Crossref & DataCite
joint data citation webinar

31 January 2019, 4pm UTC

Today's webinar will cover

- Introducing data citation (Rachael Lammey, Crossref)
- Data citation & data repositories (Helena Cousijn, DataCite)
- Citing data in publication metadata (Patricia Feeney, Crossref)
- Using data citation information (Robin Dasler, DataCite)
- Q&A

What is data citation?

- References to data, the same way researchers routinely provide a bibliographic reference to other scholarly resources
- Data are often shared, but they are not often cited the same way as journal articles or other publications.
- Let's change that!

Data citation: let's do this

Rachael Lammey, Helena Cousijn – 2018 October 04
In Data, Citation, Datacite

Data citation is seen as one of the most important ways to establish data as a first-class scientific output. At Crossref and DataCite we are seeing growth in journal articles and other content types citing data, and datasets making the link the other way. Our organizations are committed to working together to help realize the data citation community's ambition, so we're embarking on a dedicated effort to get things moving.
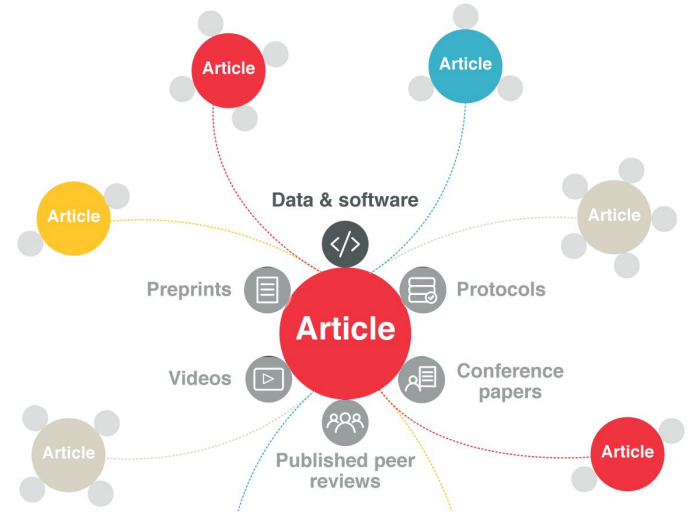
Why cite data?

- Access
- Transparency and reproducibility
- Reuse
- Track, measure, and count
- Credit
- Mandates
  - Funders
  - Publishers

Created by priyanka
from Noun Project

Data citation for publishers

- Supports scholarship
  - Extends research

- Data cited consistently provides
  - Transparency
  - Context

"eLife is committed to ensuring researchers get credit for all their outputs, and data is a major component of this." -- Melissa Harrison, eLife

Data citation for publishers

Step 1
- Develop a data policy that includes data citation - not too scary

Step 2
- Explain to authors how they should be citing data - not too scary

Step 3
- Update internal workflows, including your DTD and instructions to suppliers - quite scary but others have done it and are happy to help

Step 4
- Include these citations in your Crossref metadata - Crossref is making this as easy as possible! (Patricia will talk about this later in the webinar)

Data citation for repositories

➢ Many data repositories recognize the importance of establishing links between datasets and articles

'Data citation makes data visible to the research community. Without it, data cannot be accessed for re-use or reproduced for transparency' - ICPSR

'Data publishers efforts often risk going unnoticed, and the true impact of sharing data remains invisible' - GBIF

https://doi.org/10.5438/tyey-k867

Data citation for repositories - how?

Step 1

- Ask researchers to include associated publications in dataset metadata (many repositories do additional curation work to establish these links)

Step 2

- Include these links in your DataCite metadata (instructions on next slide)

Step 3

- Publishers can access this information and link back from the article to the dataset

