PID Graph and ROR

RDC Webinar

27 February 2020

Robin Dasler, DataCite

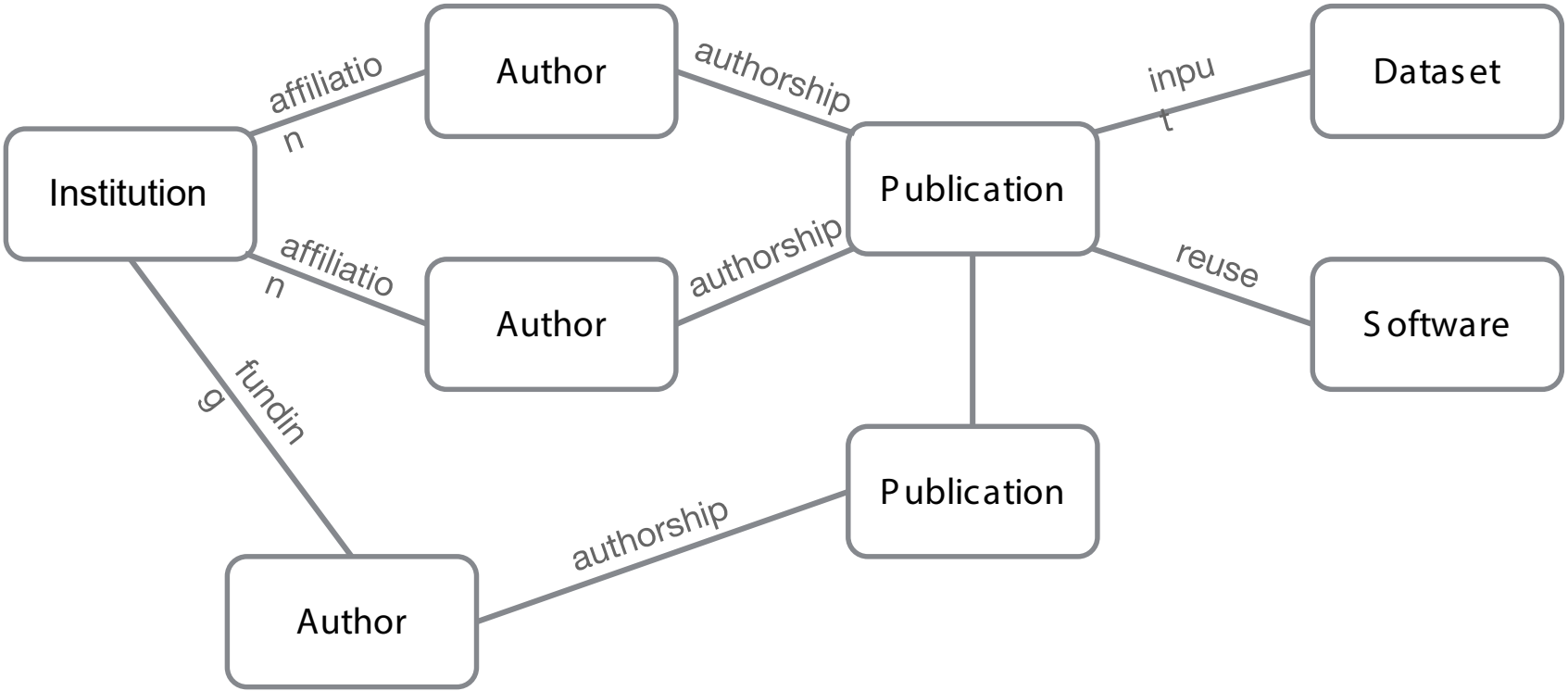
What is the PID Graph?

Research is already a graph

Researchers, institutions, publications, datasets, and more are interconnected.

Entities and the relationships between them form a conceptual graph of the connected research landscape.

