The pan-European Data Infrastructure

Research data services, expertise & technology solutions

www.eudat.eu

EUDAT is …

- a pan-European initiative building a sustainable **cross-disciplinary** and **cross-national** data infrastructure providing a set of shared services for accessing and preserving research data

- supporting **multiple research communities** by working closely with them to deliver these technical services as part of the EUDAT Collaborative Data Infrastructure (CDI)
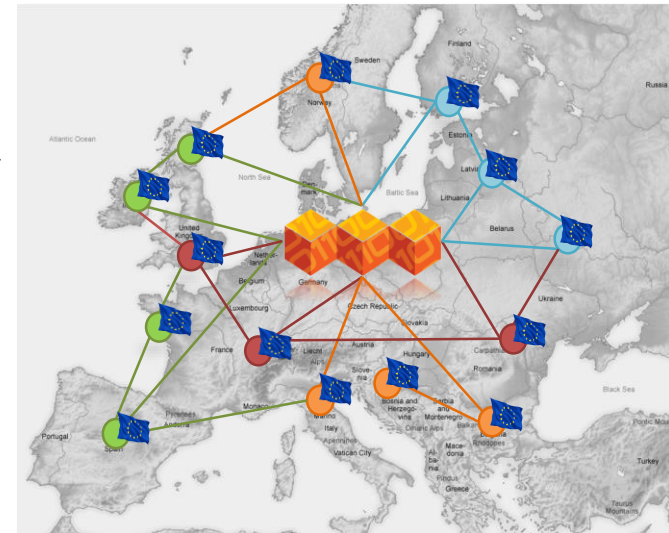
Why Does EUDAT Need RDA?
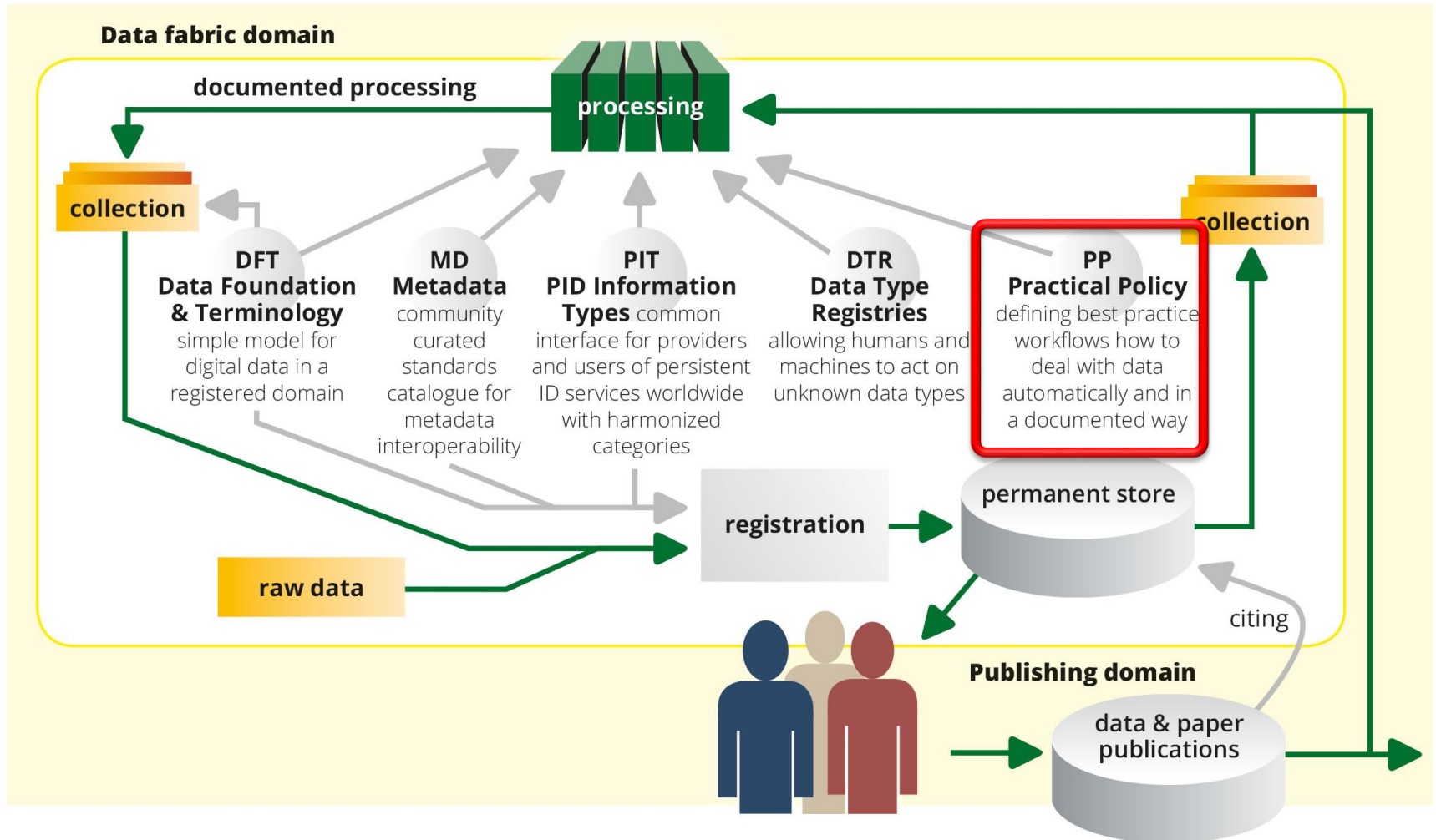
B2 SERVICE SUITE

