



# To document, or not to document?

How to write a readme file and what is metadata.



NCR DAYS 2018  
The future river  
Celebrating 20 years NCR

# Why documenting data?

- Integrity
- Traceability
- Unambiguous Understanding
- Reuse

# Why documenting data?

- Content
  - Data files
- Context
  - Data collection
  - Data processing








## What does documenting data mean?

- Metadata:
  - discovery, citation, terms of use
- Documentation:
  - understanding and reuse

## Dataset: Pan-European data sets of annual number of situations conducive to thunderstorm formation under present and future climate

▶▶▶▶▶ Link/cite as <https://doi.org/10.4121/uuid:425f1f99-dee7-4a32-b5c6-8ad820af5f69> | full citation


▼ go to DATA section ▼

title	?	Pan-European data sets of annual number of situations conducive to thunderstorm formation under present and future climate
creator	?	Pucik, T. (Tomas)
creator	?	Groenemeijer, P. (Pieter)
contributor	?	European Severe Storms Laboratory
contributor	?	info.eu-repo/grantAgreement/EC/FP7/608166
date accepted	?	2016-05-02
date created	?	2014 through 2016
date published	?	2016
description	?	NetCDF files containing gridded annual number of situations conducive for thunderstorm formation for present day and the future climate. The fields are multi model means of 15 regional climate model simulations (CORDEX).
language	?	en
publisher	?	European Severe Storms Laboratory
subject	?	Climate ◊ Climate change ◊ Climate hazards ◊ CORDEX (Coordinated Regional Climate Downscaling, Experiment) ◊ Future climate ◊ Meteorology ◊ Risk analysis ◊ Thunderstorm ◊ Thunderstorms probability
▲ in collection	?	RAIN: Pan-European gridded data sets of extreme weather probability of occurrence under present and future climate
spatial coverage	?	<a href="#">Europe</a>
time coverage	?	<a href="#">years 1971-2100</a>
licence	?	 <a href="#">General terms of use</a>

### DATA






[readme\\_thunderstorms.pdf](#) (application/pdf)

Access on our OPeNDAP server: [xml](#), [html](#)

 **Direct access through OPeNDAP**


Total 643 KiB

[Move over links for options and info]

-  [tstm\\_1971\\_2000.nc](#)
-  [tstm\\_rcp45\\_2021\\_2050.nc](#)
-  [tstm\\_rcp45\\_2071\\_2100.nc](#)
-  [tstm\\_rcp85\\_2021\\_2050.nc](#)
-  [tstm\\_rcp85\\_2071\\_2100.nc](#)

<b>Creator*</b>	<b>Main researchers involved in producing the data.</b>
<b>Title*</b>	<b>Name or title by which the dataset is known.</b>
<b>Contributor</b>	<b>Institution where the data was created or collected. A person or organization responsible for making contributions to the dataset.</b>
<b>Publisher*</b>	<b>A holder of the data (including archives appropriate) or institution which submitted the work. Any others may be listed as contributors.</b>
<b>Publication year*</b>	<b>The year when the data was or will be made publicly available.</b>
<b>Date created*</b>	<b>Date the resource itself was put together; this could be a date range or a single date.</b>
<b>Description*</b>	<b>Concise description of the contents of the dataset. Describe the research objective, type of research, method of data collection and type of data.</b>
<b>Subject</b>	<b>Subject, keyword, or key phrase describing the resource.</b>
<b>Temporal coverage</b>	<b>Indicate the dates to which the data refer. Enter the year, or beginning and ending dates.</b>
<b>Spatial coverage</b>	<b>Describe the geographic area to which the data refer (e.g. municipality, town/city, region, country). The geographic coordinates of the area may be included, if desired.</b>
<b>Identifier</b>	<b>4TU.ResearchData automatically assigns a DOI to a dataset once the entire deposit procedure has been completed. In some cases, a dataset may be known by one or more other (persistent) identifiers.</b>
<b>Language*</b>	<b>The primary language of the resource. When no language is added, 4TU.ResearchData will automatically assign 'English'.</b>
<b>Link to publication</b>	<b>Include the web addresses or DOIs for any publication, important internal reports or other datasets that are related to your dataset.</b>