It could look like this



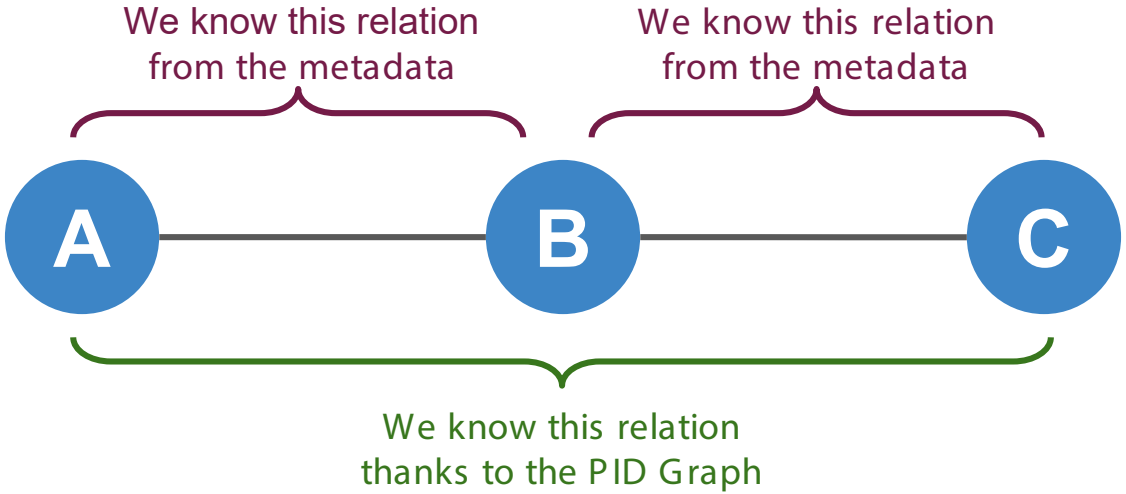
PIDs are the backbone of connected research

Having unique persistent identifiers for researchers and their outputs is crucial to connecting pieces of the research landscape together.

PIDs already have the potential to enable the connected research graph, but we’re not yet taking full advantage of their connecting powers.

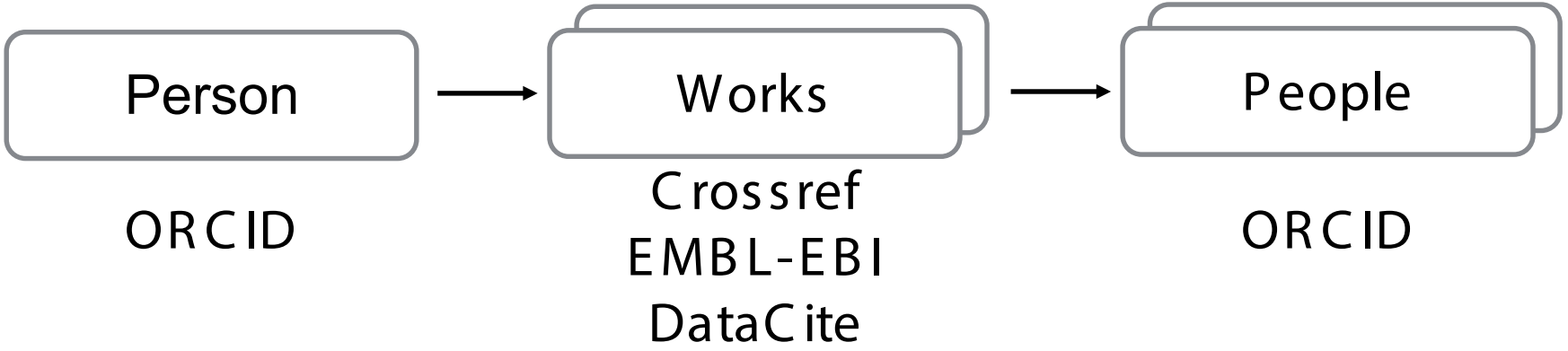
Enter the PID Graph

We can link PIDs together via relations in their metadata to enable the discovery of connections at least two “hops” away.

