

“Module data sharing archiving and publishing”

All about archiving

Cees H.J. Hof

Data Archiving and Networked Services (DANS-KNAW)

With the help of Jasmin Böhmer (UMCU)

 @CeesH_DANS

29 MAY 2019, Helis Academy, FAIR Data Stewardship Course

Incubator Darwin, Niel, Belgium

About the DANS long term repository EASY

- **For long term storage**
- **Certified according to Core Trust Seal (<https://www.coretrustseal.org>)**
- **Discipline agnostic**
Biased towards humanities – social sciences – life sciences
- **> 80.000 datasets**
- **Team of data custodians**
- **Below 1 TB free service**
- **Ongoing developments**

<https://easy.dans.knaw.nl/>



About the DANS long term repository EASY

RDNL-clip depositing your data in EASY:

https://www.youtube.com/watch?v=KBusZUM_W0s

In Dutch....

DANS-EASY



1. Prepare your data

Select the relevant data files. Check them for privacy aspects and file format against the guidelines issued by DANS.

- **License Agreement and Terms of Use**
 - You are getting into a legal agreement...
- **Intellectual property rights**
 - Do you have permission?
 - Consider copyrights, database and patent rights.
- **Personal data** (allow for the identification of a living person)
 - Informed consent needed?
 - Acceptance of responsibility by you as depositor
 - Anonymizing personal data

DANS-EASY



1. Prepare your data

Select the relevant data files. Check them for privacy aspects and file format against the guidelines issued by DANS.

● File formats!

As a general guideline, DANS believes that the file formats best suited for long-term sustainability and accessibility:

- Are frequently used.
- Have open specifications.
- Are independent of specific software, developers or vendors.

Preferably the formats are as basic as possible and unzipped.....

DANS-EASY

Preferred file formats policy...

Type	• Preferred format(s)	• Non-preferred format(s)
Text documents	<ul style="list-style-type: none">• PDF/A (.pdf)	<ul style="list-style-type: none">• ODT (.odt)• MS Word (.doc, .docx)• RTF (.rtf)• PDF (.pdf)
Plain text	<ul style="list-style-type: none">• Unicode text (.txt)	<ul style="list-style-type: none">• Non-Unicode text (.txt)
Markup language	<ul style="list-style-type: none">• XML (.xml)• HTML (.html)• Related files: .css, .xslt, .js, .es	<ul style="list-style-type: none">• SGML (.sgml)
Spreadsheets	<ul style="list-style-type: none">• ODS (.ods)• CSV (.csv)	<ul style="list-style-type: none">• MS Excel (.xls, .xlsx)• PDF/A (.pdf)• OOXML (.docx, .docm)
Databases	<ul style="list-style-type: none">• SQL (.sql)• SIARD (.siard)• DB tables (.csv)	<ul style="list-style-type: none">• MS Access (.mdb, .accdb) (v. 2000 or later)• dBase (.dbf)• HDF5 (.hdf5, .he5, .h5)
Statistical data	<ul style="list-style-type: none">• SPSS Portable (.por)• STATA (.dta)• DDI (.xml)• data (.csv) + setup (.txt)	<ul style="list-style-type: none">• SPSS (.sav)• SAS (.7dat; .sd2; .tpt)• R (* under examination)
Raster images	<ul style="list-style-type: none">• JPEG (.jpg, .jpeg)• TIFF (.tif, .tiff)	<ul style="list-style-type: none">• DICOM (.dcm) (by mutual agreement)

DANS-EASY



2. Go to EASY

Log in at <http://easy.dans.knaw.nl>. If you are new to EASY, you will have to register for an account first.

- **Open to individual users**
- **Possibility of dedicated communities**
- **Equipped for machine – machine interaction**
 - Using SWORD protocol for automatic data ingestion.
 - iRODS data bridges.
 - Preservation service for other repositories like:



DANS-EASY



3. Start the deposit procedure

Go to 'deposit your data', select your discipline and click 'start deposit'.

- **Selecting your discipline means.....**
 - Activate domain specific metadata fields.
 - Activate domain specific vocabularies.
 - Allow for digital linking to other / similar datasets.
 - Easy repetition of identical information.
 - Connecting to external portals (under development....).

DANS-EASY



4. Documentation and access level

Describe the dataset and indicate whether it is open access or whether access restrictions apply.

- **Describe the dataset means adding metadata!**

- Creator
- Title
- Description
- Keywords
- Identifiers
- Discipline(s)
- Time span
- Locality
- Language
- Access rights

**This is the gateway to
your primary data!**

DANS-EASY

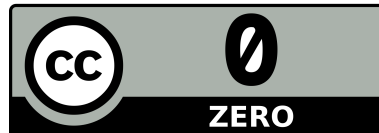


4. Documentation and access level

Describe the dataset and indicate whether it is open access or whether access restrictions apply.

● Access restrictions:

- Preferably



But also:

- CCBy
- Only accessible to logged-in users
- Restricted (request permission)
- Under license agreement
- Under embargo
- Delayed (2 years maximum)

DANS-EASY



5. Upload your data files

Select your data files and click 'upload dataset'.



6. Submit your data files

Accept the licence agreement and send your dataset to DANS by clicking the 'submit' button.

- **Automatically receive a DOI**
- **Receive a confirmation e-mail**

DANS-EASY



7. **Publication by DANS**

DANS will verify the dataset and publish the description you made. Your data have now been sustainably archived and will be accessible to others on a permanent basis under the conditions you specified.

