DataCite and DOIs

Presentation at CLARIN workshop, Soesterberg, October 23rd, 2014

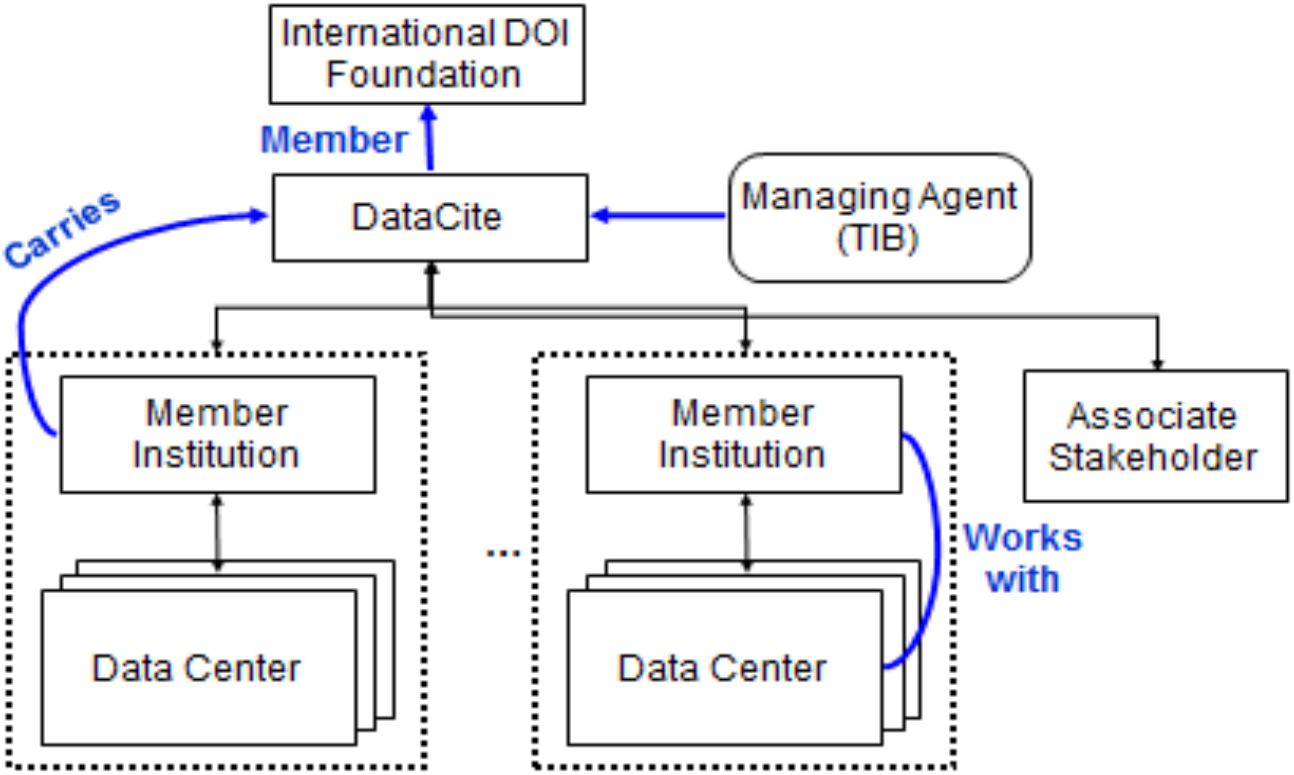
Madeleine de Smaele - m.m.e.desmaele@tudelft.nl

DataCite

- Global consortium carried by local institutions
- Focused on improving the scholarly infrastructure around datasets and other non-textual information
- Focused on working with data centres and organisations that hold data
- Providing standards, workflows, best practice
- Founded Dec. 1st 2009 in London

http://www.datacite.org/

DataCite structure



Responsibilities

The DataCite registration agency:

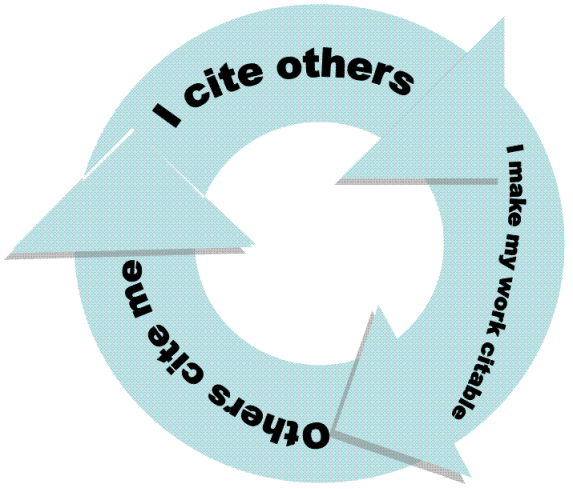
- Maintains the resolution infrastructure
- Maintains a searchable database of metadata
- Manages the identifiers over the long term
- Establishes and shares best practice

Publishing agents (data centres, research institutions):

- Content storage and access
- Quality assurance
- Creating the identifier
- Creating and updating metadata

Why DOIs ?

- Easy and permanent access to research data via the internet
- enhanced discovery, retrieval and management of data to enable data reuse and verification of research results



What is a DOI ?

- *What you see*: alphanumeric string (never changes)
- *Associated with*: location (such as URL)
- *Accompanied with*: who, what, when… (metadata)

Example DOI

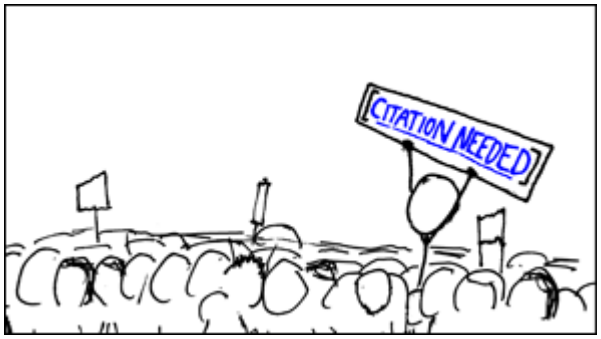
- String: 10.4121/uuid:57acdc8d-5c86-478a-9ada-8c075cc30b0a
- Html version: http://dx.doi.org/10.4121/uuid:57acdc8d-5c86-478a-9ada-8c075cc30b0a
- Location: http://data.3tu.nl/repository/uuid:57acdc8d-5c86-478a-9ada-8c075cc30b0a
- Metadata
  - Creator: Westhoff, M.C.(Martijn)
  - Title: High resolution temperature observations to identify different runoff processes
  - Publisher: TU Delft
  - Publication year: 2011

DataCite and metadata

- Metadata make data discoverable.
- Long-term maintenance of metadata is an important part of the persistence of an identifier.
- Schema is inspired by Dublin Core.
- Core value of the DataCite Metadata Schema: Linking between data and related objects.
- Future vision: Links between all related publications and objects

Benefits of data citation

- Including citable data in related publications increases the citation rate of those publications
- Routine citation of data will assist in gaining acknowledgement of data as a first class research output
- Citations for published data can be included in CVs along with journal articles, reports and conference papers
- Only cited data can be counted and tracked (in a similar manner to journal articles) to measure impact



Data citation (example 1)

PANGAEA®
Data Publisher for Earth & Environmental Science

Data Description

Citation: Bartzke, G et al. (2013): Hydraulic conductivity and experiments of sediment beds. doi:10.1594/PANGAEA.821648,
Supplement to: **Bartzke, Gerhard; Bryan, Karin R; Pilditch, Conrad A; Huhn, Katrin (2013):** On the stabilizing influence of silt on sand beds. *Journal of Sedimentary Research*, **83(8)**, 691-703, doi:10.2110/jsr.2013.57