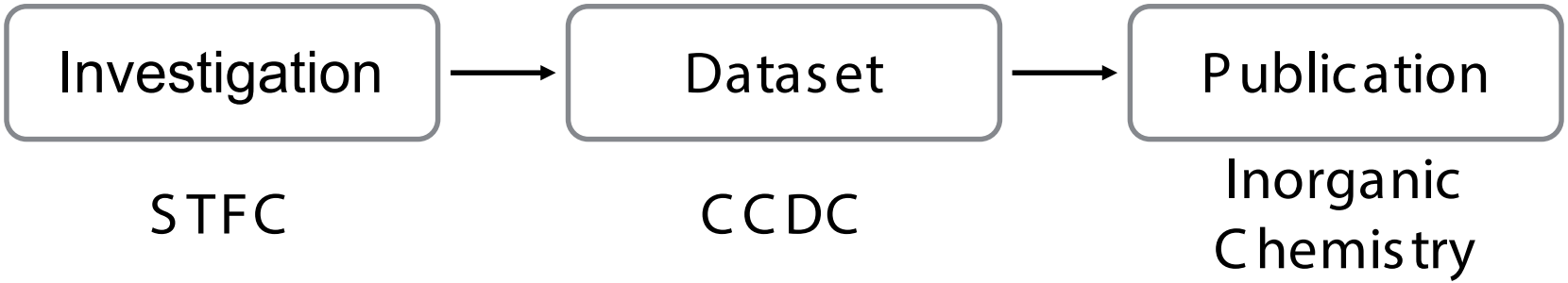


What questions could the PID Graph answer?

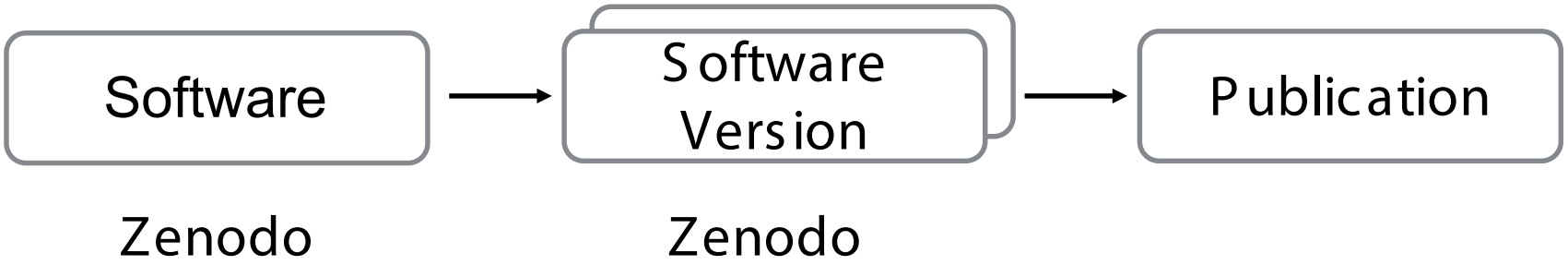
Who are all the co-authors of a particular researcher?



Which publications used STFC investigations for the underlying data?



Which publications cite any version of a piece of software?



Enabling the PID Graph at DataCite

DataCite GraphQL API

DataCite has developed a GraphQL API. This is the most convenient way to consume the PID Graph with DataCite metadata as a starting point.

GraphQL is a query language that's specially built for graphs. It lets you specify exactly which information you want and in what structure you'd like to receive it.

The DataCite GraphQL API is currently in pre-release, with a stable release planned for around RDA in March.

GraphQL Endpoint https://api.datacite.org/graphql Method POST Edit HTTP Headers

GraphiQL Prettify History Docs

```
{
  publications(query: "creators.name:dasler") {
    id
    titles {
      title
    }
    descriptions {
      description
    }
    creators {
      id
      name
      familyName
    }
    fundingReferences {
      funderIdentifier
      funderName
      awardTitle
      awardNumber
    }
  }
}
```

