PIDs and metadata: bringing rigor to the scholarly record

Matt Buys, DataCite
ExPaNDS workshop
October 22, 2021

DataCite
FIND, ACCESS, AND REUSE DATA
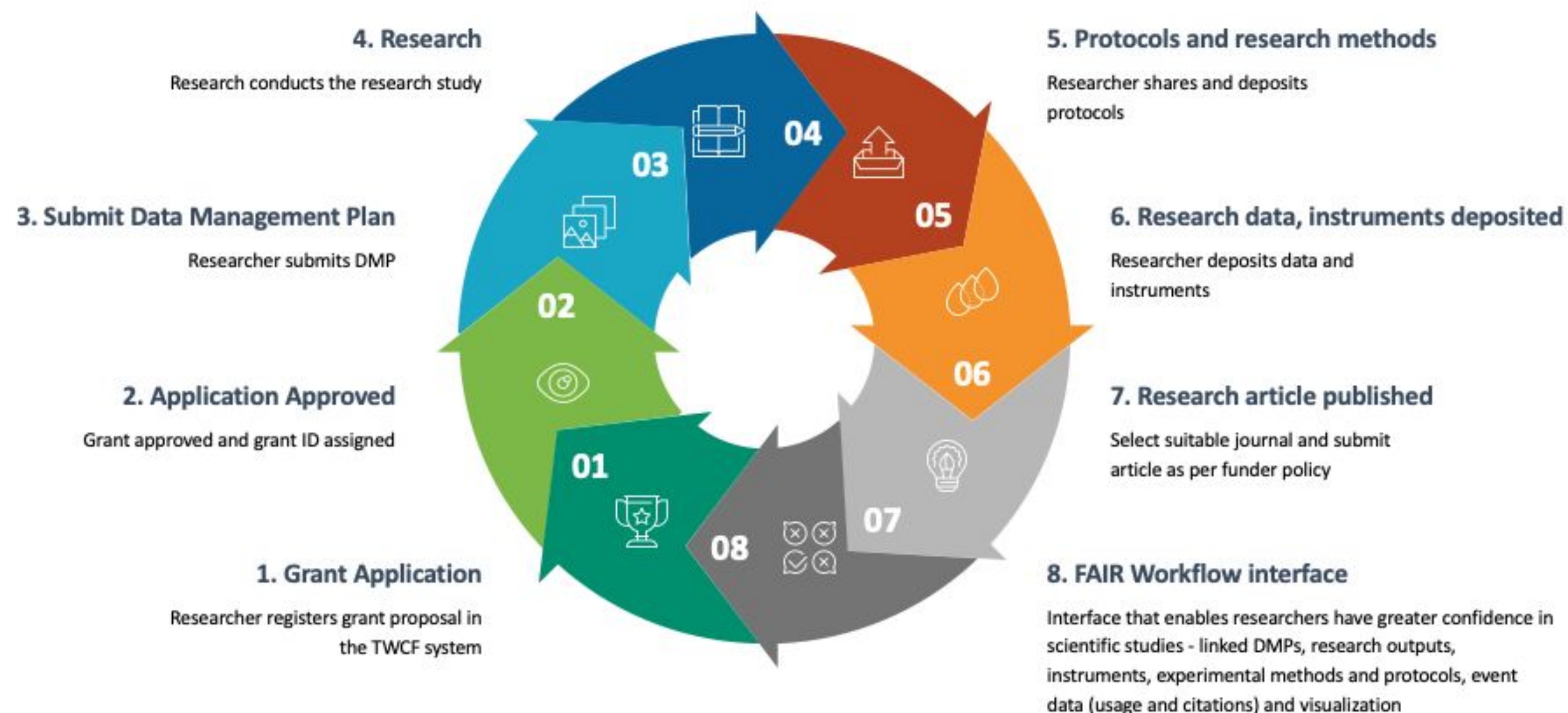
@mjbuys
@datacite

Let's start with some context

DataCite

At DataCite our vision is **"*Connecting Research, Identifying Knowledge*"**

We know that articles provide a succinct description of the methods and work carried out as well as the conclusions drawn based on that work. Most of the time, there is no underlying information available and no mechanism to easily link to the experimental design, the research data, and the analytical tools that were used to generate the reported outcomes. **This challenge prevents the research community from being able to fully understand the results of the research, to replicate its results, and to decisively evaluate,and reuse existing research.**

The research lifecycle

We need to focus on the entire research lifecycle, bringing together disparate pieces of the research study together: using PIDs and metadata

