



Persistent Identifiers: Consolidated Assertions

ASOV Workshop 2020.

Carlo Zwölf, Peter Wittenburg, Margareta Hellström,
Szuzsanna Szeredi

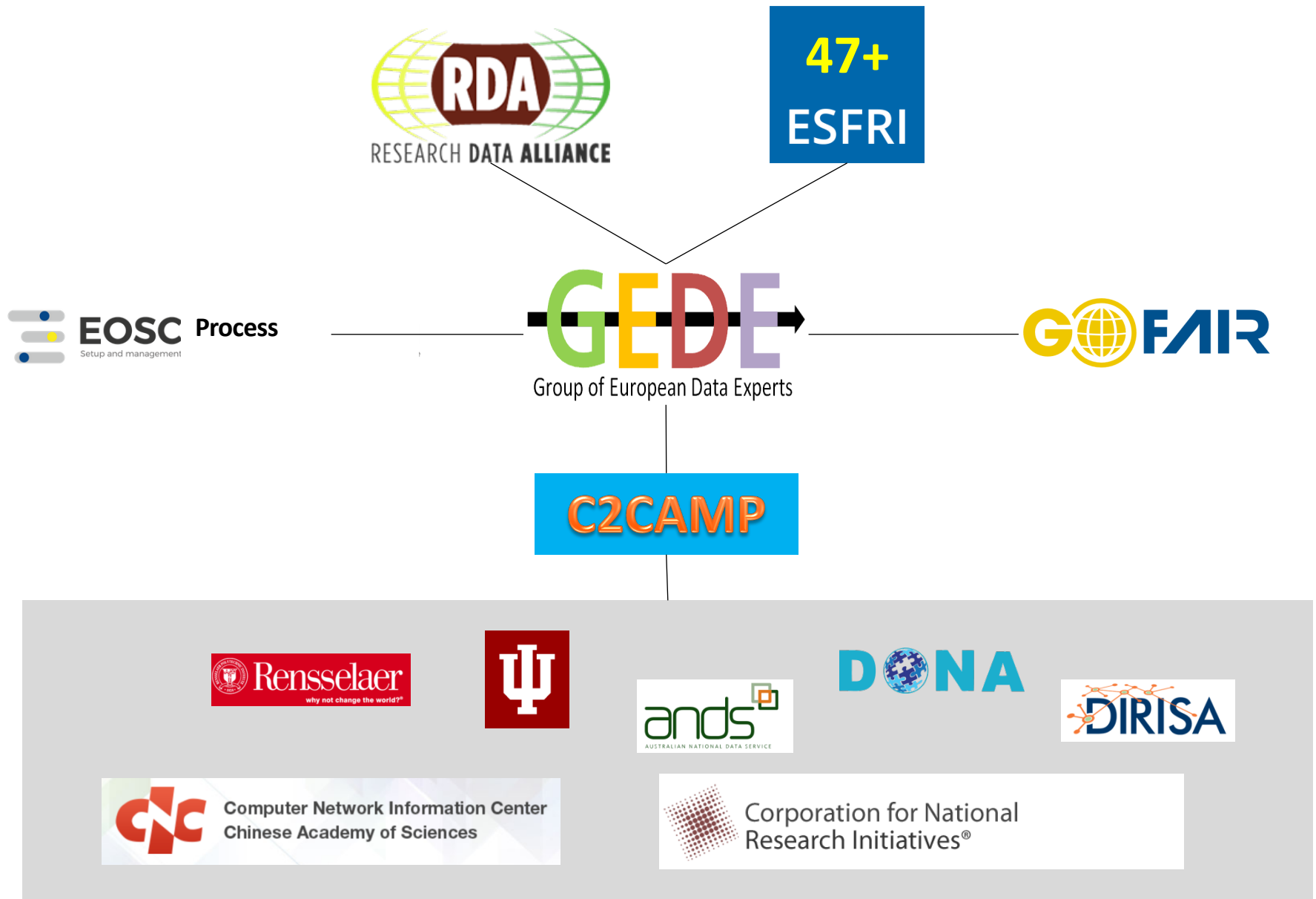


What is **GEDE**?

