# Machine-Actionable DMPs

Exposing DMPs WG

Robin Dasler & John Chodacki
RDA 13th Plenary - Philadelphia
April 2, 2019

If DMPs get DOIs...

With a DataCite DOI, DMPs can get:

- Persistent and actionable identifier
  - Access to DataCite's supporting DOI services
- Unbreakable link between data plan to the data assets
- Open metadata available to all
- Programmatic access for humans and machines
- Born natively integrated into open scholarly infrastructure (DataCite/Crossref/ORCID/etc.)

DMP

DataCite
FIND, ACCESS, AND REUSE DATA

CDL DataCite
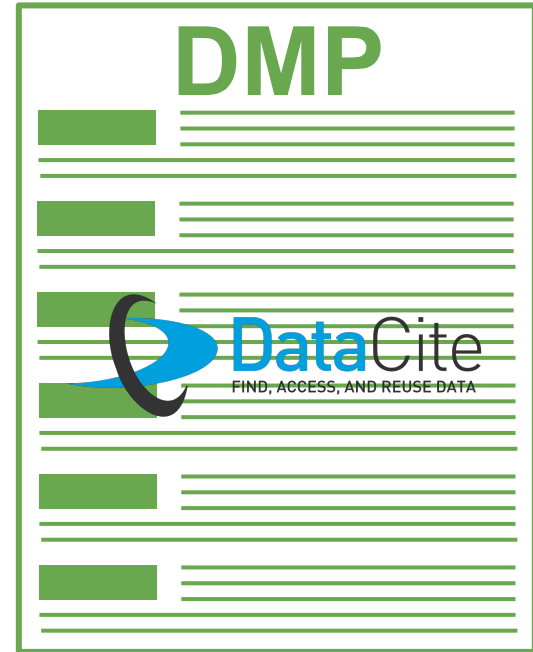
# Still the current problem: DMPs are static objects
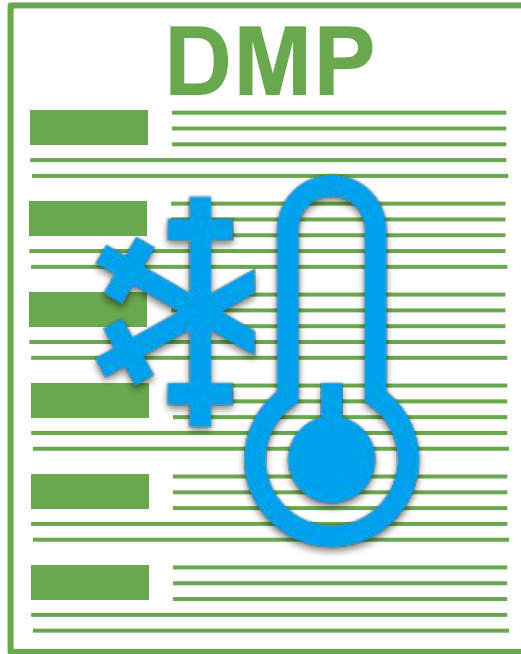
1. Created as part of a grant application BEFORE the project starts

2. Forgotten and buried in the paperwork once the project begins

3. Exchanged between P.I. and program officer

   - numerous groups involved in the research process are left out (e.g., data and software repository, field stations collecting data, curators cleaning the data, etc.)

While research is constantly changing aka 'time'

Today's DMPs are an ill fit for our ever-evolving research process.

The DMP Common Standards give a framework for description, BUT

- not all information is available at the same time

- information needs to be updated over time
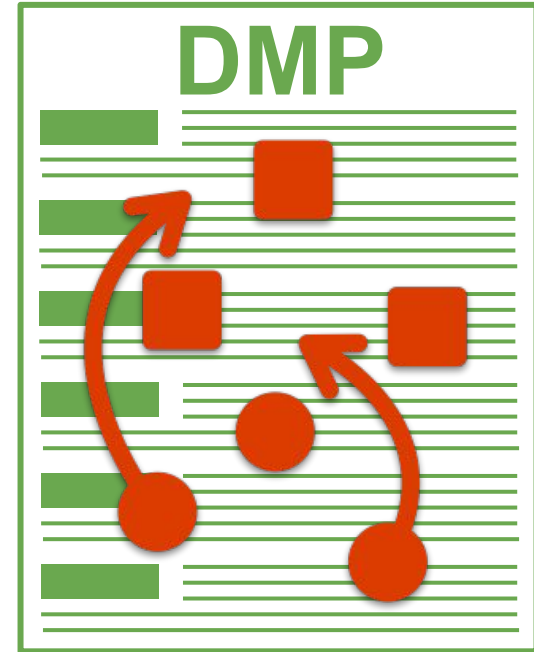
DMPs become a active document

