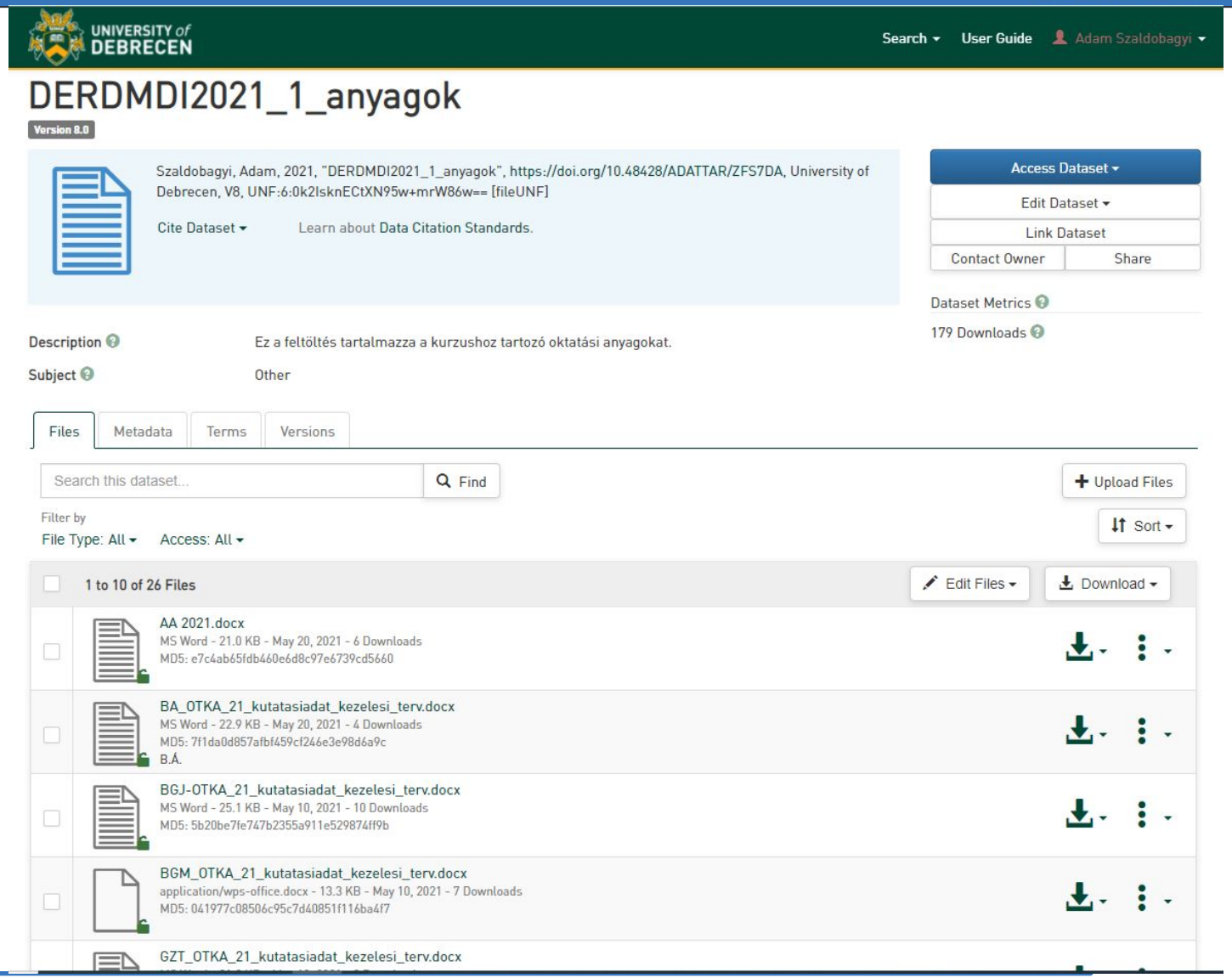
2, Monitoring the open access policy of Horizon 2020: https://op.europa.eu/en/publication-detail/-/publication/56cc104f-0ebb-11ec-b771-01aa75ed71a1/language-en

The importance of data repositories in FAIR research

How does a data repository help FAIR research?

