

DataCite Services: Promoting Open Science Data

データサイトサービス: オープンサイエンスデータの推進

By Kristian Garza

Open Science Data Promotion Workshop, Kyoto 2018

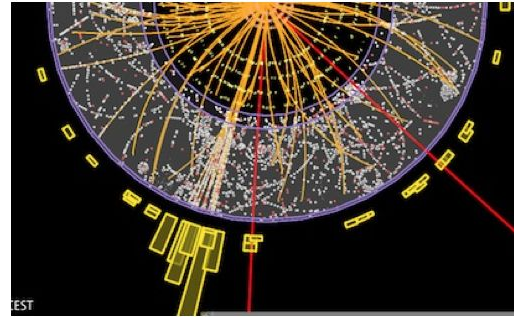
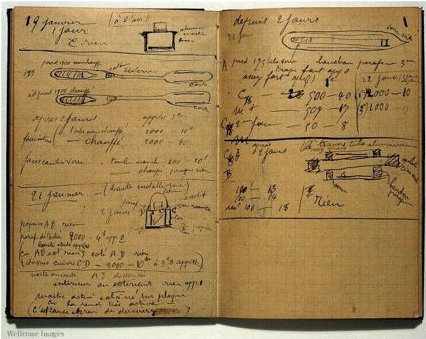
Agenda

発表内容

1. DataCite Introduction データサイトの紹介
2. Services and a Framework for Crosslinking Data, Authors and Publications データ・著者・論文の相互リンクのためのサービス及びフレームワーク
3. Services to Globally Identify Data Authors データ作成者をグローバルに特定するサービス
4. Guidelines for Research Data Metrics 研究データメトリックのためのガイドライン

The problem

- Data form the foundation of research
 - yet, they are treated as ‘second class’ citizens!
- We need reliable and unambiguous access to data:
 - collaboration
 - reuse
 - reproducibility
 - attribution
 - grant credit
 - faster progress
 - consistent funding
 - feed future researchers



Curie Notebook → ATLAS Experiment

How to tackle it



Technical infrastructure

- building a reliable underlying layer



Support services

- applying technology to the real world



Integration

- making it useful & usable for researchers



Metrics

- tracking impact and making it visible



Funding

- working with funding agencies



Training

- sharing best practices

What is DataCite?



- DataCite provides **technical infrastructure**:

- Assign DOIs to research data
- Build and adopt services for data sharing
- Integrate with other community services



- DataCite provides **community infrastructure**:

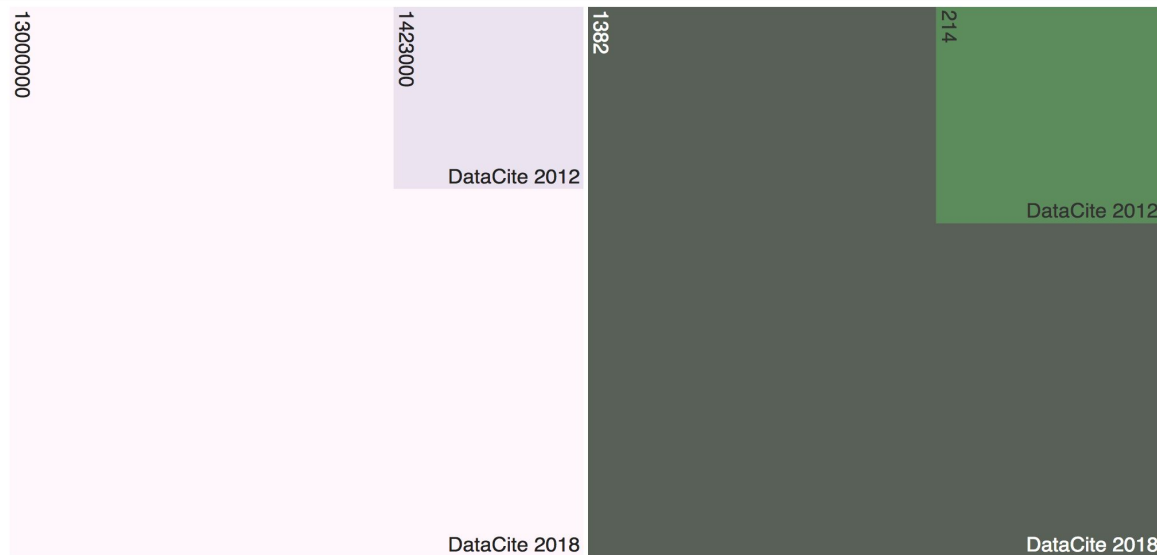
- Advocate and communicate about the importance of data
- Provide training and collaborate with other organisations



DataCite a quick snapshot

- Not-for-profit Global initiative since 2009
- > 60 Members worldwide (Large institutions)
- > 1300 Datacenters
- >13,000,000 DOIs registered
- > 6,000,000 resolutions per month

DataCite's growth



*Proportional Area Chart for DOI registrations by 2012 superposed on **total of number of DOI Registrations** by 2018*

*Proportional Area Chart for datacenters with DataCite by 2012 superposed on **total of number of datacenters with DataCite** by 2018*

Persistent Identifier Systems 永続的 ID システム

Persistent Uniform Resource Locator (PURL) 永続的 URL

Life Science Identifier (LSID)

ライフサイエンス ID

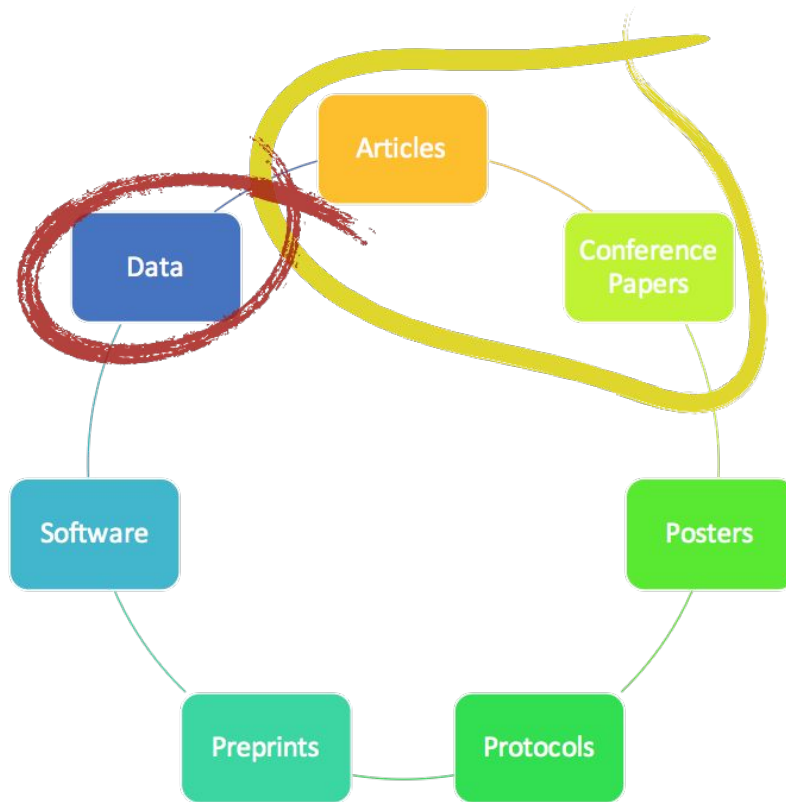
Digital Object Identifier (DOI)

