

# Project THOR

Persistent identifiers, everywhere

Tom Demeranville - ORCID EU

# Introduction

Quick THOR intro

About persistent identifiers

The THOR partners

Some examples in action

More on THOR

Sum-up



# Project THOR

Aims to embed persistent identifiers (PIDs) at all stages of the research lifecycle, improve links between them and promote uptake of PID infrastructure.

(which would be a really good thing)

EC funded, international, cross disciplinary, cross sector effort



# What are persistent identifiers?

Persistent Identifiers (PIDs) uniquely identify entities within the research ecosystem and help define relations between contributors, research artefacts, and organizations.

# Why do we need them?

Names are not enough to identify grants, people, organisations and research outputs, even with affiliations or other supplementary data. Names are not machine actionable.



# Why are persistent identifiers important?

When multiple systems wish to operate on or exchange information about a 'thing' they need to be sure that they're talking about the same 'thing'

PIDs enable entities to be linked across infrastructures, disciplines, borders and other boundaries.

PIDs are used to aid discovery, disambiguation, metadata resolution, attribution, metrics gathering and more. They help build a clearer picture of how research is generated



# What makes a good persistent identifier?

The ODIN project defined a 'good' persistent identifier as:

- Unique on a global scale
- Resolvable as URIs by machines (as metadata) and humans (as web pages)
- Are searchable and come with metadata that describe their most relevant properties
- Interoperable with other identifiers through metadata describing relationships
- Issued and managed by an organization that focuses on that goal as its primary mission, has a sustainable business model and a critical mass of member organizations that have agreed to common procedures and policies, has a governing body, and is committed to using open technologies



# How is THOR helping?

Research - Identifying challenges, supporting standards, designing workflows

Development - Building tools, setting up services, connecting platforms

Outreach - Running bootcamps, providing training, aligning communities

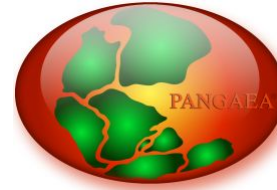
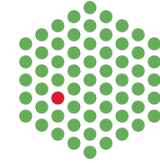
Evaluation - gauging sustainability, developing metrics, offering feedback



# THOR Partners

ORCID

EMBL-EBI





# About ORCID

- ORCID provides identifiers to people involved in generating research outputs such as data, articles and books. (There are over 2.2 million ORCID iDs!)
- ORCID connects people with their articles, data, software, education, funding, employment, peer reviews and more using persistent identifiers
- ORCID is user centric - records are controlled by their owners
- All public data is available, for free, through the website, API and data dump
- There is a member api for updating records, receiving notifications and other value added services
- Funders and publishers are increasingly requiring ORCID iDs on submission - national funders in the UK, Sweden and Austria all require them.



# DOIs, Datacite and Crossref

DOIs are used to identify scholarly outputs. DOIs are issued for the majority of journal articles and increasingly for other outputs such as data, software, music and more

A DOI can be resolved to a URL or metadata, for free, by anyone.

Crossref and Datacite are registration agencies that supply DOIs to the community and provide a variety of services, such as search, to the public and their members.

There are several services that enable you to allocate a DOI to an entity for free.

As of February 2016, DataCite had minted over 7 million identifiers!



# Examples in action - ORCID autoupdate

Enables ORCID iDs to seamlessly propagate from publishers to ORCID and beyond

When the publisher creates a DOI and supplies an ORCID to Datacite/Crossref, then a notification can be sent to the user asking them for permission to update their ORCID record

Being taken up beyond the DOI registrars - publishers, datacentres, repositories

Once in ORCID, notifications propagate to other systems, such as institutional CRIS and funders



# Examples in action - Data publishing workflows

Capturing ORCID iDs early is the key to providing value later on and it's not limited to journal articles

EBI have been working with ORCID to capture identifiers when ingesting datasets in the life sciences. They are looking at ways of incorporating their many layers of life sciences identifiers into the wider PID ecosystem

CERN are doing the same with high energy physics, where ORCID iDs and DOIs are quickly becoming the norm rather than the exception

Pangaea are doing similar work with data publishing workflows in the earth sciences



# More on THOR

Working with many other research infrastructure projects, including EuroCRIS, AARC and more to ensure that PIDs underpin efforts in other areas

Raising awareness, producing documentation and running workshops (see you at open repositories?)

We are hosting a one day event in Amsterdam on the 7th of July. Come along if you'd like more detail on data publishing workflows and what we've been up to



# The take home slide

Researchers are using PIDs for their works and their activities across the whole research landscape, *even if they don't know about it.*

Institutions should be using persistent identifiers as their point of connection to the author's information - *especially information created outside of the institution.*

Actively working with ORCID iDs for people and DOIs for data and publications is the first step towards realising the potential that these PID infrastructures offer.

See the blog for more info: <https://project-thor.eu/>