Dataset landing page contains most of the necessary information:

- globally unique and persistent identifier
- metadata fields
- license

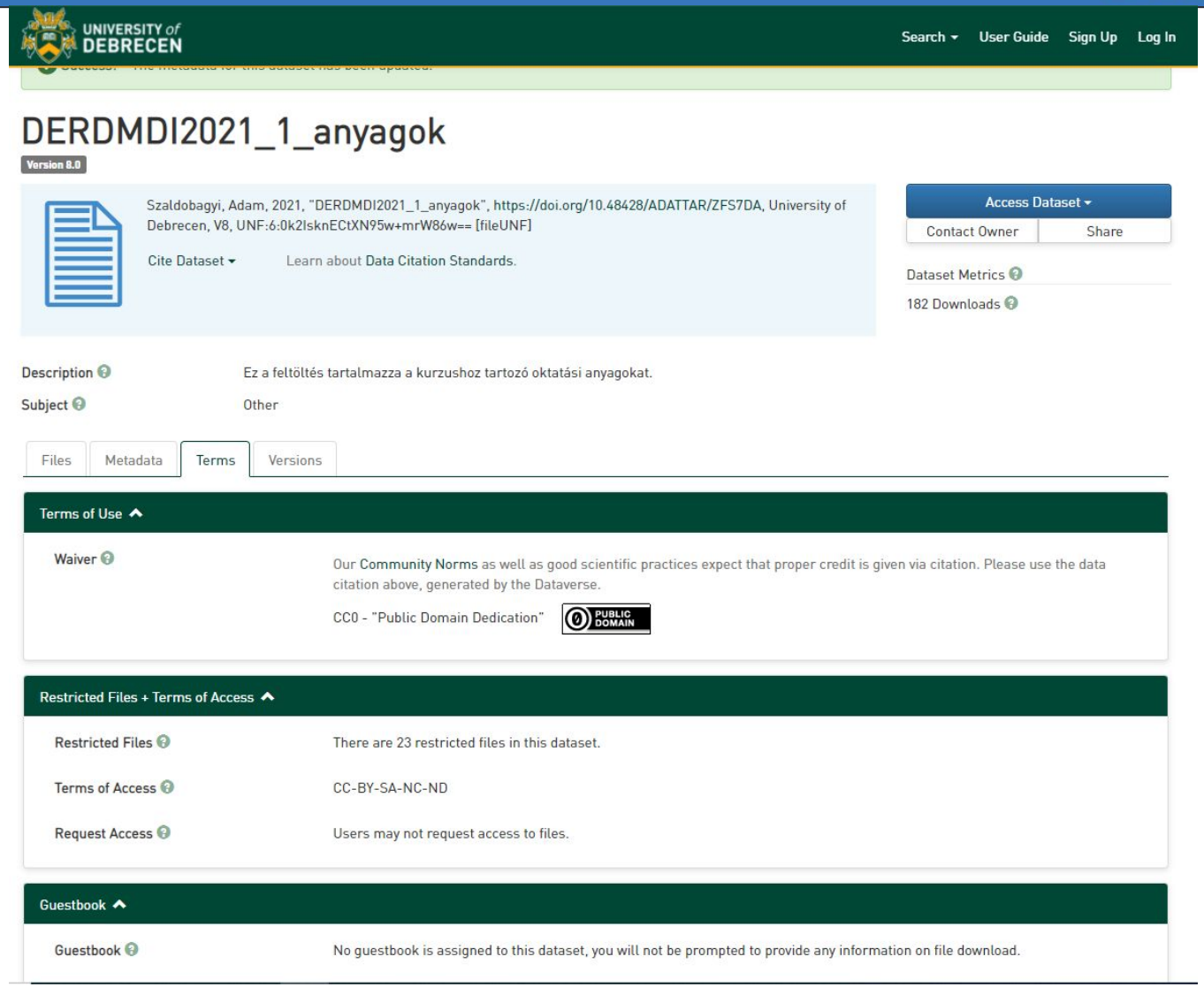


The importance of data repositories in FAIR research

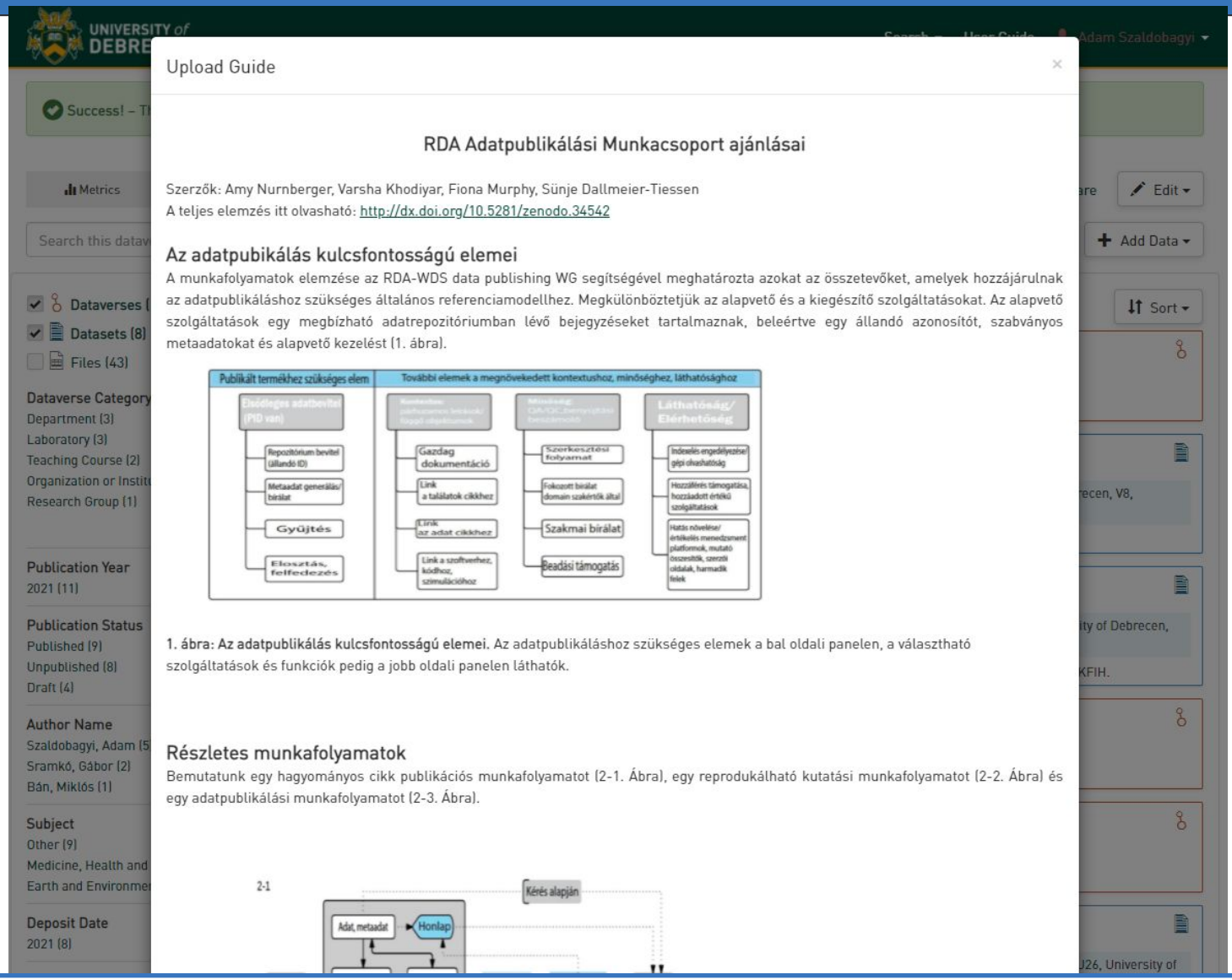
How does a data repository help FAIR research?

Dataset landing page contains most of the necessary information:

- globally unique and persistent identifier
- metadata fields
- license



Upload guide for researcher



Metadata for different disciplines

Metadata Fields

Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

- ☐ Use metadata fields from University of Debrecen
- ☑ Citation Metadata (Required) [+] View fields + set as hidden, required, or optional
- ☐ Geospatial Metadata [+] View fields
- ☐ Social Science and Humanities Metadata [+] View fields
- ☐ Astronomy and Astrophysics Metadata [+] View fields
- ☐ Life Sciences Metadata [+] View fields
- ☐ Journal Metadata [+] View fields

Browse/Search Facets

Choose the metadata fields to use as facets for browsing datasets and dataverses in this dataverse.