デジタルオブジェクトID

Handles ハンドラ

Archival Resource Key (ARK) アーカイブ資源キー



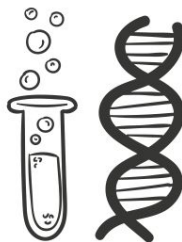
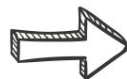


Advantages of Persistent Identifiers for research data

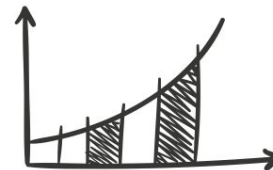
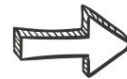
- Unique and long-lasting references to research outputs



Ideas



Research



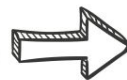
Publish results

- PIDs make it easier to find datasets

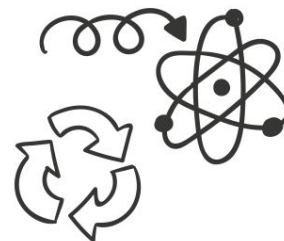
- PIDs make it easier to cite data and track its impact



Search



Find!



Reuse and evolve

What is a DOI?



- DOI stands for Digital Object Identifier, the PID that has been used for many years to reference research articles in scholarly communication
- DOIs:
 - uniquely identify/name digital content
 - are a stable, persistent location on the web
 - are well known by researchers



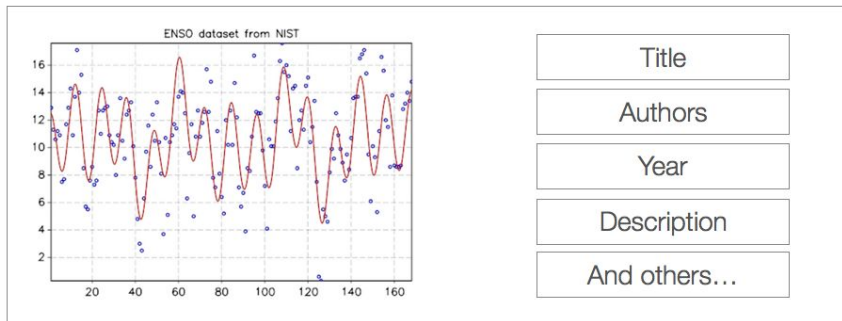
Name

URL

Metadata

Down in the weeds

1. Take a dataset



2. Describe it

3. Assign a DOI



4. Reuse and reference!

ATLAS Collaboration, "Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC: $H \rightarrow \gamma\gamma$,"
<http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44>

5. Enjoy the benefits!

Findability

Track citations

Reusability

Measure impact

Unique

Persistent



The benefits can be enjoyed worldwide



Agenda

発表内容

1. DataCite Introduction
2. **Services and a Framework for Crosslinking
Data, Authors and Publications** データ・著者
・論文の相互リンクのためのサービス及びフ
レームワーク
3. Services to Globally Identify Data Authors
4. Guidelines for Research Data Metrics

データ・著者・論文を相互に リンクするためのサービス及び フレームワーク

Services and a Framework for
Crosslinking

Data, Authors and Publications



PIDs Cross-linking PIDの相互リンク

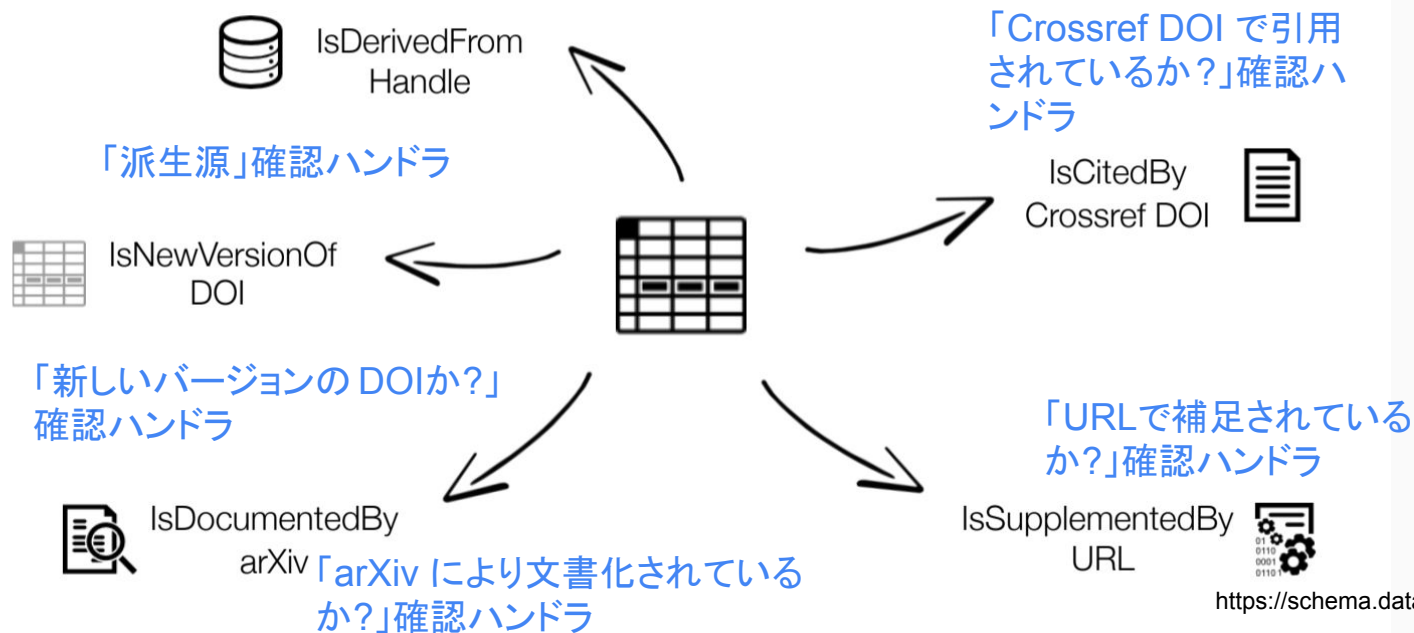
Cross-linking refers to the explicit relationship between two or more PIDs. We worked on identifying the best practices to link different types of identifiers and data types in the context the services that provide the PID, whether this be connecting data to publications, to researchers, to organizations, to funders, or some combination thereof.

