FAIRSFAIR

Fostering Fair Data Practices in Europe

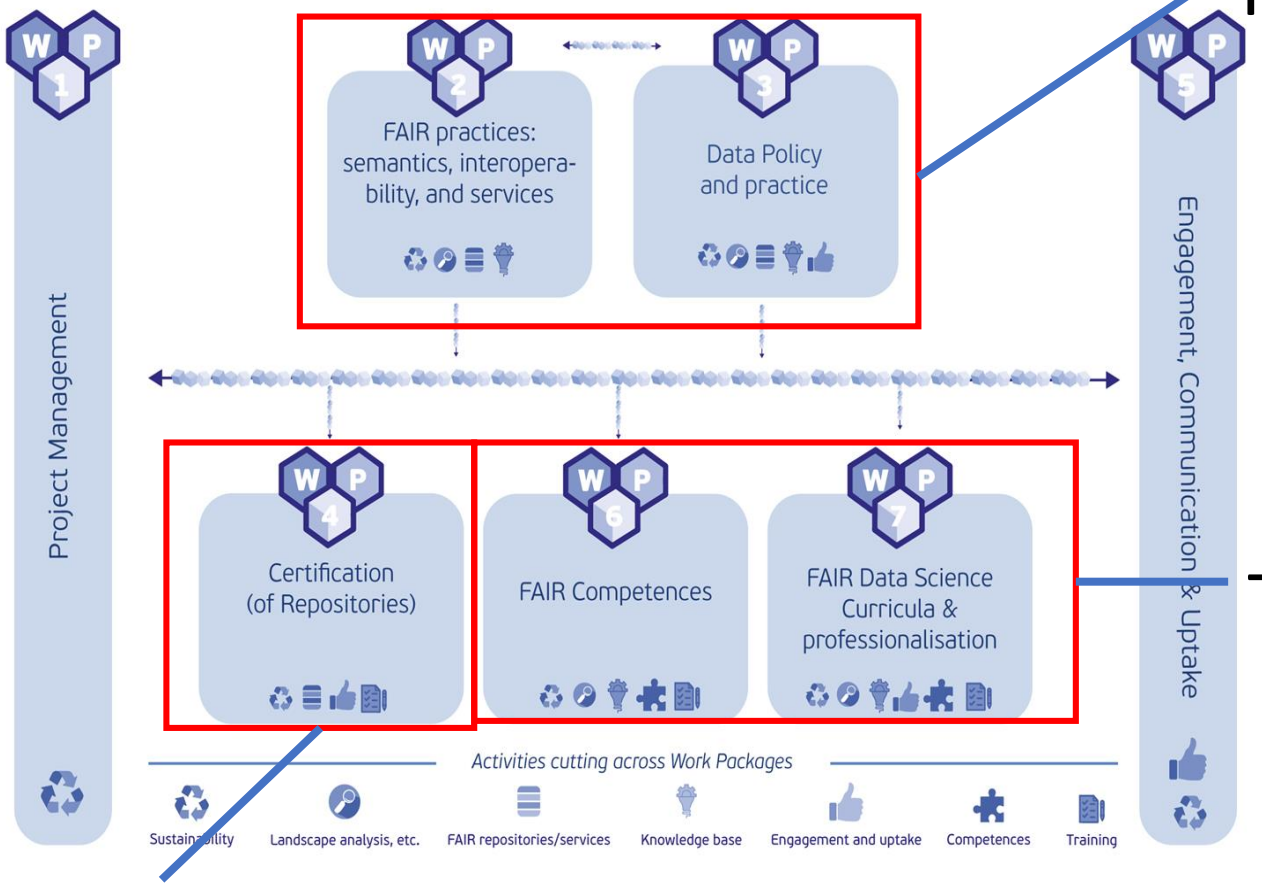
FAIR REQUIREMENTS FOR PERSISTENCE AND INTEROPERABILITY 2019

Jessica Parland-von Essen

Leah Riungu-Kalliosaari

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

Work Packages in FAIRsFAIR



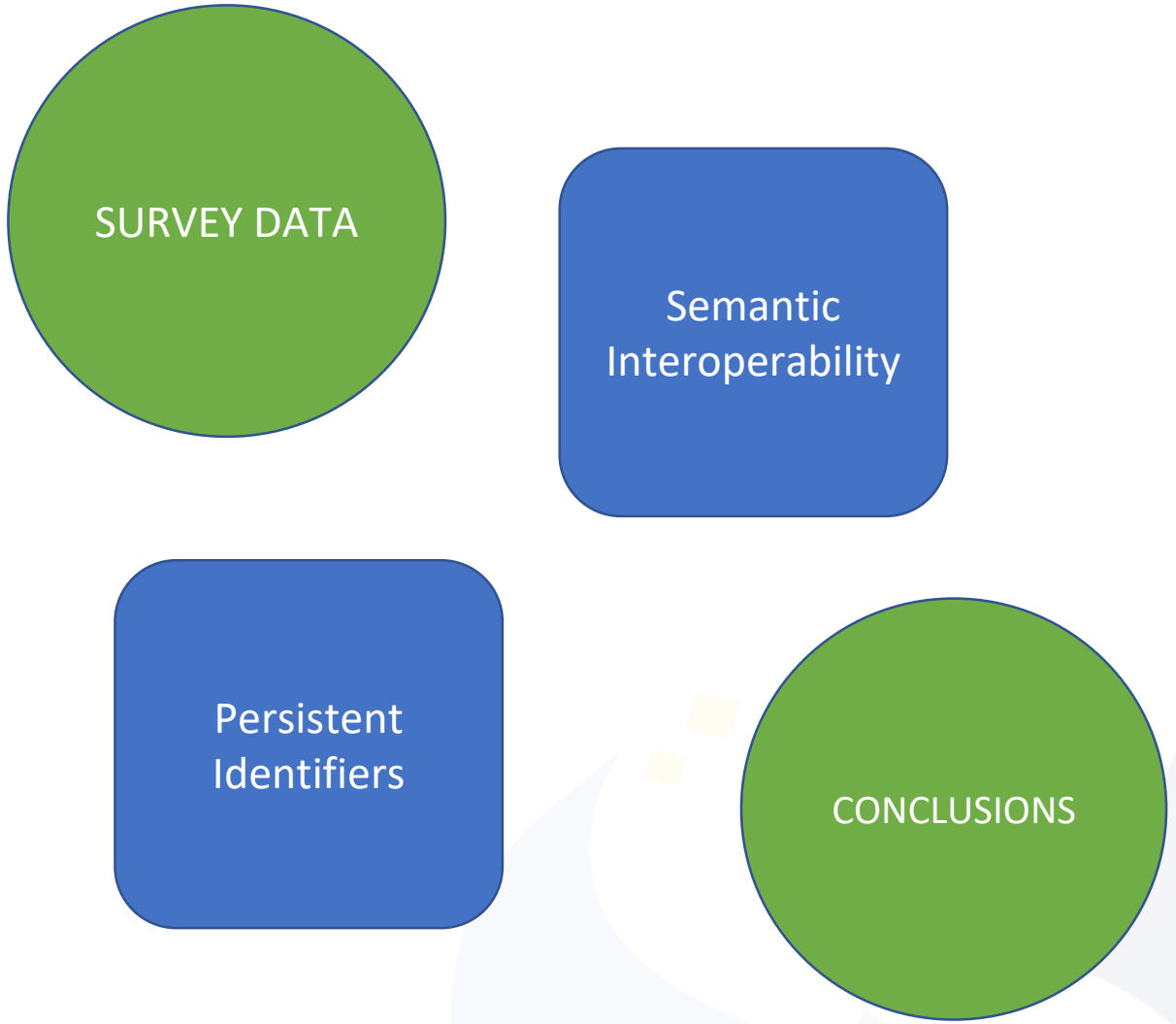
