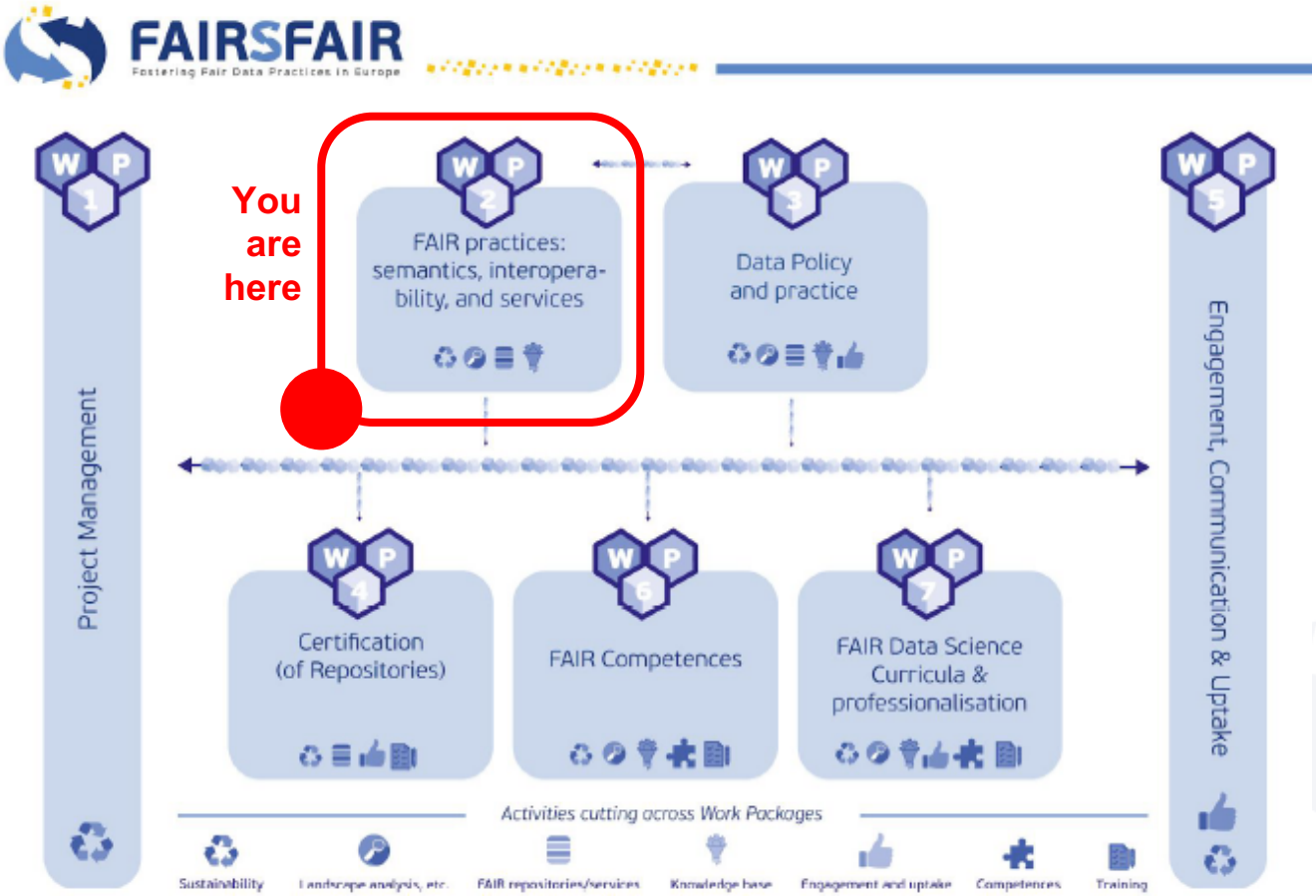
FAIRSFAIR

Fostering Fair Data Practices in Europe

FAIR assessment for data services

April 8, 2020

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
Fostering Fair Data Practices in Europe
You are here
WP 1
Project Management
WP 2
FAIR practices: semantics, interopera-bility, and services
WP 3
Data Policy and practice
WP 5
Engagement, Communication & Uptake
WP 4
Certification (of Repositories)
WP 6
FAIR Competences
WP 7
FAIR Data Science Curricula & professionalisation
Activities cutting across Work Packages
Sustainability
Landscape analysis, etc.
FAIR repositories/services
Knowledge base
Engagement and uptake
Competences
Training