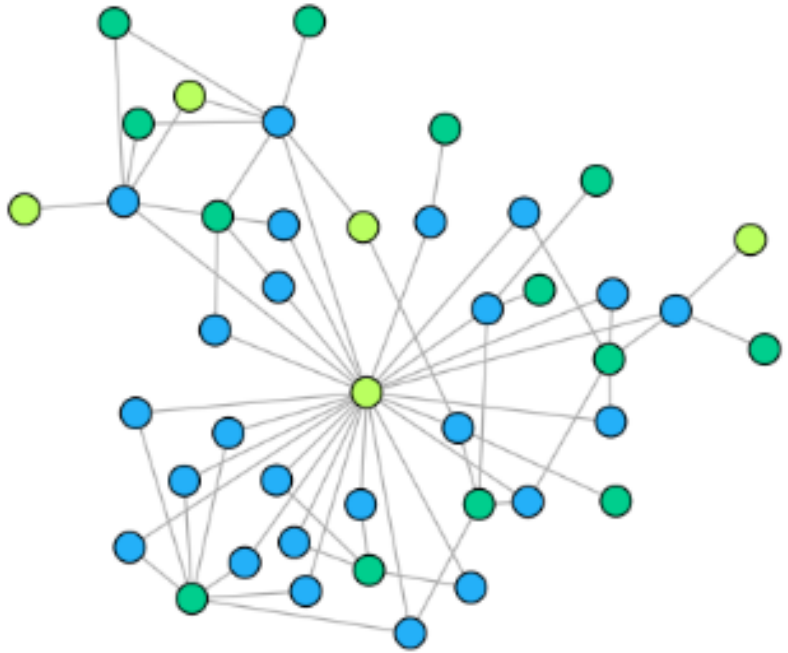

```
{
  "data": {
    "publications": [
      {
        "id": "https://doi.org/10.5281/zenodo.1064000",
        "titles": [
          {
            "title": "Pid Service Adoption"
          }
        ],
        "descriptions": [
          {
            "description": "This presentation describes how the uptake of persistent identifiers can be measured and gives an overview of the main results of the ORCID adoption study."
          }
        ],
        "creators": [
          {
            "id": null,
            "name": "Dasler, Robin",
            "familyName": "Dasler"
          }
```

Surfacing connections

Graph of all the publications associated with DataCite, plus all the researchers and organizations associated with those publications.

Answers the question :

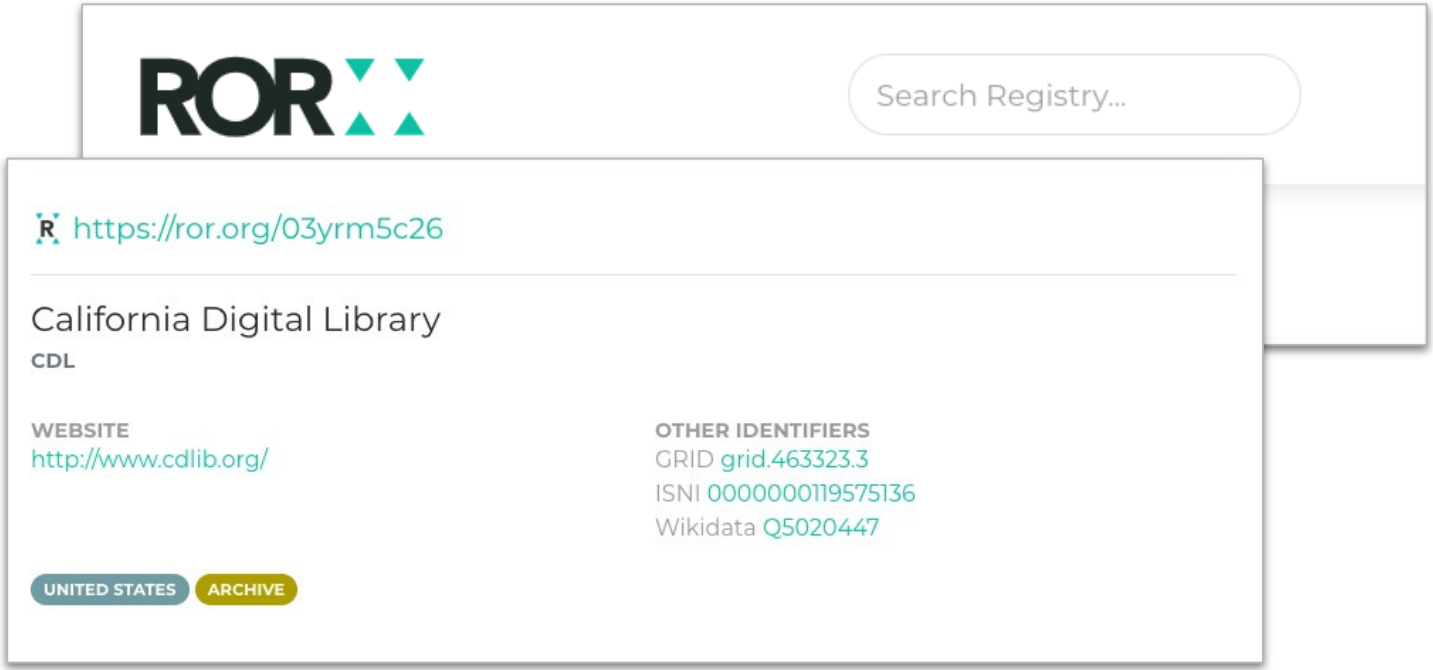
With whom has DataCite collaborated on their publications?