4. Research

Research conducts the research study

3. Submit Data Management Plan

Researcher submits DMP

2. Application Approved

Grant approved and grant ID assigned

1. Grant Application

Researcher registers grant proposal in the TWCF system

5. Protocols and research methods

Researcher shares and deposits protocols

6. Research data, instruments deposited

Researcher deposits data and instruments

7. Research article published

Select suitable journal and submit article as per funder policy

8. FAIR Workflow interface

Interface that enables researchers have greater confidence in scientific studies - linked DMPs, research outputs, instruments, experimental methods and protocols, event data (usage and citations) and visualization

Technology + Engagement

As a consequence, there is currently a big push to make science Open and FAIR to increase reproducibility and reusability of scientific results. Recognizing the importance of better management of research entities has led to critical advances concerning development of infrastructure for preregistration of studies, data repository platforms, standards for data sharing and ontologies.

Our approach at DataCite is to bring **technology and engagement** in order to provide sustainable PID infrastructure services globally, this in turn helps bring rigor to the scholarly record.

Technology

We can talk about building things, APIs and the magic that makes it happen.

Our focus and efforts across the domains, disciplines and borders vary, although we see a common goal in creating (/enhancing) technology to make Open Science easy, possible and normative.
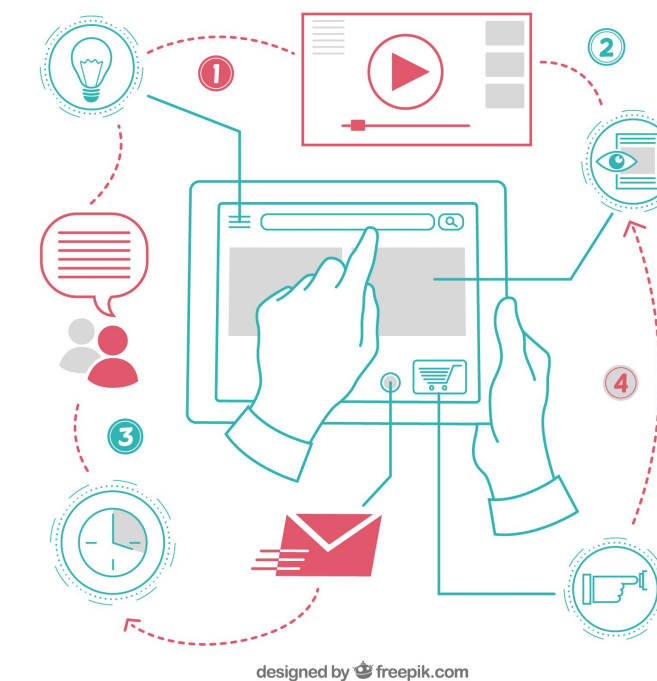
Engagement

