

# Adding value and facilitating data reuse: the case of the 4TU.Centre for Research Data

**Maria Cruz**, Egbert Gramsbergen

TU Delft Library / 4TU.Centre for Research Data, The Netherlands

**PV2018, Harwell UK, 16 May 2018**





An aerial photograph of the Delft University of Technology campus. The image shows various university buildings, green spaces, and a large orange arrow pointing towards a specific building in the middle ground. The background shows a dense urban area of Delft, Netherlands.

## Delft University of Technology

176 years old – largest and oldest of the four Dutch technical universities.

World-class research with focus on science, engineering and design.



## 4TU.Centre for Research Data

Hosted by TU Delft  
Library



## 4TU.Centre for Research Data

- 2008: Started by the libraries of 3 (of the 4) Dutch technical universities as a national service for archiving science & engineering data.
- 2010: Up and running as a fully operational data archive.
- Front offices in Delft, Eindhoven and Twente.
- Researchers from other (national and international) institutions can upload data (with size restrictions for free upload).

## 4TU.Centre for Research Data



- Certified and trusted repository for engineering and applied sciences.
- At least 15 years long-term curation and access to open research data.
- Each dataset assigned a DOI and released with a license.
- Metadata checked and improved; documentation validated before publication.

An aerial photograph of a coastal area. In the foreground, there's a wide, sandy beach with some tracks and a few small structures. The ocean is a deep blue, with white foam from waves breaking onto the shore. In the background, a city skyline is visible across the water, with several tall buildings. The sky is a pale blue.

## Special interest in netCDF data

90% of data in netCDF format

- 6523 out of a total of 7582 datasets.\*
- 30.3 TB out of 32.6 TB.

\*As of 1 May 2018.





## NetCDF – a brief introduction

- Network Common Data Form (1987).
- Created and maintained by Unidata, part of the University Corporation for Atmospheric Research (USA).
- Particularly useful for gridded, multidimensional data.
- Mainly and widely used in climate, ocean, and atmospheric sciences.







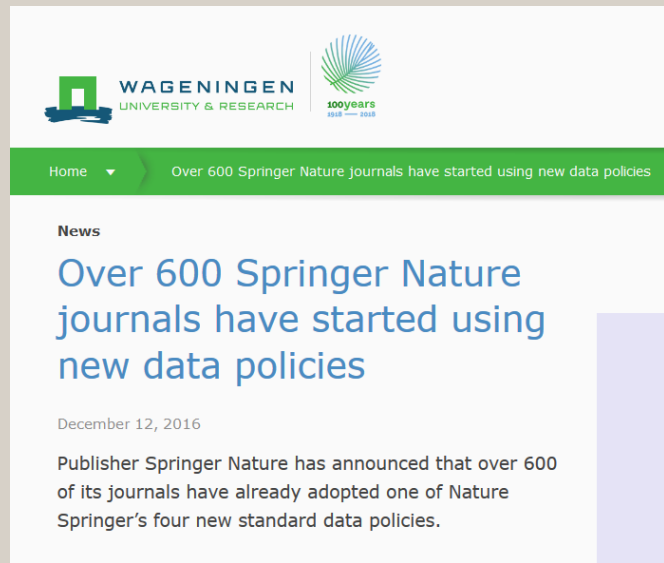
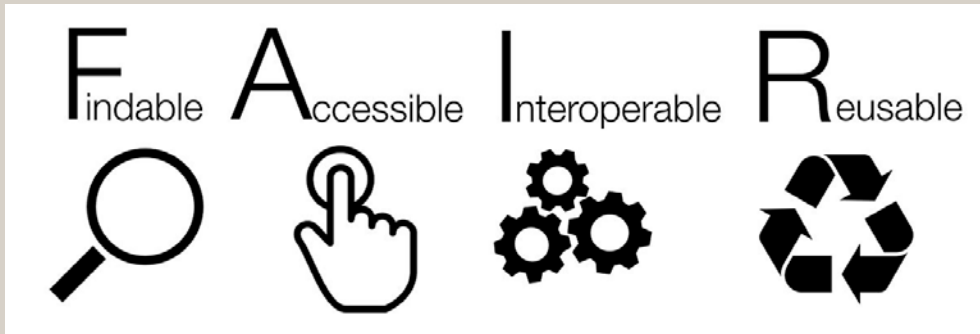
**Repositories need to have a subject or format focus to remain relevant and be successful in the long term.**

IDCC 2018, Barcelona

Paper:

[10.17605/OSF.IO/JGRKB](https://doi.org/10.17605/OSF.IO/JGRKB)

# A lot has changed since 2008





A woman with long brown hair, wearing a dark jacket and dark trousers, is walking up a modern staircase with wooden steps and metal railings. The staircase is set within a large library with floor-to-ceiling bookshelves filled with books. The background is a dense wall of books, and the lighting is bright and even.