Practice and policy

Training and education

Certification

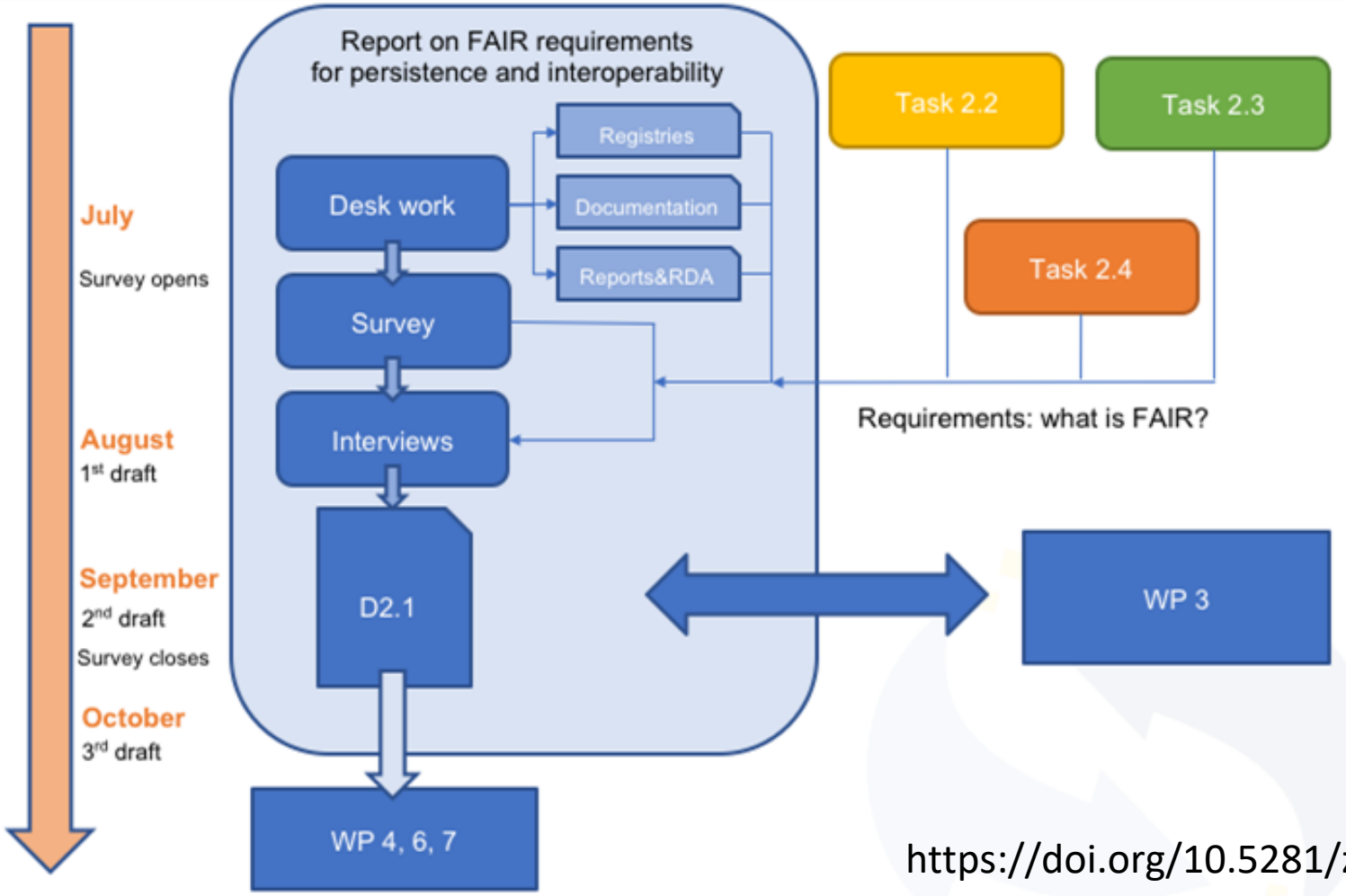
Content of the report

- FAIR technologies and methods
  - Semantic interoperability
  - Semantic artefacts
  - PIDs and PID services
- FAIR in the context of the data life cycle
  - Repositories
  - Evolving datasets and data citation
- ESFRI RI's