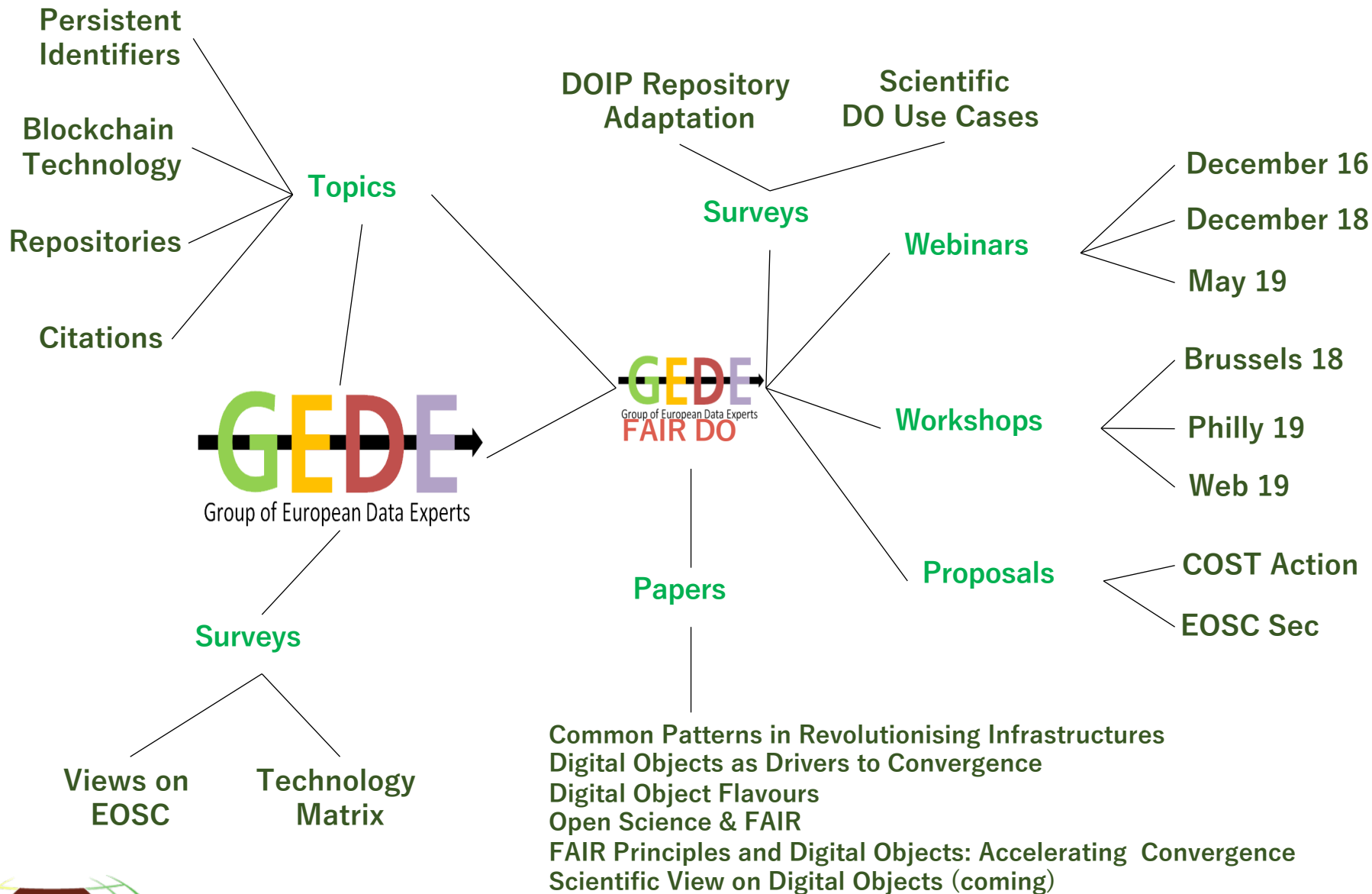
Group of European Data Experts

- Founded in 2015 by Leif Laaksonen and Peter Wittenburg within the RDA Europe 3 Project and is recognised as an RDA Coordination Group
- Aim at improving the interactions with the “practitioners” who are in charge of building infrastructures in the ESFRIs and ERICs and who have the deep experience.
- DOERs Network: high reputation, Fast reaction, cross-domain,
- GEDE is working with small funds – now being supported by RDA Europe 4. Most of the work in GEDE is done on a voluntary basis.
- Delegates of 47 + research infrastructures (mostly ESFRI/ERIC)
- Elected co-chairs
 - Carlo Maria Zwölf (VAMDC)
 - Peter Wittenburg (GEDE, RDA)
 - one to be elected in November (Maggie Hellström left in 2019)
- Topic groups dependent on interests
- Management Support:
 - secretariat: Zsuzsanna Szeredi
 - policy support: Edit Herczog

