THOR - Introduction

15-11-2017

Dr Adam Farquhar
Project Coordinator
Head of Digital Scholarship
British Library

Technical and Human infrastructure for Open Research

Our goal is to ensure that every researcher, at any phase of their career, or at any institution, will have seamless access to Persistent Identifiers (PIDs) for their research artefacts and their work will be uniquely attributed to them

Perhaps it used to be easier

- Figure 1. Two pages (scan) from Galilei's Sidereus Nuncius ("The Starry Messenger" or "The Herald of the Stars"), Venice, 1610.
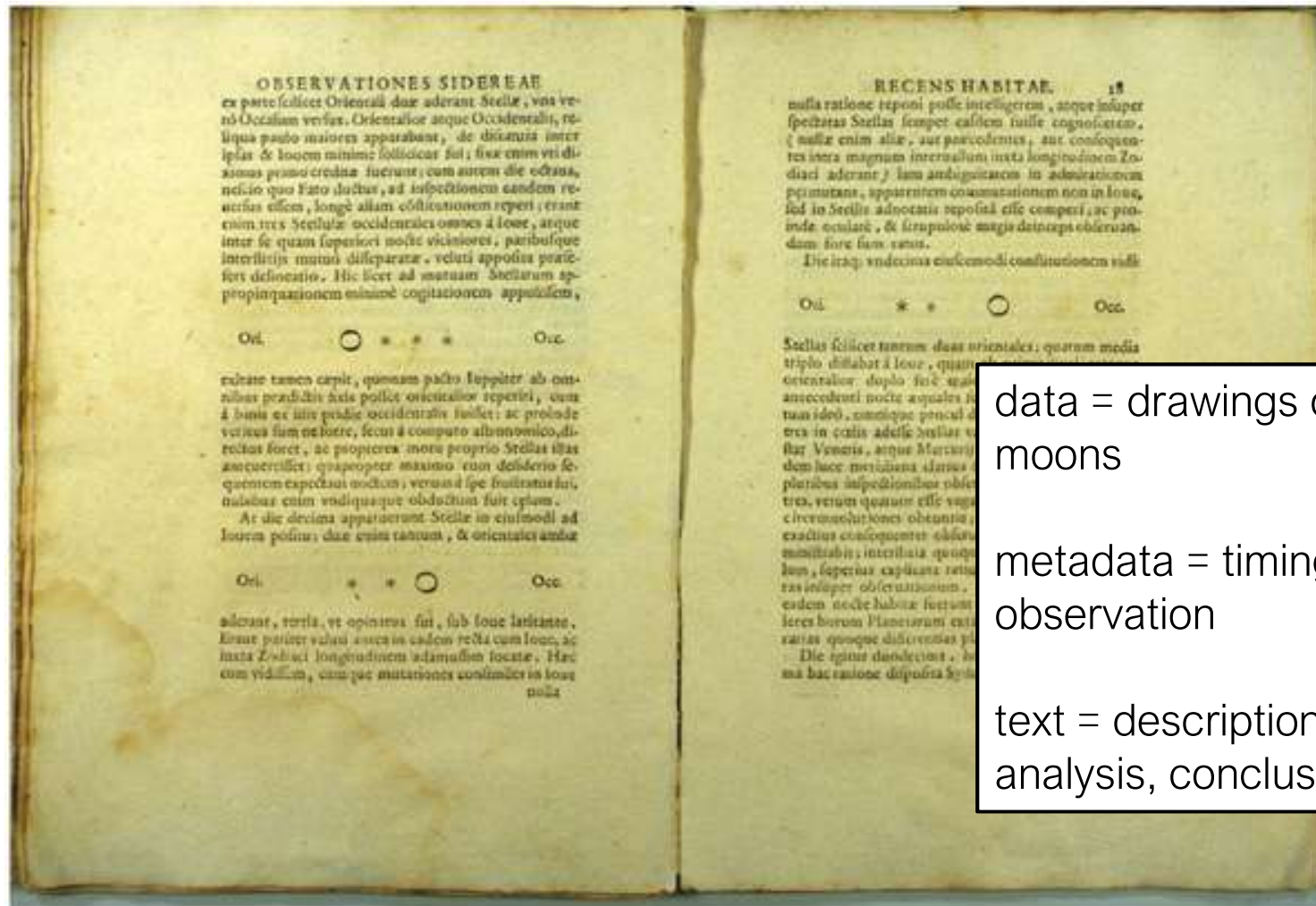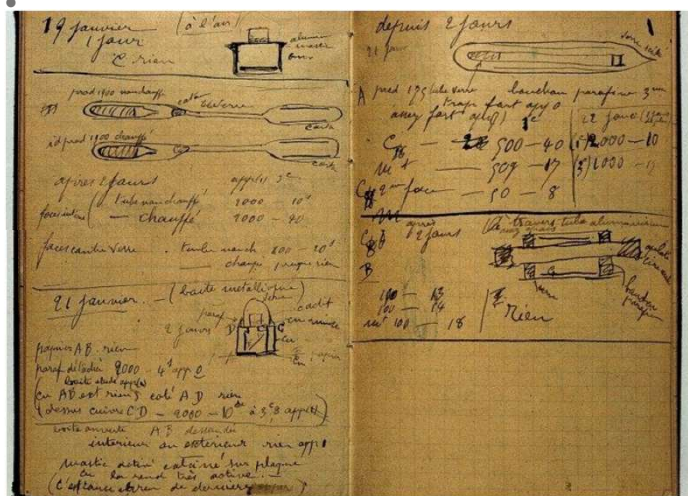