https://doi.org/10.5281/zenodo.3557381

D2.1 REPORT ON FAIR REQUIREMENTS FOR PERSISTENCE AND INTEROPERABILITY



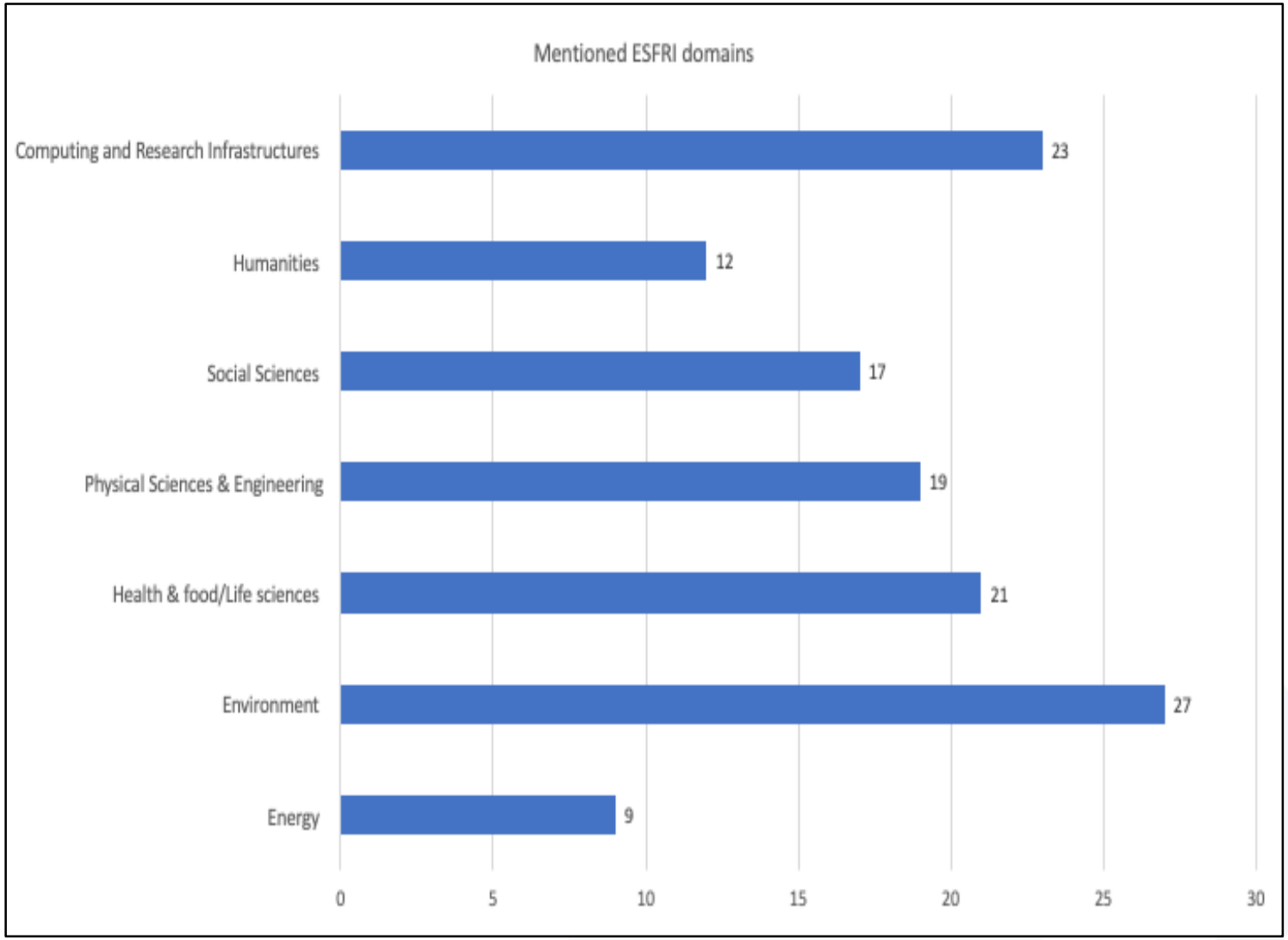
https://doi.org/10.5281/zenodo.3557381

Survey data https://doi.org/10.5281/zenodo.3518922

Roles of the respondents

- Research support staff 28
- Repository staff 19
- Research infrastructure operator 17
- Researcher 22
- Policy maker 5
- Other 5

Number of researchers in organization	Number of responses (N=64)
< 100	16
100 - 500	13
500 - 1000	7