FAIRSFAIR
Fostering Fair Data Practices in Europe

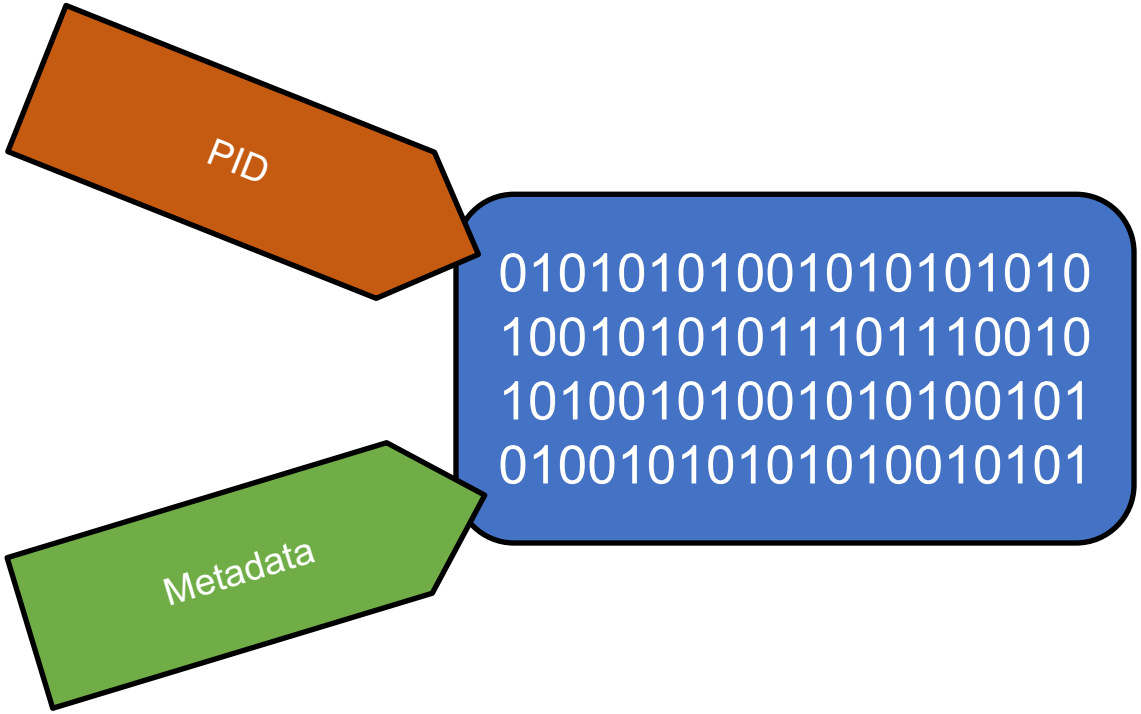
What does it take …

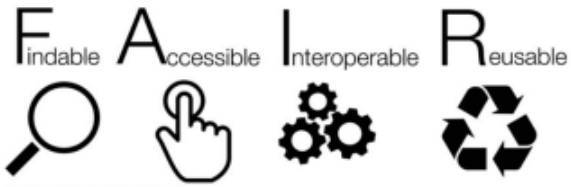
FAIRSFAIR
Fostering Fair Data Practices in Europe

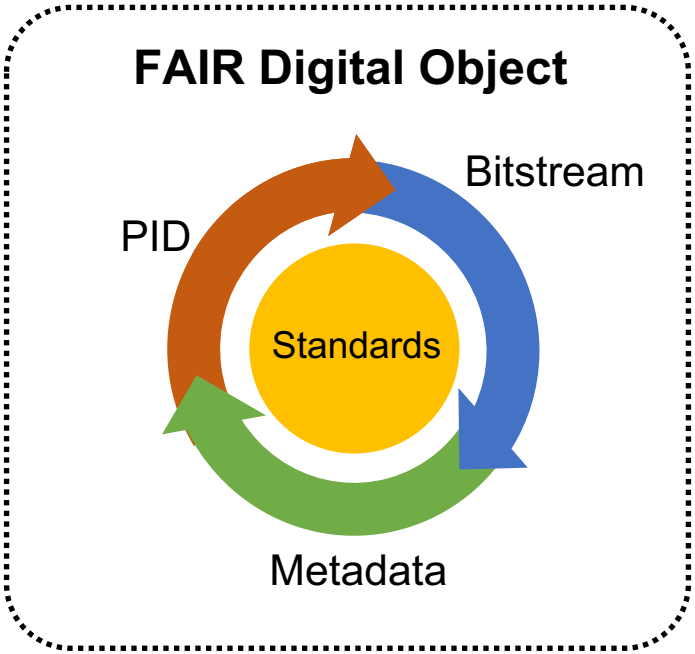
… for a service to be ‘FAIR?’

(Is that even a good question?)

0101010100101010101010
1001010101110110010
1010010100101010100101
0100101010101010010101

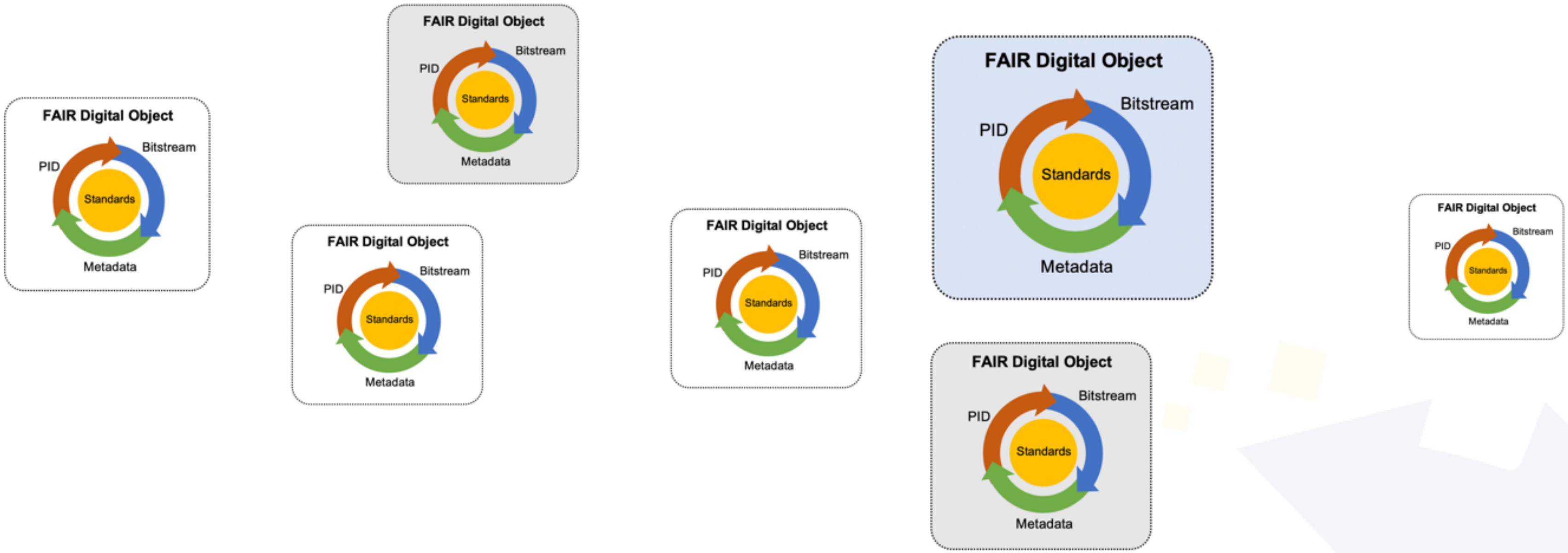

PID
Metadata
01010101001010101010
10010101011101110010
10100101001010100101
01001010101010010101


Findable
Accessible
Interoperable
Reusable

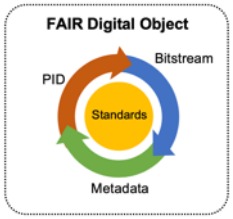


FAIR Digital Objects include many different types of academic outputs:

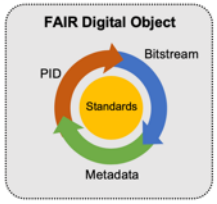
- Data sets
- Research software
- Methods
- Ontologies
- …