Expanding the PID Graph with ROR

The Research Organization Registry
is a **community-led project**
to develop an **open**, sustainable,
usable, and unique identifier
for every **research organization**
in the world.

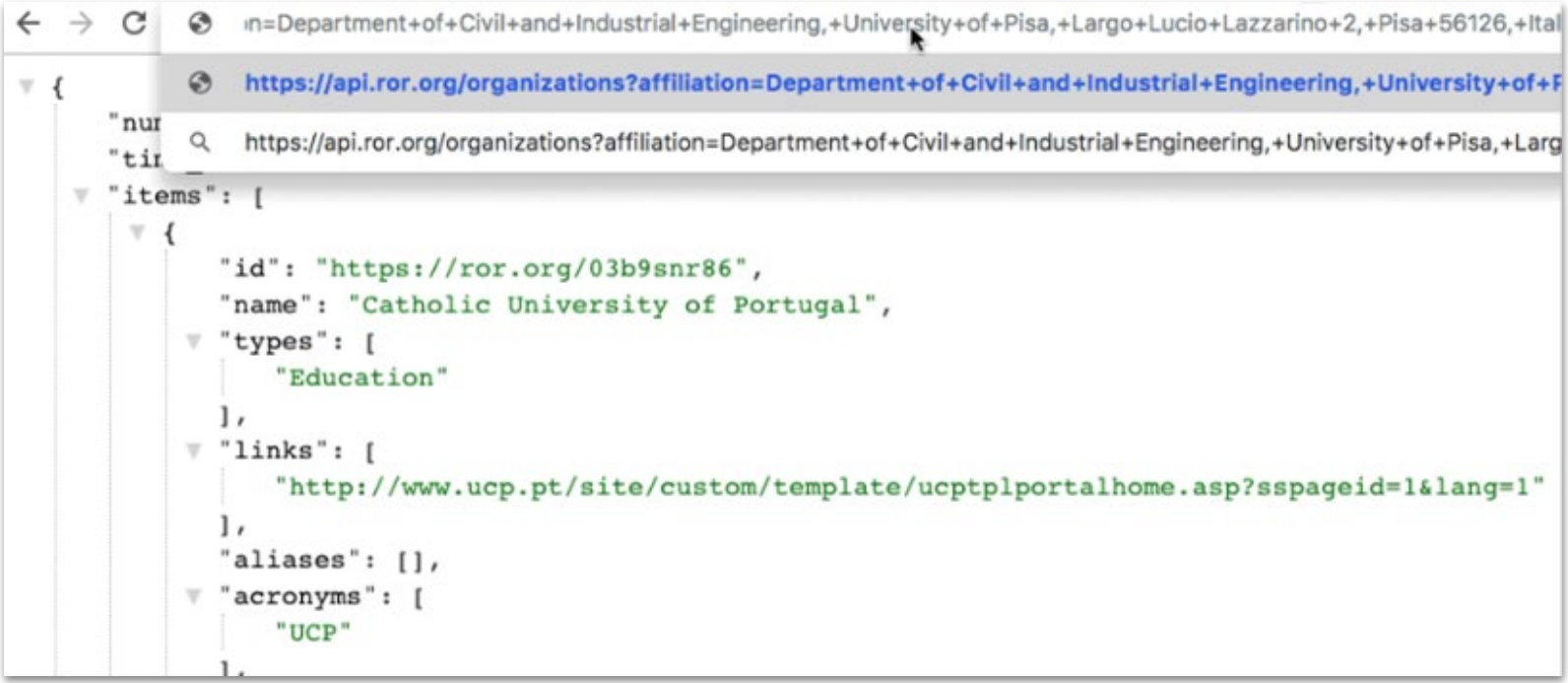
ROR registry: https://ror.org/search



ROR API

https://api.ror.org/organizations

Affiliation matching (via the API)



Submitting to DataCite

```
xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="http://datacite.org/schema/kernel-
4 http://schema.datacite.org/meta/kernel-4.3/metadata.xsd">
  <identifier identifierType="DOI">10.5072/example-full</identifier>
 ▼<creators>
  ▼<creator>
     <creatorName nameType="Personal">Miller, Elizabeth</creatorName>
     <givenName>Elizabeth</givenName>
     <familyName>Miller</familyName>
     <nameIdentifier schemeURI="http://orcid.org/" nameIdentifierScheme="ORCID">0000-0001-5000-
     0007</nameIdentifier>