1000 <	28

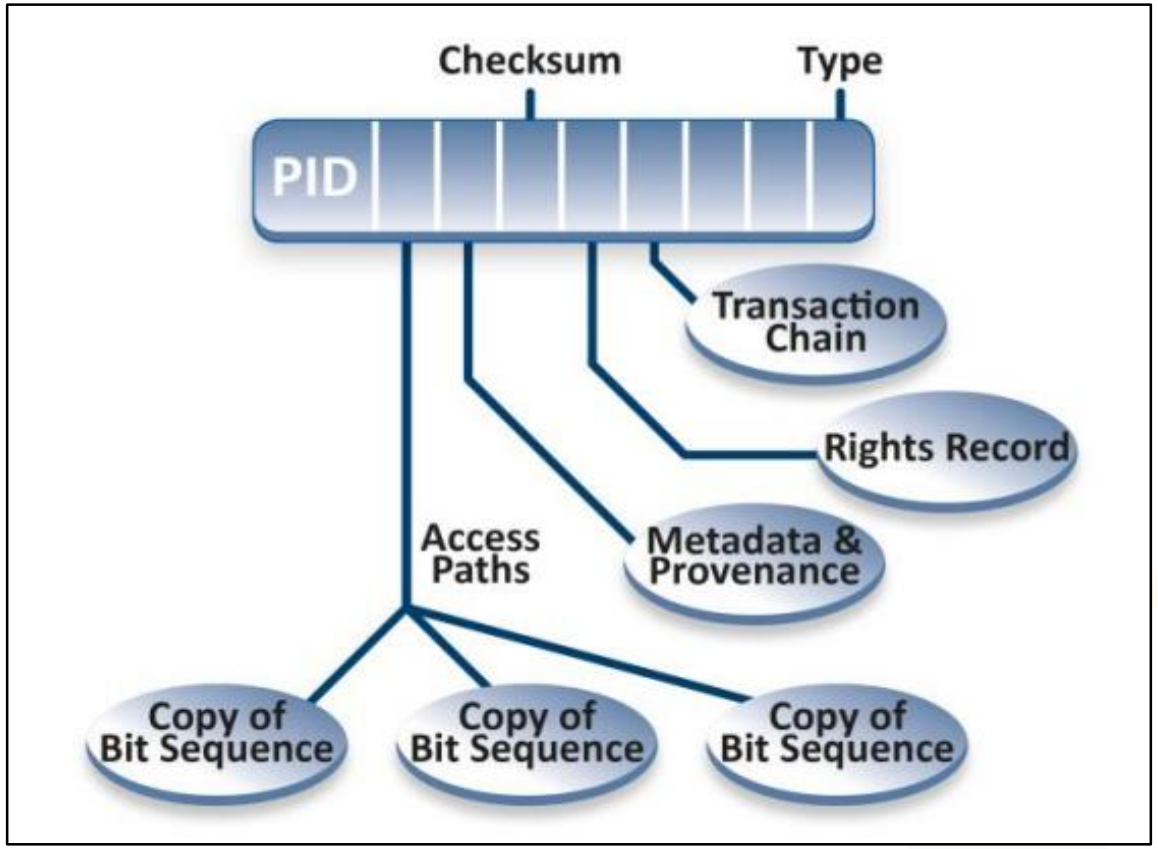


Survey data https://doi.org/10.5281/zenodo.3518922

Mentioned infrastructures

ACTRIS
ADC
AnaEE (2)
BBMRI ERIC
CESSDA ERIC (5)
CLARIN (4)
DARIAH (3)
DiSSCo (3)
EISCAT_3D
ELIXIR (3)
eLTER (2)
EMBRC ERIC
EMODnet
EMPHASIS (2)
EPOS
ESS ERIC
EU-SOLARIS
EURO-ARGO ERIC (2)
European XFEL
FAIR (3)
Go FAIR Initiative
IAGOS (2)
INSTRUCT ERIC (2)
IODE
IS-ENES (2)
LifeWatch ERIC
ODP
OpenAIRE
PRACE (3)
SCADM
SeaDataNet/SeaDataCloud (2)
SHARE ERIC
SKA
SOOS