Abstract: In marine environments, sediments from different sources are stirred and dispersed, generating beds that are composed of mixed and layered sediments of differing grain sizes. Traditional engineering formulations used to predict erosion thresholds are however, generally for unimodal sediment distributions, and so may be inadequate for commonly occurring coastal sediments. We tested the transport behavior of deposited and mixed sediment beds consisting of a simplified two-grain fraction (silt (D50 = 55 µm) and sand (D50 = 300 µm)) in a laboratory-based annular flume with the objective of investigating the parameters controlling the stability of a sediment bed. To mimic recent deposition of particles following large storm events and the longer-term result of the incorporation of fines in coarse sediment, we designed two suites of experiments: (1) "the layering experiment": in which a sandy bed was covered by a thin layer of silt of varying thickness (0.2 - 3 mm; 0.5 - 3.7 wt %, dry weight in a layer 10 cm deep); and (2) "the mixing experiment" where the bed was composed of sand homogeneously mixed with small amounts of silt (0.07 - 0.7 wt %, dry weight). To initiate erosion and to detect a possible stabilizing effect in both settings, we increased the flow speeds in increments up to 0.30 m/s. Results showed that the sediment bed (or the underlying sand bed in the case of the layering experiment) stabilized with increasing silt composition. The increasing sediment stability was defined by a shift of the initial threshold conditions towards higher flow speeds, combined with, in the case of the mixed bed, decreasing erosion rates. Our results show that even extremely low concentrations of silt play a stabilizing role (1.4% silt (wt %) on a layered sediment bed of 10 cm thickness). In the case of a mixed sediment bed, 0.18% silt (wt %, in a sample of 10 cm depth) stabilized the bed. Both cases show that the depositional history of the sediment fractions can change the erosion characteristics of the seabed. These observations are summarized in a conceptual model that suggests that, in addition to the effect on surface roughness, silt stabilizes the sand bed by pore-space plugging and reducing the inflow in th bed, and hence increases the bed stability. Measurements of hydraulic conductivity on similar bed assemblages qualitatively supported this conclusion by showing that silt could decrease the permeability by up to 22% in the case of a layered bed and by up to 70% in the case of a mixed bed.

Project(s): **Center for Marine Environmental Sciences** (MARUM)

Coverage: *Median Latitude:* -37.353397 * *Median Longitude:* 176.009699 * *South-bound Latitude:* -37.695272 * *West-bound Longitude:* 175.866119 * *North-bound Latitude:* -37.011522 * *East-bound Longitude:* 176.15
Date/Time Start: 2010-04-20T15:46:38 * *Date/Time End:* 2011-04-30T18:55:21

Size: 11 datasets

Data citation (example 2)

Journal of Petroleum Science and Engineering
Volume 105, May 2013, Pages 76–83

A data-constrained modelling approach to sandstone microstructure characterisation

Y.S. Yang[a], K.Y. Liu[b, c], S. Mayo[a], A. Tulloh[a], M.B. Clennell[b], T.Q. Xiao[d]
[a] CSIRO Materials Science & Engineering, Private Bag 33, Clayton South, Victoria 3169, Australia
[b] CSIRO Earth Science and Resource Engineering, PO Box 1130, Bentley, WA 6102, Australia
[c] Research Institute of Petroleum Exploration and Development, PetroChina., Beijing, 100083, PR China
[d] Shanghai Synchrotron Radiation Facility, Shanghai Institute of Applied Physics, 239 Zhangheng Road, Shanghai 201204, PR China

Abstract

This paper outlines the data-constrained microstructure modelling (DCM) approach to determine microscopic distributions of pores (voids), quartz and ca... two computed-tomography (CT) data sets as constraint... 45 keV. The DCM predicted 3D microstructure is consiste... sandstone as aggregates comparable to the size of the ...