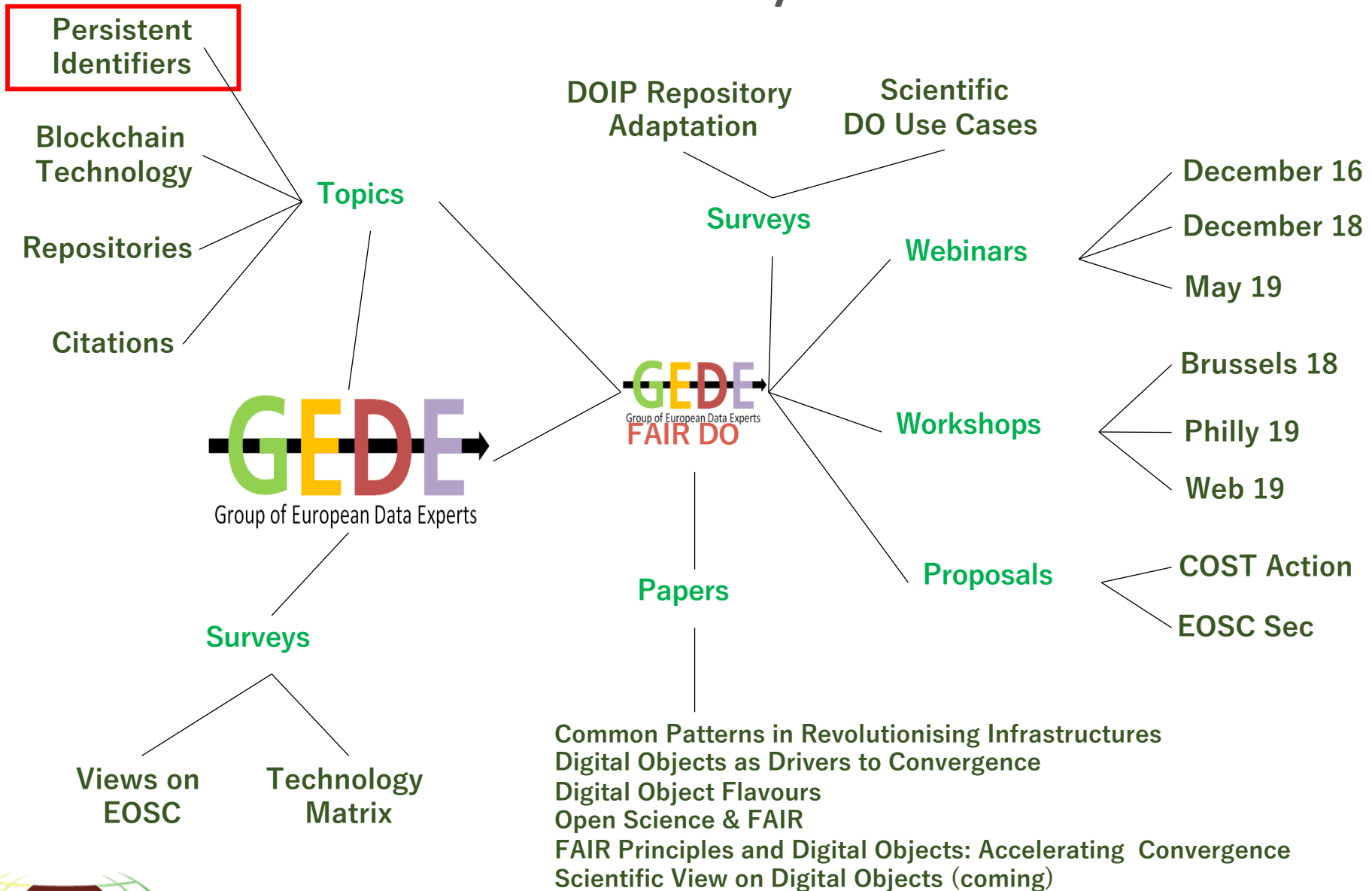
The GEDE Interaction Chart



The GEDE Activity Chart



The GEDE Activity Chart

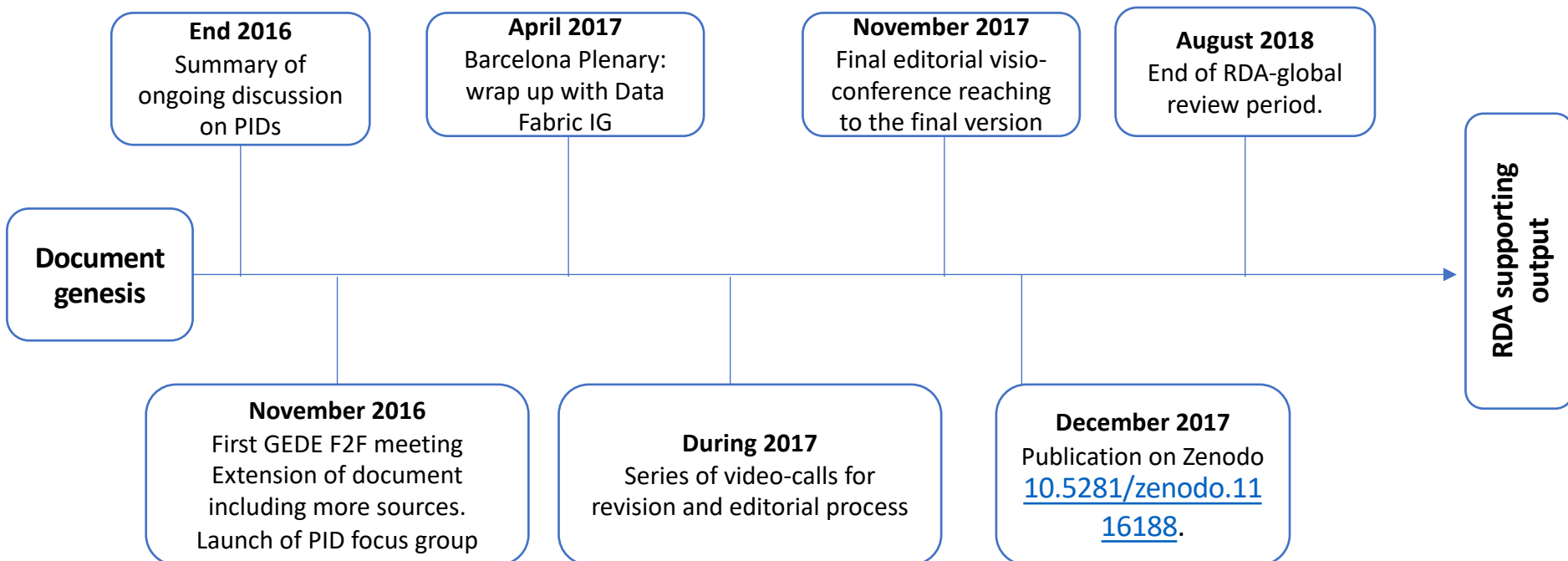


Persistent Identifiers: consolidated assertions

- Bring more clarity about persistent identifiers assigned to data in research communities
 - Many documents exist on this topic in many research communities, using different terminologies
- GEDE goal :
 - not produce another document with new ideas, but a compilation of common of relevant assertions (extracted statements from original sources), with context information leading up to a summary of agreements and disagreements reflecting the convergence of understanding.
- Work carried out in a focus group common to GEDE and RDA-Data Fabric Interest Group