The FAIR digital object



Wittenburg P, Strawn G, Mons B, Bonino L, Schultes E. Digital Objects as Drivers towards Convergence in Data Infrastructures. [presentation] RDA. 2018. Available at: https://www.rd-alliance.org/sites/default/files/Digital%20Objects%20as%20Drivers%20towards%20Convergence%20in%20Data.pdf

RDF

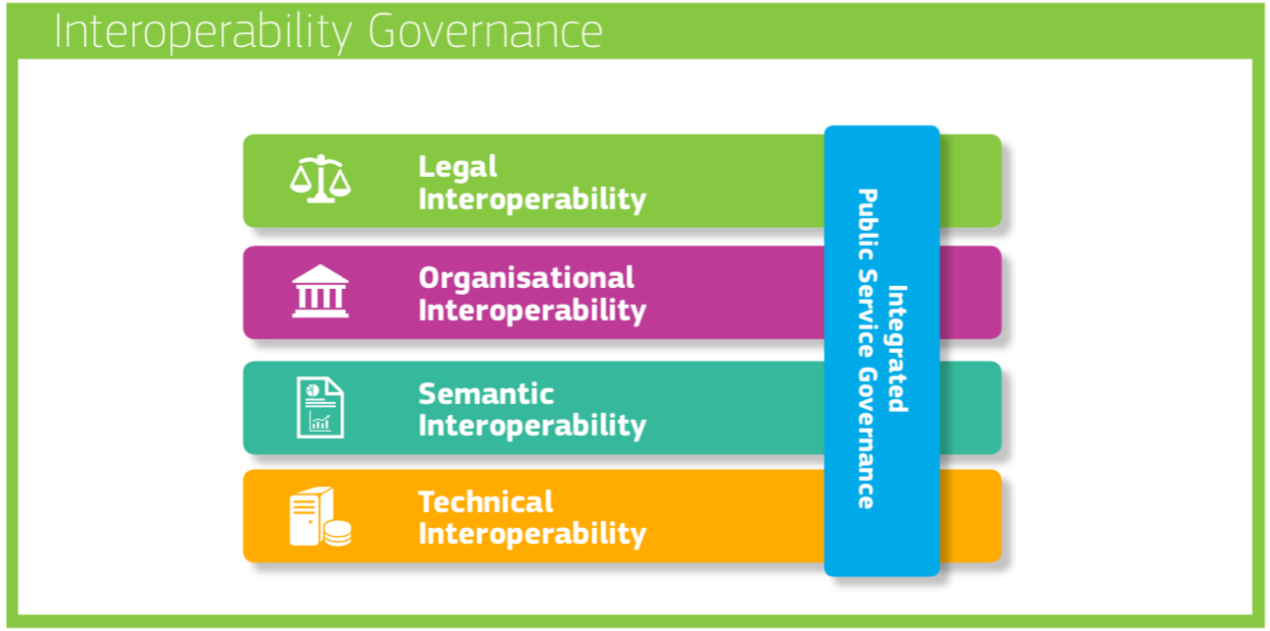
RDF
SKOS
Data Cube Vocabulary
R2RML
RDFS
OWL
ICV
SPIN/SPARQL
ShEx
SHACL

