GROUP ON EARTH OBSERVATIONS

Data Citation: DOI-Enabling GEOSS Discovery and Access

4th GEOSS Science and Technology Stakeholder Workshop
March 27, 2015 -- Norfolk, Virginia

David K Arctur, University of Texas at Austin
Joan Starr, California Digital Library
Stefano Nativi, Enrico Boldrini, and **Mattia Santoro**,
ESSI Lab, Italy Center for National Research
Robert Arko, Lamont-Doherty Earth Observatory

GROUP ON EARTH OBSERVATIONS

Abstract

Data citation is enabling much more than just discovery of published resources; it is enabling direct linking of data in a way that enables its use in technical papers, websites, presentations, and other data sets. Furthermore, it represents a contract for permanent availability and access to the referenced data.

From Wikipedia: "A DOI name differs from standard identifiers such as the ISBN. The purpose of an identifier registry is to manage a given collection of identifiers, whereas the primary purpose of the DOI system is to make a collection of identifiers actionable and interoperable."

There is considerable current discussion on roles and conventions for using DOIs with Earth and space science data. An important aspect of this is the metadata schema used with DOIs. This schema, based on the indecs Content Model, has some overlap with the ISO 19115 content model. DataCite, a consortium of leading research libraries and scientific data centers, has a registry of over 5 million DOIs and their metadata. The DataCite DOI metadata schema has added new features that support the geoscience community, such as a GeoLocation element and the ability to supply extra disciplinespecific metadata. Numerous geoscience data providers routinely publish DOIs using the DataCite schema. By mapping the DOI metadata to ISO 19115, and harvesting DataCite's DOI registry, the GEO Discovery and Access Broker (DAB) can add an immensely valuable resource to its alreadysignificant distributed catalogs.

This presentation reviews the current discussion of DOI usage in geoscience research, and the role that GEOSS can now play in connecting this with a broader community.

GROUP ON EARTH OBSERVATIONS

Data Citation: DOI-Enabling GEOSS

Outline

- Introduction to Digital Object Identifiers and DataCite
  (Joan Starr)
- Initial registration of DataCite.org's DOI registry with GEO Discovery and Access Broker (DAB)
  (Stefano Nativi, Enrico Boldrini, Mattia Santoro)
- Questions raised, next steps
  (All)

DataCite Metadata and ISO 19115

Joan Starr
California Digital Library

University of California
CDL
California Digital Library

DataCite
FIND, ACCESS, AND REUSE DATA

What is a DOI?

Springer Link

Search

Home • Contact Us

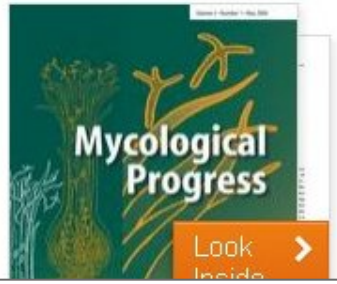
Access old SpringerLink

» Download PDF (518 KB) » View Article

Mycological Progress
August 2012, Volume 11, Issue 3, pp 827-833

Lepidostroma vilgalysii, a new basidiolichen from the New World

Brendan P. Hodkinson, Jessie K. Uehling, Matthew E. Smith



Hodkinson BP, Lendemer JC, Esslinger TL (2010) *Parmelia barrenoae*, a macrolichen new to North America and Africa. North American Fungi 5(3):1–5

Hodkinson BP, Uehling JK, Smith ME (2011) Data from: *Lepidostroma vilgalysii*, a new basidiolichen from the New World. Dryad Digital Repository. doi:10.5061/dryad.j1g5dh23

Honegger R (1996) Mycobionts. In: Nash TH III (ed) The biology of lichens. Cambridge University Press, Cambridge, pp 25–36

from the nuclear ribosomal LSU locus are used to confirm the placement of the holotype in *Lepidostroma* and to evaluate the molecular distinctiveness of the species from all other described species in the family and genus.

Within this Article

DataCite
FIND, ACCESS, AND REUSE DATA

What is a DOI?