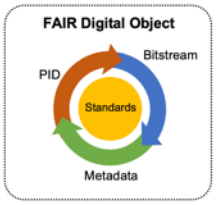

FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata

Data that's just been created (and FAIRified)

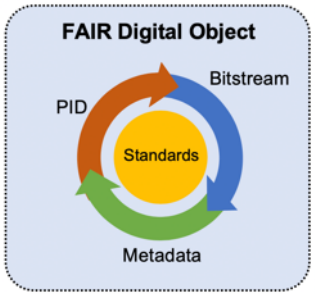


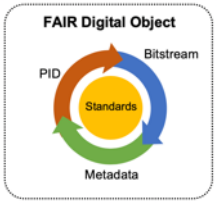
Data that's being analyzed

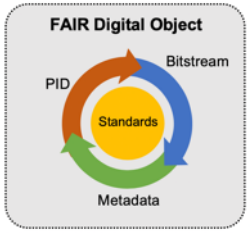




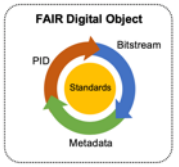
Data that's been published in a repository



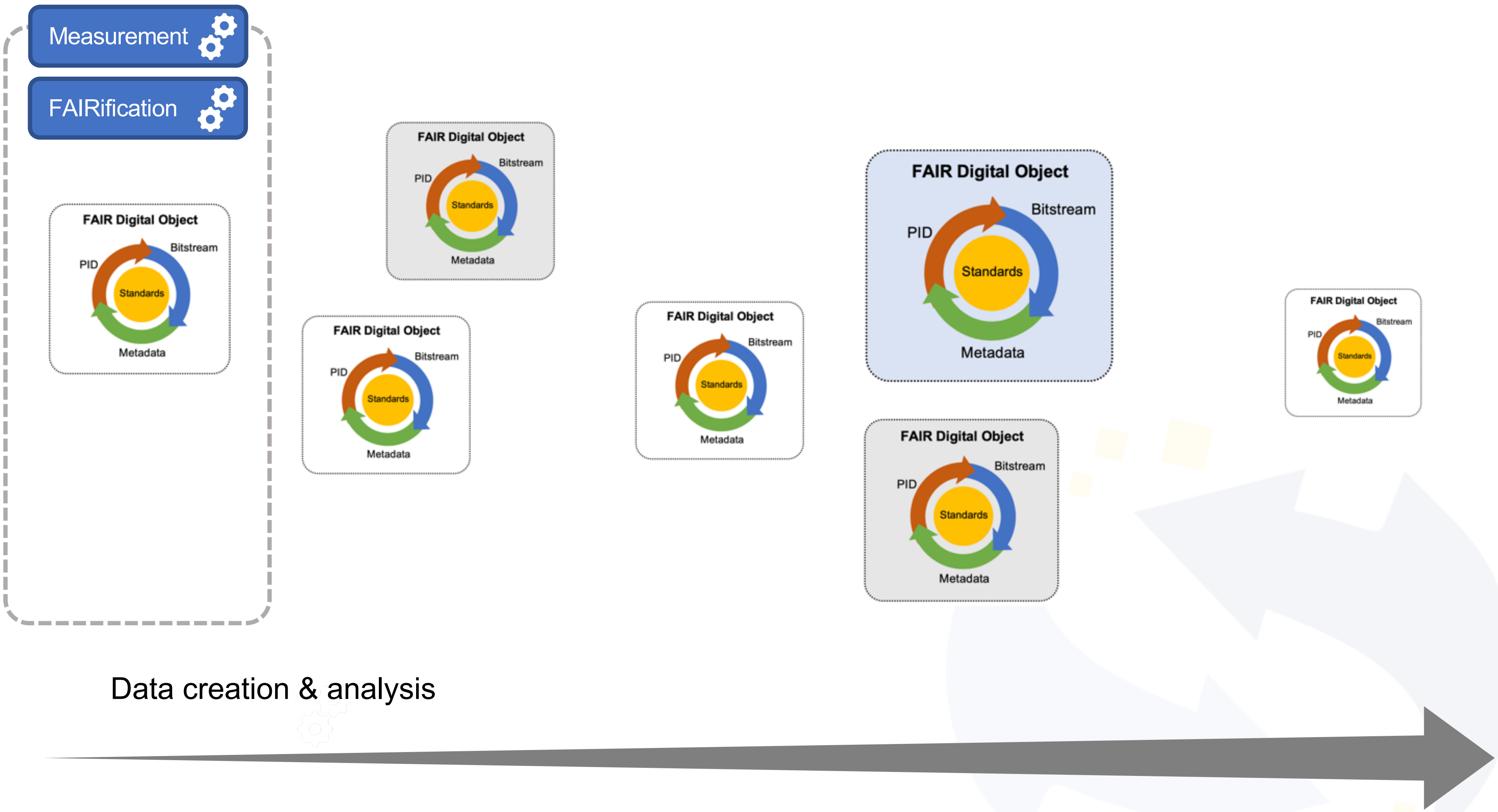


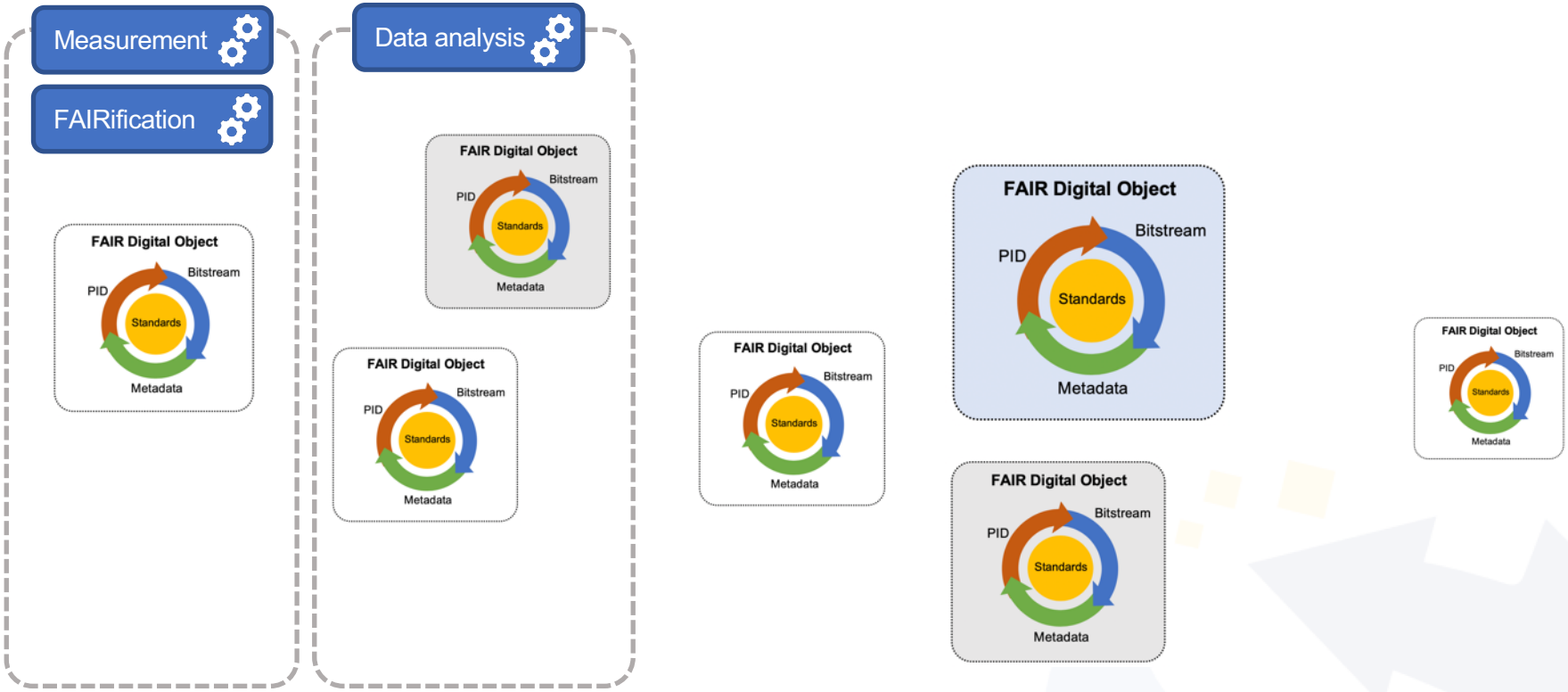


Data that's being re-used

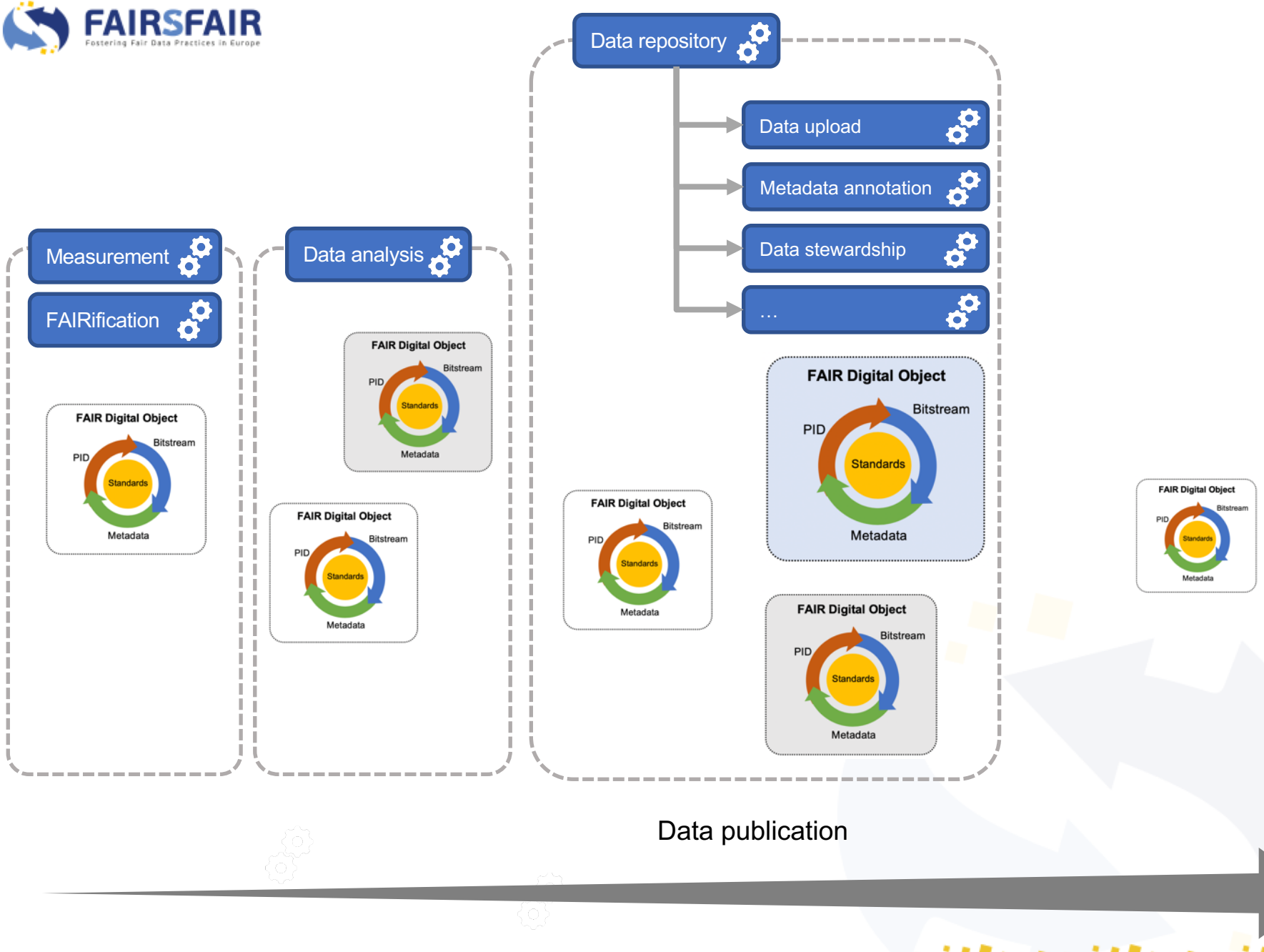


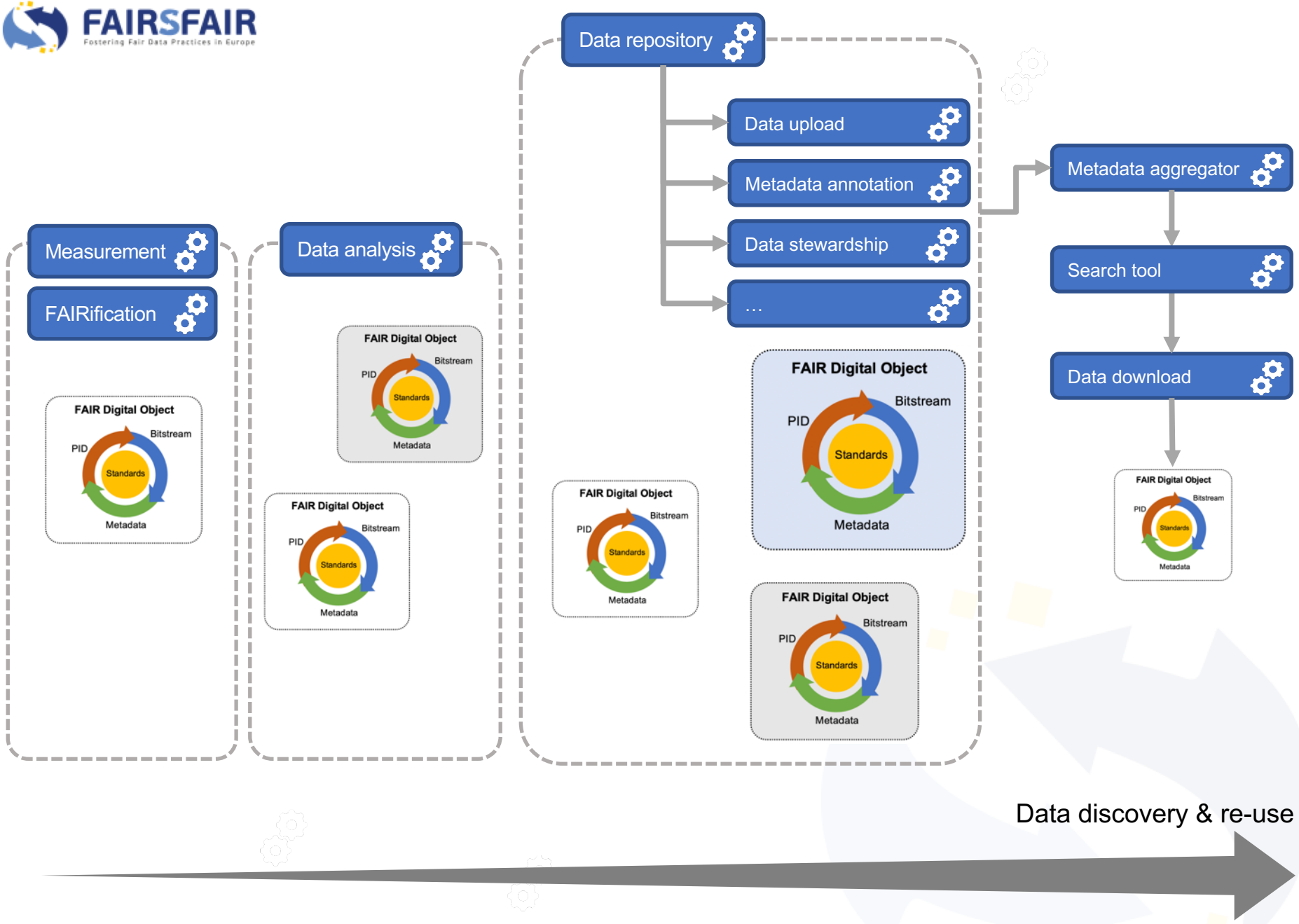
Data life cycle

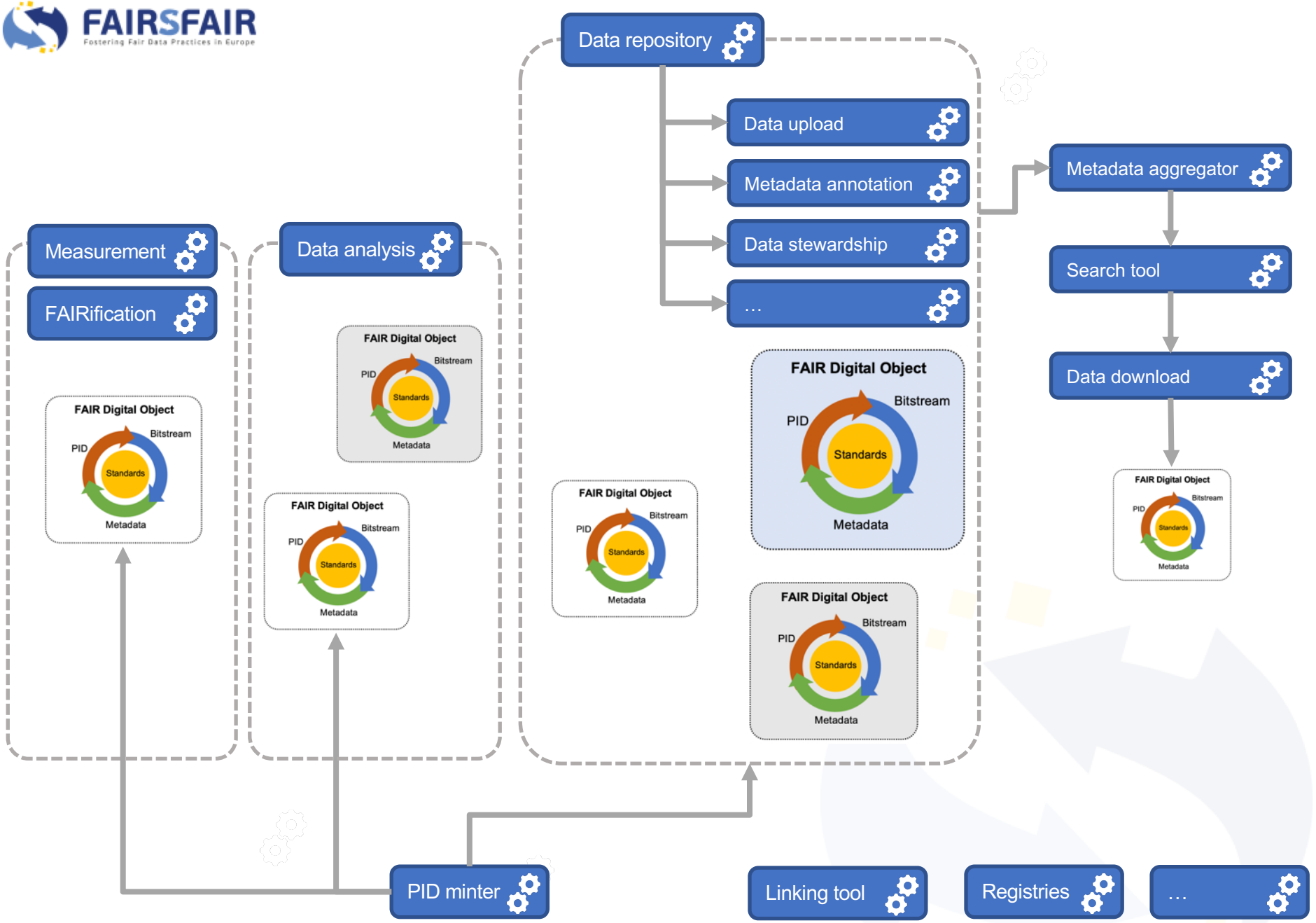

Measurement