     <affiliation affiliationIdentifier="https://ror.org/04wxnsj81"
     affiliationIdentifierScheme="ROR">DataCite</affiliation>
   </creator>
  ▼<creator>
     <creatorName nameType="Personal">Carberry, Josiah</creatorName>
     <givenName>Josiah</givenName>
     <familyName>Carberry</familyName>
     <nameIdentifier schemeURI="http://orcid.org/" nameIdentifierScheme="ORCID">0000-0002-1825-
     0097</nameIdentifier>
     <affiliation affiliationIdentifier="https://ror.org/05gq02987"
     affiliationIdentifierScheme="ROR">Brown University</affiliation>
     <affiliation affiliationIdentifier="grid.268117.b" affiliationIdentifierScheme="GRID"
     schemeURI="https://grid.ac/institutes/">Wesleyan University</affiliation>
   </creator>
```

So what?

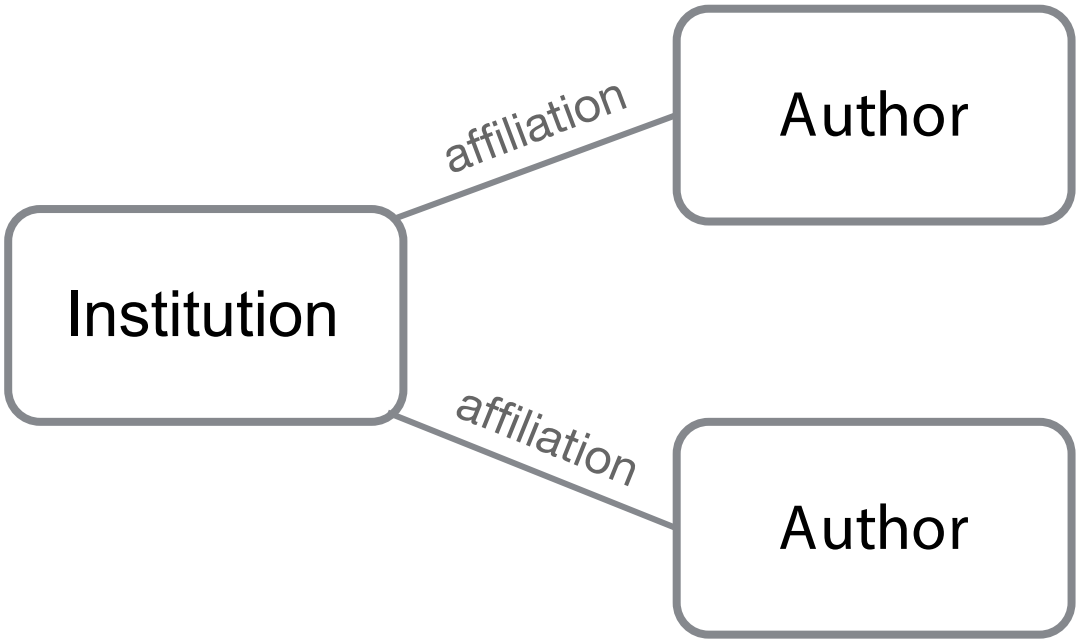
As a university administrator, I want to get a list of all datasets and software published by our researchers, so that I can get a comprehensive view of our research outputs.