Research data for this article

PANGAEA
Data Publisher for Earth & Environmental Science
Data associated with the article:
PanMap - a Mini-GIS (Geographical Information System) to draw point and vector data in maps including geographical resources ↗
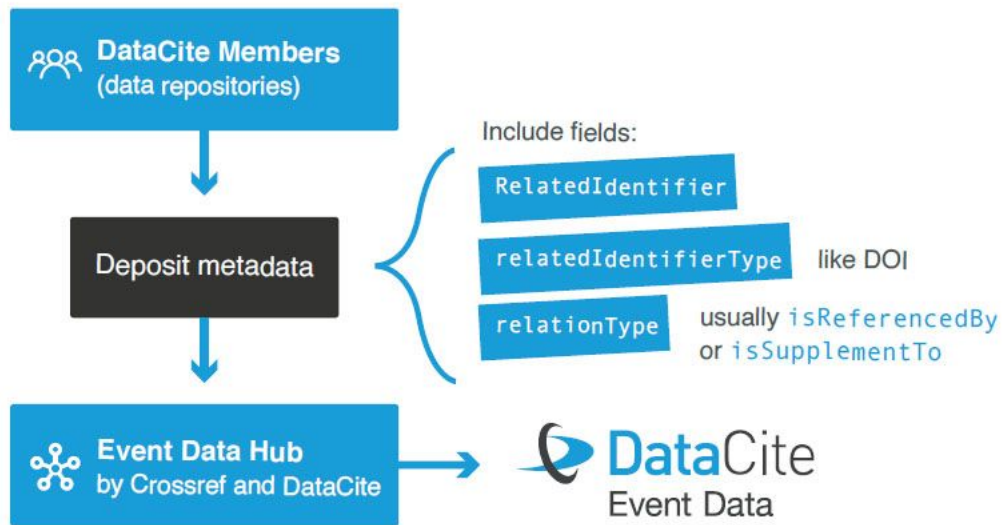
https://doi.org/10.1016/S0098-3004(02)00039-0

Adding links to your metadata deposit

Work with us on data citation! | Want to share your data citations following the Scholix framework? Here is what you need to do:

DataCite Members (data repositories)

Deposit metadata

Include fields:

`RelatedIdentifier`

`relatedIdentifierType` like DOI

`relationType` usually `isReferencedBy` or `isSupplementTo`

Event Data Hub by Crossref and DataCite → **DataCite** Event Data

Interested in using this information? Find out more at: https://support.datacite.org/docs/eventdata-guide

Citing data in publication metadata - 2 choices

1. References - include data citations in the citations you register with Crossref

2. Relations - include relationships between DOIs and other items in your Crossref metadata records

Why 2 methods?

References: ★best practice

Data citations are citations / we currently support with DataCite DOIs / our limited citation markup support makes identifying other data citations difficult

Relations:

Citations are relations ('references')/ allows you to include a more granular type of relationship as well as a description / can append to existing record without touching citations

Tagging data citations as references

Just the DataCite DOI

```
<citation key="ref2">
  <doi>10.5061/dryad.684v0</doi>
</citation>
```

Citation data plus a DOI

```xml
<citation key="ref3">
    <unstructured_citation>Morinha F, Dávila JA, Estela B, Cabral JA, Frías Ó, González JL, Travassos P, Carvalho D, Milá B, Blanco G (2017) Data from: Extreme genetic structure in a social bird species despite high dispersal capacity. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.684v0</unstructured_citation>
<doi>10.5061/dryad.684v0</doi>
</citation>
```

Future plans

Expanded citation markup with publication types

```xml
<citation key="ref3" publication_type="data">
  <author>Morinha F</author>
  <cYear>2017</cYear>
  <institution>Dryad Digital Repository</institution>
  <title>Data from: Extreme genetic structure in a social bird species despite high dispersal capacity</title>
  <doi>10.5061/dryad.684v0</doi>
  <identifier type="accession">ABC123</identifier>
   <unstructured_citation>Morinha F, Dávila JA, Estela B, Cabral JA, Frías Ó, González JL, Travassos P, Carvalho D, Milá B, Blanco G (2017) Data from: Extreme genetic structure in a social bird species despite high dispersal capacity. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.684v0</unstructured_citation>
</citation>
```

Tagging data citations as relations

A dataset is cited by the article being registered using a URI as an identifier

```
<program xmlns="http://www.crossref.org/relations.xsd">
  <rel:related_item>
    <rel:description>Human and mouse DRIP-seq and
    DRIPc-seq</rel:description>
    <rel:inter_work_relation identifier-type="uri"
    relationship-type="references">
      https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70189
    </rel:inter_work_relation>
  </rel:related_item>
</program>
```

Tagging data citations as relations

A dataset with accession number GSE44902 is cited by the article being registered. The dataset was generated as part of the article being registered.

```xml
<program xmlns="http://www.crossref.org/relations.xsd">
  <rel:related_item>
   <rel:description>NKX2-5 mutations causative for congenital heart
   disease retain and are directed to hundreds of
   targets</rel:description>
   <rel:inter_work_relation identifier-type="accession"
   relationship-type="isSupplementedBy">GSE44902</rel:inter_work_relatio
   n>
  </rel:related_item>
</program>
```

References

- Citations are available via metadata APIs *if references are set to 'open'*

- Citations with DataCite DOI are sent to Event Data

- Future: citations flagged as datasets are events in Event Data

Relations

- Relations are available via metadata APIs

- Future: will be events in Event Data

- Can specify 'references' or 'isSupplementedBy' as relation type

Data Citation in Three Steps (Simple Version)

1

Reference to dataset included in article metadata (Crossref) and/or reference to article included in dataset metadata (DataCite).