What you see: alphanumeric string (never changes)

Associated with: location of object (such as a URL)

And: who, what, when, etc (i.e. metadata)

Hodkinson BP, Lendemer JC, Esslinger TL (2010) *Parmelia barrenoae*, a macrolichen new to North America and Africa. North American Fungi 5(3):1–5

Hodkinson BP, Uehling JK, Smith ME (2011) Data from: *Lepidostroma vilgalysii*, a new basidiolichen from the New World. Dryad Digital Repository. doi:10.5061/dryad.j1g5dh23

Honegger R (1996) Mycobionts. In: Nash TH III (ed) The biology of lichens. Cambridge University Press, Cambridge, pp 25–36

DataCite
FIND, ACCESS, AND REUSE DATA

DOI example

string: **doi:10.9999/FK40K2GTV**

html version: http://dx.doi.org/10.9999/FK40K2GTV

location: http://www.bologna.edu/biology/xfg/123.xls

metadata

creator: Dr. Felix Kottor

title: Data for chromosomal study of catfish (Ictalurus punctatus)

publisher: University of Bologna

date: 8/31/2012

DataCite
FIND, ACCESS, AND REUSE DATA

DOI example

string: **doi:10.9999/FK40K2GTV**

html version: http://dx.doi.org/10.9999/FK40K2GTV

location: http://www.state.edu/ecology/783sdr/123.xls

metadata

creator: Dr. Felix Kottor

title: Data for chromosomal study of catfish (Ictalurus punctatus)

publisher: Dryad Data Repository

date: 10/01/2013

DataCite
FIND, ACCESS, AND REUSE DATA

Why are DOIs important?

The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the httpd.apache.org home page, and then look for links to the information you want.
- Click the Back button to try another link.
- Click Search to look for information on the Internet.

HTTP 404 - File not found
Internet Explorer

DataCite
FIND, ACCESS, AND REUSE DATA

Why are DOIs important?

Example:
Sidlauskas, B. 2007. Data from: Testing for unequal rates of morphological diversification in the absence of a detailed phylogeny: a case study from characiform fishes. Dryad Digital Repository. doi:10.5061/dryad.20

Allow readers to find data products

Get credit for data **and** publications

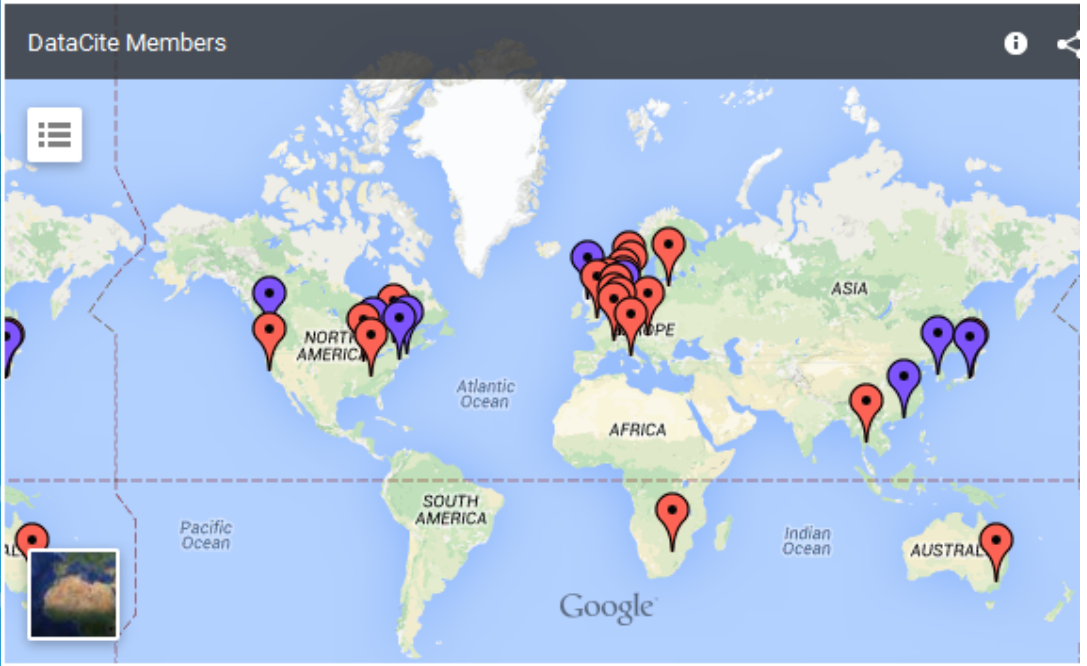
Promote reproducibility

Better measure of research impact

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite

Creating a global citation framework for data



View DataCite Members in a larger map

Key: DataCite Member | DataCite Affiliated Member

DataCite Services

1. DOIs for data!
2. Local service & support
3. Usage stats
4. Citation formatter
5. Content negotiation
6. Metadata search
7. OAI provider
8. DataCite-to-ORCID hookup

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite Metadata

- Mandatory elements support data citation
- Recommended elements support indexing, discovery
- Optional elements support additional description, access, management, etc.

