



OpenAIRE

Open Access Infrastructure for Research in Europe

OpenAIRE/plus Technical Infrastructure

Donatella Castelli

CNR-ISTI

Paolo Manghi

CNR - ISTI



Past, Present, OpenAIREplus



OpenAIREplus Guidelines for Content Providers



Metadata



Dataset repositories,
CRIS Systems,
Registries

6,300,000 OA articles
340 validated repositories

OpenAIRE Guidelines v2.0

FP7 publications
National funding publications

Driver Guidelines

OpenAIRE Guidelines v1.0



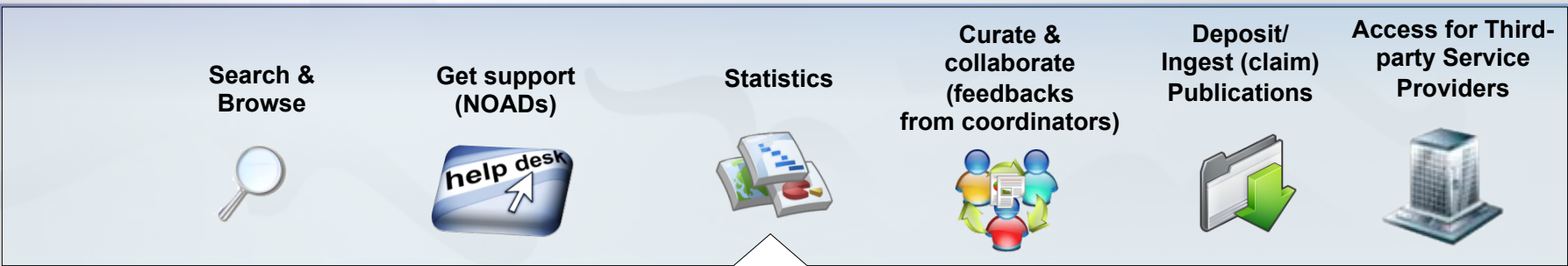
**Publication
repositories network**
Institutional & Thematic