**http://www.eudat.eu/services**
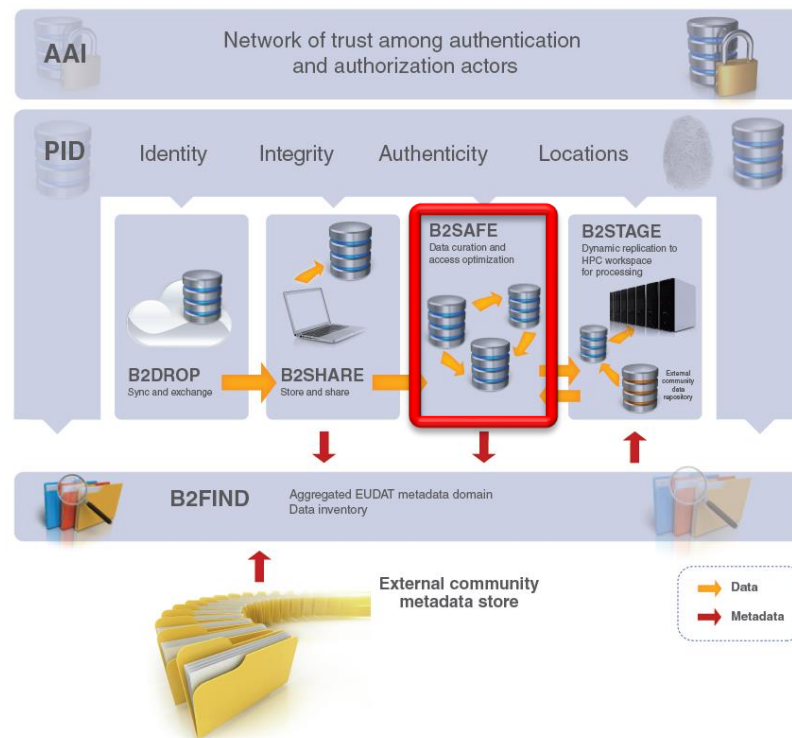
Practical Policy Working Group

Computer actionable policies are used to

- enforce data management

- automate administrative tasks

- validate compliance with assessment criteria

- automate scientific data processing and analyses

Practical Policy
Assertion or assurance that is enforced about a (data) collection (data set, digital object, file)