We need DMPs to:

- expose up-to-date information as a project progresses over time

- make info available to the right parties at the right time (i.e., respecting privacy until it can be made public)

- be update-able over time by multiple parties in a decentralized fashion

With access to DataCite's Event Data service, we can tap into the potential of the PID Graph.
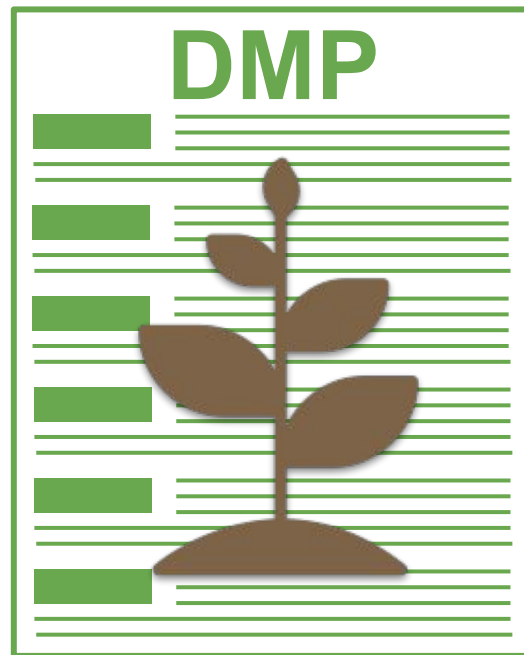
Imagine a world where...

we have a dynamic manifest of all resources involved at every point in the research process

Funders

Researchers

Data repositories

maDMP

Publishers

Institutional repositories
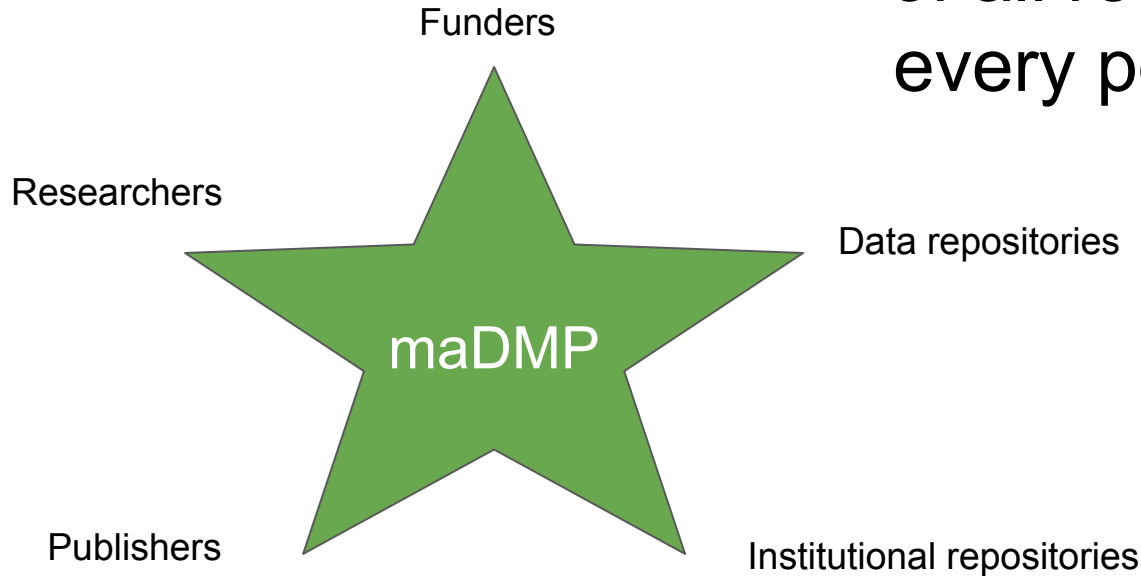
Event Data: a foundational layer

Event Data is an open scholarly infrastructure service run jointly by:

Why is Event Data useful for active DMPs?