Benefits of Cross-linking Data and Publications

- Increased visibility, discovery and retrieval of both literature and data
- Facilitating reuse, reproducibility and transparency of research
- Enabling better attribution of credit for published data, providing an additional incentive for researchers to share their data.

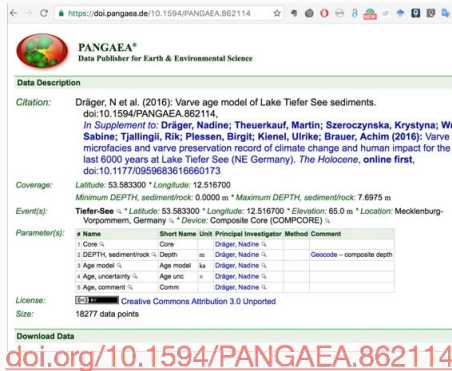
Using Related-Identifier field in the Metadata

メタデータの関係するIDフィールドを利用



Discovery of Links between Data and Scholarly Articles

Dataset



PANGAEA®
Data Publisher for Earth & Environmental Science

Data Description

Citation: Dräger, N et al. (2016): Varve age model of Lake Tiefsee sediments. doi:10.1594/PANGAEA.862114.
In Supplement to: Dräger, Nadine; Theuerkauf, Martin; Szeroczynska, Krystyna; Wu Sabine; Tjallingii, Rik; Plessen, Birgit; Kienel, Ulrike; Brauer, Achim (2016): Varve microfacies and varve preservation record of climate change and human impact for the last 6000 years at Lake Tiefsee (NE Germany). The Holocene, online first, doi:10.1177/0959683616660173

Coverage: Latitude: 53.583300 * Longitude: 12.516700
Minimum DEPTH_sedimentrock: 0.0000 m * Maximum DEPTH_sedimentrock: 7.8975 m

Event(s): Tiefsee % * Latitude: 53.583300 * Longitude: 12.516700 * Elevation: 65.0 m * Location: Mecklenburg-Vorpommern, Germany % * Device: Composite Core (COMPCORE) %

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	Core			Dräger, Nadine %		
2	DEPTH_sedimentrock % Depth	m		Dräger, Nadine %		GeoCODE - composite depth
3	Age model %	Age model	ka	Dräger, Nadine %		
4	Age uncertainty %	Age unc		Dräger, Nadine %		
5	Age comment %	Comen		Dräger, Nadine %		

License: CC BY Creative Commons Attribution 3.0 Unported

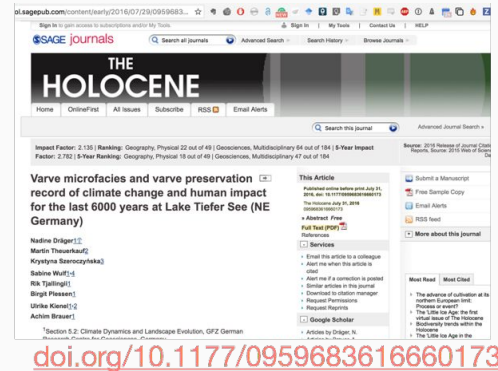
Size: 18277 data points

Download Data

doi.org/10.1594/PANGAEA.862114



Publication



SAGE Journals

THE HOLOCENE

Home OnlineFirst All Issues Subscribe RSS Email Alerts

Impact Factor: 2.135 | Ranking: Geography, Physical 22 out of 48 | Geosciences, Multidisciplinary 64 out of 104 | 5-Year Impact Factor: 2.782 | 5-Year Ranking: Geography, Physical 18 out of 48 | Geosciences, Multidisciplinary 47 out of 104

Varve microfacies and varve preservation record of climate change and human impact for the last 6000 years at Lake Tiefsee (NE Germany)

This Article
Published online before print July 28, 2016, as 10.1177/0959683616660173

Abstract Free Full text (open access) References

References

- Email this article to a colleague
- Alert me when this article is cited
- Alert me if a correction is posted
- Social articles in this journal
- Download this article manager
- Request Permissions
- Request Reprints
- Google Scholar
- Articles by Dräger, N.

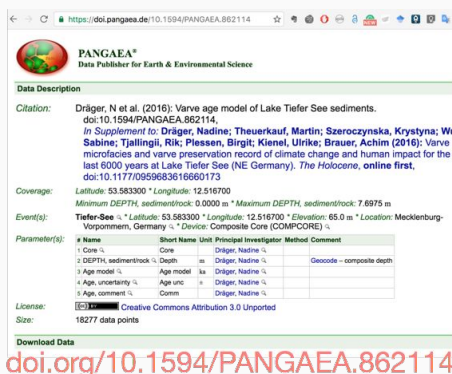
Most Read | **Most Cited**

- The Holocene of Subarctic and Arctic Siberia
- The Life in Age: the first millennium BC in the Mediterranean
- The Life in Age in the

Section 5.2: Climate Dynamics and Landscape Evolution, GFZ German Research Data Archive

doi.org/10.1177/0959683616660173

Discovery of Links between Data and Scholarly Articles



PANGAEA®
Data Publisher for Earth & Environmental Science

Data Description

Citation: Dräger, N et al. (2016): Varve age model of Lake Tiefer See sediments. doi:10.1594/PANGAEA.862114.
In Supplement to: Dräger, Nadine; Theuerkauf, Martin; Szeroczynska, Krystyna; Wu Sabine; Tjallingii, Rik; Plessen, Birgit; Kienel, Ulrike; Brauer, Achim (2016): Varve microfacies and varve preservation record of climate change and human impact for the last 6000 years at Lake Tiefer See (NE Germany). The Holocene, online first, doi:10.1177/0959683616660173

Coverage: Latitude: 53.583300 ° Longitude: 12.516700
Minimum DEPTH_sedimentrock: 0.0000 m * Maximum DEPTH_sedimentrock: 7.8975 m