The triple

Semantic Interoperability

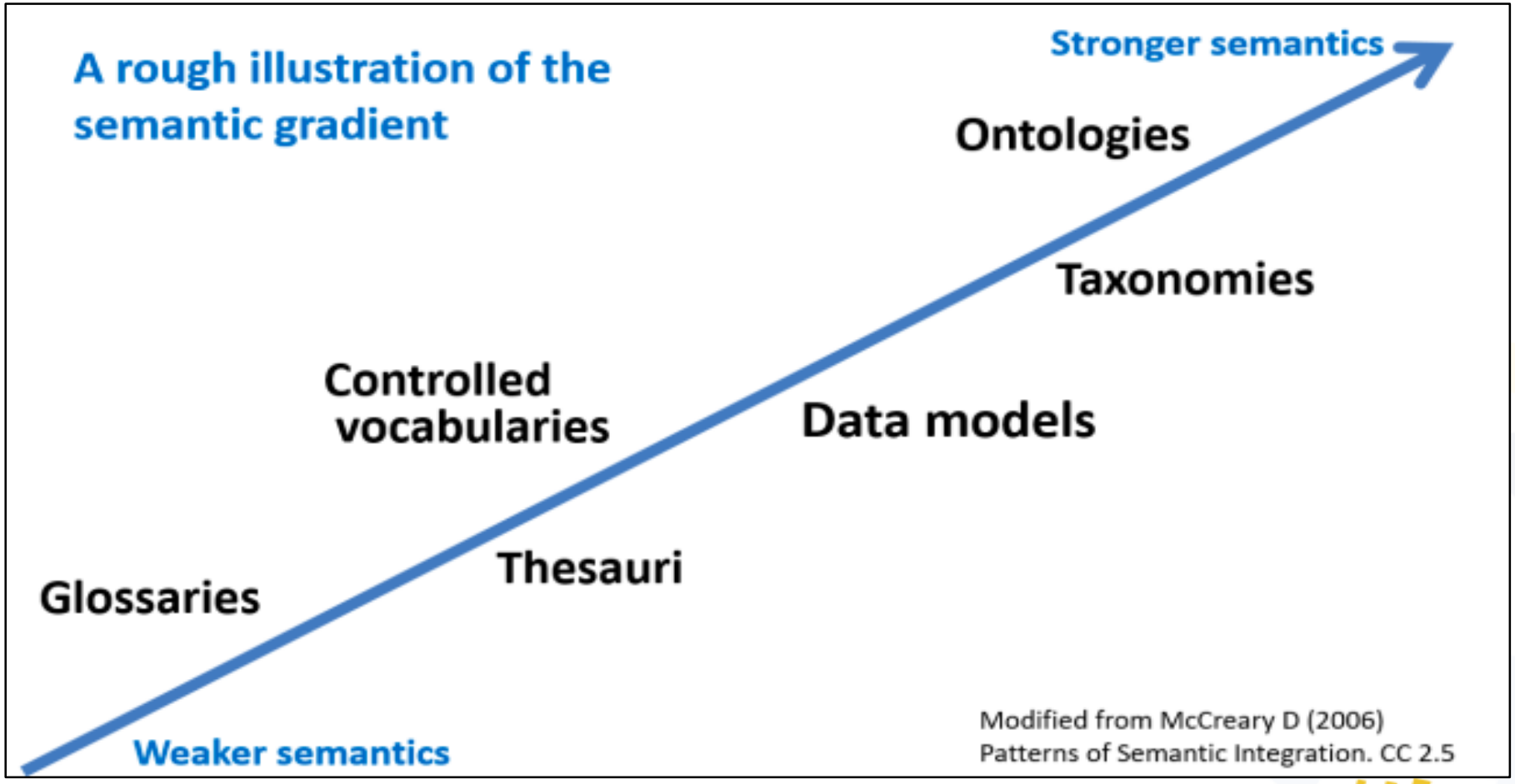
Interoperability is the ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems. Semantic interoperability is achieved when the information transferred has, in its communicated form, all of the meaning required for the receiving system to interpret it correctly, even when the algorithms used by the receiving system are unknown to the sending system. Syntactic interoperability is a prerequisite to semantic interoperability.

CASRAI. https://dictionary.casrai.org/Semantic_interoperability



Trans-language interoperability

Semantic Artefacts



Semantic artefacts adoption

	Characteristics	Indicators
Coverage in field	widely approved and adopted	used within community, acknowledged mandate
Coverage of content	sufficient amount of the terminology	coverage, completeness, coherence
	structure corresponds to the ontology of the domain	certification, quality, community approval
Governance	usable and fit for purpose	compatibility, format, granularity, workflow
	actively maintained by a trusted, authoritative party	curation, versioning, persistence
	(re)usability	open access, documented