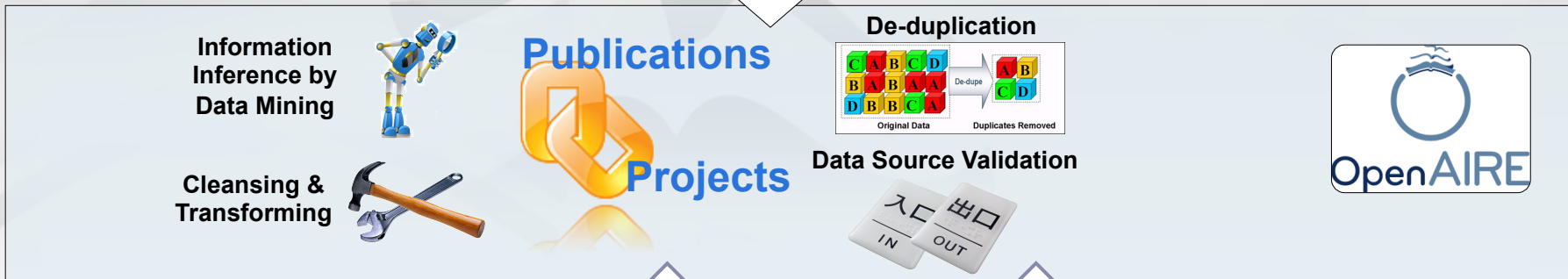
*EC Project
metadata*

*National
Project
metadata*

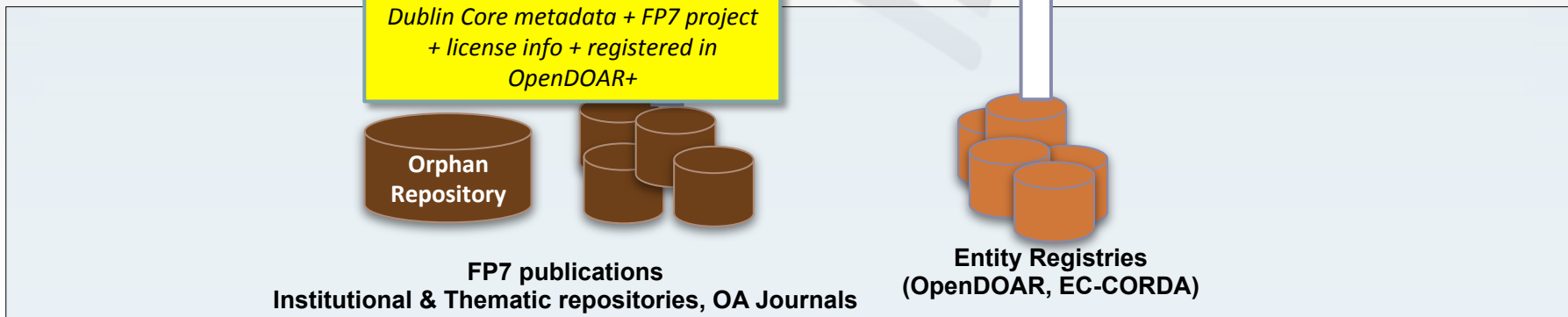
+ OpenAIRE Production Infrastructure in November 2012



Usage policies

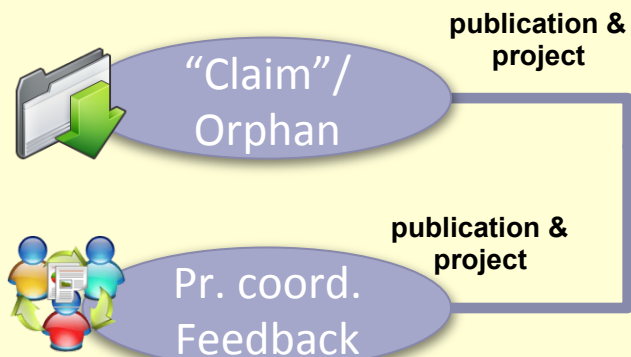


OpenAIRE guidelines



+ Measuring research impact “European Commission”

Manual provision:
Tools to link publication
metadata to EC projects



Publications
Projects



publications & projects



Inference
by mining

Automated inferred provision:
EC project references embedded in publication text

Any relationships &
entities



Collection

Automated provision:
Supported by compliance
to “Guidelines
for Repository
Providers”

Publication Repositories,
entity registries (OpenDOAR, EC-CORDA)



Technical activity performed

○ e-infastructures

- Development (CNR)
- Production (ICM)
 - 37 block upgrades
 - 105 package upgrades

(number of unscheduled downtimes= 0)

○ Approx. 16 full time developers & designers + 3 infras administrators (2 for production, 1 for development)

○ Other numbers: 285 modules, 358,786 lines of java code



Technological adoptions & alliances

f2f Training

- SIU-Argentina (24-28 Sept. 2012, Pisa)
- FEYCET (Spain) & CNR (Italy) (10-14 Dec. 2012, Pisa)

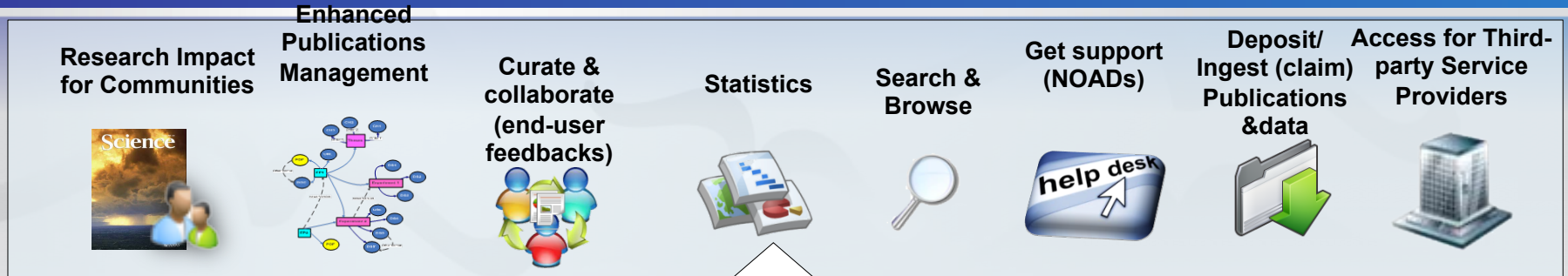
D-NET other enabled e-Infrastructures

- EFG (EFG1914) HOPE, ESPAS, Recolecta (Spain), National Repository Infrastructure (Poland), National Repository Infrastructure (Argentina)