Event(s): Tiefer-See °, * Latitude: 53.583300 ° Longitude: 12.516700 ° Elevation: 65.0 m ° Location: Mecklenburg-Vorpommern, Germany ° * Device: Composite Core (COMPCORE) °

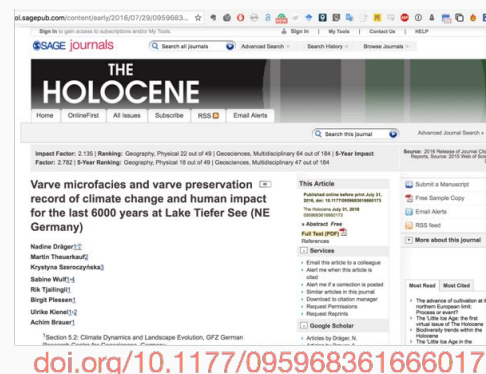
#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	Core			Dräger, Nadine °		
2	DEPTH_sedimentrock °_Depth		ka	Dräger, Nadine °		Geocode - composite depth
3	Age model °			Dräger, Nadine °		
4	Age uncertainty °			Dräger, Nadine °		
5	Age comment °			Dräger, Nadine °		

License: CC BY Creative Commons Attribution 3.0 Unported

Size: 18277 data points

Download Data

doi.org/10.1594/PANGAEA.862114



SAGE Journals

THE HOLOCENE

Home OnlineFirst All Issues Subscribe RSS Email Alerts

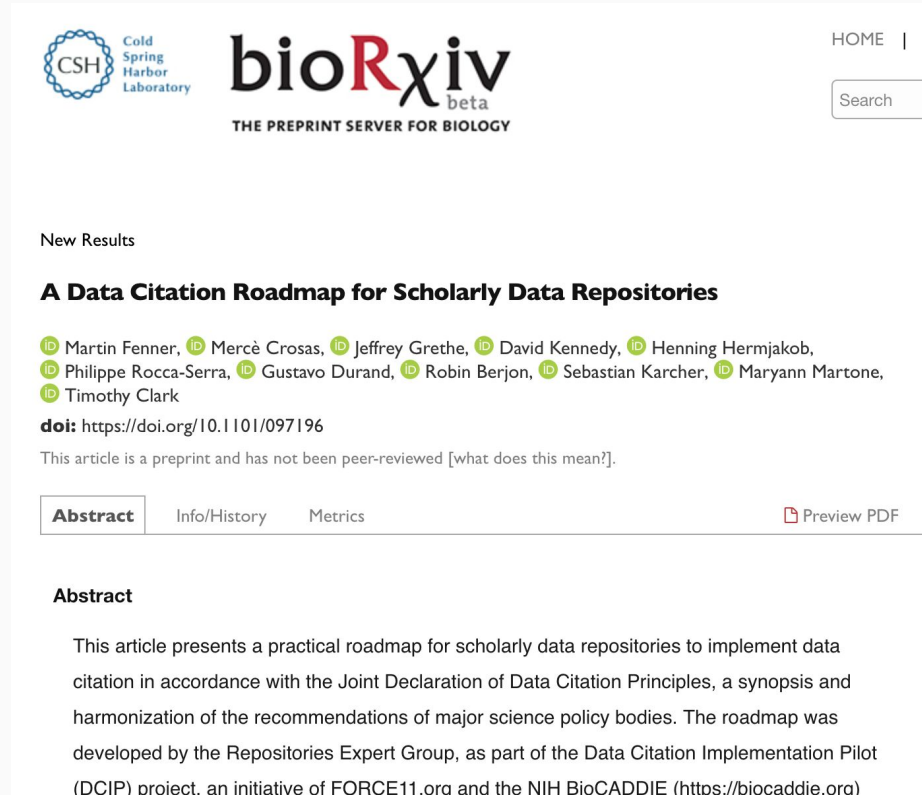
Impact Factor: 2.135 | Ranking: Geography, Physical 22 out of 48 | Geosciences, Multidisciplinary 64 out of 104 | 5-Year Impact Factor: 2.782 | 5-Year Ranking: Geography, Physical 18 out of 48 | Geosciences, Multidisciplinary 47 out of 104

Varve microfacies and varve preservation record of climate change and human impact for the last 6000 years at Lake Tiefer See (NE Germany)

Nadine Dräger[1]
Martin Theuerkauf[2]
Krystyna Szeroczynska[164]
Sabine Wu[1-4]
Rik Tjallingii[1]
Birgit Plessen[1]
Ulrike Kienel[2]
Achim Brauer[1]

doi.org/10.1177/0959683616660173





The screenshot shows the bioRxiv preprint server interface. At the top left is the Cold Spring Harbor Laboratory (CSH) logo. The bioRxiv logo is prominently displayed in the center, with the text 'beta' and 'THE PREPRINT SERVER FOR BIOLOGY' below it. In the top right corner, there are navigation links for 'HOME' and 'A', and a search bar. The main content area features a 'New Results' section with the title 'A Data Citation Roadmap for Scholarly Data Repositories'. Below the title, the authors are listed with ORCID iD icons: Martin Fenner, Mercè Crosas, Jeffrey Grethe, David Kennedy, Henning Hermjakob, Philippe Rocca-Serra, Gustavo Durand, Robin Berjon, Sebastian Karcher, Maryann Martone, and Timothy Clark. The DOI is provided as <https://doi.org/10.1101/097196>. A note indicates that the article is a preprint and has not been peer-reviewed. Below the text, there are navigation tabs for 'Abstract', 'Info/History', and 'Metrics', along with a 'Preview PDF' button. The abstract text begins with 'This article presents a practical roadmap for scholarly data repositories to implement data citation in accordance with the Joint Declaration of Data Citation Principles, a synopsis and harmonization of the recommendations of major science policy bodies. The roadmap was developed by the Repositories Expert Group, as part of the Data Citation Implementation Pilot (DCIP) project, an initiative of FORCE11.org and the NIH BioCADDIE (<https://biocaddie.org>)'.