When we talk about engagement, we need to focus on people, communities and how we rally stakeholders to effect change.
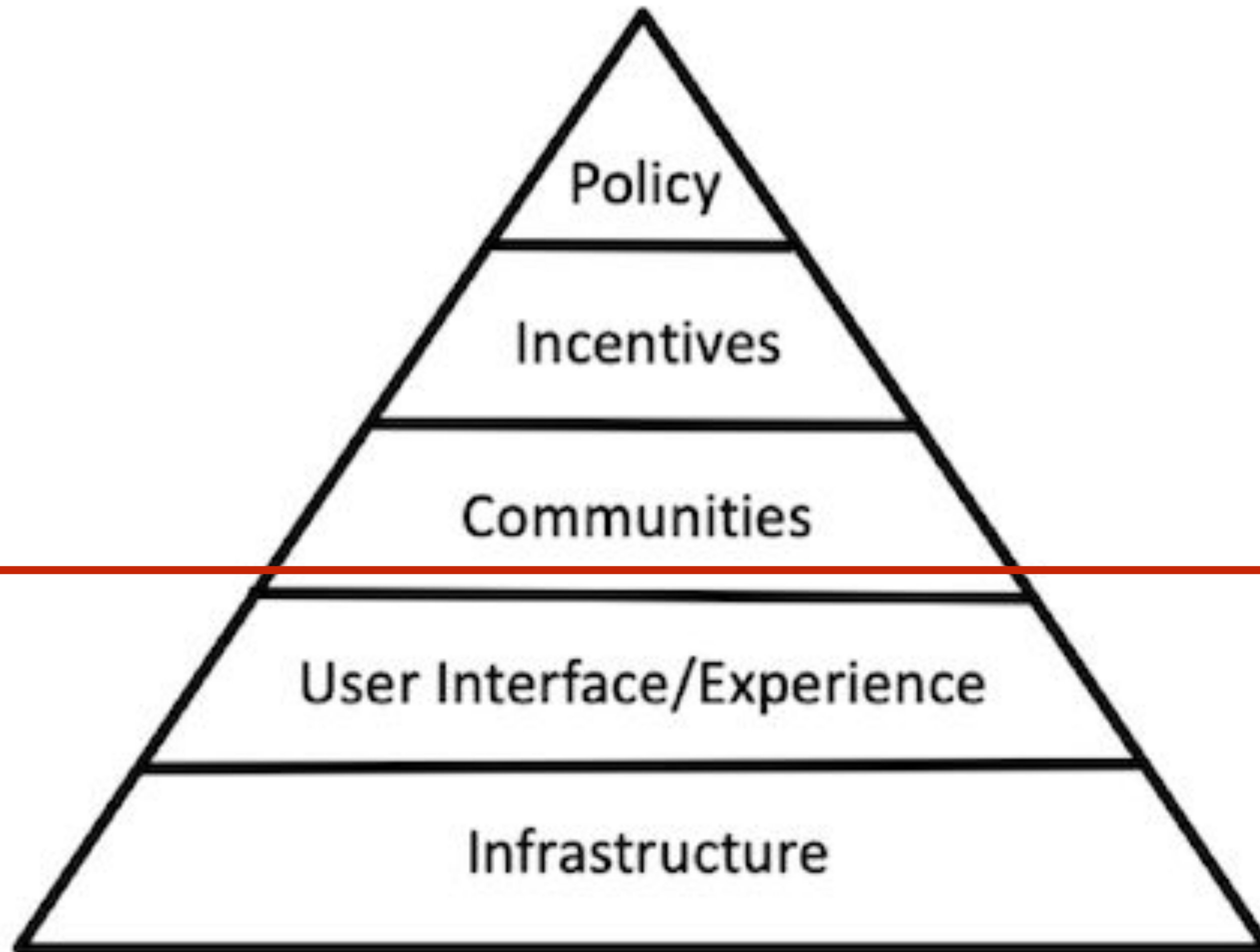
Source: https://www.cos.io/blog/strategy-for-culture-change

DataCite
Create, Discover, Integrate

Our community
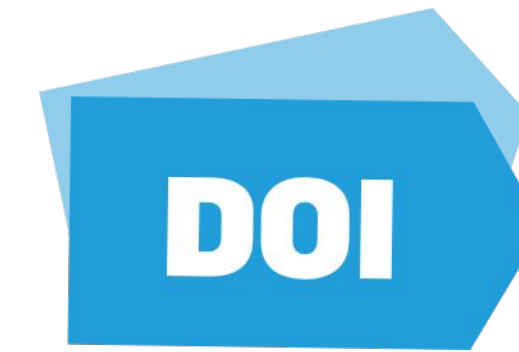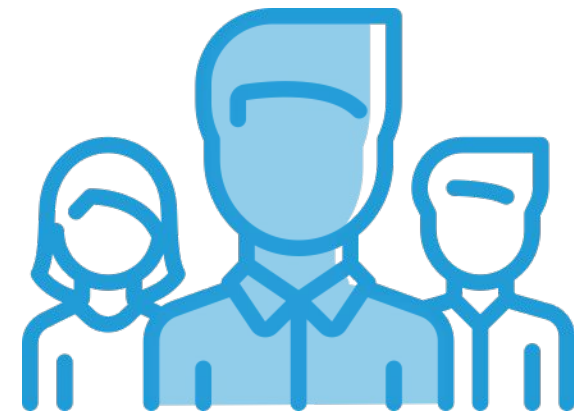
DataCite

2370+	250+	48	28m+	9990+
Repositories	Members	Countries	DOIs	Organizations

DataCite PIDs

DataCite's mission is to be a world leading provider of persistent identifier services to help make research outputs and resources findable, citable, connected and reused globally. We seek to create value for our members through community-driven, innovative, open, integrated, usable, and sustainable services for research.

As such, we have over the years supported various identifier communities in scaling their efforts to a global community.

This includes PIDs across **28 resource types** (supporting datasets, samples, preprints, software, DMPs, events etc etc.)

But as we know it is not only about the PID...

the services and metadata are crucial

DataCite Services

- We group our services into three distinct categories; Create, Discover & Integrate.

