

# DataCite Services

ODIN Webinar  
June 3<sup>rd</sup> 2014

Sebastian Peters

DataCite Tech Team Lead

German National Library of  
Science and Technology

[tech@datacite.org](mailto:tech@datacite.org) / [sebastian.peters@tib.uni-hannover.de](mailto:sebastian.peters@tib.uni-hannover.de)

# What is DataCite? What is a DOI?

# What is DataCite?

- Global consortium carried by local institutions
- focused on improving the scholarly infrastructure around datasets and other non-textual information
- focused on working with datacentres and organizations that hold data
- Providing standards, workflows and best-practice
- **Make research data citable** (data publication)
- Initially, but not exclusively based on the DOI system

<http://www.datacite.org>

# DataCite History

- Founded December 1st 2009 in London
- Forerunner (since 2004), founder member and managing agent: German National Library of Science and Technology (TIB)
- 1st DOI for a dataset registered by TIB in 2004
- Happy 10<sup>th</sup> Anniversary:  
<http://www.datacite.org/node/121>

# Facts & Figures

- 22 full members, 9 associated; from 18 countries
- ~ 3.4 Mio. DOIs
- ~ 3.2 Mio. Metadata objects
- > 300 DOI prefixes
- > 300 datacentres (publishing agents)

# DOI – Digital Object Identifier

- *Digital Identifier of an Object*
- Maintained by the International DOI Foundation (IDF)
- Based on Handle System
- Actionable, interoperable, persistent
- ISO standard 26324:2012
- 9 (?) registration agencies (e.g. CrossRef)
- ~ 100 Mio. DOIs registered
- Example: doi:10.5524/100001, <http://doi.org/10.5524/100001>
- Resolves to Landing Pages

<http://www.doi.org>

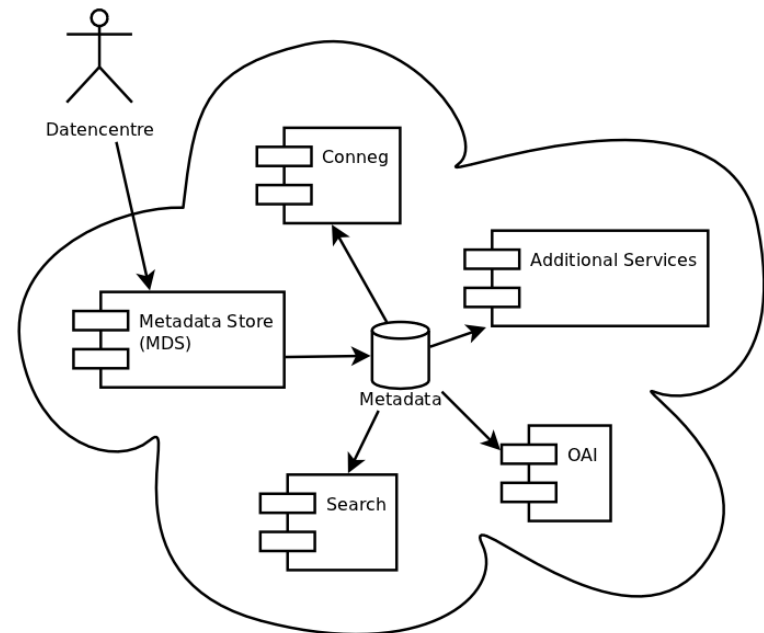
# DataCite Specific Services

**TIB** | GERMAN NATIONAL LIBRARY OF  
SCIENCE AND TECHNOLOGY



# DataCite Services

- Metadata Store
- Search
- Metadata Export (OAI)
- Stats
- Content Negotiation
- Citation Formatter



- Open Source: Code and Tickets on github

<http://github.com/datacite>



# Metadata Store

MDS is DataCite's central infrastructure for DOI Management:

- Minting/Updating of DOIs
- Storing of metadata

Interfaces:

- User interface (UI)
- Programmatic interface (API)

MDS is only accessible by DataCite members (and their datacentres):