Providing services for
crosslinking and
embedding metadata
in landing pages

→ [Secure https://search.datacite.org/works/10.6084/M9.FIGSHARE.1](https://search.datacite.org/works/10.6084/M9.FIGSHARE.1)

DataCite Search

Framing the Community Data System Interface

Kristian Garza

Dataset published 2015 via Figshare



<https://doi.org/10.6084/M9.FIGSHARE.1300051> Cite

Elements Console Sources Network Timeline Profiles Application Security

```
</div>
<div class="container-fluid">
  ::before
  <div class="content">
    <div class="row">
      ::before
      <div class="col-md-9 id="content-work">
        <script type="application/ld+json">
          {
            "@context": "http://schema.org",
            "@type": "Dataset",
            "@id": "https://doi.org/10.6084/M9.FIGSHARE.1300051",
            "name": "Framing the Community Data System Interface",
            "author": {
              "@type": "Person",
              "givenName": "Kristian",
              "familyName": "Garza"
            },
            "datePublished": "2015",
            "dateModified": "2016-03-31T17:10:34Z",
            "description": "",
            "license": "https://creativecommons.org/licenses/by/4.0/us/"
          }
        </script>
      </div>
    </div>
  </div>
</div>
```

DataCite Search

Framing the Community Data System Interface

Kristian Garza

Dataset published 2015 via Figshare



<https://doi.org/10.6084/M9.FIGSHARE.1300051> Cite

🔍 Elements Console Sources Network Timeline Profiles Application Security

```
</div>
<div class="container-fluid">
  ::before
  <div class="content">
    <div class="row">
      ::before
      <div class="col-md-9" id="content-work">
        <script type="application/ld+json">
          {
            "@context": "http://schema.org",
            "@type": "Dataset",
            "@id": "https://doi.org/10.6084/M9.FIGSHARE.1300051",
            "name": "Framing the Community Data System Interface",
            "author": {
              "@type": "Person",
              "givenName": "Kristian",
              "familyName": "Garza"
            },
            "datePublished": "2015",
            "dateModified": "2016-03-31T17:10:34Z",
            "description": "",
            "license": "https://creativecommons.org/licenses/by/4.0/us/"
          }
        </script>
      </div>
    </div>
  </div>
</div>
```

Using Content-Negotiation

```
$(document).ready(function() {
  var identifier =
  $("meta[name='DC.identifier']").attr("content");
  if (identifier === undefined) {
    return;
  }
  var doi = new URL(identifier);
  var url = $('#site-title').attr('data-conneg');
  url += '/application/vnd.schemaorg.ld+json/' +
  doi.pathname;

  $.ajax({
    url: url,
    dataType: 'text', // don't convert JSON to
    // Javascript object
    success: function(data) {
      $('<script>')
        .attr('type', 'application/ld+json')
        .text(data)
        .appendTo('head');
    },
    error: function(error) {
      console.log(error.responseJSON);
    }
  });
});
```

DEMO デモ

<https://doi.pangaea.de/10.1594/PANGAEA.760907>

How to embed metadata in web pages

DataCite exposes webpages with crawl-able structured data using schema.org markup.

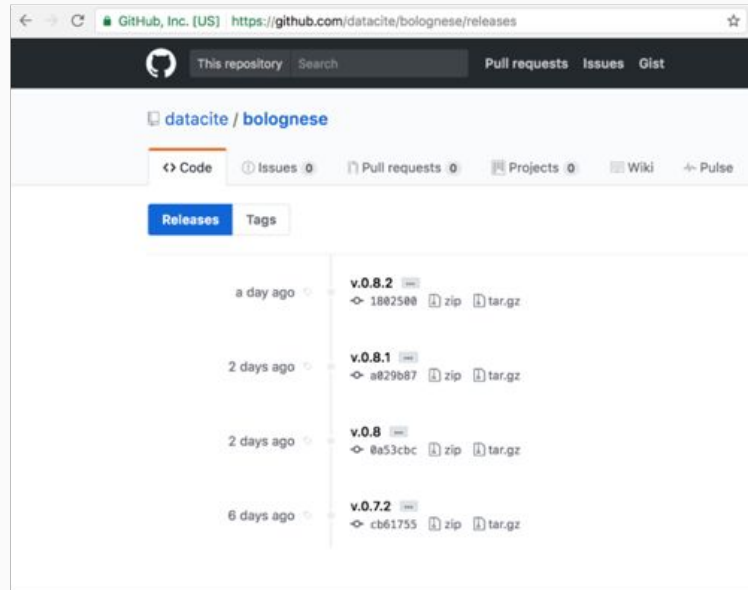
Using the dataset class from schema.org

Mapping from DataCite Schema (v4.0, 2016) (which has only 6 mandatory fields)



Facilitating schema conversion and formatting.

- DataCite provides libraries for programmatically provide datasets metadata in schema.org format.
- Repositories can use these libraries to add schema.org markup in their webpages for each dataset.



Uptake

Repositories with metadata embedded in Schema.org. Some examples:

- Dryad
- Dataverse
- Zenodo
- And others...
- Full list:

https://github.com/datacite/freya/blob/master/schema_org.csv



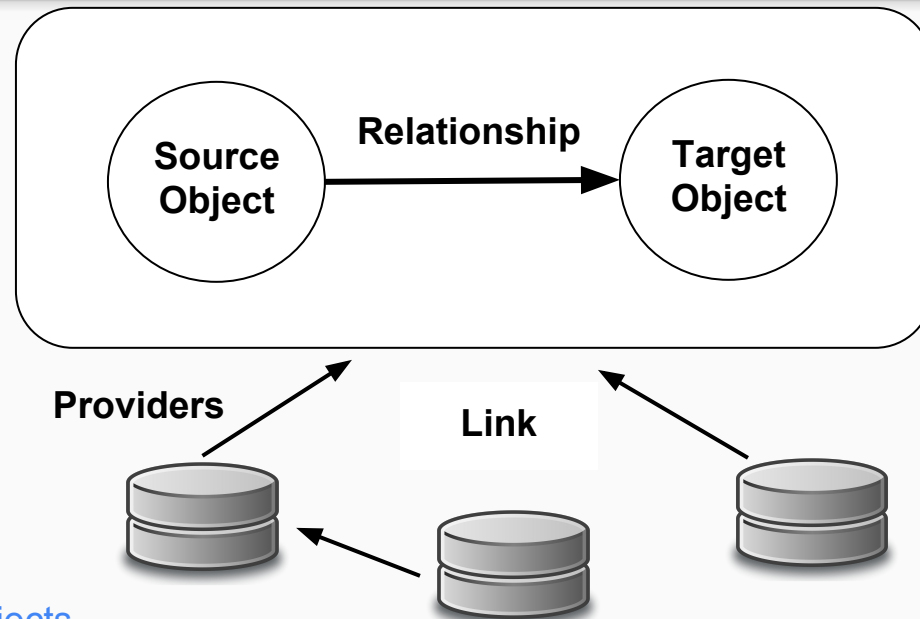
28

Cross-linking and the Scholix WG

A framework for exchanging information about the links between **scholarly literature** and **data**.