☐ Use browse/search facets from University of Debrecen

All Metadata Fields

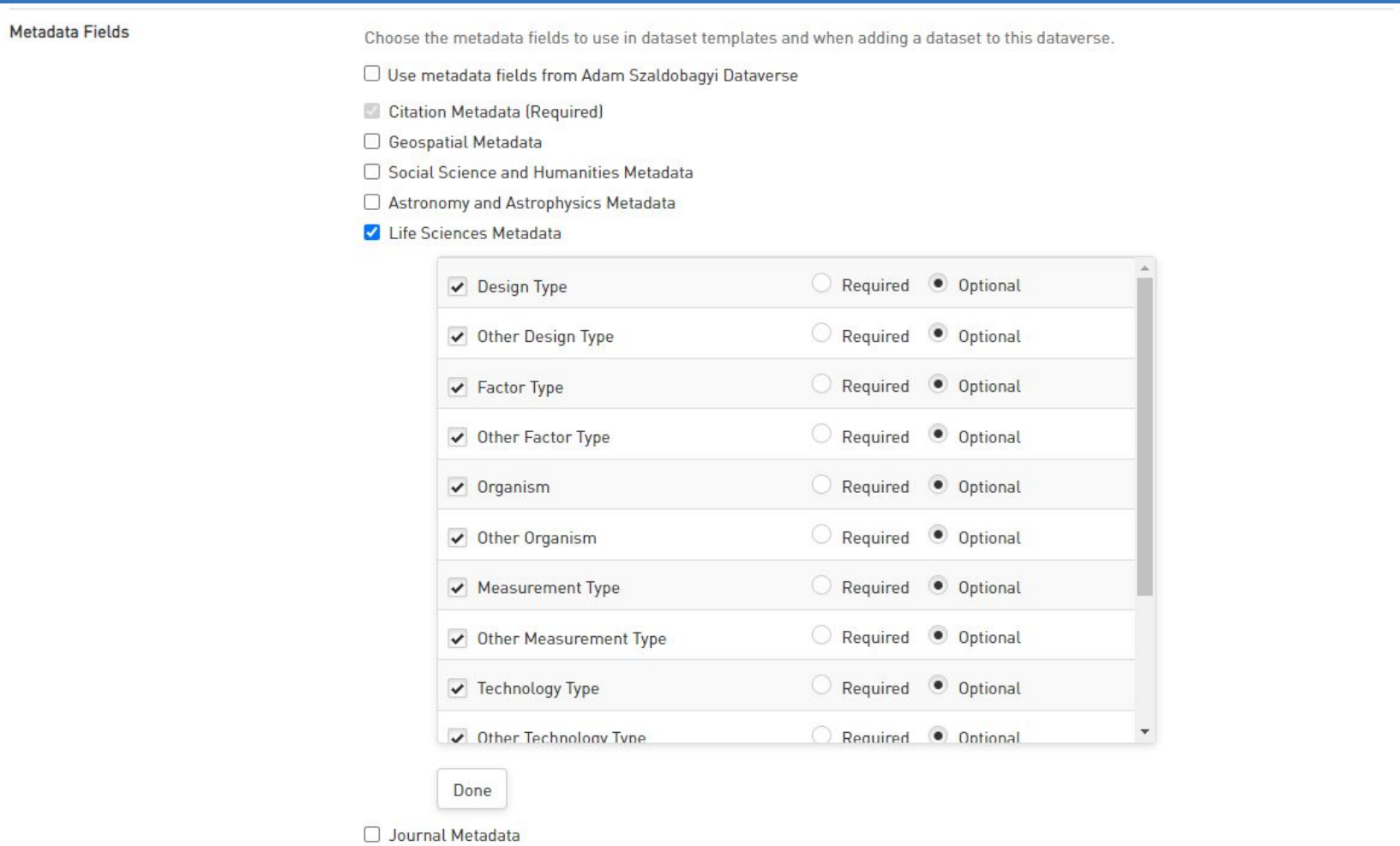
- Author Affiliation
- Topic Classification Term
- Language
- Producer Name
- Production Date
- Contributor Type
- Contributor Name
- Grant Information Grant