Persistent Identifiers: consolidated assertions

- Bring more clarity about persistent identifiers assigned to data in research communities
 - Many documents exist on this topic in many research communities, using different terminologies
- GEDE goal :
 - not produce another document with new ideas, but a compilation of common of relevant assertions (extracted statements from original sources), with context information leading up to a summary of agreements and disagreements reflecting the convergence of understanding.
- Work carried out in a focus group common to GEDE and RDA-Data Fabric Interest Group



Persistent (Resolvable) Identifiers

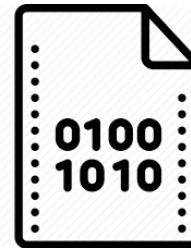
MyIDProvider/5afaa6c3-9f46-45d3-8c39-3d3d4c70536b

[10.5281/zenodo.1116188](#)

Global resolver



Landing Page



Underlying Bit
sequence



Metadata state
information

Persistent (Resolvable) Identifiers

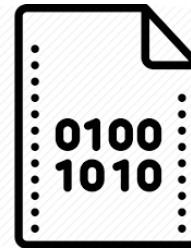
MyIDProvider/5afaa6c3-9f46-45d3-8c39-3d3d4c70536b

[10.5281/zenodo.1116188](#)

Global resolver



Landing Page



Underlying Bit
sequence



Metadata state
information

Persistence of

PID

PID resolution

PID management

Access to data

Structure and content of the document

Sources:

~ 30 documents
and whitepapers
from many
research
communities
dealing with data

Structure and content of the document

Consolidates assertions

- Building common semantic background
- Focusing on core message
- **63 consolidated assertions**

Definition

Best practices

Strategy

granularity

Provenance

versioning

Policy

PID-record

Community-specific

semantics

Sources:

~ 30 documents
and whitepapers
from many
research
communities
dealing with data

Structure and content of the document

Convergence of definitions and understanding

- Summarize the essential from the collected assertions

Nature of PID and PID systems

Relevance of PID and PID systems

Assigning PIDs

Using PIDs

Handles and DOIs

Consolidates assertions

- Building common semantic background
- Focusing on core message
- **63 consolidated assertions**

Definition

Best practices

Strategy

granularity

Provenance

versioning

Policy

PID-record

Community-specific

semantics

Sources:

~ 30 documents and whitepapers from many research communities dealing with data

Structure and content of the document

Convergence of definitions and understanding

- Summarize the essential from the collected assertions

Nature of PID and PID systems

Relevance of PID and PID systems

Assigning PIDs

Using PIDs

Handles and DOIs

Consolidates assertions

- Building common semantic background
- Focusing on core message
- **63 consolidated assertions**

Definition

Best practices

Strategy

granularity

Provenance

versioning

Policy

PID-record

Community-specific

semantics

Sources:

~ 30 documents and whitepapers from many research communities dealing with data

Appendixes :

- PIDs forms
- Use of semantics
- PID, landing pages
- Considerations around persistence

Example of assertion: PID definition

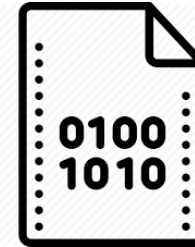
MyIDProvider/5afaa6c3-9f46-45d3-8c39-3d3d4c70536b

[10.5281/zenodo.1116188](https://doi.org/10.5281/zenodo.1116188)

Global resolver



Landing Page



Underlying Bit
sequence



Metadata state
information

PID-1. [RDA DFT 1.2] {definition}

A persistent identifier is a long-lasting ID represented by a string that uniquely identifies a DO and that is intended to be persistently resolved to meaningful state information about the identified DO.

In this document we are using the definition as given by the RDA DFT Core Term and Model¹⁰ document. Note that in our context DO stands for digital object, which may be either a document, dataset, piece of software, a service or other similar non-physical entities, or the digital representation of an analogue entity, such as a device, a sample, a person or an organisation. Also see the PID-3 definition below, and Appendix 3, for an elaboration on "state information".