Scholix Information Model



- Information about the two objects
- Information about the nature of the link and the link package itself

Scholix Link Schema

- This schema represents the broad community consensus after two in-person workshops and lots of discussion
- The schema has intentionally very few required elements
- Focus is on the link, only required properties for source and target object are PID and type (either literature or dataset)
- Optional metadata for objects where those are not available via PID
- Small controlled vocabulary for relation type: isReferencedBy, isSupplementTo, isRelatedTo

Scholix Link Package

Link Information Package

Link Publication Date (1)

Link Provider (1..N)

Relationship Type (1)

License URL (0..1)

Source Object

Object Identifier (1)

Object Type (1)

Object Title (0..1)

Object Publisher (0..1)

Object Creator (0..N)

Object Publication Date (0..1)

Target Object

Object Identifier (1)

Object Type (1)

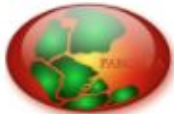
Object Title (0..1)

Object Publisher (0..1)

Object Creator (0..N)

Object Publication Date (0..1)

Figure 2 - Link Information Package: high-level information model



Scholix Schema 3.0

her | https://zenodo.org/record/1120265#_WldUVJ0dXd0

November 21, 2017

Software documentation Open Access

Scholix Metadata Schema for Exchange of Scholarly Communication Links

Burton, Adrian; Fenner, Martin; Haak, Wouter; Manghi, Paolo

Editor(s)
Brady, Catherine

Researcher(s)
Diepenbroek, Michael; Schindler, Uwe; Ariani, Amir; Bilder, Geoffrey; Ryder, Gerry; Graef, Florian; Demeranville, Torry; Bruno, Ian; Cousijn, Helena; Shaw, Phil; Wass, Joe; La Bruzzo, Sandro; Rauber, Andreas; Nielsen, Lars Holm

Sponsor(s)
World Data System/CSU

The goal of the Scholix initiative is to establish a high level interoperability framework for exchanging information about the links between scholarly literature and data. It aims to enable an open information ecosystem to understand systematically what data underpins literature and what literature references data. The DLI Service is the first exemplar aggregation and query service fed by the Scholix open information ecosystem. The Scholix framework together with the DLI aggregation are designed to enable other 3rd party services (domain-specific aggregators, integrations with other global services, discovery tools, impact assessments etc).

Scholix is an evolving lightweight set of guidelines to increase interoperability. It consists of: (i) a consensus among a growing group of publishers, datacentres, and global/domain service providers to work collaboratively and systematically to improve exchange of data-literature link information, (ii) an information model: conceptual definition of what is a Scholix scholarly link, (iii) Link metadata schema: metadata representation of a Scholix link. Options for exchange protocols (forthcoming) Scholix is the "wholesaler to wholesaler" exchange framework, to be implemented by existing hubs or global aggregators of data-literature link information such as DataCite, CrossRef, OpenAIRE, or EMBL-EBI. These hubs in turn work with their natural communities of data centres or literature publishers to collect the information through existing community-specific workflows and standards. Scholix thus enables interoperability between a smaller number of large hubs and leverages the existing exchange arrangements between those hubs and their natural communities (eg between CrossRef and journal publishers). Scholix is a technical solution to wholesale information aggregation; it will need to be complemented by other policy, practice and cultural change advocacy initiatives. This approach could be extended over time to other types of research objects in and beyond research (e.g. software, tweets, etc)

6 Tweeted by 8

Indexed in
OpenAIRE

Publication date:
November 21, 2017

DOI:
10.5281/zenodo.1120265

Keyword(s):
Publishing Data Citation Links
Scholarly communication Open science

Grants:
European Commission

- OpenAIRE2020 - Open Access Infrastructure for Research in Europe 2020 (643410)
- RDA Europe - Research Data Alliance - Europe 3 (653194)

Related identifiers:
Supplementary material:
10.5281/zenodo.1120275

データ作成者をグローバルに特定
するためのサービス

Services to Globally Identify Data
Authors





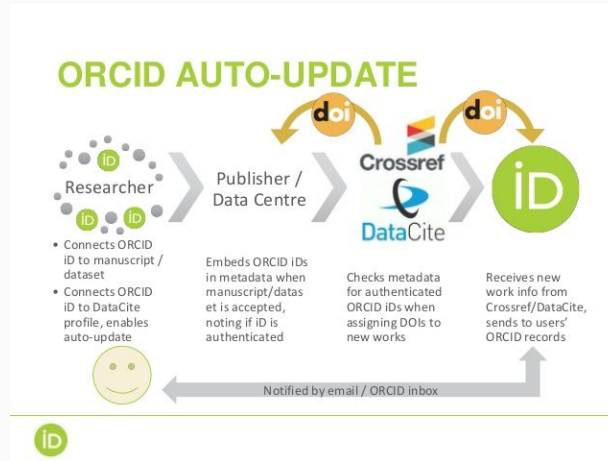
Claiming Dataset to your ORCID record from DataCite Search

The screenshot shows a web browser window with the address bar displaying "Sicher | https://search.datacite.org/data-centers/jalc.jalc?year=2015". The main content area features a search result for "Manually scaled parameters of Ionogram at Wakkanai, Japan" by Ishii Mamoru. Below the title and author, there is a brief description of the work. At the bottom of this result, there are links for "https://doi.org/10.17594/5704641f8b11d", "Cite", and "Add to ORCID record". A modal dialog box is open over the "Add to ORCID record" button, with the title "Add work to ORCID" and the question "Are you sure you want to add this work to your ORCID record?". The dialog has "Cancel" and "Ok" buttons. Below this result, another result for "Ionogram at Wakkanai, Japan" is partially visible, followed by a third result for "Automatically scaled parameters of Ionogram at Wakkanai, Japan".

Support datasets claiming in multiple workflows.



Figure 2: Metadata, in particular relations between data publication DOIs and ORCIDs, flow from PANGAEA to ORCID via DataCite.

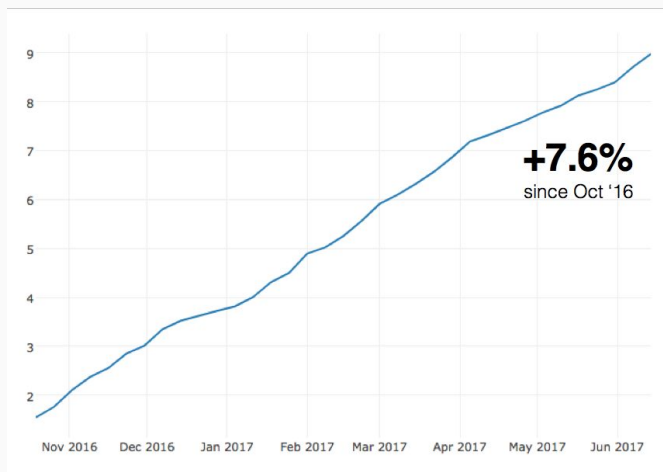


Support datasets claiming in multiple workflows.