FAIRification
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
FAIR Digital Object
PID
Bitstream
Standards
Metadata
Data creation & analysis



Data creation & analysis


FAIRSFAIR
Fostering Fair Data Practices in Europe
Data repository
Data upload
Metadata annotation
Data stewardship
...
Measurement
FAIRification
Data analysis
FAIR Digital Object
PID
Bitstream
Standards
Metadata
Data publication


FAIRSFAIR
Fostering Fair Data Practices in Europe
Measurement
FAIRification
Data analysis
Data repository
Data upload
Metadata annotation
Data stewardship
...
Metadata aggregator
Search tool
Data download
FAIR Digital Object
PID
Bitstream
Standards
Metadata
Data discovery & re-use


FAIRSFAIR
Fostering Fair Data Practices in Europe
Data repository
Data upload
Metadata annotation
Data stewardship
…
Metadata aggregator
Search tool
Data download
Measurement
FAIRification
Data analysis
FAIR Digital Object
PID
Bitstream
Standards
Metadata
PID minter
Linking tool
Registries
…

FAIRSFAIR
Fostering Fair Data Practices in Europe

FAIR Digital Objects & Services:
A perfect pair for a FAIR data ecosystem

The FAIR data ecosystem

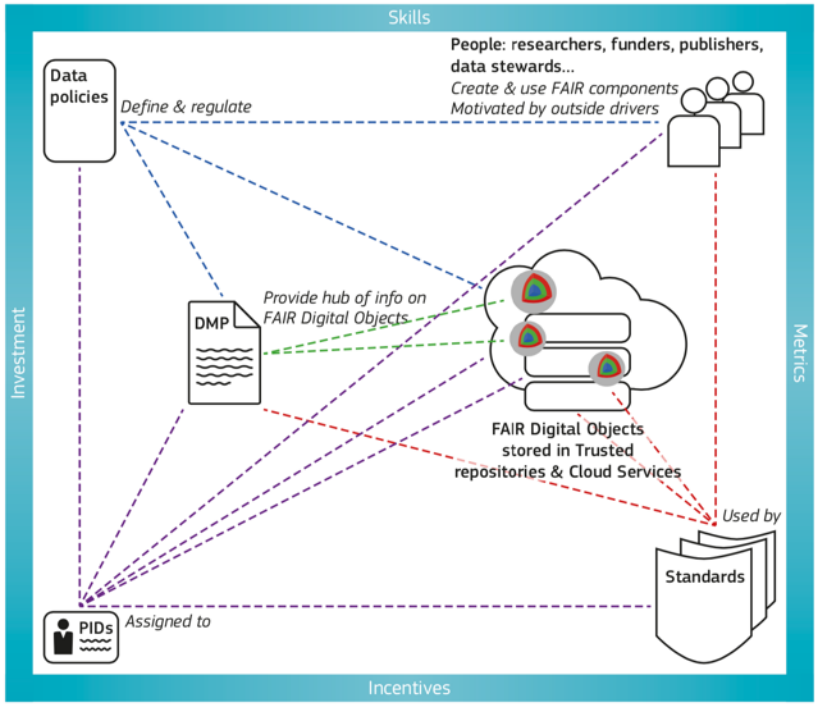


Figure 9. The interactions between components in the FAIR data ecosystem. Notes on this figure:

So.. what is the issue?

- FAIR is not an absolute, but rather a set of guiding principles that need further interpretation and definition by research communities to become truly actionable
- For ‘FAIR data’ there has been a lot of work on this recently, leading to check-lists, assessments, certification, etc.
- However, for services that is not the case → *There is little guidance for service owners on how to make their service fit in the FAIR data ecosystem*

Enter FAIRsFAIR Task 2.4: ‘FAIR services & software)

Core Objective: To deliver an assessment framework for data services that will enable and stimulate such an interplay and thus help realize the full potential of a truly FAIR ecosystem.