Key points

- Is this a good Identifier?
 - [MyDomain/results/VO-Table1.4-format/Query123](#)

Key points

- Is this a good Identifier?
 - [MyDomain/results/VO-Table1.4-format/Query123](#)
 - **No! Do not use PID encoding semantics. Put instead all the relevant information as metadata information of the landing page. NOT IN THE PID.**

Key points

- Is this a good Identifier?
 - [MyDomain/results/VO-Table1.4-format/Query123](#)
 - No! Do not use PID encoding semantics. Put instead all the relevant information as metadata information of the landing page. NOT IN THE PID.
- What is the good granularity for assigning PID?
 - A good PID system is designed to react to each applicable loads.
 - Assign PID with the granularity required by your science-case.
- Assign PID to each element: data, authors (ORCID), services, documents of standards, etc...
 - Metadata state information will contain PID for defining authors, references, used standard, codes, ...
 - Thus PID are the basis of Digital Objects architectures (as Fair Digital Objects Framework)
 - PIDs are a key element in data citation
- Live as your data were to die tomorrow. Preserve as your data were to live forever
 - When a PID is assigned to data, prepare its tombstone.
 - Prepare the information that will be returned while resolving a PID, when the underlying data will be removed.
- DOIs are just another PID mechanisms (subset of Handle)
 - Don't be DOI-centric. There are others PID out of DOIs.
 - Example: Life science IDs en.wikipedia.org/wiki/LSID

Conclusions and further works

- PID work was the launch pad for the ongoing GEDE activities, which may be seen as a natural extension
 - PID are strictly linked with Digital Object and FAIR-Digital Objects;
 - PID are used for citation, since we always need an anchor to refer to;
 - Repositories rely on PID and assure the proper PID administration and maintenance.

Data Fabric Interest Group

Group co-chairs: Jianhui Li, Tobias Weigel

Supporting Output title: Persistent identifiers: Consolidated assertions

Editors: Peter Wittenburg, Margareta Hellström, and Carlo-Maria Zwölf

Co-authors: Hossein Abroshan, Ari Asmi, Giuseppe Di Bernardo, Danielle Couvreur, Tamas Gaizer, Petr Holub, Rob Hooft, Ingemar Häggström, Manfred Kohler, Dimitris Koureas, Wolfgang Kuchinke, Luciano Milanesi, Joseph Padfield, Antonio Rosato, Christine Staiger, Dieter van Uytvanck and Tobias Weigel

DOI: [10.15497/RDA00027](https://doi.org/10.15497/RDA00027)

Citation: Peter Wittenburg, Margareta Hellström, and Carlo-Maria Zwölf (eds.), Hossein Abroshan, Ari Asmi, Giuseppe Di Bernardo, Danielle Couvreur, Tamas Gaizer, Petr Holub, Rob Hooft, Ingemar Häggström, Manfred Kohler, Dimitris Koureas, Wolfgang Kuchinke, Luciano Milanesi, Joseph Padfield, Antonio Rosato, Christine Staiger, Dieter van Uytvanck and Tobias Weigel (2017): Persistent identifiers: Consolidated assertions. DOI: [10.15497/RDA00027](https://doi.org/10.15497/RDA00027).

Summary:

Experts from 47 European research infrastructure initiatives and ERICs have agreed on a set of assertions about the nature, the creation and the usage of Persistent Identifiers (PIDs). This work was done in close synchronisation with the RDA Data Fabric Interest Group (DFIG) ensuring a global validation of the assertions. The intention of this cross-disciplinary report is to overcome still existing confusions about PIDs and the lack of detail knowledge in many disciplines. It is not meant to produce yet another comprehensive document on PIDs, but to identify agreements across documents that have been suggested to be included by experts. With this document GEDE is happy to help demystify PIDs, overcome confusion and create bridges between the various disciplines.

Supporting Output file:

https://rd-alliance.org/system/files/PID-report_v6.1_2017-12-13_final.pdf

Data Fabric IG

Status: Recognised & Endorsed

Chair(s): Jianhui Li, Tobias Weigel, Robert Quick

Public - accessible to all site users

[Join Group](#)

[Index](#)

[Add new content](#)

- [Group Wiki](#)
- [Group Mailing list Archive](#)

Group sessions at RDA Plenaries

Governance and Implementation of the Intelligent Data Fabric

By [Tobias Weigel](#) On 26, Jun 2019

Case Statement

Data Fabric IG Case Statement
02 July 2014

[Comments 2](#)