- **Check by DANS and thereafter activation of DOI**
- **Data frozen and permanently archived**
- **Dataset visible in EASY and metadata harvested by national scholarly catalogue NARCIS**

DANS-EASY

Metadata are the key...

- Increase discoverability of your data
- Are interoperable

A short exercise(I):

Study the 2 most common and generic metadata standards:

DublinCore <http://www.dublincore.org/specifications/dublin-core/dces/>

DataCite <https://schema.datacite.org/meta/kernel-4.2/>

Note the primary **differences**

What would you do with the property “**Description**”?

If you want to know more for specific domains:

<http://www.dcc.ac.uk/drupal/resources/metadata-standards>

Archive developments

Archives of the future:

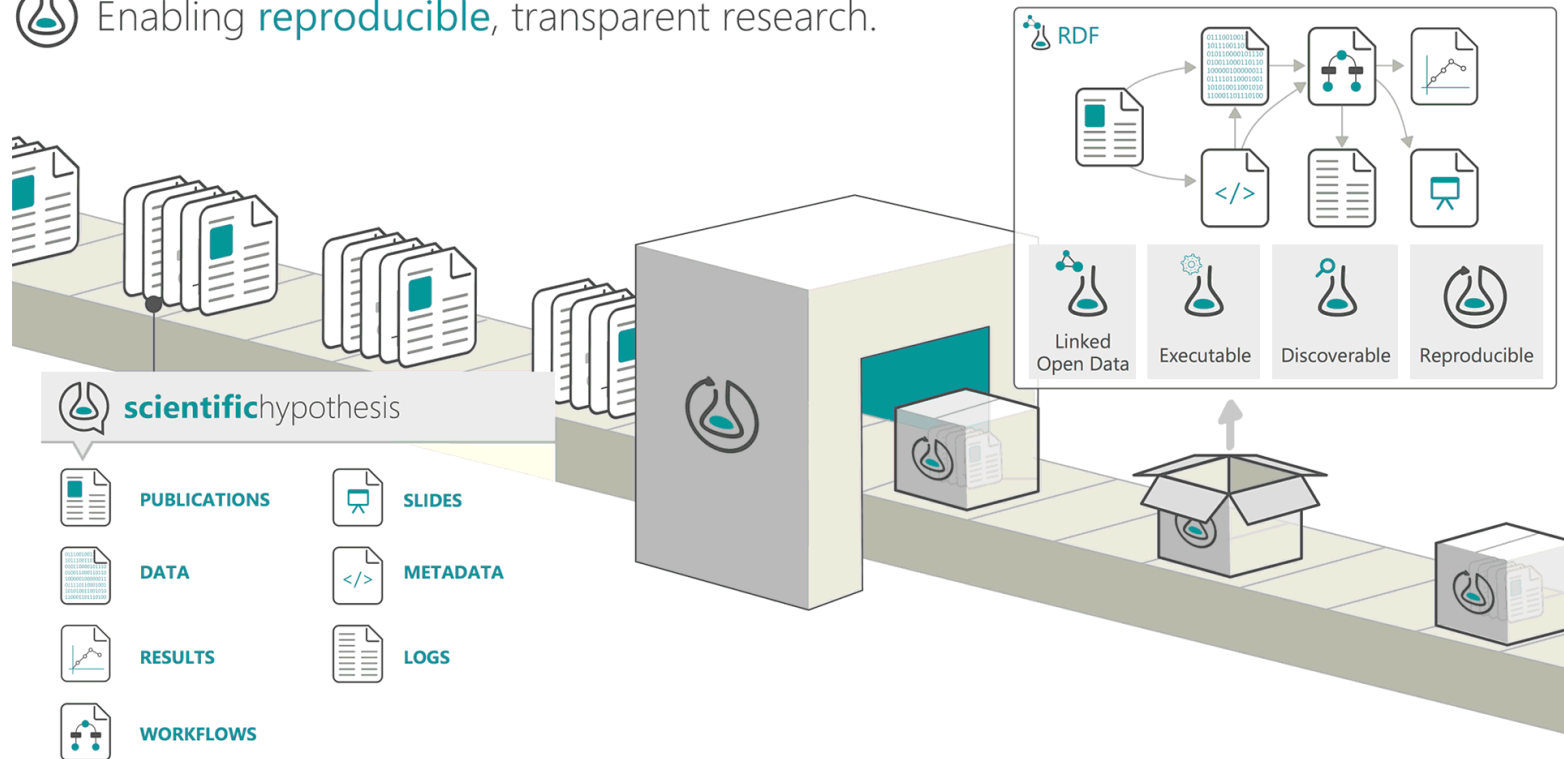
1. Exposing Identifiers
2. Declaring Licenses at a Resource Level
3. Discovery through Navigation
4. Interacting with Resources (Annotation, Commentary and Review)
5. Resource Transfer
6. Batch Discovery
7. Collecting and Exposing Activities
8. Identification of Users
9. Authentication of Users
10. Exposing Standardized Usage Metrics
11. Preserving Resources



Archive developments

Archives of the future: dealing with Research Objects

 Enabling **reproducible**, transparent research.



Hands-on!

Exercise II: Defining your requirements => Finding your archive

In this exercise you will, based on real or fictive research, try to find the best digital archive to store your research output. First you will define the requirements for your ultimate long-term archive. After defining your requirements, you will search for your top 3 most suitable repositories using Re3data. After this we will assess the “FAIRness” of your choice(s).

Thanks for your attention!

Cees Hof

cees.hof@dans.knaw.nl



@CeesH_DANS