Policy Categories
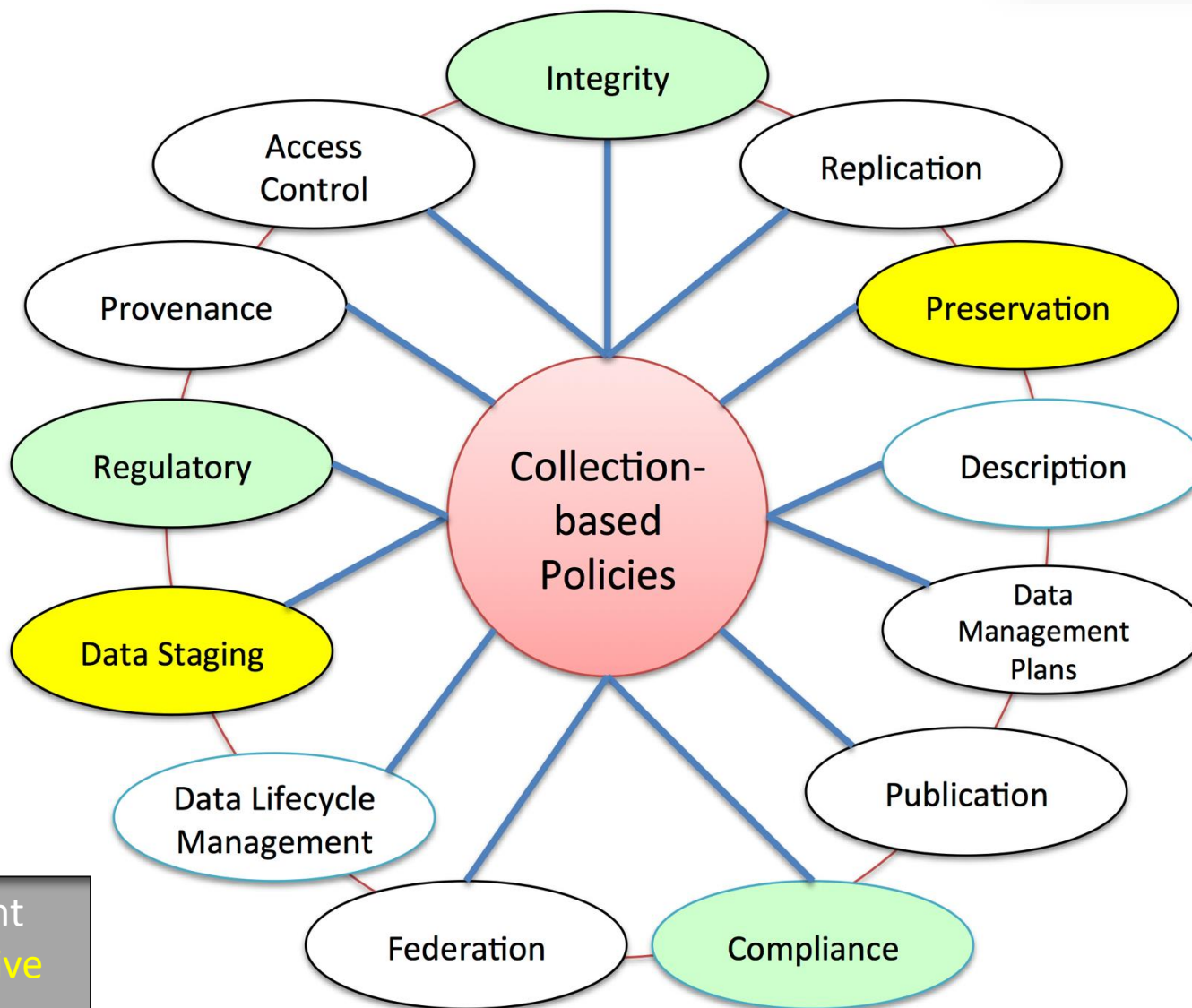
EUDAT Community requirements

Community managers (CM) can specify M replicas to be stored for N years.

CM can specify which objects/collections need to be copied.

CMs want to know whether the replicas are identical with the source.

CMs want to execute regular checks across all object instances and give indications in case of differences.

CMs want the replicas to be accessible by users while maintaining the access permissions as defined by the originating community center.

EUDAT Objectives

- allow a Community Manager (CM) to specify data management policies

- allow a Community Manager (CM) to manage those policies via a web-portal

Replication

1839/abc
29db…279b4a
1.10.13 00:00

CM wants to specify the type of replication, the replica sites and how often data shall be replicated