References

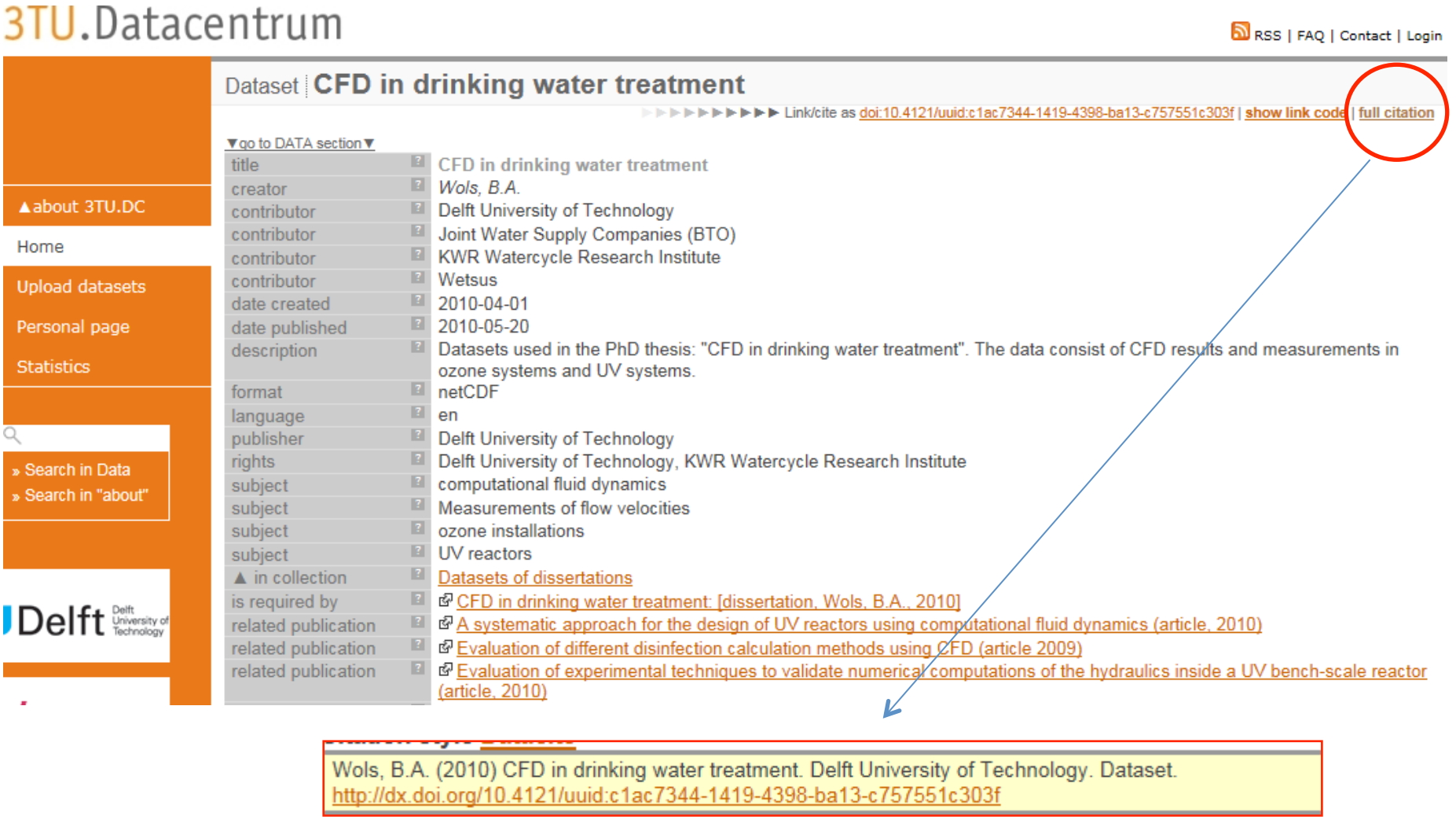
Arns et al., 2005 C.H. Arns, F. Bauget, A. Ghous, A. Sa... Sok, W.V. Pinczewski, J.C. Kelly, M.A. Knackstedt
Digital core laboratory: petrophysical analysis from 3D imaging of reservoir core fragments
Petrophysics, 46 (2005), pp. 260–277

Yang and Taylor, 2010 Yang, S., Taylor, J., 2010. Model and Data Work Together to Reveal Microscopic Structures of Materials. SPIE Newsroom, 29 September, 2010. doi: 10.1117/2.1201009.003099 〈http://spie.org/x42065.xml?highlight=x2406&ArticleID=x42065〉.