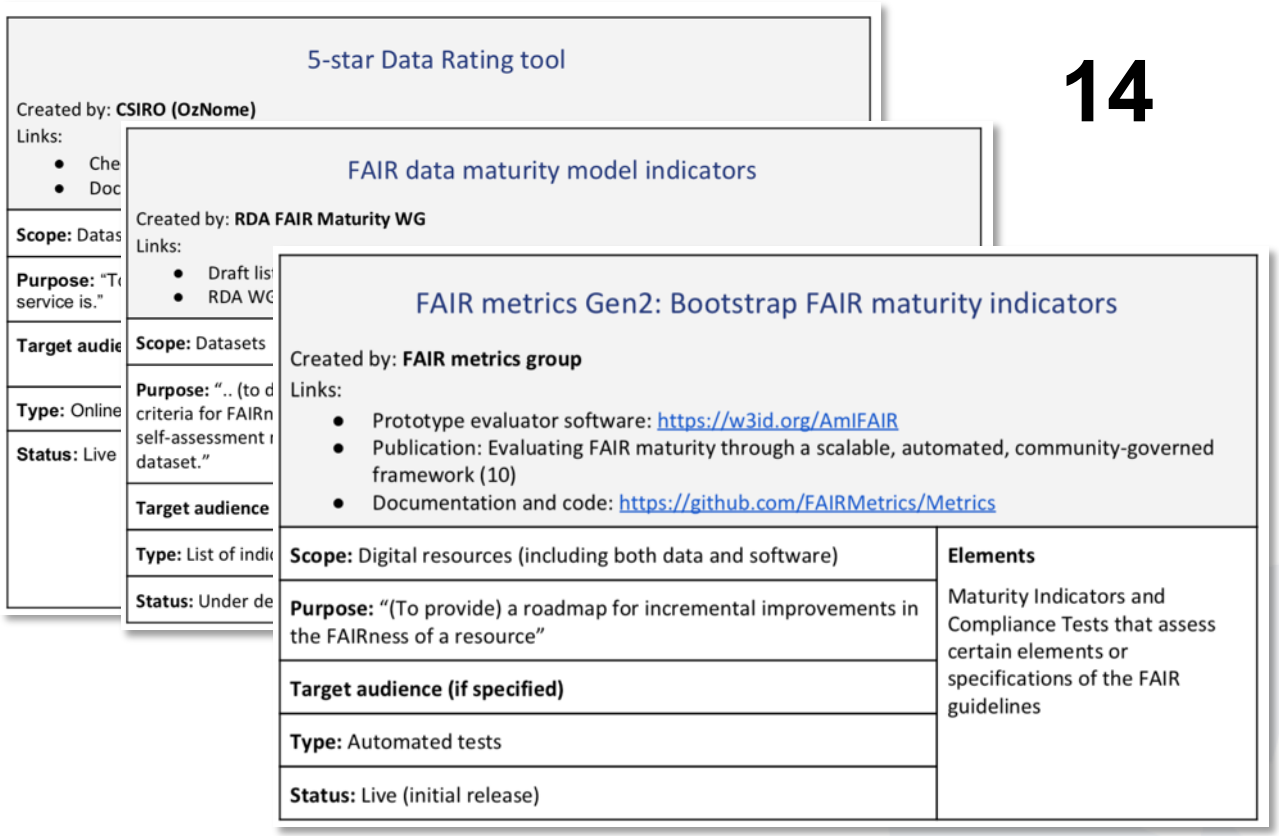
(+ FAIR assessment framework for software)

What we have done so far

- Review of FAIR assessment frameworks for data (14 in total)
- Review existing assessment frameworks for services (not necessarily 'FAIR')
- Case studies: 'How is this service *enabling FAIR'* ?
- Formulate guiding principles for the assessment framework (so, still 'meta')

And, of course, lots of interactions and discussions with stakeholders and related working groups & projects

Review of FAIR assessment frameworks for data



14

Review of FAIR assessment frameworks for services

- Assessment frameworks, including more formal certification schemes, for data repositories, notably CoreTrustSeal
- "Rules of Engagement" and other lists with criteria for services to be included in catalogues such as the European Open Science Cloud (EOSC) or the ELIXIR Core Data Resources
- General service quality requirements stemming from IT service management, for example the FitSM framework.

Case Study: How is B2FIND helping to make data FAIR?

For each FAIR principle:

- **Enable (color-coded in green)**: the service actively helps to realize this particular FAIR principle — for example by adding metadata or enabling discoverability;
- **Respect (in blue):** the service does not actively enable this particular FAIR principle, but also does not interfere with it — it can be said to respect the "FAIR-in-FAIR-out" principle;
- **Reduce (in red):** the service actually makes data less FAIR — at least for a particular principle — for example by detaching metadata or a PID when it acts on a digital object;
- N/A (not clear or not applicable; in white): This particular FAIR principle is not relevant for the service, or there was insufficient information to determine if the FAIR principle applies.

4.2. Case Study 1: B2FIND

Service Summary

B2FIND[21] is a metadata aggregator. The service harvests metadata from different community repositories and harmonises them such that users and services can search through the combined metadata. B2 FIND offers a rich faceted graphical search interface and a HTTP REST API that has been implemented in python for EUDAT's B2FIND Training[22]

URL: http://b2find.eudat.eu/ **EOSC:** https://marketplace.eosc-portal.eu/services/b2find

Users

The service targets two types of user groups:

- Scientific communities that can provide their metadata and integrate via the B2FIND service with other metadata
- Scientists who can employ the service to search for interesting research data across different communities simultaneously.

Services

- Metadata harvesting and harmonisation to communities with a tool to search across the metadata for scientists.
- The relevant metadata of a DO is shown and a link to the metadata provenance is provided.

Target Digital Objects

- Metadata entries

Examples

- B2FIND entry (KONTROL 1984[23])
- OAI-PMH dataset's metadata[24]

Purpose

B2FIND is a metadata aggregator. It gathers metadata from communities and repositories and integrates the different types of metadata. It provides a graphical user interface and an API to present the metadata and allows faceted searches across the metadata corpus.

Adoption

By now B2FIND hosts 824566 metadata entries harvested from 22 communities. We were unable to establish from the documentation how many users use B2FIND.

Documentation

EUDAT provides guidelines on how to use the B2FIND services[25] as well as detailed guidelines for harvesting and mapping metadata[26]

FAIR enablement mapping (see Annex C for details)

F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3

Case Study: How is B2FIND helping to make data FAIR?

F2. data are described with rich metadata.

ENABLE: B2FIND relies on the provided metadata by the harvested repositories. B2FIND does not further enrich metadata. **Enhances** F2 with citation metadata

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

RESPECT: B2FIND relies on the harvested repositories to attach a PID to its records and expose that as part of the metadata. If the provided metadata contains a PID B2FIND represents this PID and uses it to link the harvested metadata to the original data object in the repository.

4.2. Case Study 1: B2FIND

Service Summary

B2FIND[21] is a metadata aggregator. The service harvests metadata from different community repositories and harmonises them such that users and services can search through the combined metadata. B2 FIND offers a rich faceted graphical search interface and a HTTP REST API that has been implemented in python for EUDAT's B2FIND Training[22]

URL: http://b2find.eudat.eu/ **EOSC:** https://marketplace.eosc-portal.eu/services/b2find

Users

The service targets two types of user groups:

- Scientific communities that can provide their metadata and integrate via the B2FIND service with other metadata
- Scientists who can employ the service to search for interesting research data across different communities simultaneously.