- **Our team focus on working with the community to define our roadmap priorities.**
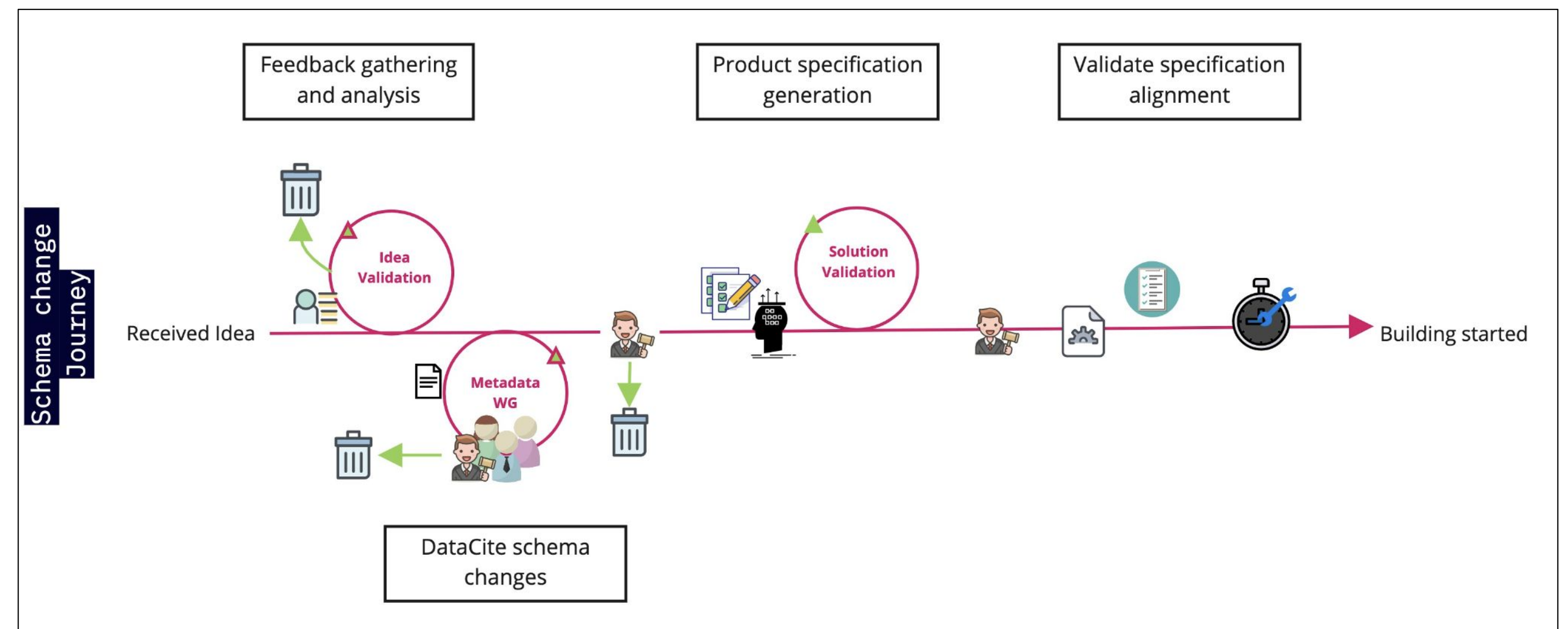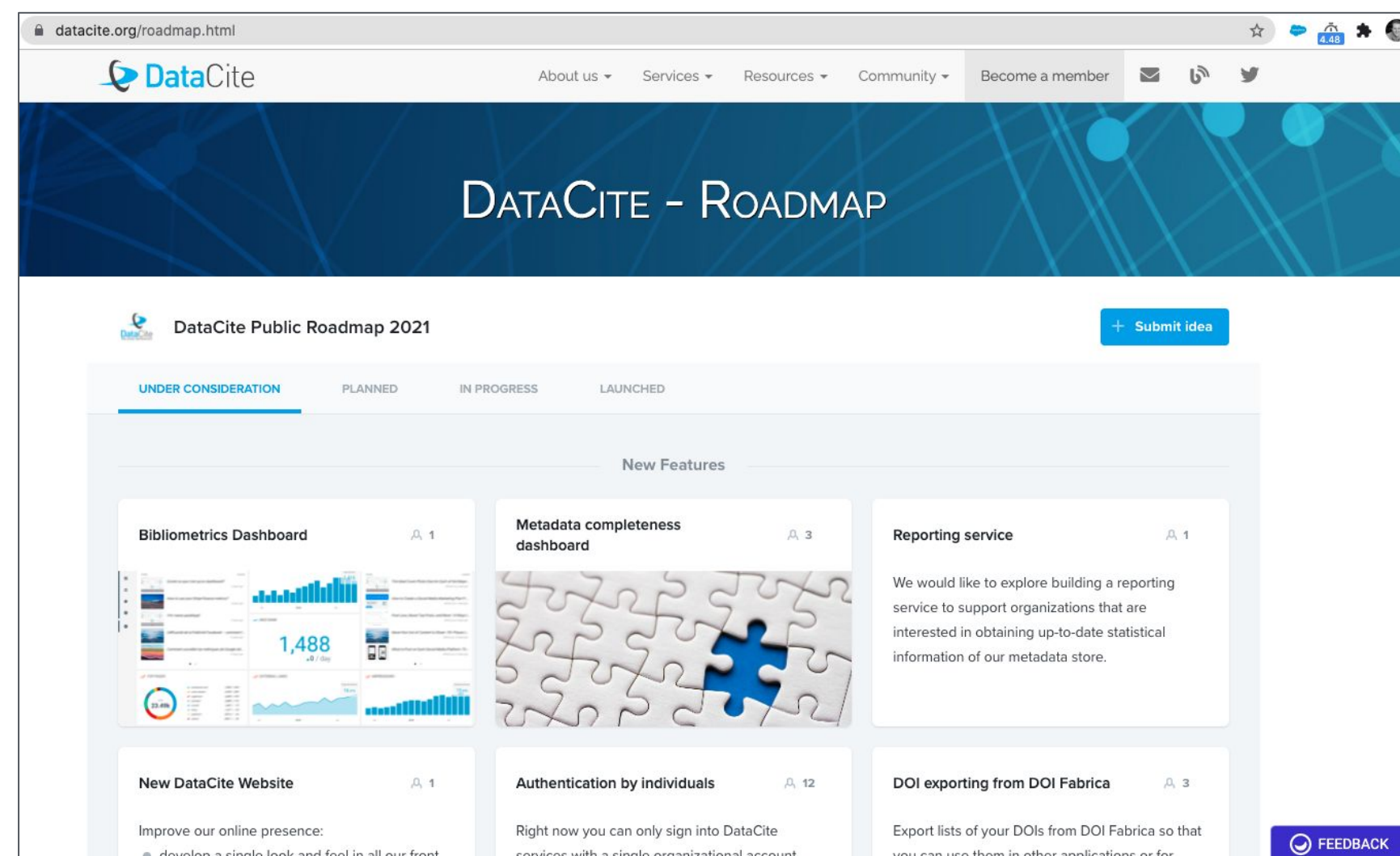
Create
and Manage DOIs