- Give your feedback, questions, input here:

bit.ly/1I6F3Im

DataCite
FIND, ACCESS, AND REUSE DATA

For more information

- DataCite metadata scheme
- DataCite metadata search
- All about DataCite

DataCite
FIND, ACCESS, AND REUSE DATA

GROUP ON EARTH OBSERVATIONS

Data Citation: DOI-Enabling GEOSS

Outline

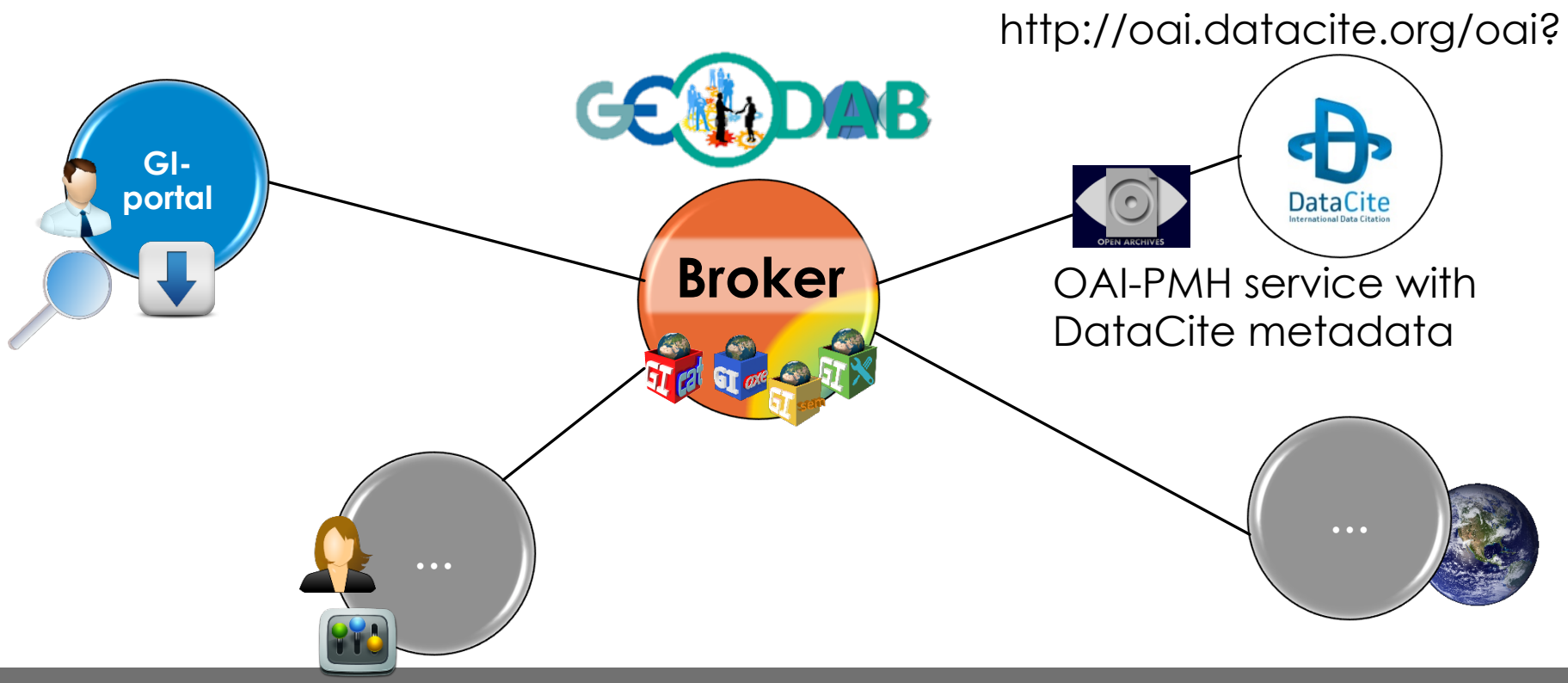
- Introduction to Digital Object Identifiers and DataCite
  (Joan Starr)
- Initial registration of DataCite.org's DOI registry with GEO Discovery and Access Broker (DAB)
  (Enrico Boldrini, Mattia Santoro, Stefano Nativi)
- Questions raised, next steps
  (All)