Creator*	Main researchers involved in producing the data
Title*	Name or title by which the data is known
Contributor	Institution where the data is held, or the person responsible for making contact
Publisher*	A holder of the data (including the data publisher) responsible for the work. Any others may be listed
Publication year*	The year when the data was published
Date created*	Date the resource itself was created
Description*	Concise description of the data, including the type of research, methods used, and any other relevant information
Subject	Subject, keyword, or key phrase
Temporal coverage	Indicate the dates to which the data applies
Spatial coverage	Describe the geographic area covered by the data (e.g. region, country). The geographic coordinates should be included where possible
Identifier	4TU.ResearchData automatic identifier (persistent) id
Language*	The primary language of the data. If the data is multilingual, the language(s) should be listed
Link to publication	Include the web addresses of the data and other datasets that are related to it


because good research needs good data

Home Digital curation About us News Events Resources Training Projects Community Tailored support

Home > Resources > Metadata Standards

In this section

Disciplinary Metadata

Briefing Papers  
How-to Guides & Checklists  
Developing RDM Services  
Curation Lifecycle Model  
Curation Reference Manual  
Policy and legal  
Data Management Plans  
Tools  
Case studies  
Repository audit and assessment  
Standards  
Disciplinary Metadata  
DIFFUSE  
Publications and presentations  
Roles  
Curation journals  
Informatics research  
External resources  
Online Store

## Hydrogeology

## Metadata Standards

### ISO 19115

An internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

### Repository-Developed Metadata Schemas

Some repositories have decided that current standards do not fit their metadata needs, and so have created their own requirements.

## Extensions

### EDMED Metadata Profile

The European Directory of Marine Environmental Datasets metadata scheme, which is a profile of [ISO 19115](#).

### INSPIRE Metadata Regulation

A profile of [ISO 19115:2003](#), adopted in 2007 as the common metadata standard for the Infrastructure for Spatial Information in the European Community (INSPIRE). The other profiles of ISO 19115 in use in European Member States have been made compliant with INSPIRE.

## Use Cases

### BAS PDC - British Antarctic Survey Polar Data Centre

An organisation coordinating the management of data collected by UK-funded scientists in the polar regions, using an application profile that is harmonious with both [ISO 19115](#) and [DIF](#).

### BGS NGDC - British Geological Survey National Geoscience Data Centre

The world's oldest national geological survey and the United Kingdom's premier centre for earth science information and expertise, with information access via an [ISO 19115](#)-compliant discovery metadata database.

# What to document?

- **Variable names** and **descriptions**
- **Definition of codes** and classification schemes
- Codes of, and reasons for, missing values
- Definitions of **specialty terminology** and acronyms



## What to document?

- Algorithms used to transform data
- **File format** and **software used**
- Structure and organization of data files

# How to document?

- **Context** of data collection
- Data collection **methodology**
- Data validation and quality assurance
- Data manipulations through data analysis from raw data
- Data confidentiality, access and use conditions



# Hands-On

## Metadata and Data Documentation

Go to DANS (<https://dans.knaw.nl/en/>),  
4TU.Centre for Research Data (<https://data.4tu.nl/repository/>)  
or Registry of Research Data Repositories  
(<http://www.re3data.org/>)

Search for a dataset which roughly fits in your field of research and contains a documentation or readme file.

**Assess  
the metadata**

**Assess  
the documentation**



## Assess the metadata

**Has been made use of a metadata scheme/standard?**  
If yes, which one?

**Are the metadata and values clear?**

**Is there a description of the dataset and is it clear?**  
If not, what information is missing or unclear?

## Assess the documentation

**Is documentation enough for understanding AND/OR reusing the data?**  
If not, what information is missing or unclear?

**Evaluate your answers with one or two workshop participants.**

# How to write a readme-file?



A readme file provides information about a dataset and is intended to help ensure that the data can be correctly interpreted, by yourself at a later date or by others when sharing or publishing data.

A readme file must be submitted along with the dataset file(s).

The outline below should be completed with information relevant to the submitted dataset.

#### Best practices

- **Create one readme file for each dataset**
- **Name the file README;** not readme, read\_me, ABOUT, etc.
- **Write your readme document as a plain text file;** save as README.txt or README.md when writing in [Markdown](#). Or use README.pdf when text formatting is important for your file.