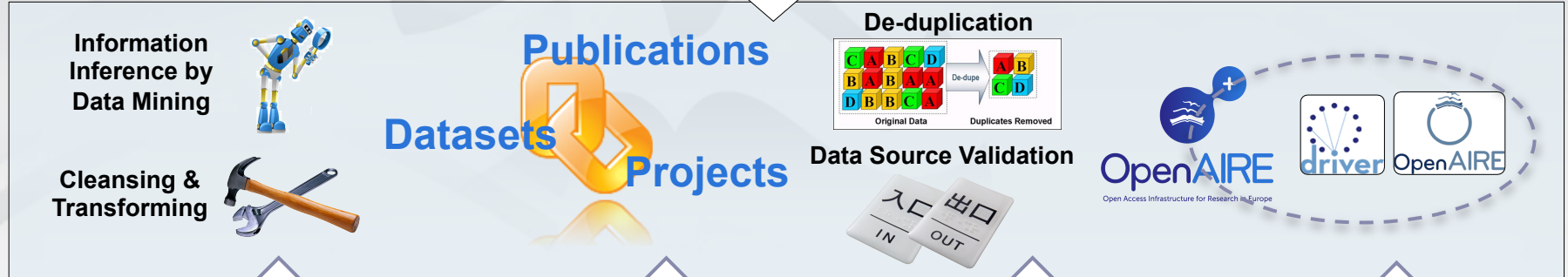
Remark1: The majority of these initiatives are contributing to the D-Net open source enrichment

Remark2: These initiatives demand for D-Net sustainability

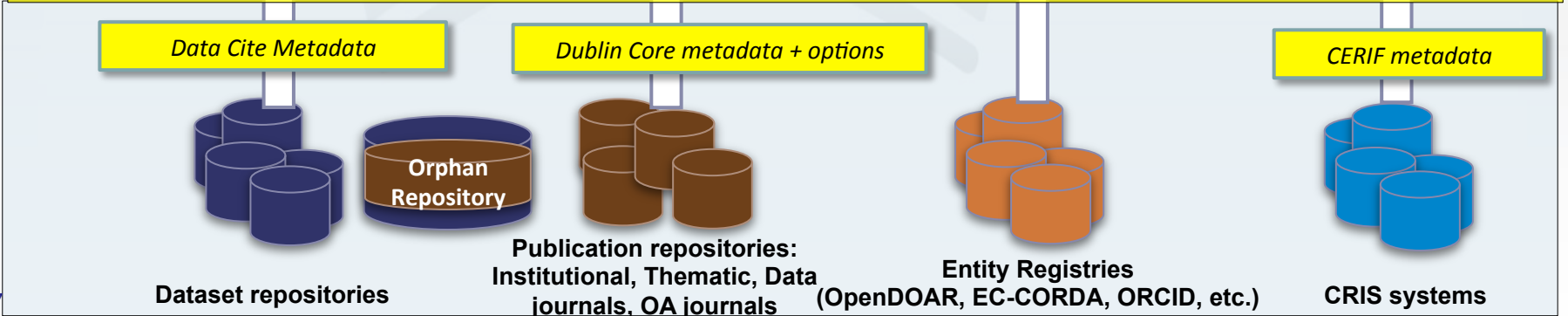
OpenAIRE Production Infra by the end of OpenAIREplus