Yang et al., 2012 Yang, S., Liu, K., Mayo, S., Tulloh, A., 2012. CIPS Sandstone Microstructure. CSIRO Data Collection 〈http://dx.doi.org/10.4225/08/5045B5990B44E〉.

Yang, S., Lui, K., Mayo, S., Tullon, A. (2012) CIPS Sandstone Microstructure. CSIRO Data Collection. http://dx.doi.org/10.4225/08/5045B5990B44

Data citation (example 3)

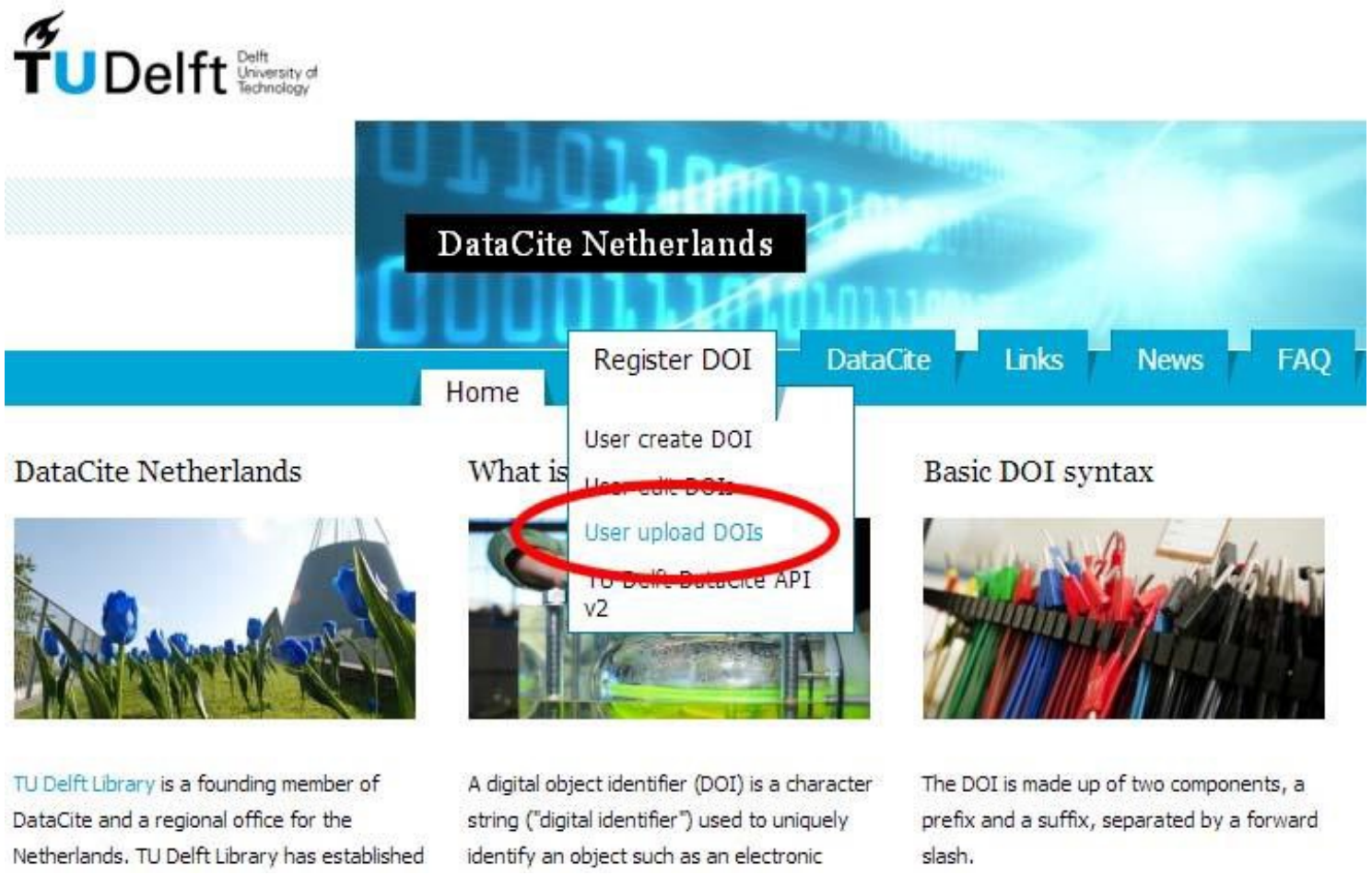


DataCite Netherlands

- Enables research organisations to assign DOIs to research datasets or collections
- Machine to machine service
- Clients embed the service within their data management workflows
- No accessibility for individual researchers
- User interface for clients to list and update DOIs

http://datacite.tudelft.nl/

Creating DOIs



```
<resources>
   <DOIdata>
      <DOI>10.5072/unique_identifier_001</DOI>
      <URL>http://example.org/resolve/unique_identifier_001</URL>
      <metadata>
         <resource
            xmlns="http://datacite.org/schema/kernel-3"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:schemaLocation=
                  "http://datacite.org/schema/kernel-3
                   http://schema.datacite.org/meta/kernel-3/metadata.xsd"
            >
            <identifier identifierType="DOI">10.5072/unique_identifier_001</identifier>
            <creators>
               <creator>
                  <creatorName>Name of the Creator</creatorName>
               </creator>
            </creators>
            <titles>
               <title>Title of the DOI</title>
            </titles>
            <publisher>Publisher of the DOI</publisher>
            <publicationYear>2014</publicationYear>