Event Data connects our PIDs together according to relational "events", such as **A references B**.
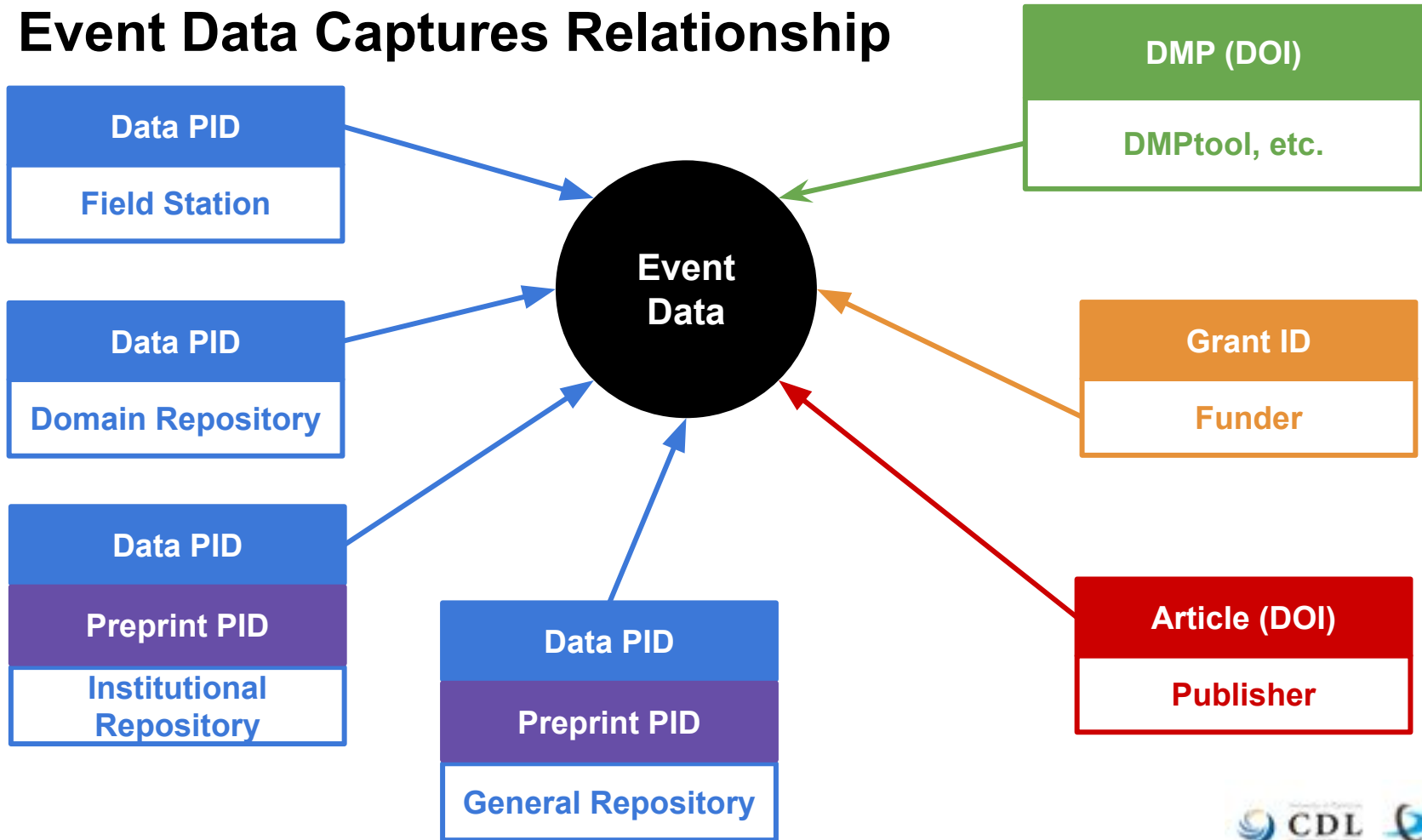
From the DataCite side, submitting a PID as a `relatedIdentifier` in the DOI metadata pipes that "event" into Event Data.