DataCite Archive Interoperability

Enrico Boldrini, Mattia Santoro and Stefano Nativi

CNR-IIA, UOS di Firenze

Brokering DataCite archive



DataCite results from the GI-portal

GEOSS
EuroGEOSS
A EUROPEAN APPROACH TO GEOSS

GI-CAT BROKER

DataCite TIB.WDCC

- Partial harvest of DataCite catalog (2588 records)
- Full catalog: > 5 million records

Start search

Map

Log In

Close Tab

Search results: 2568 - Elapsed time: 1 minutes and 19 seconds

	CCSM4 model run for CMIP5 future projection (2006-2300) forced by Representative Concentration Pathway 8.5, served by ESGF
	CCSM4 model run for CMIP5 future projection (2006-2300) forced by Representative Concentration Pathway 4.5, served by ESGF
	NOAA GFDL GFDL-HIRAM-C180, amip experiment output for CMIP5 AR5, served by ESGF
	NOAA GFDL GFDL-HIRAM-C180, sst2030 experiment output for CMIP5 AR5, served by ESGF
	NASA-GISS: GISS-E2-H model output prepared for CMIP5 historicalGHG, served by ESGF
	NASA-GISS: GISS-E2-H model output prepared for CMIP5 historicalMisc, served by ESGF
	NASA-GISS: GISS-E2-H model output prepared for CMIP5 historicalNat, served by ESGF

<< first < prev 248 249 250 251 252 253 254 255 256 257 next > last >>

Query by text « cmip5 »

Search results - All: 644 | GEOSS Category: Websites and Documents | GEOSS Category: Monitoring and Observation Systems

<< first < prev 1 2 3 4 5 6 7 8 9 10 next > last >>

Access/Use Constraints			Title
			CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-LR model: The amipFuture experiment, served by ESGF
			CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-MR model: The amipFuture experiment, served by ESGF
			cmip5 output1 NCC NorESM1-M sstClim4xCO2, served by ESGF
			NOAA GFDL GFDL-ESM2G, esmrcp85 experiment output for CMIP5 AR5, served by ESGF
			NOAA GFDL GFDL-ESM2G, rcp45 experiment output for CMIP5 AR5, served by ESGF
			NOAA GFDL GFDL-ESM2G, esmControl experiment output for CMIP5 AR5, served by ESGF

DataCite metadata (ISO 19115)

NOAA GFDL GFDL-ESM2G, esmrcp85 experiment output for CMIP5 AR5, served by ESGF

ISO 19115 Overview (see raw metadata)

Powered by GEOSS EuroGEOSS

Record	
File identifier	oai:oai.datacite.org:4257169
Hierarchy level	Digital
Date stamp	2014-04-03

Identification Information	
Title	NOAA GFDL GFDL-ESM2G, esmrcp85 experiment output for CMIP5 AR5, served by ESGF

ESSI Lab Florence Division

GEOSS Search for « esm2g AND esmrcp85 »

GROUP ON EARTH OBSERVATIONS | GEOSS Portal

Discover, Access, Contribute
Earth Observations, Information and Services

HOME

cmip5 output1 NOAA-GFDL GFDL-ESM2G esmrcp85

	Record
File identifier	de.dkrz.wdcc.iso3205549
Hierarchy level	series

	Identification Information
Title	cmip5 output1 NOAA-GFDL GFDL-ESM2G esmrcp85
Creation date	2014-04-03

How can we know these are the same datasets?

DataCite

FIND, ACCESS, AND REUSE DATA

GEOSS Portal

Discover, Access, Contribute
Earth Observations, Information and Services

GROUP ON EARTH OBSERVATIONS

Data Citation: DOI-Enabling GEOSS

Outline

- Introduction to Digital Object Identifiers and DataCite
  (Joan Starr)
- Initial registration of DataCite.org's DOI registry with GEO Discovery and Access Broker (DAB)
  (Stefano Nativi, Enrico Boldrini, Mattia Santoro)
- Questions raised, next steps
  (All)

GROUP ON EARTH OBSERVATIONS

Questions raised, next steps…

- Key benefit of supporting DOI within GEOSS:
  - ***Enabling permanent, unique, resolvable links to data within documents***
- Questions about metadata mapping between DataCite and ISO 19115
  - ***Handling of fields optional for DOI but required for ISO 19115 by INSPIRE***
- Possible next steps

DataCite to ISO 19115 Mapping

DataCite Requirements	ISO 19115 support
Identifier	identificationInfo[1]/*/citation/*/identifier[1]
Creator	identificationInfo[1]/*/pointOfContact, role=originator
Title	identificationInfo[1]/*/citation/*/title
Publisher	identificationInfo[1]/*/pointOfContact, role=publisher
PublicationYear	identificationInfo[1]/*/citation/*/date, dateType=publication