2

References Crossref DOI <-> DataCite DOI extracted from metadata, stored in Event Data service and made available via APIs.

3

New services and updates of existing services, e.g. DOI Search, that integrate Event Data information.

Data Citation in Four Steps (Simple Version)

4

Work with bibliometrics community to understand data citation data provided via Event Data, and develop data metrics.

Event Data

- Service jointly developed by Crossref and DataCite to capture references, mentions and other events around DOIs that are not provided via DOI metadata.

- Also includes references between different DOI registration agencies.

- Events is used as a broad term to for example also include social media mentions and usage statistics.

- Data citations are a small subset of the events captured by the Event Data service.

Event Data Service is crucial for Data Citation

Before Event Data

- Data citations needed to be found in DOI metadata, separate for Crossref and DataCite.
- Doesn't scale well, limiting adoption of data citation.

After Event Data

- Data citations are extracted into a separate service and can easily be found.
- Scales well, will lead to adoption of data citation.

Event Data Service is crucial for Data Citation

Additional Data Citation Sources

Event Data can be extended to include other types of events outside of the usual article-to-dataset relations.

- Crossref is pulling social media information into Event Data (like Twitter and Wikipedia).
- DataCite is sending data repository usage reports to Event Data.

Event Data APIs

1

Crossref Event Data Query API

Query Event Data using a large number of parameters.

https://www.eventdata.crossref.org/guide/

2

DataCite REST API

Query Event Data that include a DataCite DOI. Integrated into the DataCite REST API, includes basic DOI metadata, and aggregations.

https://support.datacite.org/docs/eventdata-guide

3

Scholix-compatible API

Data citations in Event Data using the Scholix metadata standard.

Event Data APIs

- Both the Crossref and DataCite API support queries by DOI, DOI prefix, date range, relation type, and source (e.g. Crossref to DataCite, DataCite to Crossref and/or DataCite usage statistics).

- To better support data citation, the DataCite API also supports filtering by content type (e.g. only show events that include a DOI for a dataset)

Example from DataCite REST API

(note: results truncated)

```
"id": "ebac9644-73c2-484e-8311-e9abd146333a",

"type": "events",

"attributes": {

    "subj-id": "https://doi.org/10.1594/wdcc/coastdat-1_hamsom",

    "obj-id": "https://doi.org/10.1016/j.jmarsys.2011.02.001",

    "source-id": "datacite-crossref",

    "relation-type-id": "cites",

    "occurred-at": "2015-05-12T13:11:01.000Z"

},
```

Example from Crossref API

```
"license": "https://doi.org/10.13003/CED-terms-of-use",
"obj_id": "https://doi.org/10.5281/zenodo.54705",
"source_token": "8676e950-8ac5-4074-8ac3-c0a18ada7e99",
"occurred_at": "2016-11-23T00:00:00Z",
"subj_id": "https://doi.org/10.1111/jeb.13002",
"id": "71f04fe8-5c15-4b5e-90ab-25e825ae820f",
"terms": "https://doi.org/10.13003/CED-terms-of-use",
"message_action": "create",
"source_id": "crossref",
"timestamp": "2017-05-19T13:30:17Z",
"relation_type_id": "references"
```

Future Work: Services based on Event Data

1

Crossref and DataCite

Integration of Event Data information into other services, e.g. DOI search.

2

Other Organizations

Third parties build services based on the information provided by Event Data.

3

PID Graph

DataCite and other partners in the EC-funded FREYA project use Event Data to build the PID Graph, connecting persistent identifiers for articles, data, people, etc.

Thank you!
Questions?

Useful links and contacts:

- Patricia Cruse, Data Citations and why they matter, Crossref LIVE18: https://www.youtube.com/watch?v=4GPOf_29zkg&index=5&list=PLe_-TawAqQj2W3Lc_hMxz1CDZckbOEYnO&t=0s
- Crossref Data & Software Citation Deposit Guide for Publishers: https://support.crossref.org/hc/en-us/articles/215787303-Crossref-Data-Software-Citation-Deposit-Guide-for-Publishers
- Crossref Event Data: https://www.crossref.org/services/event-data/
- A Data Citation Roadmap for Scientific Publishers: https://doi.org/10.1038/sdata.2018.259
- support@crossref.org/support@datacite.org