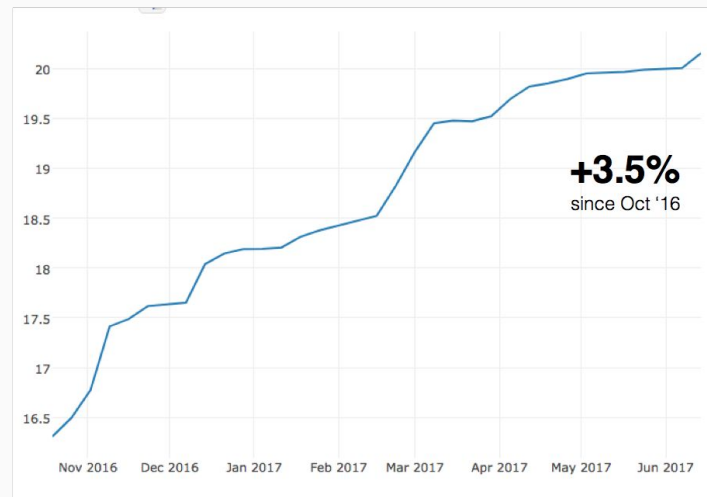
Figure 2: Metadata, in particular relations between data publication DOIs and ORCIDs, flow from PANGAEA to ORCID via DataCite.

PANGAEA users having ORCID iDs, relative to the total number of users [%]



Total Users: 7,644

Datasets with at least one author having ORCID iD, relative to the total number of datasets [%]



Total Datasets: 356,627

研究データメトリックのための ガイドライン

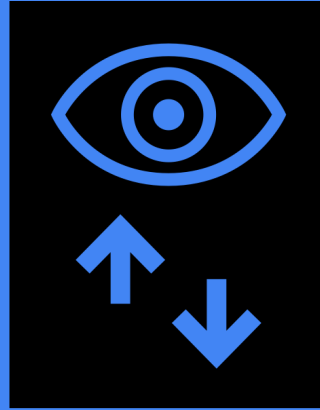
Guidelines for Research Data Metrics



There has been previous efforts on counting citations

but

- Does not include usage
- How can we do metrics?

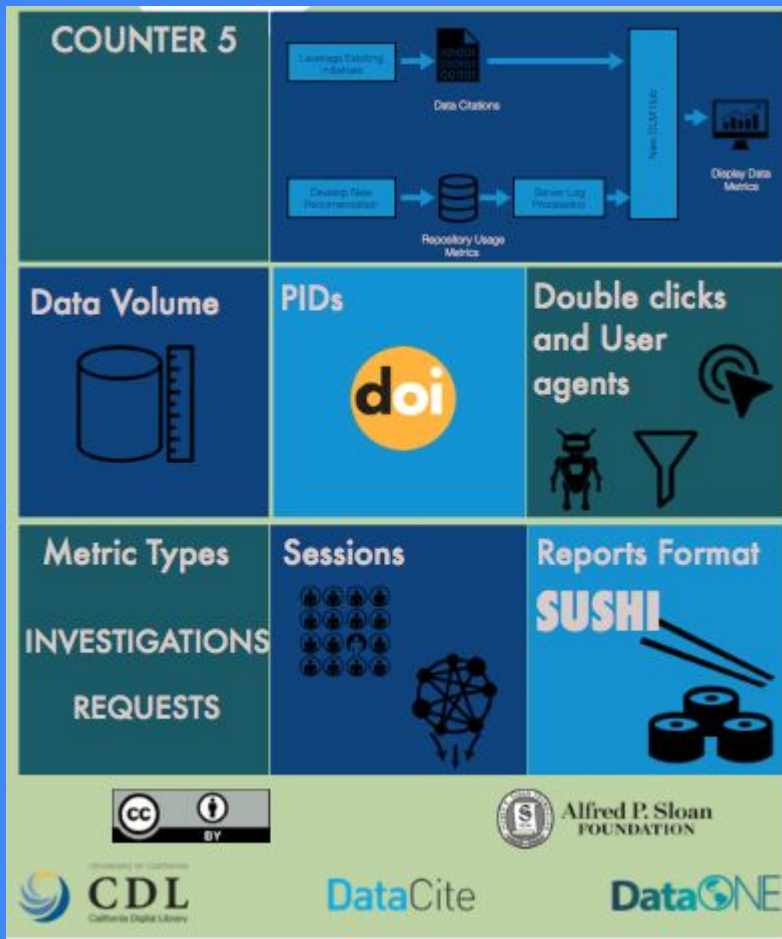


Created by AlfredoCreates.com/Icons
from Noun Project

?