         </resource>
      </metadata>
   </DOIdata>
   <DOIdata>
      <DOI>10.5072/unique_identifier_002</DOI>
      <URL>http://example.org/resolve/unique_identifier_002</URL>
      <metadata>
         <resource
            xmlns="http://datacite.org/schema/kernel-3"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:schemaLocation=
                  "http://datacite.org/schema/kernel-3
                   http://schema.datacite.org/meta/kernel-3/metadata.xsd">
            <identifier identifierType="DOI">10.5072/unique_identifier_002</identifier>
            <creators>
               <creator>
                  <creatorName>Name of the Creator</creatorName>
               </creator>
            </creators>
            <titles>
               <title>Title of the DOI</title>
            </titles>
            <publisher>Publisher of the DOI</publisher>
            <publicationYear>2014</publicationYear>
         </resource>
      </metadata>
   </DOIdata>
</resources>
```

User registration form

Username *

Password *

Password confirmation *

First name: *

Last name: *

Email: *

Organization: * DataCite Netherlands Test

* This field is required

Register

Granularity of data

▲3TU.DC info

Home

Upload datasets

Personal page

Statistics

» Search in Data

» Search in "info"

Delft University of Technology

Technische Universiteit Eindhoven University of Technology

RSITY OF TWENTE.

Dataset **IDRA weather radar measurements - day 2010-10-11**

Link/cite as doi:10.4121/uuid:d0bcc7c5-8e28-4985-bde6-11ea9f29ec5e | **show link code** | **full citation**