DataCite membership allows you to create and manage DOIs for all of your repositories. You can do this through:

· Our primary REST API that supports JSON and enables automated DOI registration

· Our manual interface that enables you to register DOIs in less than a minute.

· Registered DataCite Service providers that provide a platform where you can register DOIs.

Discover and Connect Research

Find Research
with Data Cite Commons

PID Graph
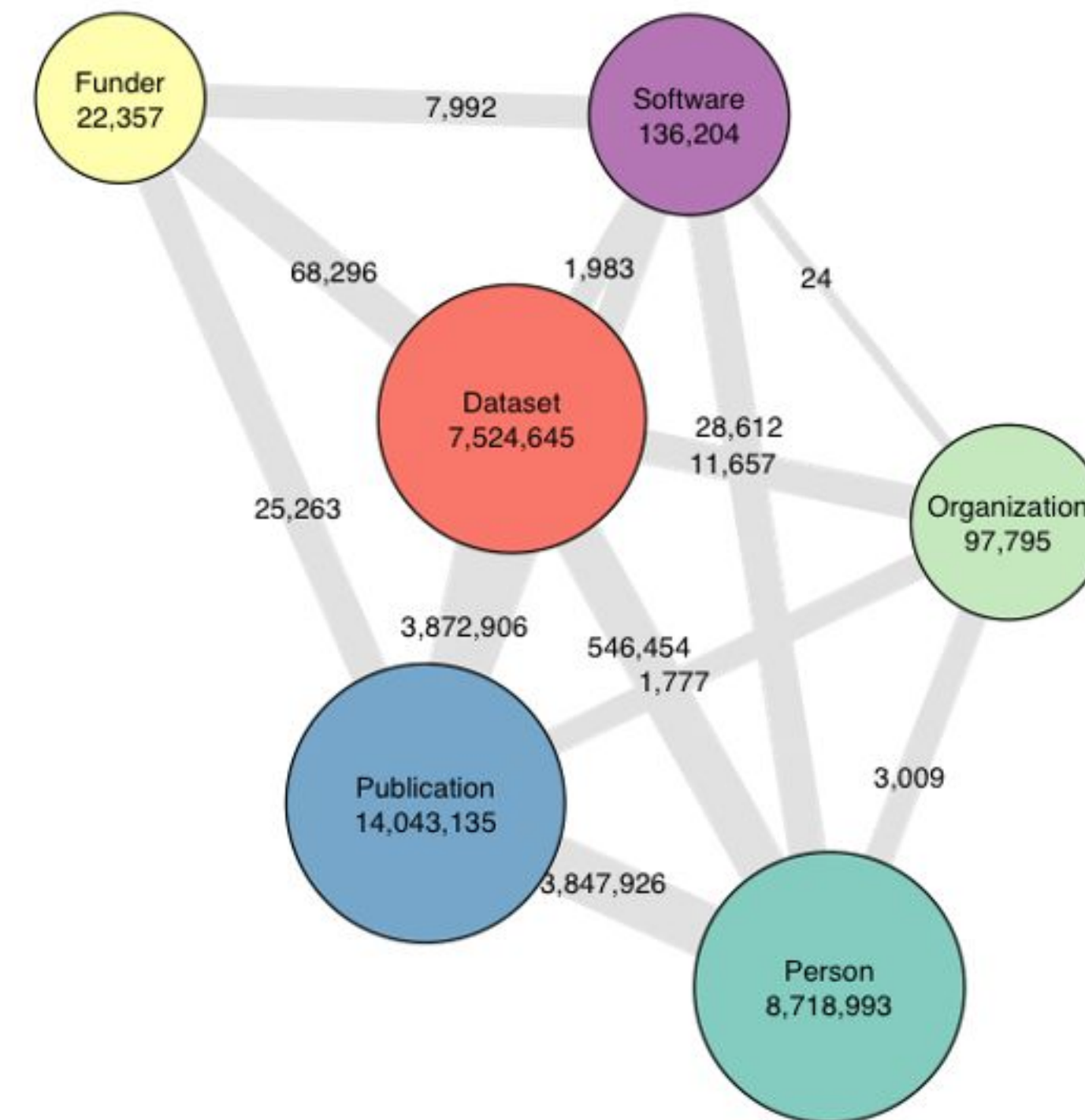
Integrate with our APIs

DataCite · 📖 Feedback · Home

🏠 Home · 📖 Guides · `<>` API Reference

REST API

REST API
- activities >
- client-prefixes >
- clients >
- dois >
- events >
- heartbeat >
- prefixes >
- provider-prefixes >
- providers >
- reports >

MDS API

MDS API
- /dois >
- /metadata >
- /media >

EZ API

EZ API
- /id >
- /shoulder >

GRAPHQL API

Getting Started With Your API

REST API

📝 SUGGEST EDITS

Version History

- v.1: June 25, 2016, first draft.
- v.1.1: October 10, 2016, follow JSONAPI spec for side-loading associations
- v.1.2: December 9, 2017, follow JSONAPI spec for pagination
- v.2.0: January 7, 2019, new API endpoints, using camelCase for attributes

Overview

The DataCite REST API returns information about DataCite content. The API is RESTFUL and returns results in JSON. It follows the JSONAPI specification. The mime-type for API results is `application/vnd.api+json` .

More information about the DataCite REST API can be found in the REST API Guide.

Metadata

- The key to making data citable, searchable and accessible is equipping outputs and resources with metadata – descriptions of and facts and figures about the data – that meets basic standards and adheres to uniform, consistent schema.

- The DataCite metadata schema has evolved over the years and our efforts have been led by the Metadata Working Group.

- Our current version 4.4 included a number of new resource types to further support our PID services throughout the research lifecycle.

- We work with domain specific experts to map schemas to the DataCite schema to enhance discoverability.

- The schema includes six (6) mandatory properties

Recommended properties (1)

DataCite

- **RelatedIdentifier** - This one may go without saying. RelatedIdentifiers are always important for linking resources together. This is the information that DataCite uses for our citation counts and to feed the PID Graph. Having this RelatedIdentifier information is also helpful for following a trail of research from dataset to article to author to institution and so on.

- **Rights** - The Rights field is where you can include information about the data's content license (like CC0). When this information is in the metadata for the dataset, it helps researchers to quickly see at a glance whether the data being described is something they'll be able to use. This information is also important for harvesters who are trying to assemble lists of readily available datasets.

Recommended properties (2)

DataCite

- **Subject** - The Subject field is where you can include information about subjects that are relevant for the item the DOI points to. Under the current implementation of the DataCite Metadata Schema, this field is free text, so you can add any keywords that might be relevant to the item, whether or not they're part of a controlled vocabulary. If you would like to follow a scheme of vocabulary, you're not restricted in the subject scheme you can specify. It can be helpful to comply with community-specific standards, such as using MeSH subject headings for medical topics or using DFG subjects for German-funded research, for example.