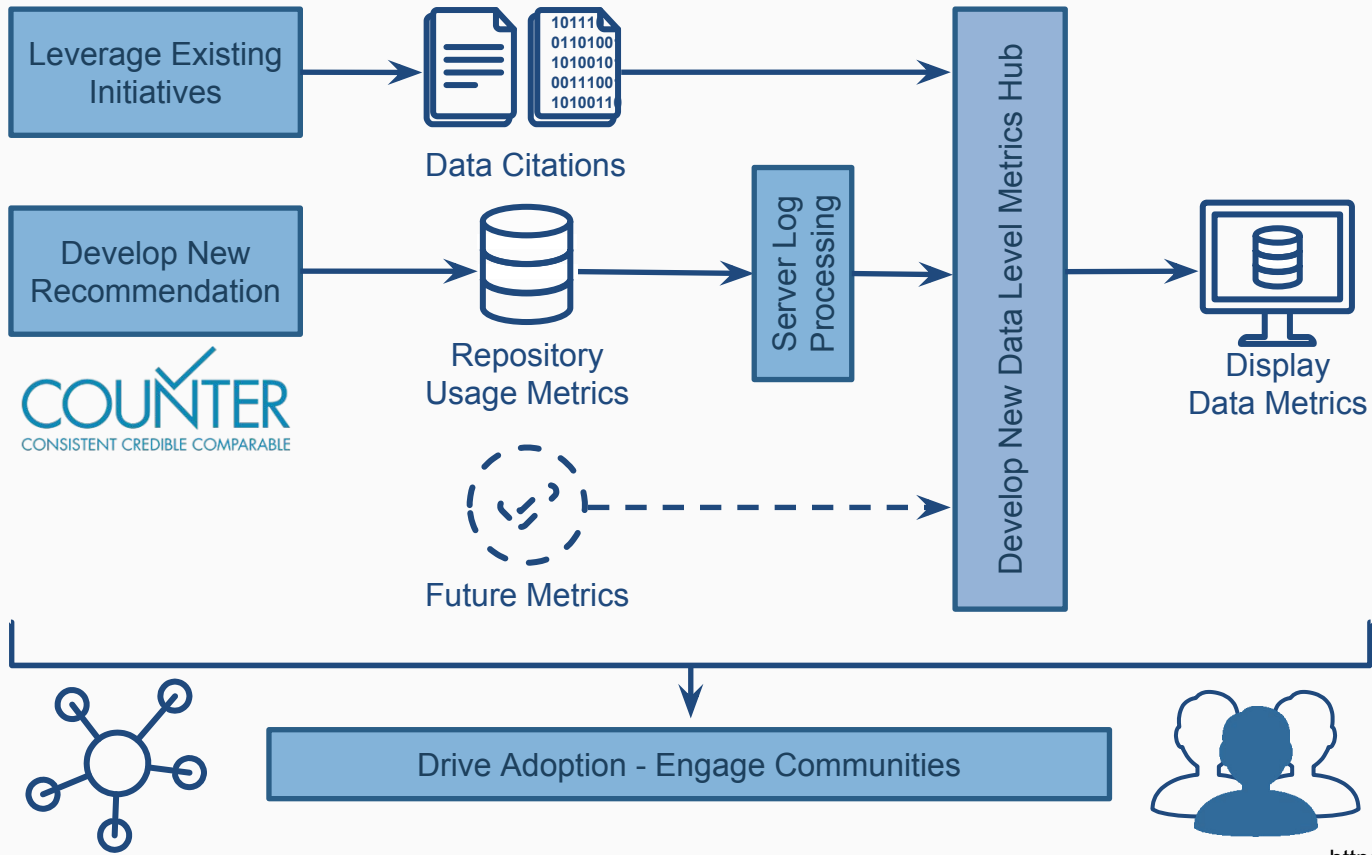
Created by SBTS
from Noun Project



Therefore...

The implementation
of a code of
practice for data
level metrics

データレベルメトリック
のための実践規約
の実装



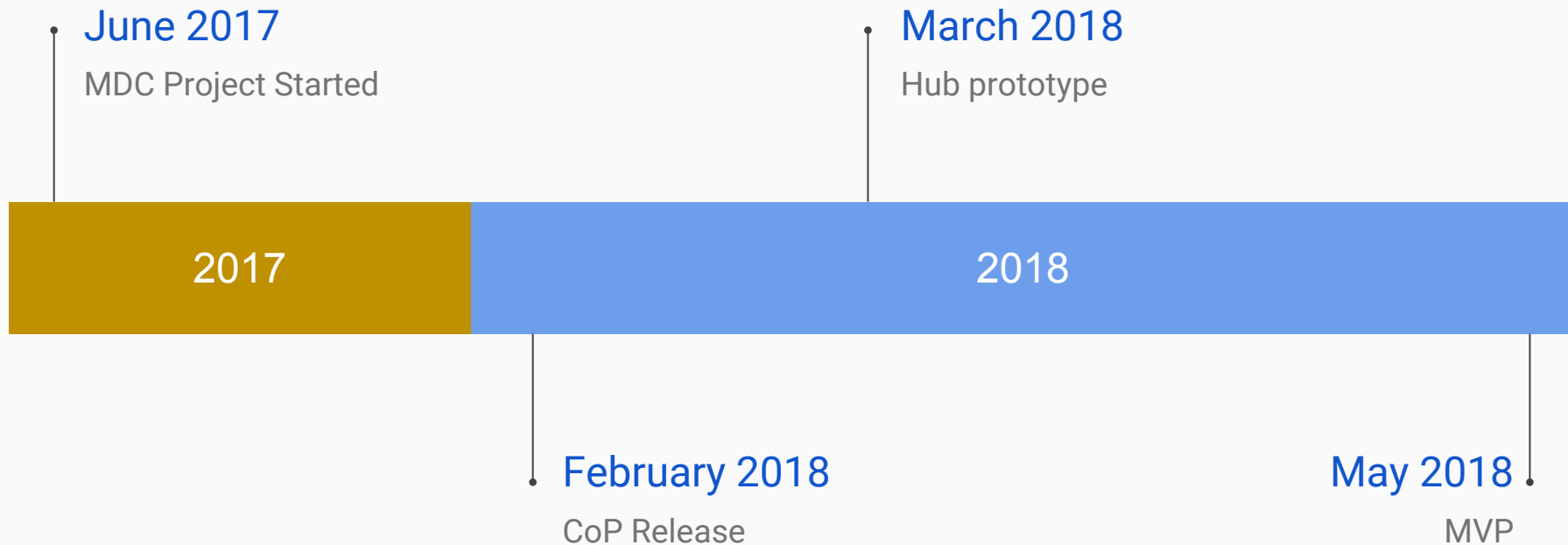
<https://makedatacount.org>

Protocol for exchanging Data Metrics

SUSHI (Standardized Usage Statistics Harvesting Initiative) **protocol** (<http://www.niso.org/workrooms/sushi/>) has greatly facilitated the handling of large volumes of usage data, and its implementation by vendors allows the automated retrieval of the COUNTER usage reports into local systems, making this process much less time consuming for the librarian or library consortium administrator.

```
},
  "Report_Datasets": [
    {
      "Dataset_Title": "Lake Erie Fish Community Data",
      "Dataset_ID": [
        {
          "Type": "DOI",
          "Value": "0931-865"
        }
      ],
      "Dataset_Contributors": [
        {
          "Type": "Name",
          "Value": "John Smith"
        }
      ],
      "Dataset_Dates": [
        {
          "Type": "Pub_Date",
          "Value": "2002-01-15"
        }
      ],
      "Dataset_Attributes": [
        {
          "Type": "Dataset_Version",
```

MDC Milestones 一里塚



Get Involved 参加方法

Repositories – contact us if you would like to be an early adopter

Publishers – support data citations & ensure your citation data and metadata are open

Institutions- contact us if you would like to support in your institutional repositories and help us spread the word

Researchers- we are looking for beta testers for our front end development of data level metrics

Funders- if you are interested in supporting this initiative, please get in touch.

Follow us on Twitter @MakeDataCount

Join our mailing list at

<https://makedatacount.org/contact/>



ご静聴ありがとうございました

Visit:
www.datacite.org