Services

- Metadata harvesting and harmonisation to communities with a tool to search across the metadata for scientists.
- The relevant metadata of a DO is shown and a link to the metadata provenance is provided.

Target Digital Objects

- Metadata entries

Purpose

B2FIND is a metadata aggregator. It gathers metadata from communities and repositories and integrates the different types of metadata. It provides a graphical user interface and an API to present the metadata and allows faceted searches across the metadata corpus.

Examples

- B2FIND entry (KONTROL 1984[23])
- OAI-PMH dataset's metadata[24]

Adoption

By now B2FIND hosts 824566 metadata entries harvested from 22 communities. We were unable to establish from the documentation how many users use B2FIND.

Documentation

EUDAT provides guidelines on how to use the B2FIND services[25] as well as detailed guidelines for harvesting and mapping metadata[26]

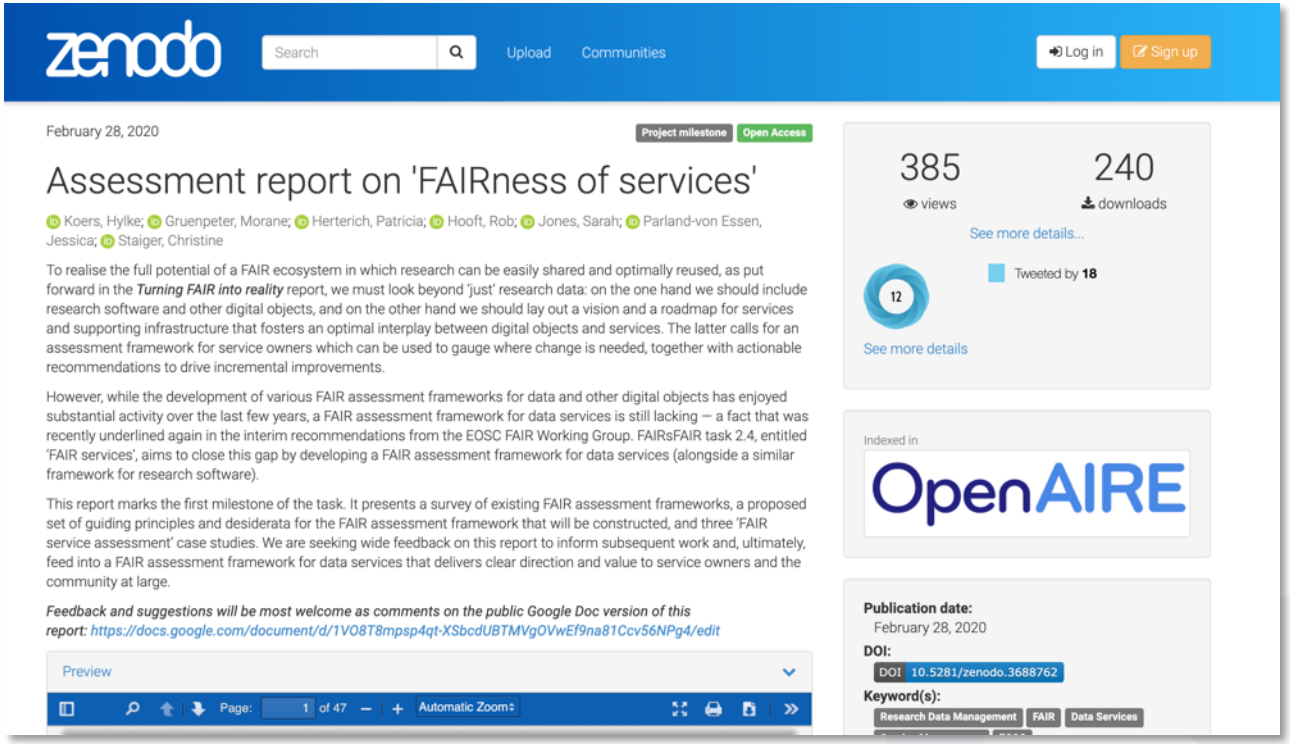
FAIR enablement mapping (see Annex C for details)

F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3

Draft guidelines for a FAIR assessment framework for services

- **Be comprehensive**, in that it applies to a broad range of functionalities across the data life cycle and across academic disciplines;
- **Be inclusive**, in that it addresses a wide array of service providers including commercial and public organizations;
- **Be rooted in FAIR data**, in that it clearly relates the FAIRness of a service to the FAIRness of the digital object that it acts on (thereby making an explicit connection to the original FAIR Data Principles);
- **Build upon existing work** as much as possible, for example extending concepts and criteria from frameworks such as CoreTrustSeal where possible;
- **Consider several dimensions of a service**, i.e. not only functional aspects ('utility' in FitSM terms) but also aspects that speak to quality, documentation, sustainability and trustworthiness ('warranty') — where human factors including capacity building and training will be critical;
- **Be actionable** and aligned with the needs of the intended audience, in that parties developing or delivering data services can use it to, very practically, know what to put on their development roadmaps;
- **Be validated** by pilots and tests, in that the framework does not just live on paper but has been tested and practice — ideally with working exemplars; and
- **Be supported by the community**, in that it may count on informal support and formal endorsement by the broader community.

First report on Zenodo



Report
https://zenodo.org/record/3688762#.XoY_W9MzbOR

Feedback very welcome through public Google Doc:
https://docs.google.com/document/d/1VO8T8mpsp4qt-XSbcdUBTMVgOVwEf9na81Ccv56NPg4/edit

Team 2.4

- Roberto Di Cosmo (INRIA)
- Morane Gruenpeter (INRIA)
- Patricia Herterich (DCC)
- Rob Hooft (DTL)
- Sarah Jones (DCC)
- Hylke Koers (SURF, task lead)
- Jessica Parland-von Essen (CSC)

And earlier contributions from:

- Christine Staiger (DTL)

Do you run a data service?

Email me at hylkek@surfsara.nl

FAIRSFAIR
Fostering Fair Data Practices in Europe

What does it take for a service to be FAIR?

FAIRSFAIR
Fostering Fair Data Practices in Europe

What does it take for a service to **enable** FAIR?

FAIRSFAIR
Fostering Fair Data Practices in Europe

A FAIR assessment framework for *data services*

FAIRSFAIR
Fostering Fair Data Practices in Europe

We’re on it

FAIRSFAIR
Fostering Fair Data Practices in Europe

And you can help!