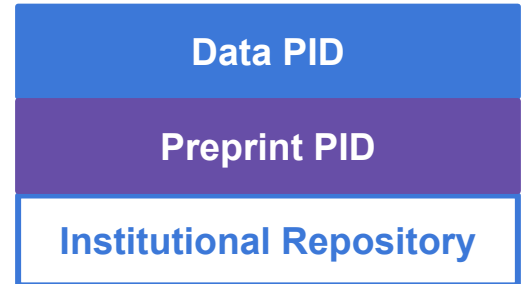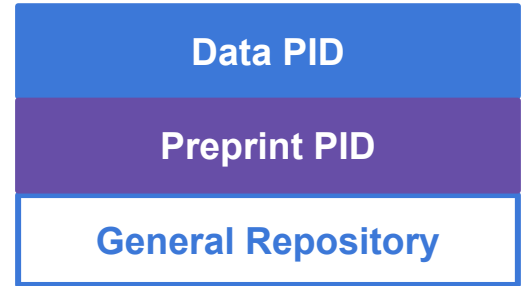
What is the PID Graph?
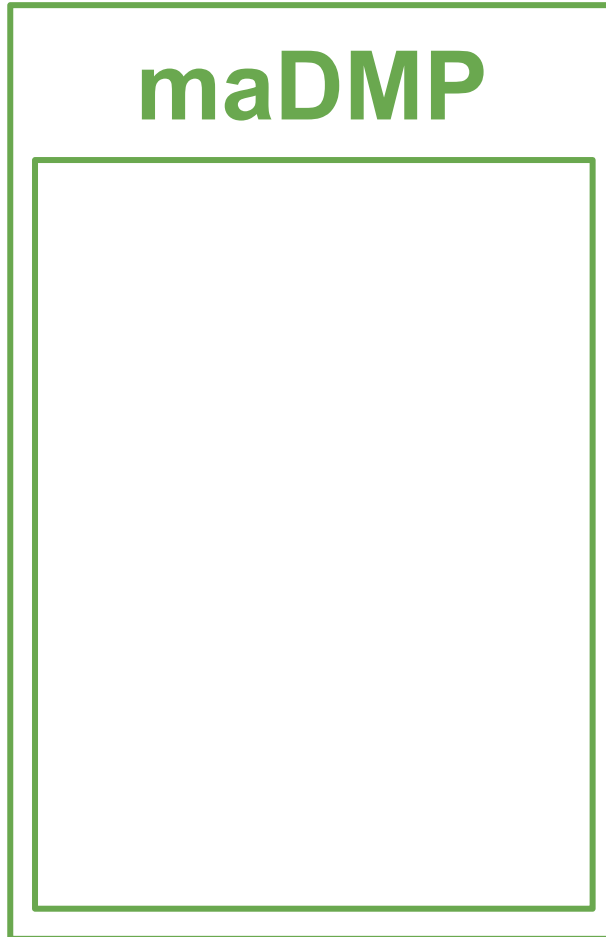
A concept of interconnected PIDs that's core to the EU-funded FREYA project (https://project-freya.eu).