<https://mds.datacite.org>

# DataCite Metadata

- XML
- DataCite Metadata Schema:  
<http://schema.datacite.org>
- Focused on citation elements (title, creator, publicationYear, etc.)
- Relations (isSupplementTo, isNewerVersionOf, etc.)
- geolocation

# DataCite Metadata Generator and ORCID

- Metadata formular with ORCID lookup
- Result from the last code sprint at CERN
- <http://www.datacite.org/node/102>
- <https://github.com/koelnconcert/datacite-metadata-gene>

## DataCite Metadata Generator - Kernel 3.0

### Mandatory Elements

DOI:  
[DOI]

Title(s): +  
[TITLE]  
[titleType] ▾

Creator(s): +  
jan brase  
Search ORCID | ORCID 0000-0002-8250-6253  
Lookup ORCID

ORCID Identifier  
0000-0002-8250-6253  
Biography

### Metadata

```
<?xml version='1.0' encoding='UTF-8'?>
<resource xmlns='http://datacite.org/schema/k
xsi:schemaLocation='http://datacite.org/schem
<identifier identifierType='DOI'></identifier>
<creators>
  <creator>
    <creatorName>jan brase</creat
    <nameIdentifier nameIdentifie
  </creator>
</creators>
<titles>
</titles>
<publisher></publisher>
<publicationYear></publicationYear>
</resource>
```

# Metadata Search

[Options](#) | [Advanced Search](#) | [About Us](#) | [Contact](#) | [Help](#)



## Metadata Search beta

# Metadata Search

- Open service to search metadata stored in the DataCite Metadata Store (MDS)
- Based on Lucene Solr
- User Interface
  - Complex boolean query language
  - Advanced Search form
  - Facets (Drilldown)
- API

<http://search.datacite.org>

# Build your own API query

- Links at the bottom of each search result
- Adjust parameters to your needs
- complex lucene query syntax, e.g.
  - `title:laser OR subject:laser`
  - `publicationYear:[1990 TO 1995]`
  - `relatedIdentifier:  
isSupplementTo\ :doi\ :*`

# Interesting Queries

- `relatedIdentifier:*github.com*`
- `nameIdentifier:ORCID*`
- `alternateIdentifier:URN\:*`

# Special Facet Search APIs

- <http://search.datacite.org/list/prefixes>
- <http://search.datacite.org/list/allocators>
- <http://search.datacite.org/list/datacentres>

Works with filters as well:

- <http://search.datacite.org/list/datacentres?fq=resourceTypeGeneral:Audiovisual>



# Metadata Export (OAI)

- Service for 3rd parties to harvest metadata stored in the DataCite Metadata Store (MDS)
- using the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH).
- Metadata formats so far: Dublin Core, DataCite Metadata Schema