FAIR Semantics

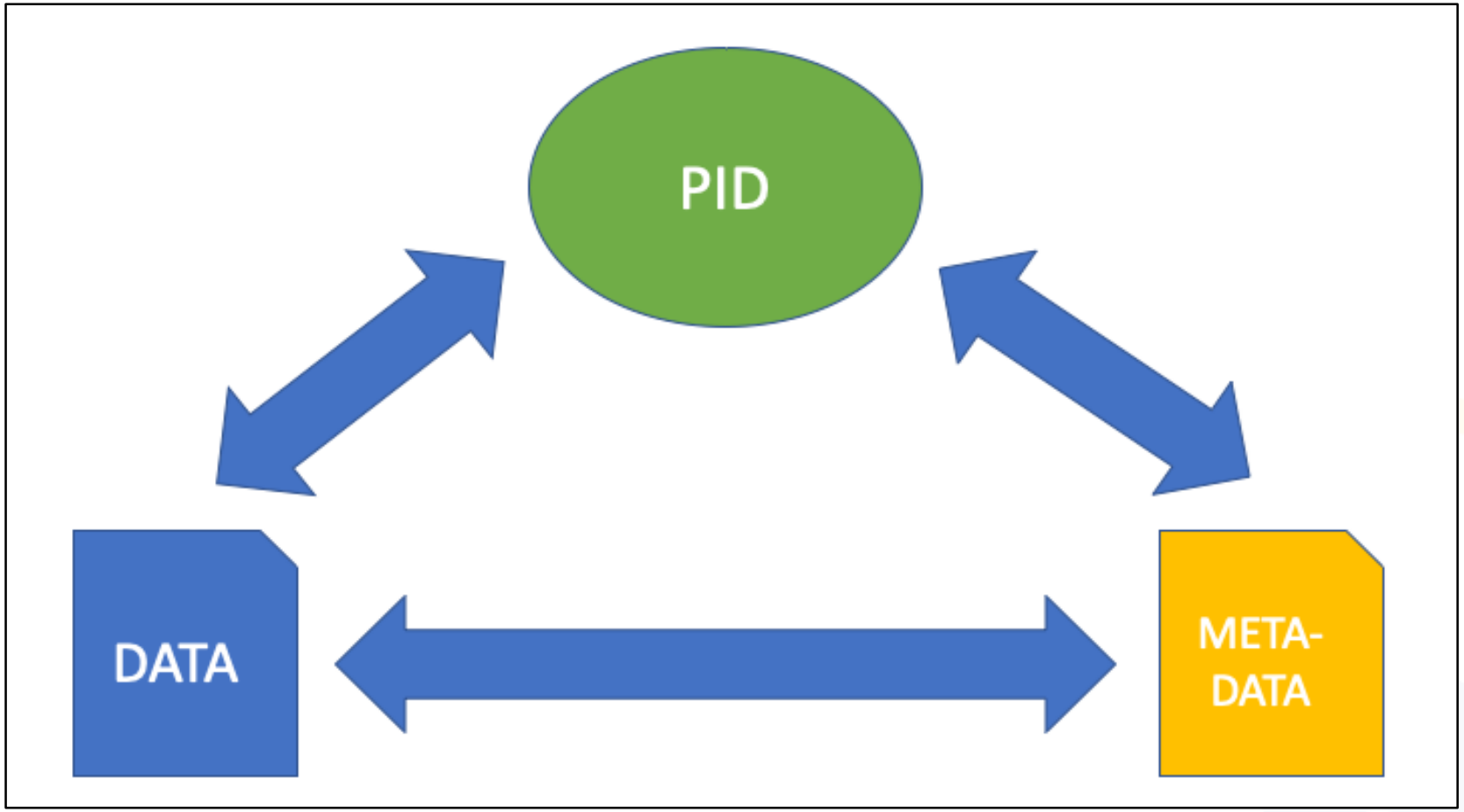
- Task in FAIRsFAIR for developing recommendations for semantic artefacts
- Semantic artefacts are often complex datasets
- Can be treated as such, but also needs further recommendations to ensure reusability

Examples of services for semantic artefacts

- FAIRsharing
- BARTOC
- BioPortal
- AgroPortal
- GEMET
- Linked Open Vocabularies (LOV)