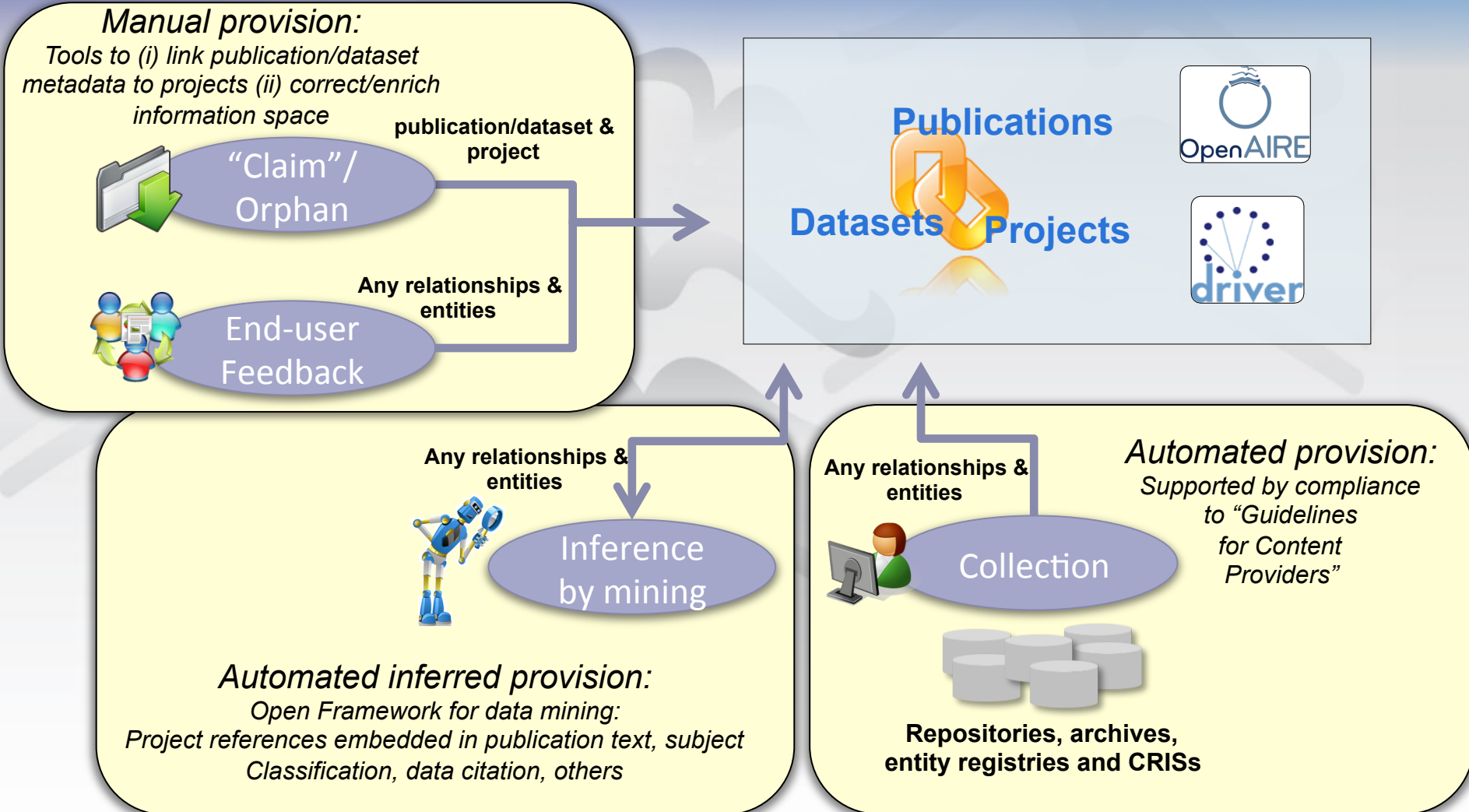
Usage policies



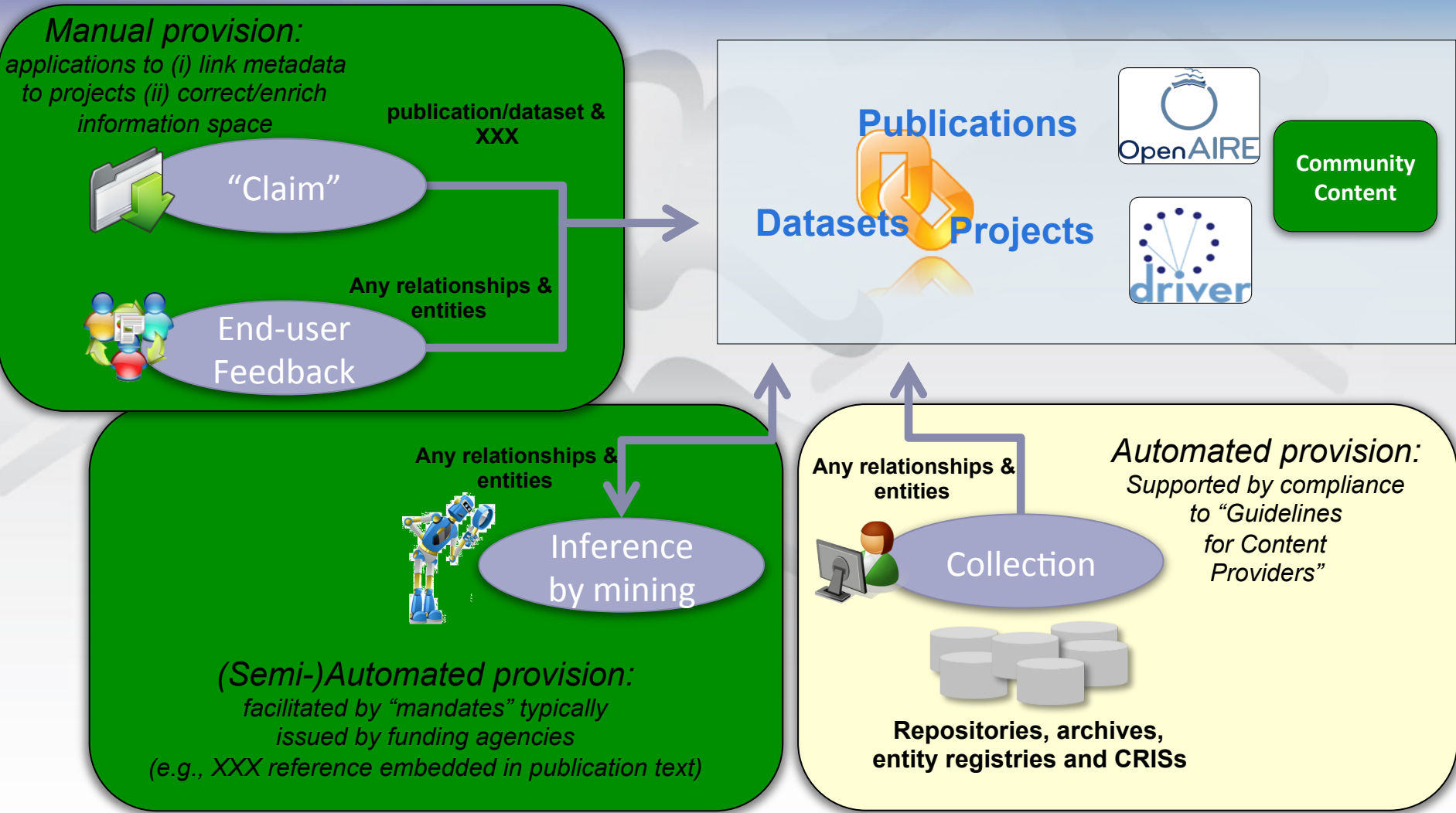
OpenAIREplus guidelines



+ Measuring research impact “European Commission” use-cases



+ Measuring research impact: “Research Impact for Communities”





Data Curation and Enrichment

Information Inference (mining framework)

- Relationship Inference: Pubs-Datasets, Citations, Similarity
- Subject Classification
- Project-Article
- De-duplication

Service Access

- SRW/CQL Publisher
- OAI-PMH Publisher
- OAI-ORE Publisher

End-user Access

- OpenAIREplus portal
- Helpdesk
- Statistics
- Enhanced Publication
- User feedback
- Deposition and Claiming

Data Storage and Indexing

HBase/Hadoop Framework

- Metadata Store
- SQL Database
- File Store
- Full-Text Index

Data mapping

- Mapping rules
- Metadata Transformation
- Metadata Cleaner
- Cleaning rules

Data mediation

- Internet protocols (FTP, HTTPs, etc)
- OAI-PMH Harvester
- Data Source Management
- Data Source Validator
- Validation rules

Data Sources

- Repositories
- Registries
- CRIS
- Data archives
- Orphan Repository



Production OpenAIREplus *Releases*

BETA production v1.0:
December 2012 (M10)

- testing from NOADs and selected users
- Integrating DRIVER infrastructure
- New content management services

Production OpenAIREplus
v2.0: February 2013 (M14)

- Advanced de-duplication

Production OpenAIREplus
v3.0: August 2013 (M20)

- Data mining open framework
- Enhanced publications visualization and interaction

Production OpenAIREplus
v4.0: February 2014 (M26)

- Refinement based on end-user feedbacks



Any other suggestions?