data = drawings of Jupiter and its moons

metadata = timing of each observation

text = descriptions of methods, analysis, conclusions

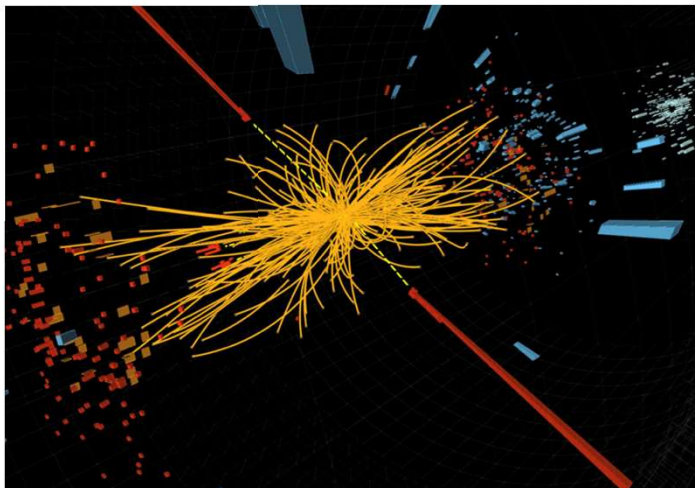PLOS | COMPUTATIONAL BIOLOGY

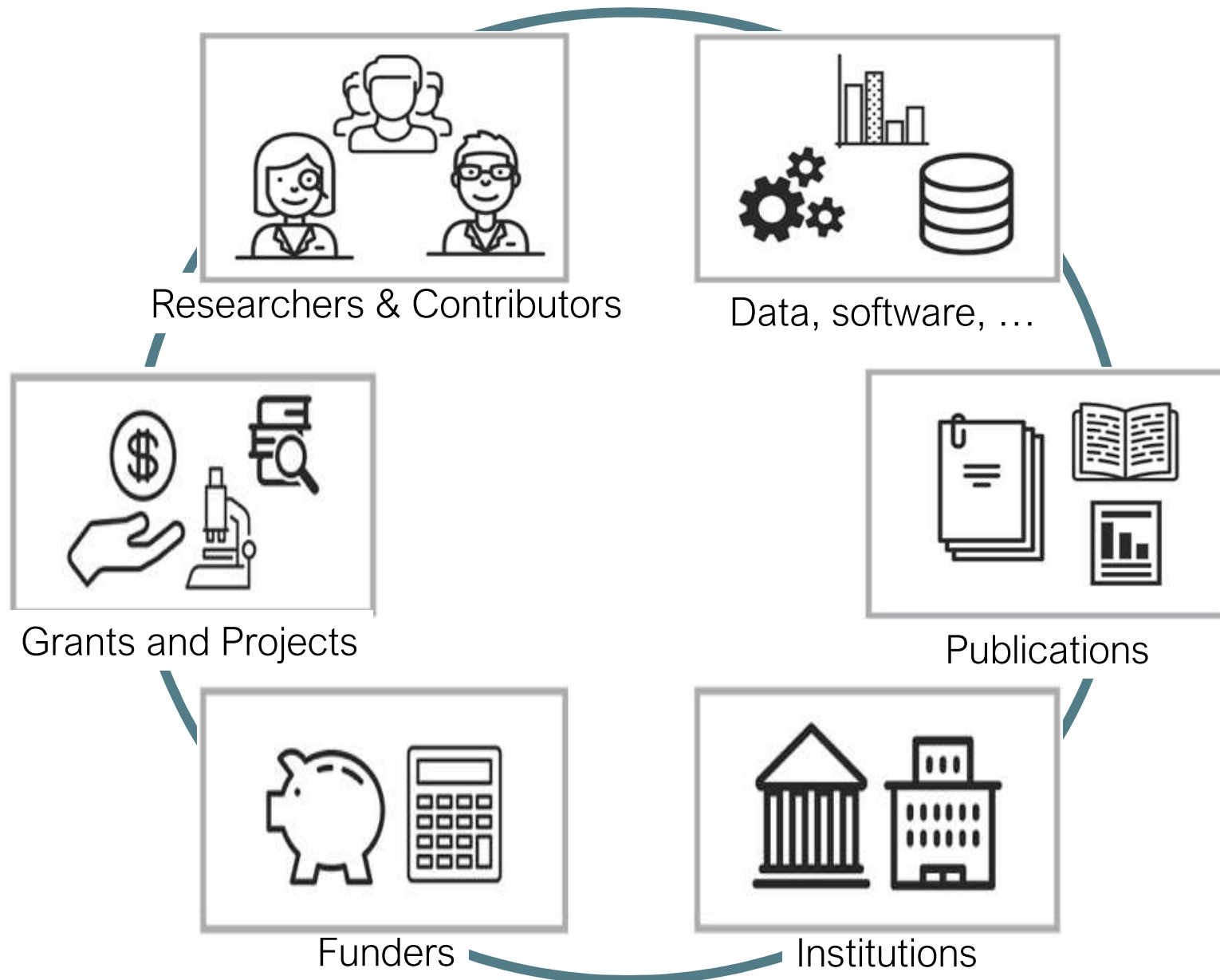# We need connected research and scholarship