▼go to DATA section▼

◄ previous	IDRA weather radar measurements - day 2010-10-10
► next	IDRA weather radar measurements - day 2010-10-12
title	IDRA weather radar measurements - day 2010-10-11
creator	*Otto, T.(Tobias)*
creator	*Russchenberg, H.W.J.(Herman)*
date created	2010
description	Radar ranges: standard, far. Max rain level: weak rain.
▲ part of	IDRA weather radar measurements - month 2010-10
▲ part of	IDRA weather radar measurements - all data
date created	2009
publisher	TU Delft
documentation	IDRA Dataset Description Document
has tool	Matlab files for analysis of the IDRA datasets.
related publication	Design of a High Resolution X-band Dopler Polarimetric Radar: [dissertation, Figueras i Ventura, J., 2009]
related publication	High-resolution polarimetric X-band weather radar observations at the Cabauw Experimental Site for Atmospheric Research (article, 2013)
measured by	IDRA atmospheric radar in CESAR observatory, Cabauw
altitude (m)	213
contributor	TU Delft - Delft University of Technology
date	2007 (installation)
description	IRCTR has built a high resolution radar system, IDRA (IRCTR Drizzle Radar), aimed at the detailed observation of the spatial and temporal distribution of rainfall and drizzle. The system was placed at the end of August 2007 on top of a 213 m high meteorological tower in the CESAR (Cabauw Experimental Site for Atmospheric Research) observatory in Cabauw, The Netherlands. This location has several advantages: In the first place, an increased sensitivity due to the reduction of the influence of ground clutter. Secondly, it allows direct observation of the horizontal distribution of low level clouds and fog. Finally, the presence of other instruments in the vicinity enhances the understanding of the physical processes in the atmosphere by synergistically combining their measurements. Data from other instruments at CESAR are available at http://www.cesar-database.nl. IDRA provides the horizontal distribution of reflectivity, mean Doppler velocity, Doppler spectrum width and polarimetric parameters like differential reflectivity, linear depolarization ratio or specific differential phase. The data collected is freely available to the scientific community.
▲ in collection	Atmospheric observations Cabauw
documentation	IDRA Dataset Description Document
located at	CESAR observatory, Cabauw (Lopik)
map	►Map of this location [kml] ►Map including all data locations within 100 km [☑+circle] [kml]

#Fragment identifiers

doi:10.4121/uuid:5f3bcaa2-a456-4a66-a67b-1eec928cae6d

http://dx.doi.org/10.4121/uuid:5f3bcaa2-a456-4a66-a67b-1eec928cae6d

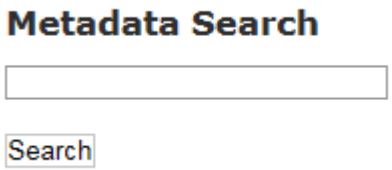
http://data.3tu.nl/repository/uuid:5f3bcaa2-a456-4a66-a67b-1eec928cae6d

http://data.3tu.nl/repository/uuid:5f3bcaa2-a456-4a66-a67b-1eec928cae6d#DATA

Will directly lead you to the DATA section of the dataset landing page

DataCite services

- DOIs for data !
- Local service & support
- Usage statistics
- Citation formatter
- Content negotiation
- Metadata search
- OAI provider
- DataCite to ORCID connection
- …….



Content negotiation

Format	Content Type	CrossRef	DataCite	mEDRA
RDF XML	application/rdf+xml	Yes	Yes	Yes
RDF Turtle	text/turtle	Yes	Yes	Yes
Citeproc JSON	application/vnd.citationstyles.csl+json	Yes	Yes	Yes
Formatted text citation	text/x-bibliography	Yes	Yes	Yes
RIS	application/x-research-info-systems	Yes	Yes	No
BibTeX	application/x-bibtex	Yes	Yes	Yes
CrossRef Unixref XML	application/vnd.crossref.unixref+xml	Yes	No	No
DataCite XML	application/vnd.datacite.datacite+xml	No	Yes	No
ONIX for DOI	application/vnd.medra.onixdoi+xml	No	No	Yes

More info: http://www.crosscite.org/cn/

ODIN Project

using identifiers to connect researchers with research

Object identifiers

digital object identifiers

DataCite DOIs & Crossref DOIs

People identifiers

Open Researcher and Contributor ID (ORCID)

International Standard Name Identifier (ISNI)