This has been unnecessarily hard for too long.

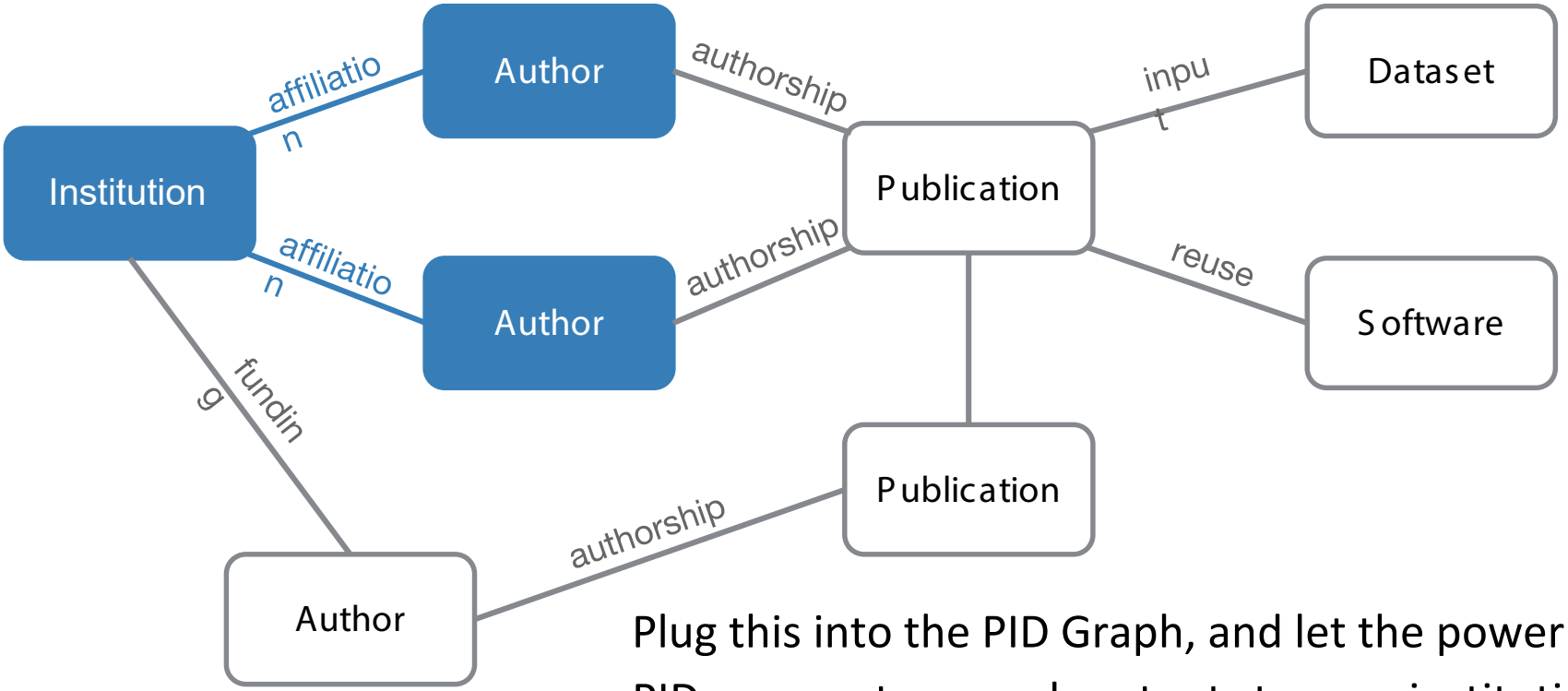
All because we couldn't definitively say who **"our researchers"** are.