Curie Notebook

Atlas Experiment

- Simpler reuse
- More reproducible research
- Faster progress
- Clear attribution
- Confident sharing
- Easier collaboration
- Help for future researchers

Connected scholarship

Researchers & Contributors

Data, software, …

Grants and Projects

Publications

Funders

Institutions

Linking articles and data

Why it matters

- Increase visibility and discovery of research data and articles

- Place research data in the right context to enable reuse

- Support credit attribution

- Challenges:

  - Data underlying findings are not always fully available

  - Data underlying findings may be available, but hidden in supplementary information

  - Data underlying the findings may are available, but not properly linked to/from article

Linking to researchers & contributors

Why it matters

- Credit and attribution

- Answers who, what, when, where

- Link one or more contributors to research output

Linking data with data

Why it matters

- provide a complete picture of the data environment

- multiple versions of the same dataset

- subsets of larger datasets or heterogeneous collections

- dynamic data

- software

- workflows

Linking data to funders and organizations

Why it matters

- The research environment is complicated (everyone wants credit)

  - Bring together scholarly output with all of the stakeholders

    - Researchers

    - Funders

    - Organizations

    - Grants

    - Projects

Key workflows

- **Researchers**
  - Discover and reuse – with direct resolution
  - Early sharing and reuse
  - Accept and give credit
  - Validate

- **Publishers and journal editors**
  - Include data citations in articles
  - Include data in peer review process
  - Interlink data and articles
  - Depend on robust accessible storage

- **Research institutions and funders**
  - Demonstrate impact through up-to-date inter-linked information
  - Manage portfolios