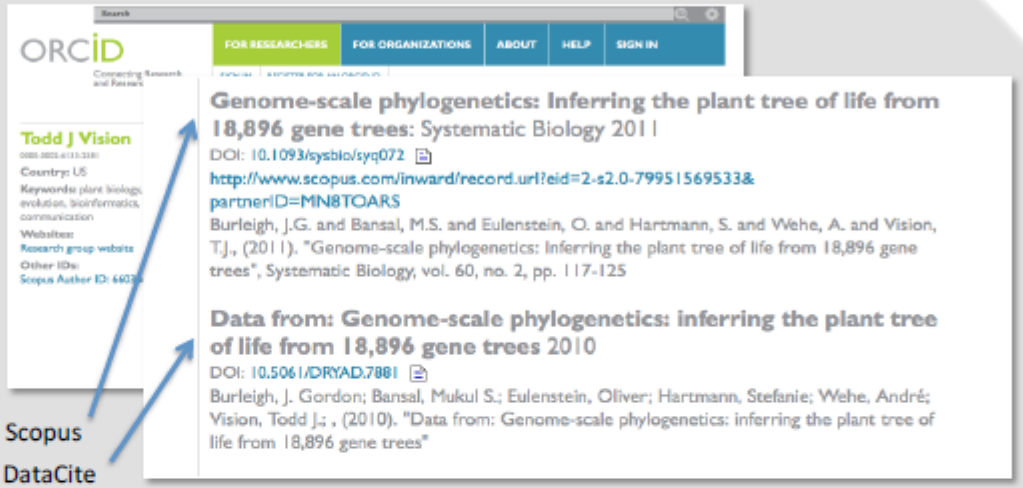
ODIN project, ORCID and DataCite Interoperability Network

ODIN

Discipline-neutral tools

To illustrate what benefits can be realized.

Import your works from DataCite to ORCID



Article from Scopus

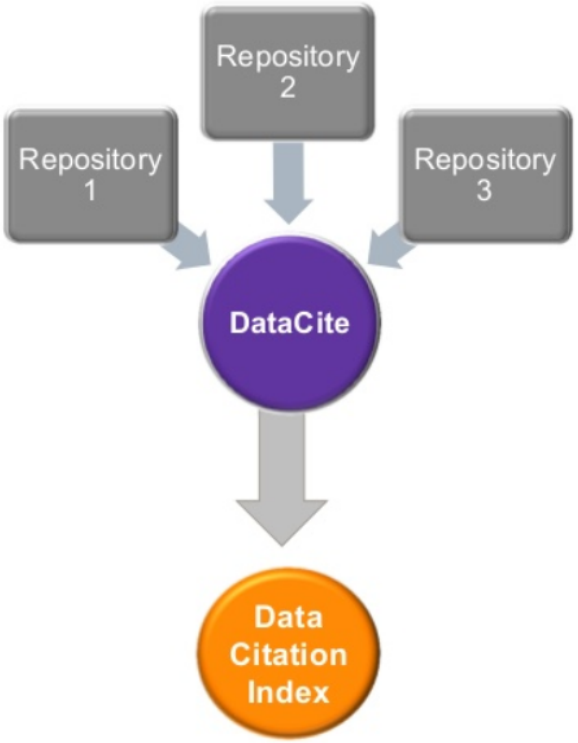
Dryad data from DataCite

In beta: add ISNIs as external identifiers to an ORCID profile

http://isni2orcid.labs.orcid-eu.org

Data Citation Index

partnership with DataCite



Repositories providing enhanced metadata to DataCite are included in the Data Citation Index

- Expanded data visibility through Web of Knowledge
- Faster and more frequent updates
- Encourage use of DataCite DOIs and citation formats
- Tracking of citations to their data in the primary research literature
- Demonstrate value of citation

Questions ?

- Contact us at: library@tudelft.nl
- website: http://datacite.tudelft.nl

- Watch our short video on persistent identifiers and data citation
  -> https://www.youtube.com/watch?v=PgqtiY7oZ6k