OBO Foundry Principles

Persistent identifiers

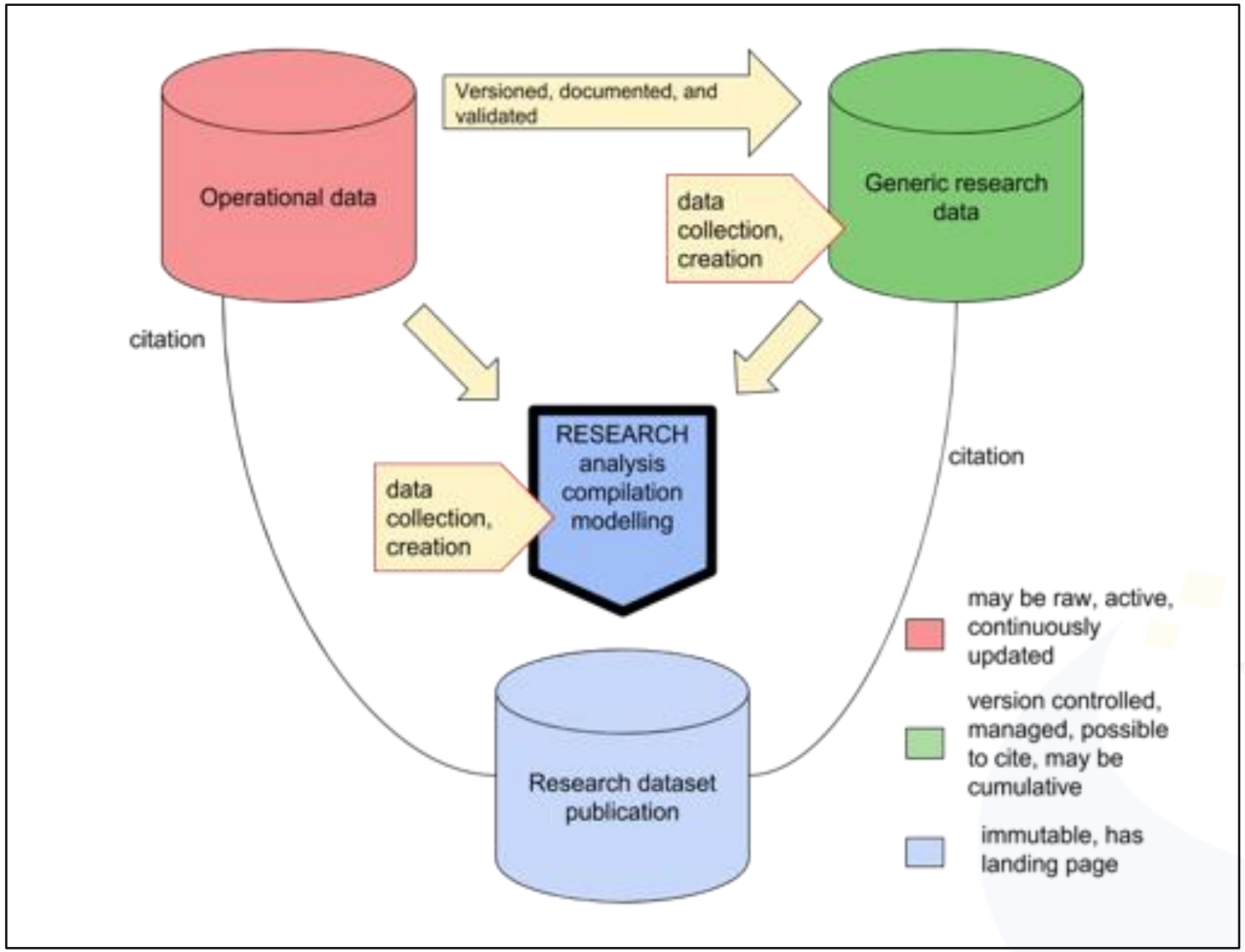


Recommendations by the RDA Data Fabric

- A persistent identifier (PID) needs to be supported by a sustainable and trustworthy resolution system that will resolve PIDs to meaningful state information for machines and humans which are metadata attributes describing essential properties of a Digital Object (DO).
- A trustworthy PID resolution system needs to fulfil quality criteria still to be defined and needs to undergo regular quality assessment.
- The persistent PID record should be used to persistently bind the context of digital objects.
- A PID should be assigned to a Digital Object when it is registered at a trustworthy repository and thus becomes part of the domain of visible and findable data.
- A DOI should be registered when Digital Objects (data) are being published and citation metadata should be associated with it.

PIDs not only for datasets, but also for

- how the various dimensions represented as variables in datasets of the form w1, d2, temp, etc., correspond to real world notions of weight, distance, temperature, etc.
- what are the measurement units associated with each of those dimensions, e.g., Kelvin, Celsius, or Fahrenheit in the case of temperature.
- how those dimensions are grouped or packed together in datasets.
- Data Type Registry. [web page] RDA. Available from: http://typeregistry.org/registrar/#


Operational data
Versioned, documented, and validated
Generic research data
data collection, creation
citation
RESEARCH analysis compilation modelling
data collection, creation
citation
Research dataset publication
may be raw, active, continuously updated
version controlled, managed, possible to cite, may be cumulative
immutable, has landing page

DataCite

- Cite a specific slice or subset
  - the set of updates to the dataset made during a particular period of time or to a particular area of the dataset
- Cite a specific snapshot
  - a copy of the entire dataset made at a specific time
- Cite the continuously updated dataset, but add Access Date and Time to the citation
  - Does not necessarily ensure reproducibility
- Cite a query, time-stamped for re-execution against a versioned database

RDA Evolving data citation

- Data Versioning: For retrieving earlier states of datasets the data needs to be versioned. Markers shall indicate inserts, updates and deletes of data in the database.
- Data Timestamping: Ensure that operations on data are timestamped, i.e. any additions, deletions are marked with a timestamp.
- Data Identification: The data used shall be identified via a PID pointing to a time-stamped query, resolving to a landing page.


FAIRSFAIR
Fostering Fair Data Practices in Europe


FREYA


COAR
Confederation
of Open Access
Repositories


re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES


OPEN KNOWLEDGE


EUROPEAN OPEN
SCIENCE CLOUD


GO FAIR


COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL


DataCite


Schema.org


OpenAIRE


RDA
RESEARCH DATA ALLIANCE


FAIRsharing.org
standards, databases, policies


OpenCitations


OECD


WORLD
DATA SYSTEM


WIKIDATA


DONA

Conclusions (1)

FAIRness on a more generic level is not ready and clearly defined.

The landscape is diverse in all aspects. Differences inside domains are often bigger than differences between domains.

Semantic artefacts are a key element in building interoperability and good quality (meta)data.

Conclusions (2)

Crosswalks, mappings and semantic application profiles should be published and registered in machine readable formats.

PID and data type registries should promote reuse rather than bulk creation of PIDs. To support interoperability, they should be considered semantic artefacts and used mindfully.

Conclusions (3)

Reuse of semantic artefacts should be promoted by publishing application profiles. Curated registries like the EOSC Hub, FAIRsharing and re3data.org are important resources.

Data citation and machine actionable solutions should be developed in parallel. Community adoption and trust are the decisive factors.

Conclusions (4)

The most popular, potentially most useful, and most complex approaches on improving FAIRness of data are based on technologies using Linked Data.

The development should be research rather than technology driven.

Solutions should be user friendly, context sensitive and transparent to the users.

Next steps

A lot of relevant work is and will be done within EOSC-related projects including our own WP. This will be included in the second report, with your help!

Please, give feedback on the report!

https://fairsfair.eu/fairsfair-deliverables-community-review

D2.1 Report on FAIR requirements for persistence and interoperability 2019
https://doi.org/10.5281/zenodo.3557381

Heikki Lehväslaiho, Heidi Laine, Claudia Behnke, Christine Staiger, Yann LeFranc