#### Recommended content

Recommended minimum content is in bold.

##### 1. *Introductory information*

- **Title of the dataset**
- **For each file or group of similar files, a short description of what data it contains**
- Explain the file naming convention, if applicable
- Format of the file if not obvious from the file name
- If the data set includes multiple files that relate to each other, the relationship between the files or a description of the file structure that holds them
- Contact information; in case users have questions regarding the data files

##### 2. *Methodological information*

- **Method description for collecting or generating the data, as well as the methods for processing data, if data other than raw data are being contributed**
- Any instrument-specific information needed to understand or interpret the data
- Software (including version number) used to produce, prepare, render, compress, analyze and/or needed to read the dataset, if applicable
- Standards and calibration information, if appropriate

## 1. Introductory Information

- Dataset title and short description

## 2. Methodological Information

- About collecting and processing

## 3. Data Specific Information

- Formats, codes, units, definitions

## 4. Sharing and Access Information

- Use license or restrictions

# Overlimiting current properties at ion exchange membranes

Joeri De Valenca

Soft matter, Fluidics and Interfaces, Faculty of Science and Technology

ORIGINAL LANGUAGE	English
AWARDING INSTITUTION	University of Twente
SUPERVISORS/ADVISORS	Lammertink, Rob G.H., Supervisor Wagterveld, R.M., Advisor
DATE OF AWARD	10 Mar 2017
PLACE OF PUBLICATION	Enschede
PUBLISHER	Universiteit Twente
PRINT ISBNs	978-90-365-4314-9
STATE	Published - 10 Mar 2017

## Fingerprint

☒ Ion Exchange Membranes

## Keywords

## Cite this

☒ APA ☐ Author ☐ BIBTEX ☐ Harvard ☐ Standard ☐ RIS ☐ Vancouver

De Valenca, J. (2017). *Overlimiting current properties at ion exchange membranes* Enschede: Universiteit Twente DOI: 10.3990/1.9789036543149

agterveld, R.M. (Martijn); Wood, J.A. (Jeffery); Lammertink, R.G.H.  
ection on electrokinetic instability in overlimiting current conditions.  
[.4121/uuid:b55184a0-930d-4b5c-940b-fd3038e05d83](#)

and Technology, Soft Ma  
e for Sustainable Water Te

ging to the publication: Inf  
Review Fluids 2, 033701

otentiometry ◊ Concentrat  
nvection

[tion on electrokinetic insta](#)  
[exchange membranes \(th](#)

[netic instability in o](#)

octet-stream, application/x-  
text/x-matlab, video/avi, vi

-----File  
types-----  
MATLAB files (for version R2013b):  
\*.mat MATLAB workspaces that contain data  
\*.m MATLAB script for calculations and data  
processing  
\*.fig MATLAB figures that display data  
Various image formats are used, which are all well known  
\*.tif high resolution images  
\*.jpg compressed images  
\*.png compressed images  
\*.eps vector image  
\*.avi movie files compressed  
\*.mp4 movie files compressed further  
Data compression  
\*.z7 compressed data using open source 7-zip  
software  
Data compression  
\*.z7 compressed data using open source 7-zip  
software  
FLIM data  
\*.fli FLIM data in commercial software LI-FLIM  
from Lambert instruments



An aerial photograph of a coastal landscape. A large, light-colored sand dune dominates the center-right of the frame, with a small, dark blue pond nestled within its folds. To the right, the ocean stretches to the horizon, with white waves breaking against the shore. In the bottom left, a sandy area contains several parked vehicles, including a red truck and a blue car. The background shows a hazy horizon with distant structures.


# Questions?

## Contact 4TU.ResearchData

**W** [researchdata.4tu.nl](https://researchdata.4tu.nl)

**T** +31 (0)15 27 88 600

**M** [researchdata@4tu.nl](mailto:researchdata@4tu.nl)

 [twitter.com/4TUTResearchData](https://twitter.com/4TUTResearchData)

 <https://openworking.wordpress.com/>

Jasmin K. Böhmer  
[j.k.boehmer@tudelft.nl](mailto:j.k.boehmer@tudelft.nl)

Maarten van Benthum  
[m.vanbentum@utwente.nl](mailto:m.vanbentum@utwente.nl)