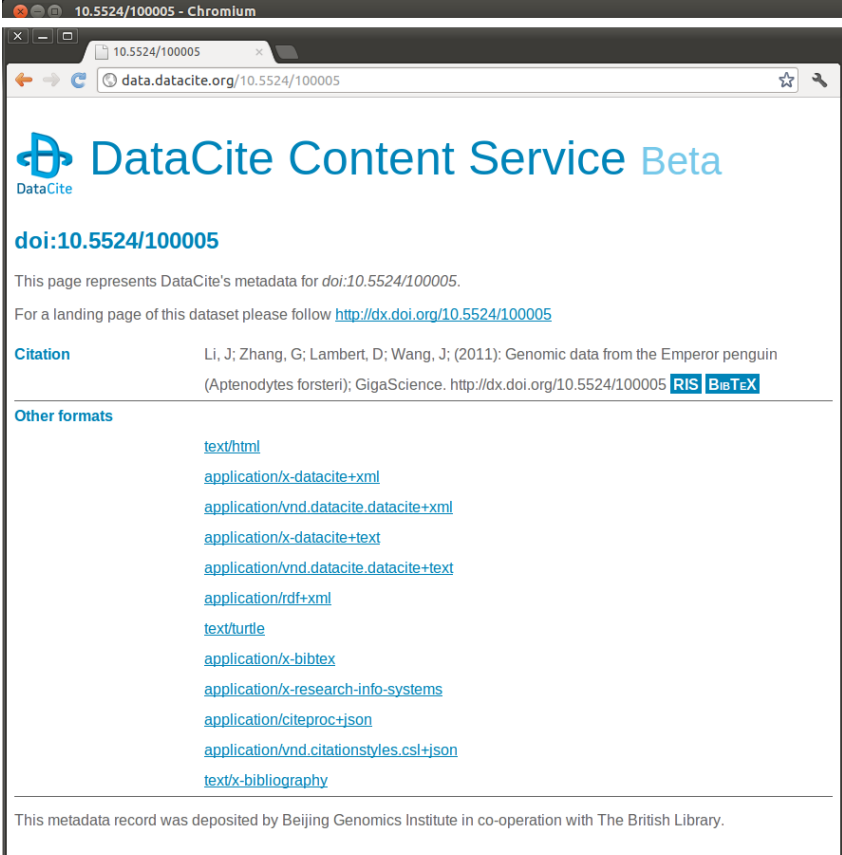
<http://oai.datacite.org>

# OAI Sets

- Sets for allocators and datacentres, e.g.
  - set=TIB.PANGAEA
- Arbitrary queries: encoded (base64url) solr query string as set name, e.g.
  - set=~cT1sYXNlciZmcT1yZXNvdXJjZVR5cGVHZW5lcmFsOmRhdGFzZXQK
  - Decodes to: q=laser&fq=resourceTypeGeneral:dataset

# Content Resolver


- Getting particular representations of a DOI's metadata
  - HTML Links
  - HTTP Content Negotiation
- <http://data.datacite.org>



10.5524/100005 - Chromium

10.5524/100005



data.datacite.org/10.5524/100005

 DataCite Content Service Beta

**doi:10.5524/100005**

This page represents DataCite's metadata for doi:10.5524/100005.

For a landing page of this dataset please follow <http://dx.doi.org/10.5524/100005>

**Citation** Li, J; Zhang, G; Lambert, D; Wang, J; (2011): Genomic data from the Emperor penguin (Aptenodytes forsteri); GigaScience. <http://dx.doi.org/10.5524/100005>  

**Other formats**

- [text/html](#)
- [application/x-datacite+xml](#)
- [application/vnd.datacite.datacite+xml](#)
- [application/x-datacite+text](#)
- [application/vnd.datacite.datacite+text](#)
- [application/rdf+xml](#)
- [text/turtle](#)
- [application/x-bibtex](#)
- [application/x-research-info-systems](#)
- [application/citeproc+json](#)
- [application/vnd.citationstyles.csl+json](#)
- [text/x-bibliography](#)

This metadata record was deposited by Beijing Genomics Institute in co-operation with The British Library.

# CrossCite Services

**TIB** | GERMAN NATIONAL LIBRARY OF  
SCIENCE AND TECHNOLOGY



# Content Negotiation – CrossRef & DataCite

- HTTP Content Negotiation with dx.doi.org proxy
- You don't need to know the actual DOI RA
- Common Formats of CrossRef & DataCite:
  - RDF, RIS, BibTeX, Citeproc, ...
- <http://crosscite.org/cn/>
- Example:
  - `curl -LH "Accept:application/rdf+xml"`  
`http://doi.org/10.5524/100001`

# Citation Formatter

- Using Content Negotiation
- More than 500 styles (<http://citationstyles.org>)
- <http://crosscite.org/citeproc/>
- Source code and API doc:  
<https://github.com/crosscite/citeproc-doi-server>



DOI:

Grobe, Hannes (2005) Data policy of the information system PANGAEA. doi:  
10.1594/PANGAEA.327791



# Summary

# Summary

- DOI & Metadata Management
  - Metadata Store, DataCite Schema
- Search
  - Powerful API & UI
- Harvesting
  - OAI-PMH, Dublin Core, DataCite Schema
- Getting Single Metadata
  - Content Negotiation ([dx.doi.org](http://dx.doi.org)), various formats, CrossCite, Citation Formatter



Thanks for your attention!