**How can the  
4TU.Centre for  
Research Data remain  
relevant in this rapidly  
evolving landscape?**





## Could we do more with netCDF?

- Explore options for providing further services, both technical and training, related to netCDF data.
- What advantages could accrue from becoming a recognised source of netCDF data?

A woman with short brown hair, wearing a white lab coat and a patterned scarf, is seated at a desk in a laboratory. She is looking at a computer monitor. The monitor displays a software interface with a flowchart and a graph. On the desk, there is a keyboard, a blue rack with several small vials, and some papers. In the background, a laboratory incubator is visible.

## Self-Describing Data Format

A netCDF file includes information about the data it contains.



Information  
about variables,  
including units,  
coordinate  
systems, etc.

Community-  
defined  
conventions  
(defining  
metadata) &  
provenance  
information

```
netcdf simple {
dimensions:
    latitude = 3 ;
    longitude = 2 ;
    time = UNLIMITED ; // (5 currently);
variables:
    double time(time) ;
        time:standard_name = "time" ;
        time:units = "minutes since 1994-01-01 00:00:00" ;
        time:long_name = "time" ;
    float latitude(latitude) ;
        latitude:standard_name = "latitude" ;
        latitude:units = "degrees_north" ;
        latitude:point_spacing = "even" ;
        latitude:long_name = "latitude" ;
    float longitude(longitude) ;
        longitude:standard_name = "longitude" ;
        longitude:units = "degrees_east" ;
        longitude:point_spacing = "even" ;
        longitude:long_name = "longitude" ;
    float temp(time, latitude, longitude) ;
        temp:standard_name = "surface temperature" ;
        temp:long_name = "Surface temperature in degrees C" ;
        temp:units = "deg_C" ;
        temp:_FillValue = 2.e+020f ;
        temp:valid_min = -80.f ;
        temp:valid_max = 60.f ;
        temp:comment = "This parameter may be erroneous." ;

// global attributes:
        :Conventions = "CF-1.0" ;
        :institute = "The British Atmospheric Data Centre." ;
        :source = "Model developed in conjunction with IPLSPSC." ;
        :history = "10 Sep 2002 - Created by hand.\n",
            "18 Mar 2003 - Modified by feet.\n" ;
        :title = "Model output from imaginary model (temperONETER)." ;
        :comment = "Not very useful data." ;
        :references = "A great report somewhere!" ;

data:

    time = 0.5, 1.5, 2.5, 3.5, 4.5 ;

    latitude = 54.2, 54.4, 54.6 ;

    longitude = 2.0, 2.5 ;

    temp = 34.5, 31.2, 23.7, 19.6, 35.8, 29.2, 24.4, 5.6, 7.2, 8.1,
        18.6, 15.2, 13.1, 4.6, 3.7, 8.2, 9.7, 34.2, 26.7, 28.7,
        2.1, 3.4, 5.6, 7.8, 9.0, 10.2, 11.2, 11.6, 11.7, 11.8 ;
}
```