EXAMPLE: Creator (PublicationYear): Title. Publisher. Identifier

A. H. Allen (1916): Juvenile Leucosticte on rocks. Museum of Vertebrate Zoology. http://dx.doi.org/10.7299/X7PV6HQ0

DataCite
FIND, ACCESS, AND REUSE DATA

Mapping of DataCite Mandatory Properties

DataCite ID	DataCite Property	ISO 19115 Property (XPath)
1	Identifier	identificationInfo[1]/*/citation/*/identifier[1]
2	Creator	identificationInfo[1]/*/pointOfContact, role=originator
3	Title	identificationInfo[1]/*/citation/*/title
4	Publisher	identificationInfo[1]/*/pointOfContact, role=publisher
5	Publication Year	identificationInfo[1]/*/citation/*/date, dateType=publication

Mapping of DataCite Recommended and Optional Properties

DataCite ID	DataCite Property	ISO 19115 Property (XPath)
6	Subject	identificationInfo[1]/*/ descriptiveKeywords/*/
7	Contributor	identificationInfo[1]/*/pointOfContact, various roles
8	Date	identificationInfo[1]/*/citation/*/date, various date types
9	Language	identificationInfo[1]/*/language
10	ResourceType	hierarchyLevel
11	AlternateIdentifier	identificationInfo[1]/*/citation/*/ identifier[2]
12	RelatedIdentifier	aggregateDataSetIdentifier

Mapping of DataCite Recommended and Optional Properties

DataCite ID	DataCite Property	ISO 19115 Property (XPath)
13	Size	distributionInfo[1]/*/transferOptions/*/transferSize
14	Format	distributionInfo[1]/*/distributionFormat/*/name
15	Version	distributionInfo[1]/*/distributionFormat/*/version
16	Rights	identificationInfo[1]/*/resourceConstraints/ otherConstraints
17	Description	identificationInfo[1]/*/abstract
18	GeoLocation	identificationInfo[1]/*/extent/*/geographicElement

Mapping DataCite to ISO 19115 & INSPIRE metadata profile

- Almost all (18) DataCite elements can be easily mapped to ISO 19115
- However ISO 19115 (and especially INSPIRE) require additional mandatory elements that are **missing** or **optional** in DataCite. E.g.
  - Metadata fileIdentifier
  - Metadata date stamp
  - Organisation email
  - Geographic bounding box
  - Metadata language
  - Lineage
  - Abstract
  - GEMET keywords
  - Conformity to INSPIRE
  - Conditions/limitations on access and use

Suggested ISO 19115 to DataCite Mapping

ISO 19115 Requirement	DataCite support
Metadata file identifier	use relatedIdentifer, relationType=HasMetadata to point to associated, additional metadata
Metadata date stamp	access via OAI-PMH
Organization email	use nameIdentifers (ORCID, FundRef, ISNI, etc.) to point to additional information for creators and contributors
Geographic bounding box	use GeoLocation and geoLocationBox
Metadata language	along with HasMetadata, use relatedMetadataScheme, schemeType and schemeURI
Lineage	use relatedIdentifier with various relationTypes to describe provenance (isSubsetOf, isPartOf, isNewVersionOf, etc.)
Abstract	use Description; descriptionType=Abstract
GEMET keywords	use Subject; subjectScheme=GEMET
Conformity to INSPIRE	use relatedIdentifer, relationType=HasMetadata to point to associated, additional metadata
Conditions on access/use	use Rights, which is repeating

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite recommended approach

- Describe relationships between research objects.
- Use name and organizational identifiers whenever available.
- Take advantage of DataCite's OAI-PMH service to retrieve GEOSS metadata in desired format (e.g. RDF).

DataCite
FIND, ACCESS, AND REUSE DATA

GROUP ON EARTH OBSERVATIONS

Next steps…

- How will GEO architect support for DOI linking and other uses within GEOSS?
  - DOI as a search qualifier
  - Generate/assign DOIs for datasets with no other ID?
- How should GEO work with DataCite to:
  - Ensure that the design of DOI-integration in GEOSS is consistent with community practice
  - Consider GEOSS-INSPIRE recommended profile for DOI metadata
- Coordination through DataCite membership? Research Data Alliance? Force11? EarthCube? …?

GROUP ON
EARTH OBSERVATIONS

Thank you

David Arctur
david.arctur AT utexas.edu