DataCite's implementation of the PID Graph is realized via Event Data.

Event Data Captures Relationship

DMP created   Dataset created   Preservation

DMP	Data PID	Data PID
DMPtool	**Field Station**	**Domain Repository**

Grant ID	Preprint PID	Article (DOI)
Funder	**Preprint Server**	**Publisher**

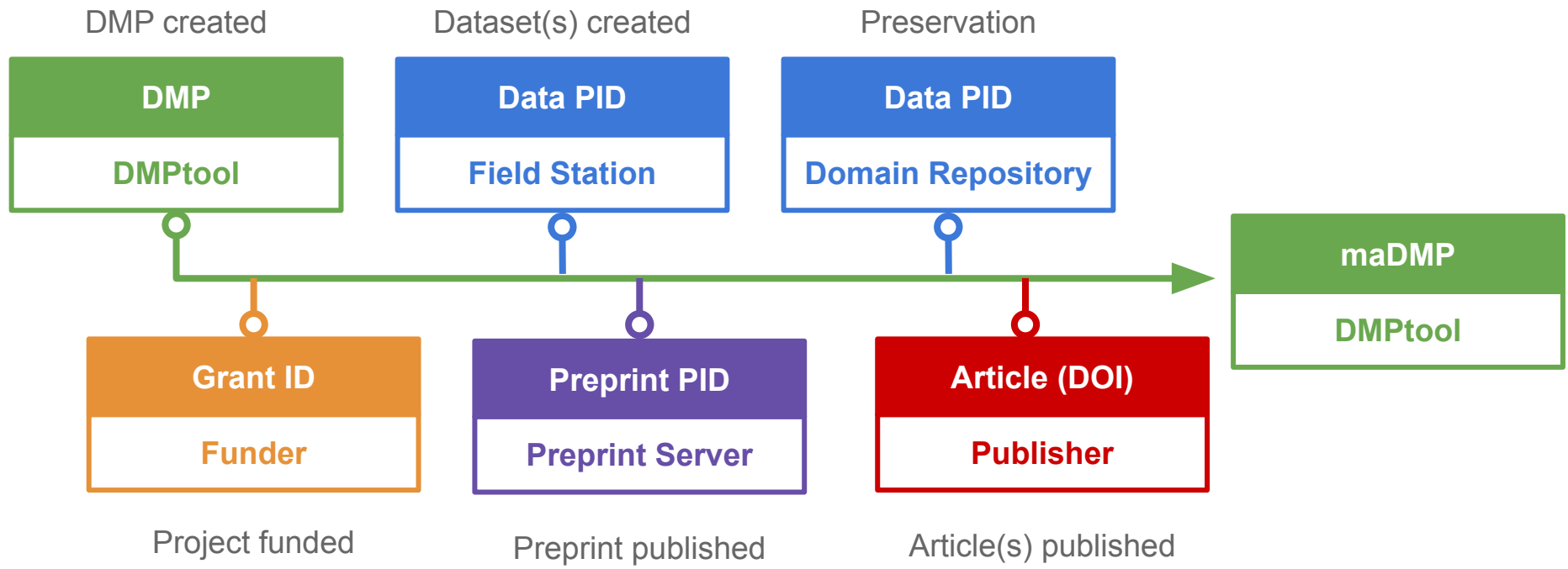Project funded   Preprint published   Article(s) published
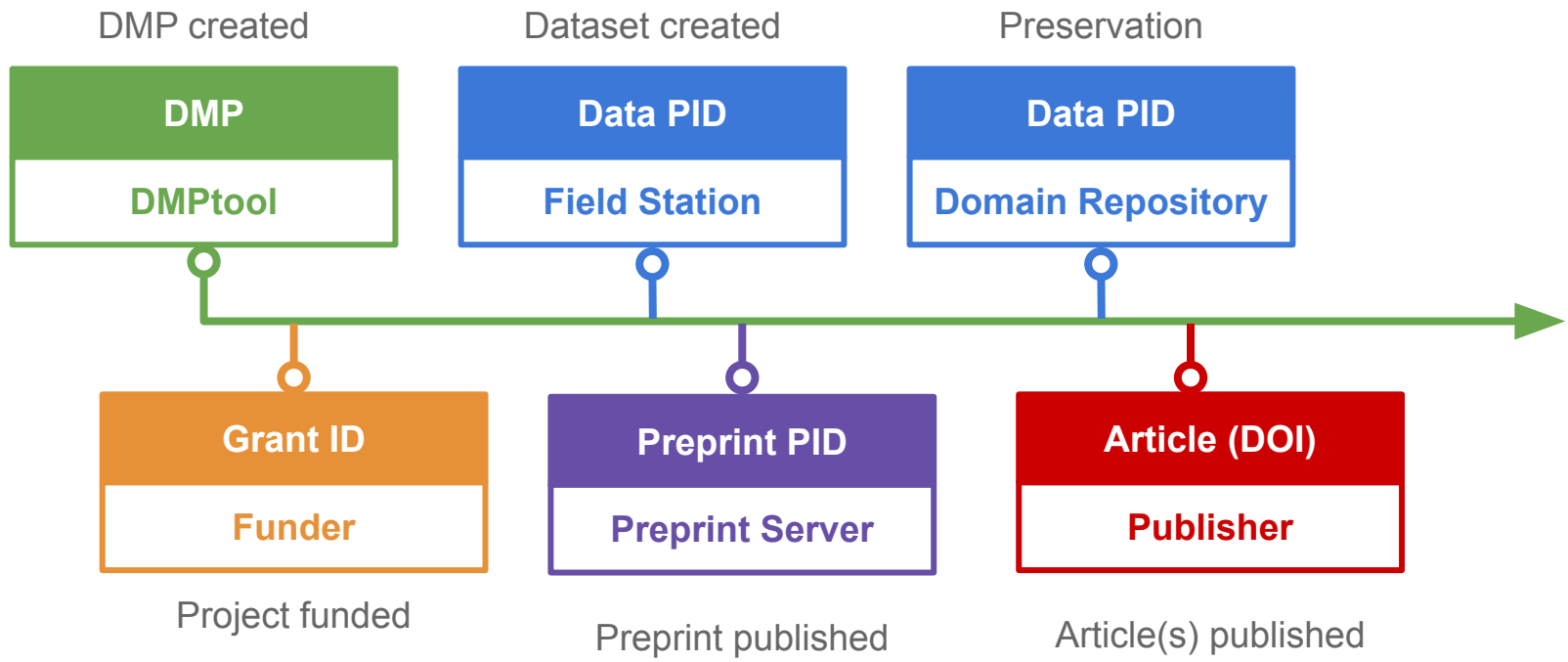
Challenges for maDMPs - getting stuff in