- **Memory institutions**
  - Research = narrative + data + software
  - Include data, plan for software
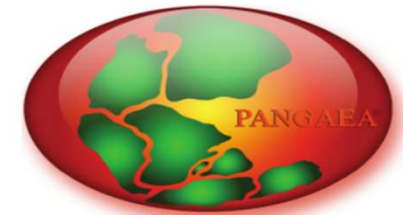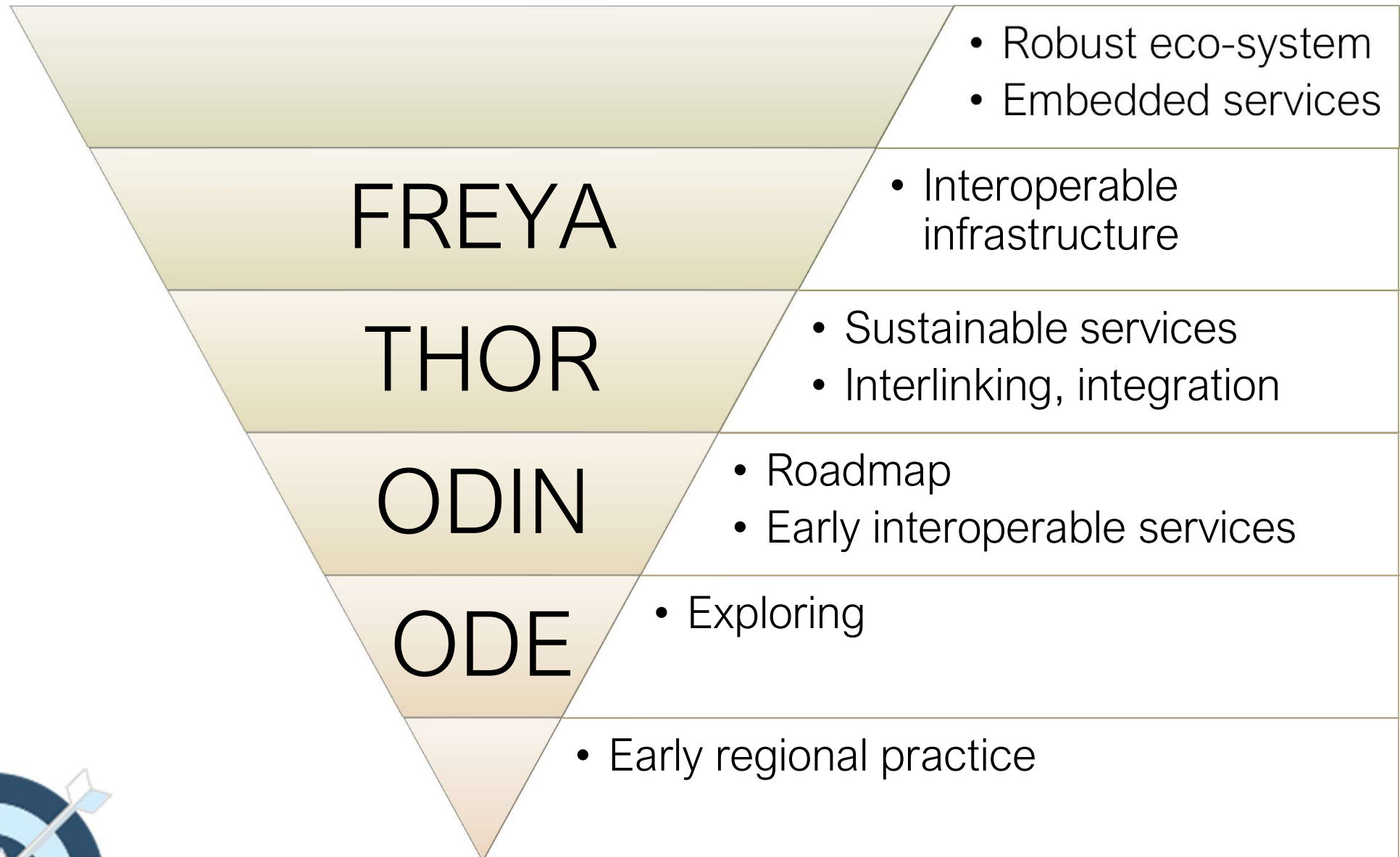 - Consider the long term

THOR

- Enable connected research by making persistent identifiers the new normal

- Integrate PIDs into services researchers already use

- Ensure PIDs are embedded in research outputs

- Provide seamless integration among articles, data, and researchers

- Focus areas

  - Biological and Medical Sciences

  - Environmental and Earth Sciences

  - Physical Sciences

  - Humanities and Social Sciences

- Funded under Horizon 2020

  - June 2015 – Nov 2017

  - 3.5m Euro

THOR partners

Historical context

FREYA
- Robust eco-system
- Embedded services
- Interoperable infrastructure

THOR
- Sustainable services
- Interlinking, integration

ODIN
- Roadmap
- Early interoperable services

ODE
- Exploring

- Early regional practice

Persistent identifiers: a key infrastructure for research

- A name not an address
- Globally unique
- Globally resolvable
- Bound to core metadata
- Interlinkable
- Interoperable
- Professionally managed
- Backed with organisational commitme
- Designed to last beyond the lifetime of any system or (most) organisation
- A digital identifier for anything