→ →| ← |←

Selected

- Author Name
- Subject
- Keyword Term
- Deposit Date

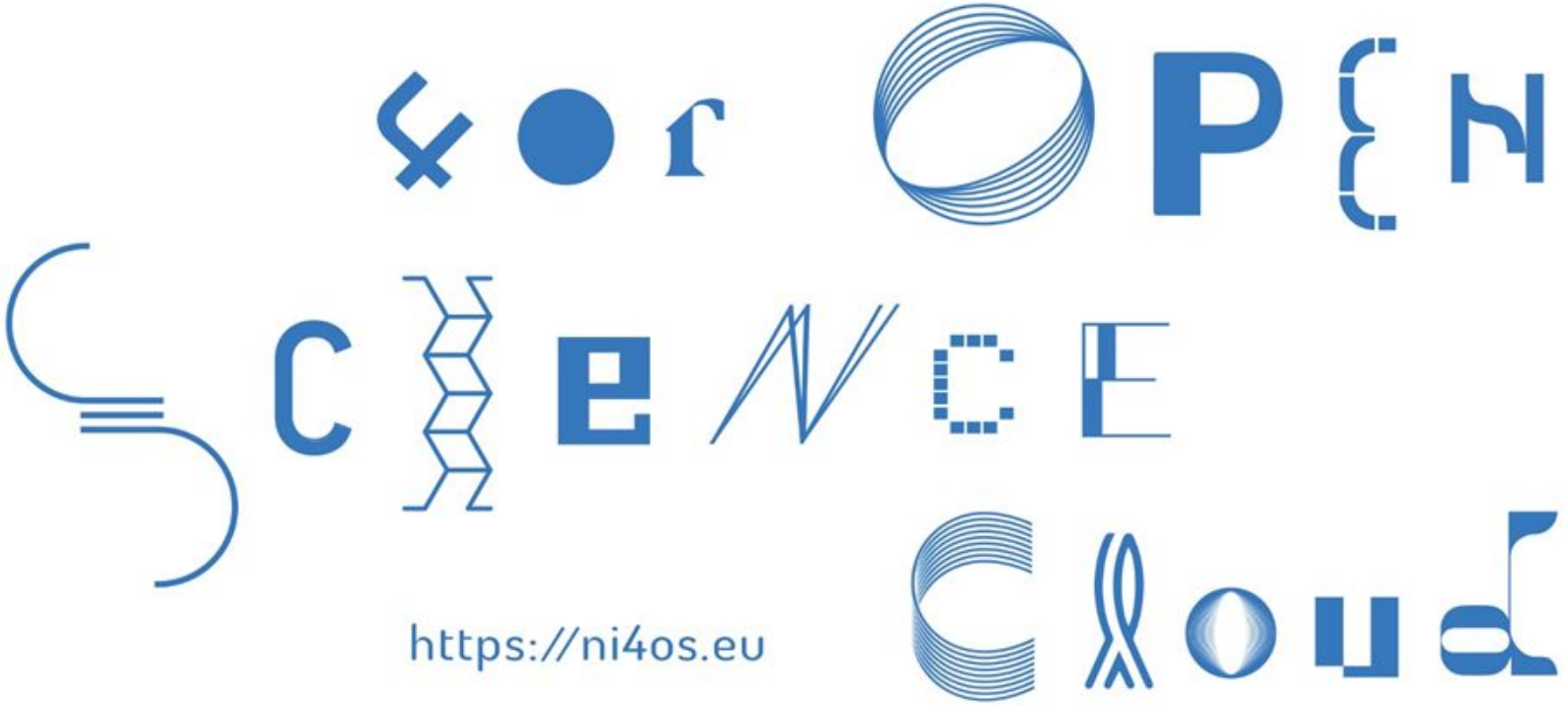
Metadata for different disciplines



Additional resources to help researchers work

We hold regular meetups and workshops focusing on the importance of open science, FAIR and data repositories.

Thank you for your attention!



OPEN SCIENCE FAIR

Speaking FAIR implementation for service providers session

Go to

www.menti.com

Enter the code

58 93 72 7

Or use QR code

THANK YOU!


FAIRSFAIR
Fostering Fair Data Practices in Europe


NI4OS
Europe