- **Description** - The Description field may at first seem not particularly critical, but abstracts and other descriptive information is often mined for emerging trends. In a case like the current epidemic, where "COVID-19" might not be a typical controlled vocabulary term, it may be possible to surface articles on this topic by mining abstracts.

Related identifiers

Orphan, Victoria
Rouse, Gregory
Levin, Lisa
Roman, Christopher
Cordes, Erik
Ashford, Oliver
People

University Of California, San Diego (Ucsd.Edu)
Organisations

Division of Ocean Sciences (nsf.gov)
Division of Ocean Sciences (nsf.gov)
Division of Ocean Sciences (nsf.gov)
Division of Ocean Sciences (nsf.gov)
Funders

Sampling information from community megafauna collected from different seep site
Mussels and associated organisms sampling information collected in th
Matrix of taxon (columns) by sample (rows) for sediment pus
Averages and standard deviation across species for
Matrix of taxon (columns) by sample (rows) fo
Datasets
Locations of sediment push cores collected
Locations of hard substrata collected durin

Publications
Fluid seepage at the continental margin offshore Co
Fluid seepage at the continental margin offshore Costa Rica
A hydrothermal seep on the Costa Rica margin: middle ground in a cor

Usage and citations

DataCite Commons

fundingReferences.awardNumber:777523

Pages ▾   Support   Sign In

📍 This Page   📖 Works   👤 People   🏛 Organizations

https://ror.org/00k4n6c32

European Commission
EC

Founded 1958

Links
Homepage
Wikipedia
Twitter

Other Identifiers
GRID grid.270680.b
Crossref Funder ID 10.13039/501100000780
Crossref Funder ID 10.13039/501100000893
Crossref Funder ID 10.13039/501100000891
Crossref Funder ID 10.13039/501100000894
Crossref Funder ID 10.13039/501100000887
Wikidata Q8880
Wikidata Q20855594

Geolocation
50° 50' 37" N, 4° 22' 58" W

Belgium   Government

🔗 https://ror.org/00k4n6c32

Share

✉ Email
🐦 Twitter
📘 Facebook

Aggregated Citations, Views and Downloads

| 65,918 Citations | 30,215 Views | 169 Downloads |

167 Works

Publication Year   Work Type   License

167   167
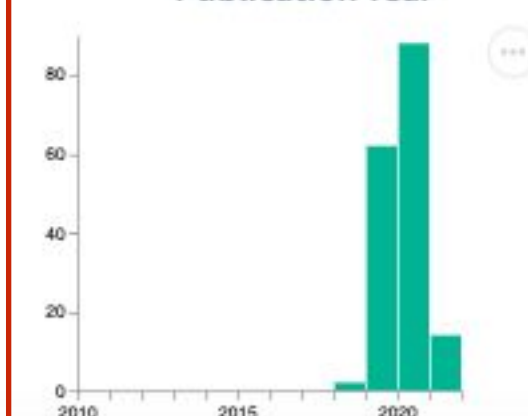
Publication Year

☐ 2021   14
☐ 2020   88
☐ 2019   62
☐ 2018   2
☐ 2005   1

Work Type

Open Science
The reason many of us get out of bed each morning

Open Science

Open Science is a collective movement, it has the promise to make science more efficient, reliable, and responsive to societal challenges.

Focussing on bringing together disparate parts of the research lifecycle, to support reproducibility and bring rigor to the scholarly record.

Our practices include activities around open access, availability and reproducibility of research results, and making it easier to communicate knowledge and advances in science.

Interpretation and application of Open Science varies across disciplines, domains and borders.