Event Data stores relations and connects two existing entities identified by PIDs. Event Data does not store the entities themselves.

→ **Not every item in a maDMP's lifecycle has a PID.**

Event Data gets its information from Crossref and DataCite metadata. Repositories can update their metadata to add relations. Third parties cannot contribute directly.

→ **Who runs the agents to pull in the relations?**

→ **Could you successfully proxy these instead? (e.g. Crossref grant IDs)**

Current relations of published DMPs

We looked up DMPs published in RIOJournal in Event Data.

- Only two (of 8) had any relations.
- Relation types are up to interpretation by the submitter

```
data : [
  {
    "id": "bbd27a23-922b-4da5-965b-a9711c445b91",
    "type": "events",
    "attributes": {
      "subj-id": "https://doi.org/10.3897/rio.2.e10708",
      "obj-id": "https://doi.org/10.6084/m9.figshare.1189330",
      "source-id": "crossref",
      "relation-type-id": "references",
      "total": 1,
      "message-action": "create",
      "source-token": "8676e950-8ac5-4074-8ac3-c0a18ada7e99",
      "license": "https://doi.org/10.13003/CED-terms-of-use",
      "occurred-at": "2016-10-04T00:00:00.000Z",
      "timestamp": "2019-02-26T11:51:07.391Z"
    },
    "relationships": {
```

```
  ],
  "citation-types": [
    {
      "id": "Dataset-ScholarlyArticle",
      "title": "Dataset-ScholarlyArticle",
      "count": 1,
      "year-months": [
        {
          "id": "2016-10",
          "title": "October 2016",
          "sum": 1
        }
      ]
    }
  ],
  "relation-types": [
```

CDL DataCite

Challenges for maDMPs - getting stuff out

If all the relevant entities are connected to a maDMP via relations in Event Data, we still only have a pile of relations.

→ **How is the maDMP represented to a user?**

→ **Does the maDMP's DOI landing page aggregate all the other information?**

Other questions/concerns

Privacy concerns. **Is some information inappropriate for aggregating and showing in a DOI landing page? Cost, etc?**

Editing and Updating. **How do we verify/trust the sources of the information? How to correct false assertions? Can we do this by exposing the source for all assertions?**

Other questions/concerns for implementation

Other people may want to hang common standards info off of other PIDs instead of the DMP (e.g. Grant IDs or Data DOIs). **But this shouldn't matter, right?**

Tool redesign. **How do we refactor tools like DMPTool to allow for Common Standards info to be pushed into Event Data as soon as it's available?**

Our next steps

- Explore this approach
- Utilize the Common Standards to represent the core info we capture
- Determine best way to expose the information in maDMPs at any given point

- CDL and DataCite NSF EAGER grant
- Gump Field Station as inputs
- WHOI / BCO-DMO as inputs

- Isolate the gaps and start to chip away at them

How do we begin to discuss this approach as a community?