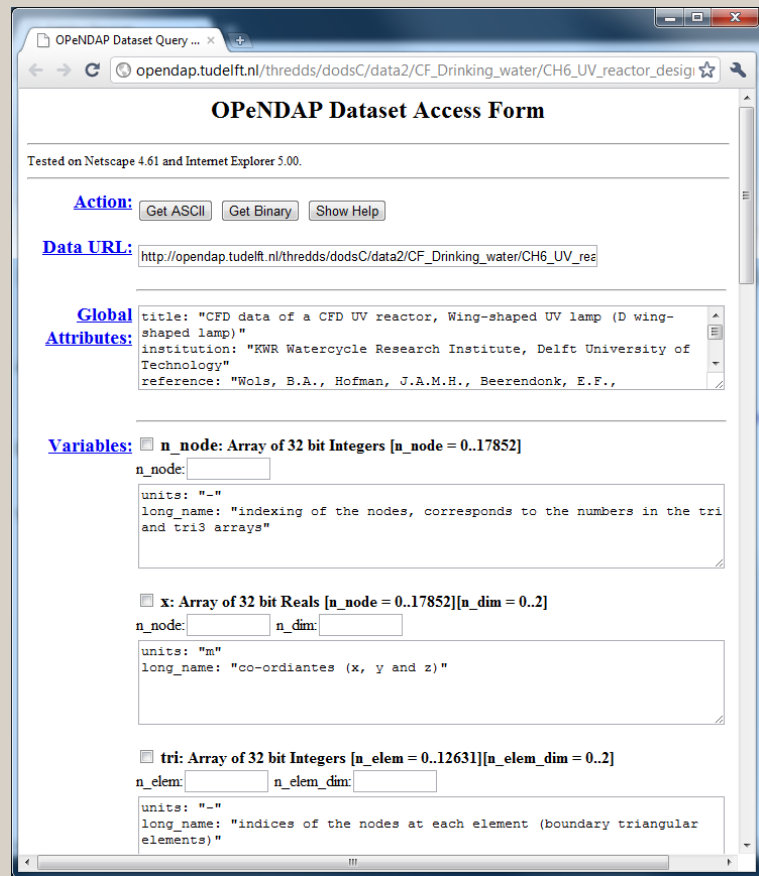
[http://cedadocs.ceda.ac.uk/72/1/nc\\_example.html](http://cedadocs.ceda.ac.uk/72/1/nc_example.html)

A male scientist with glasses and a white lab coat over a blue plaid shirt is working in a laboratory. He is holding a test tube containing a bright green liquid. In the foreground, there are several other test tubes containing liquids of different colors (orange, green, yellow) sitting on a surface. The background shows various laboratory equipment and a green gas cylinder.

## NetCDF files are scalable

A small subset of a large dataset may be accessed efficiently (without the need to download the whole file).

# THREDDS / OPeNDAP



The screenshot shows a Netscape browser window titled "OPeNDAP Dataset Query ...". The address bar shows the URL "opendap.tudelft.nl/thredds/dodsC/data2/CF\_Drinking\_water/CH6\_UV\_reactor\_desig". The main content area is titled "OPeNDAP Dataset Access Form" and includes a note "Tested on Netscape 4.61 and Internet Explorer 5.00".

**Action:**

**Data URL:**

**Global Attributes:**

- title: "CFD data of a CFD UV reactor, Wing-shaped UV lamp (D wing-shaped lamp)"
- institution: "KWR Watercycle Research Institute, Delft University of Technology"
- reference: "Wols, B.A., Hofman, J.A.M.H., Beerendonk, E.F.,

**Variables:**

☐ **n\_node:** Array of 32 bit Integers [n\_node = 0..17852]

n\_node:

units: "-"

long\_name: "indexing of the nodes, corresponds to the numbers in the tri and tri3 arrays"

☐ **x:** Array of 32 bit Reals [n\_node = 0..17852][n\_dim = 0..2]

n\_node:  n\_dim:

units: "m"

long\_name: "co-ordinates (x, y and z)"

☐ **tri:** Array of 32 bit Integers [n\_elem = 0..12631][n\_elem\_dim = 0..2]

n\_elem:  n\_elem\_dim:

units: "-"

long\_name: "indices of the nodes at each element (boundary triangular elements)"

**Access enhanced by serving netCDF data via OPeNDAP protocol.**

Make selections of the data through a web form and export to different formats.

Inspect the structure and internal metadata of netCDF data.

Programmatic access to data with APIs for Java, Python, Matlab, etc.



A man and a woman are standing in front of a glass wall covered with numerous white sticky notes. The woman, on the left, is wearing a grey cardigan over a white shirt and blue jeans. The man, on the right, is wearing a blue checkered shirt. Both are looking up at the sticky notes with interest. The background shows a modern office interior with wooden paneling and large windows.

## Adding value to data and facilitation of data use

*“Data re-use requires well-informed, sustainable, inclusive, participative development of data infrastructures.”*

Sabina Leonelli (2017), Towards the European Open Science Cloud: Five Lessons from the Study of Data Journeys [10.5281/zenodo.1043153](https://doi.org/10.5281/zenodo.1043153)



## Qualitative semi-structured interviews

- Between November 2017 and April 2018, conducted interviews **with 11 researchers based in the Netherlands who produce or use netCDF data**; most have deposited netCDF data in the 4TU archive.
- Spoke with researchers at all career stages from PhD students to Full Professors; mostly affiliated with Dutch technical universities, but also with national research facilities and industry.