ROR and the PID Graph

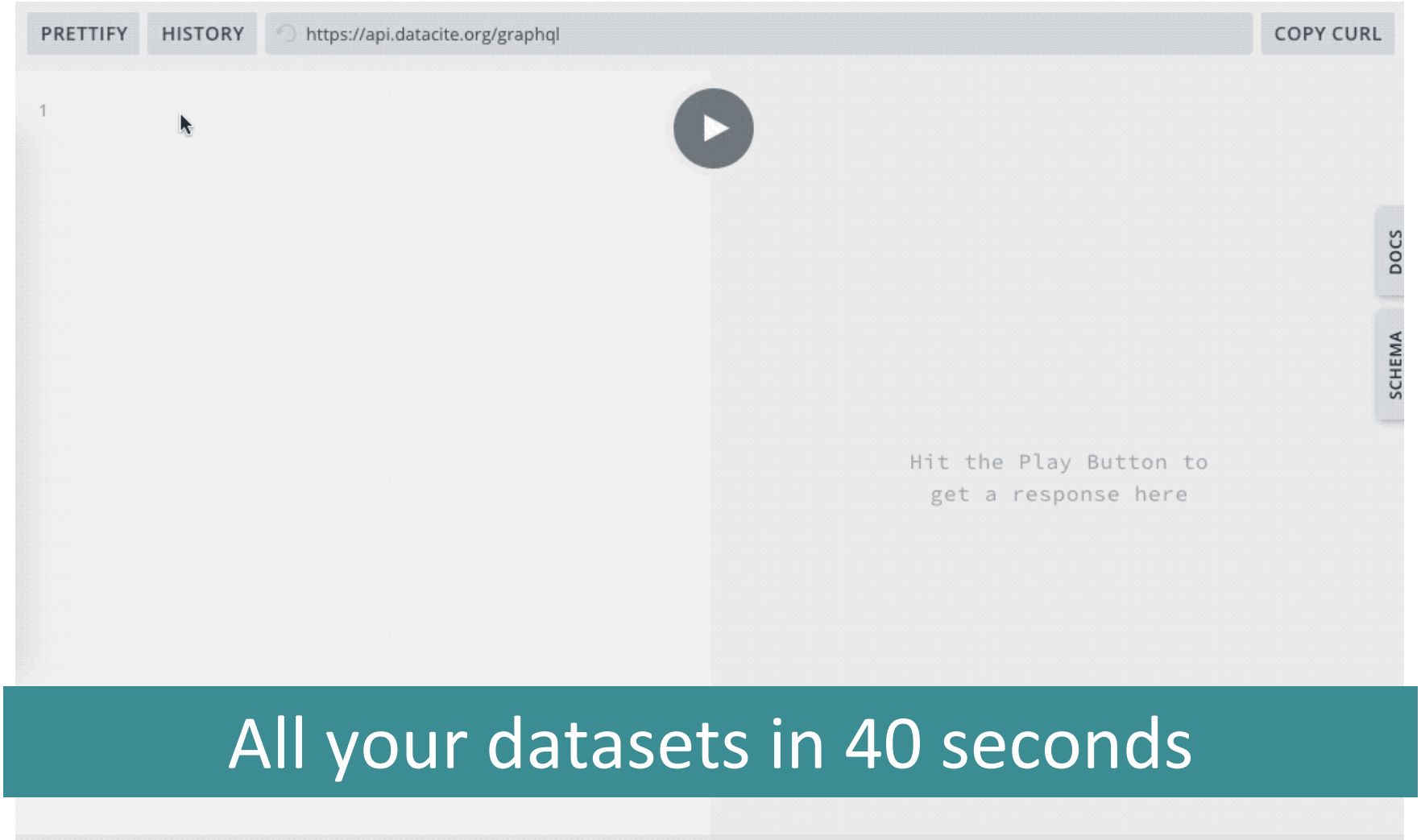
ROR definitively identifies your institution, no matter how many names it has, so you can make sure all your authors are affiliated with the right place.



ROR and the PID Graph



Plug this into the PID Graph, and let the power of PIDs connect research outputs to your institution.



All your datasets in 40 seconds