(Adapted from ODIN: http://doi.org/10.6084/M9.FIGSHARE.824314)
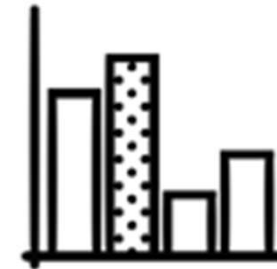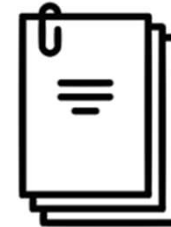
THOR contributions

- Increased the numbers of assigned PIDs

  - ORCIDs: 1.4m -> 4m

  - DataCite DOIs: 3.6m -> 9.9m

- Increased capabilities to interlink DOIs

  - Scholarly link exchange

  - DataCite / ORCID interoperability

  - New workflows to introduce PIDs early on, add retrospectively

  - Best practices for multiple records, versioning, dynamic data

- Improved robustness of services

  - Substantial re-engineering of DataCite infrastructure and service roadmap

  - Launch of DOI Fabrica service to support growth

- Increased integration into the services that researchers use

  - EBI, CERN, PANGAEA

Example: Seamless integration

Researchers: (1) use ORCID iD when submitting dataset (2) authorize DataCite to update your ORCID record.

Data centers: (1) collect ORCID identifiers during submission (2) embed iD in the work and include the iD when submitting to DataCite.

DataCite: Upon receipt of data from a data center with a valid identifier, DataCite automatically pushes information to the researcher's ORCID record.
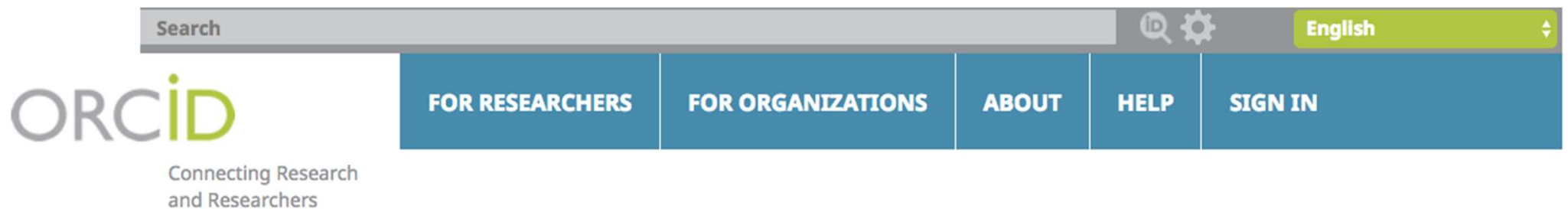
Example: Integrations at EBI
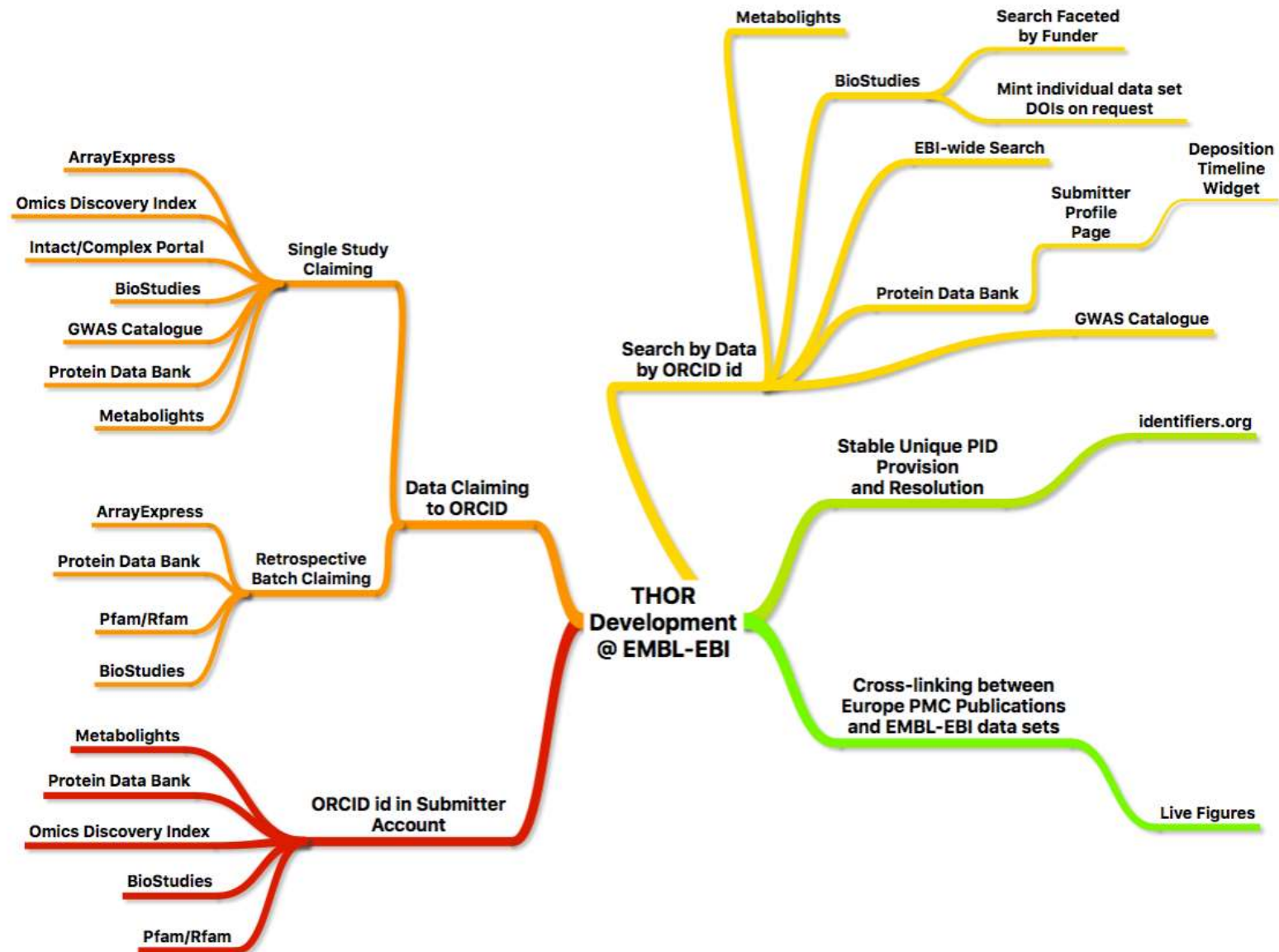
THOR contributions