THIS IS NOT A REVOLUTION

REVOLUT
MAKE T

Open Science Fundamentals

Fostering innovation and supporting knowledge transfer

In a nutshell, Open Science (has the potential to) reduce delays in the re-use of the research outputs, bring rigor through reproducibility and increase our path to innovation through open (and broad) dissemination.
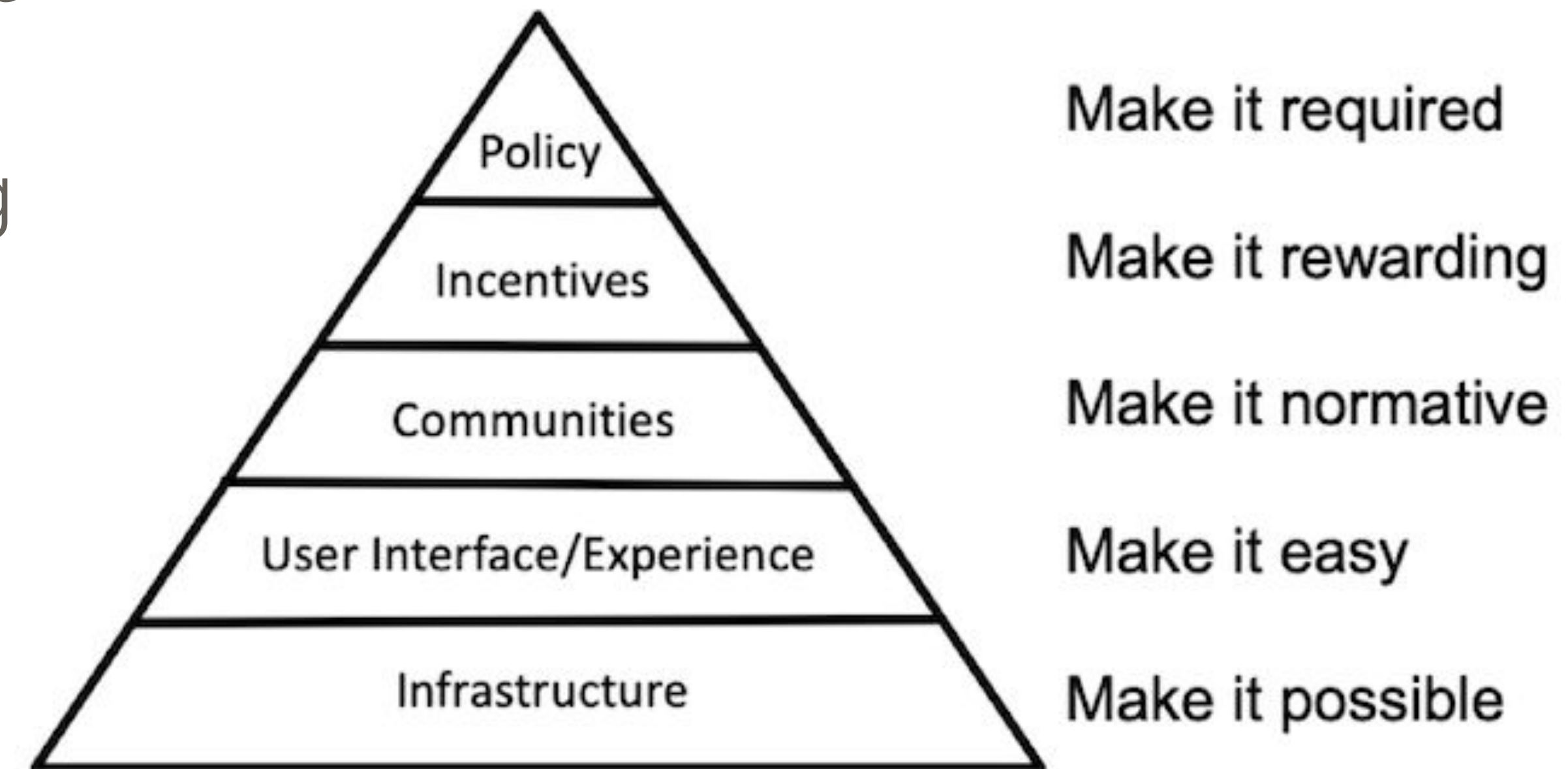
Convergence
What does this look like...

Technology + Engagement

Convergence is when we start to see Open Science becoming normative, the strong indicator for when we start to reach this point is when we start talking about Science again.

Through
…enabling + building technology
…working with people + communities
we will start to effect collective change

DataCite

Policy — Make it required

Incentives — Make it rewarding

Communities — Make it normative

User Interface/Experience — Make it easy

Infrastructure — Make it possible

Technology + Engagement = Rigor

DataCite

- **Efficiency:** better access to the disparate pieces of the research lifecycle
- **Quality and integrity:** through reproducibility and accessibility, we improve the integrity of research
- **Economic:** supporting and driving innovation with broad dissemination of knowledge
- **Innovation and knowledge transfer:** enhancing re-use and promoting further innovation
- **Recognition and attribution:** providing attribution to all involved in the research lifecycle
- **Societal:** bring innovations to the daily lives of people all around the world

Share how you are bringing #rigor to the scholarly record and how ExPaNDS makes Open Science #easy...

Let the community know how they can #collaborate with you.

@OpenScienceFAIR
@mjbuys

Get in touch!

Email us:
matt.buys@datacite.org
info@datacite.org

Follow us:
@mjbuys
@datacite

Talk to us:
pidforum.org

Read about us:
datacite.org

Get support:
support.datacite.org
support@datacite.org

Image credits

This presentation was compiled using images from the following sources:
- Freepik.com
- Wikipedia.com