Replication from A to B and from B to C
Replication from A to B and from A to C

456/abc
29db…279b4a
1.10.13 02:00

789/abc
29db…279b4a
1.10.13 02:00

Integrity

EUDAT

1839/abc

29db...279b4a

1.10.13 00:00

CM wants to specify the periodicity of integrity checking

456/abc

29db...279b4a

1.10.13 02:00

789/abc

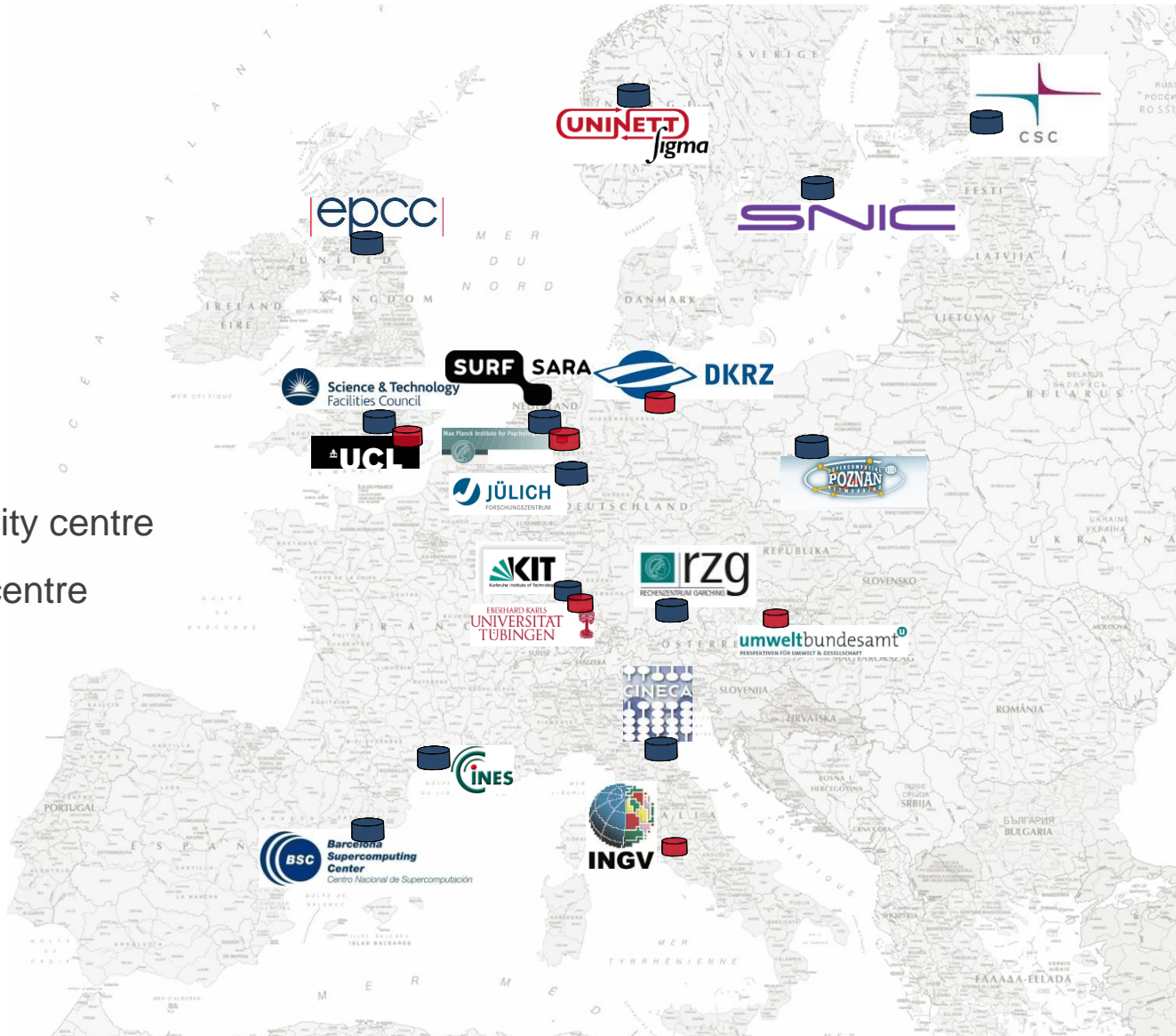29db...279b4a

5.10.13 05:00

Checksum recalculation on the physical file
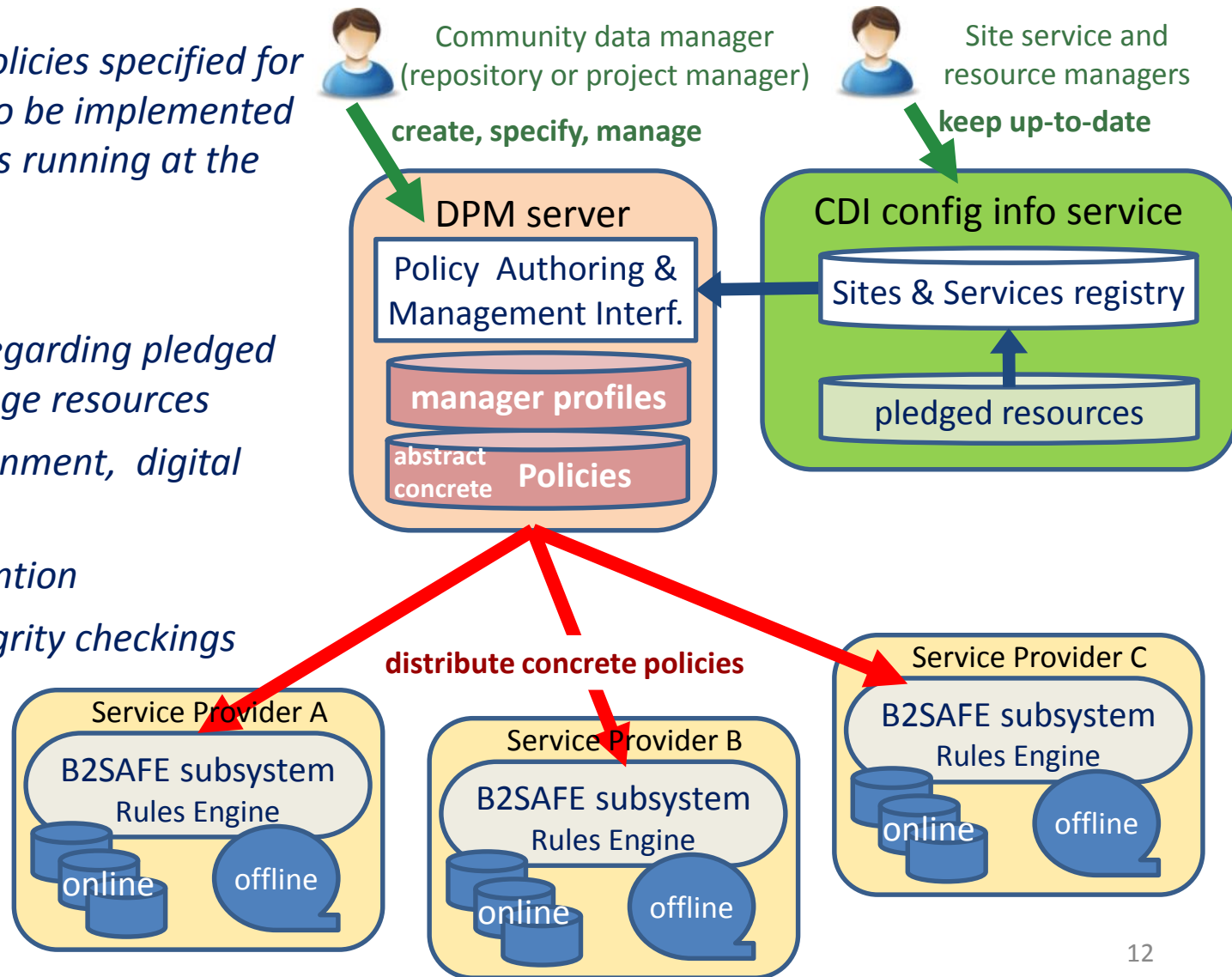
EUDAT Distributed CDI

Data Policy Manager Architecture

Data Management Policies specified for community projects to be implemented by B2SAFE subsystems running at the different sites.

Policy examples:

- *replication rules regarding pledged and required storage resources*

- *rules for PID assignment, digital object linking*

- *rules for data retention*

- *rules for data integrity checkings*

EUDAT

Community data manager (repository or project manager)

create, specify, manage

Site service and resource managers

keep up-to-date

DPM server

Policy Authoring & Management Interf.

manager profiles

abstract
concrete **Policies**

CDI config info service

Sites & Services registry

pledged resources

distribute concrete policies

Service Provider A
B2SAFE subsystem
Rules Engine
online   offline

Service Provider B
B2SAFE subsystem
Rules Engine
online   offline

Service Provider C
B2SAFE subsystem
Rules Engine
online   offline

Conclusions

- EUDAT provides building blocks for data management within a CDI

- Supporting data management policies in a distributed environment is core to the CDI infrastructure

- Harmonization and standardization of policies is important (RDA Policy templates)

- EUDAT currently supports policies on data replication and data integrity

- Extending support for data retention, data curation and data provenance