- Expanded the types of objects in scope for PIDs

    - Organization identifiers (DataCite, ORCID, CrossRef) extending to funders, researcher affiliations, projects

    - Software, instrumentation, …

- Expanded discussions to humanities

    - Sometimes similar, sometime different!

    - Long history of precise reference and citation

    - Administrative districts; historical personages, events, mythical entities

- Greater focus on sustainability and growth

    - DataCite introduced major changes over last two years

A Community Project

Search | English

ORCID
Connecting Research and Researchers

FOR RESEARCHERS | FOR ORGANIZATIONS | ABOUT | HELP | SIGN IN

ORGANIZATION ID WORKING GROUP

WORKING GROUP MEMBERS

BREAKOUT GROUP: GOVERNANCE

BREAKOUT GROUP: BUSINESS MODEL & FUNDING

BREAKOUT GROUP: REGISTRY PRODUCT DEFINITION

Organization Identifier Working Group

Summary

The Organization Identifier (OrgID) Working Group was established in January 2017 to refine the structure, principles, and technology specifications for an open, independent, non-profit organization identifier registry to facilitate the disambiguation of researcher affiliations. The scope of work includes three separate but interdependent areas: Governance, Registry Product Definition, and Business Model & Funding. The goal of the Working Group is to create an implementation plan by the end of 2017.

Sunny days or clouds ahead?

- Comprehensive coverage

  - Articles, data, contributors

  - Interlinking

  - Integration into workflows, systems, services

  - Publisher and funder mandates in force

- New types of entities

  - Software, organisations, projects, instruments, materials, methods, protocols

  - Addressing the edge cases

- FREYA

  - Starts 1 December!

  - New service ecosystem around the connected graphs of persistent identifiers

  - Deep embedding into the European Open Science Cloud

THOR

PROJECT-THOR.EU

To ensure every researcher, at any phase of their career, or at any institution, will have seamless access to Persistent Identifiers (PIDs) for their research artefacts and their work will be uniquely attributed to them.

RESEARCH

Identifying challenges
Supporting standards
Designing workflows

DEVELOPMENT

Building tools
Setting up services
Connecting platforms

OUTREACH

Running bootcamps
Providing training
Aligning communities

EVALUATION

Gauging sustainability
Developing metrics
Offering feedback

THOR Knowledge Hub @ project-thor.readme.io

PIDapalooza

PROPOSE A SESSION!

Girona, Spain

23-24 January 2018

https://doi.org/10.5438/11.0002

Thank you!

Adam.Farquhar@BL.uk

THOR