## Main findings from the interviews

- The netCDF data depositors and users of the 4TU archive represent heterogeneous research communities within the Earth sciences.
- They have very different views and attitudes to data archiving and publishing.
- They store varied types of netCDF datasets in the 4TU archive.





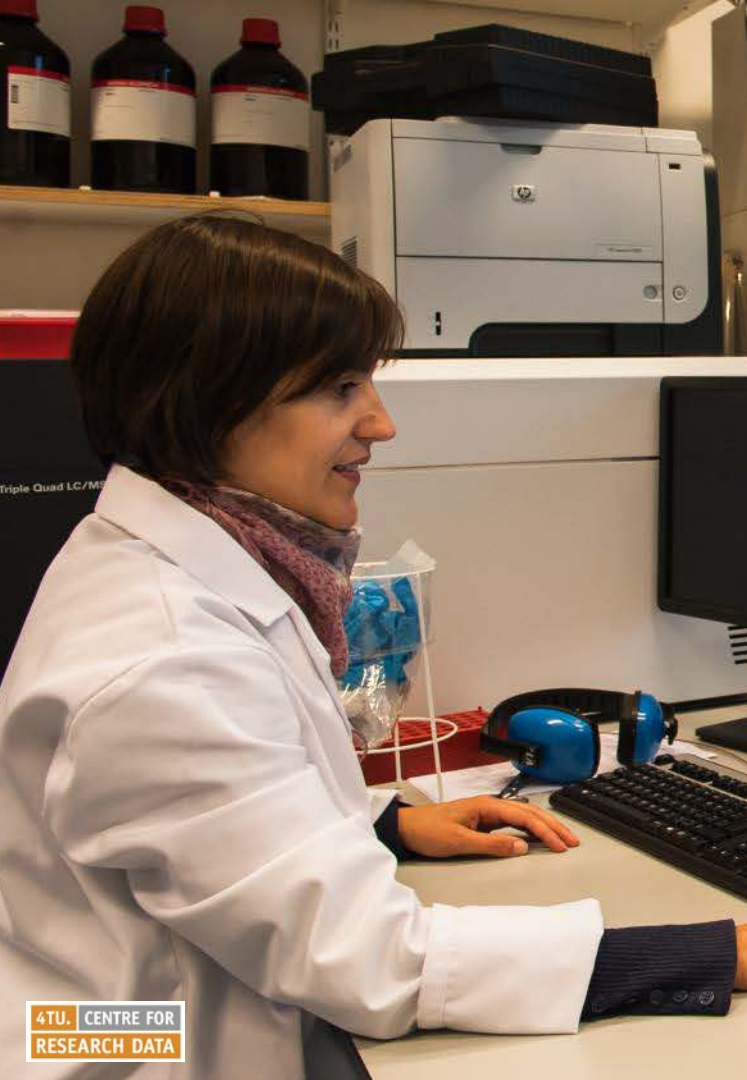
## Main findings from the interviews

- Overall, our THREDDS/OPeNDAP services did not seem to have had much influence in the choice of archive.
- Some researchers were not aware of these services and its functionalities; others were aware but did not care about them.



## Main findings from the interviews

- Many data depositors chose the 4TU archive simply because it was locally available at TU Delft.
- Most chose the 4TU archive after the recommendation of a colleague, supervisor, or data librarian.



## Conclusions from the interviews

- Ensuring that any new and current netCDF services continue to be relevant to our user communities will require taking their diversity of needs and requirements into account.
- A need for training and guidance – particularly on data management aspects related to documentation, metadata standards and conventions – may be the common thread uniting these communities.



A male scientist with short brown hair, wearing safety glasses and a dark t-shirt with 'HOLLISTER' written on it, is working in a laboratory. He is wearing blue nitrile gloves and using a pipette to transfer liquid into a small vial. The background shows various laboratory equipment, including glass tubes and a large orange balloon.

## Plans for the future

- Explore training as the way forward to build a community of data depositors and users.

## Questions, suggestions, feedback?

[M.J.MarquesdeBarrosCruz@tudelft.nl](mailto:M.J.MarquesdeBarrosCruz@tudelft.nl)

- Presentation: Zenodo DOI:10.5281/zenodo.1247903
- Open Working Blog for updates:  
<https://openworking.wordpress.com/category/netcdf/>