

What makes data last?

DANS Open Day
Open data, open science

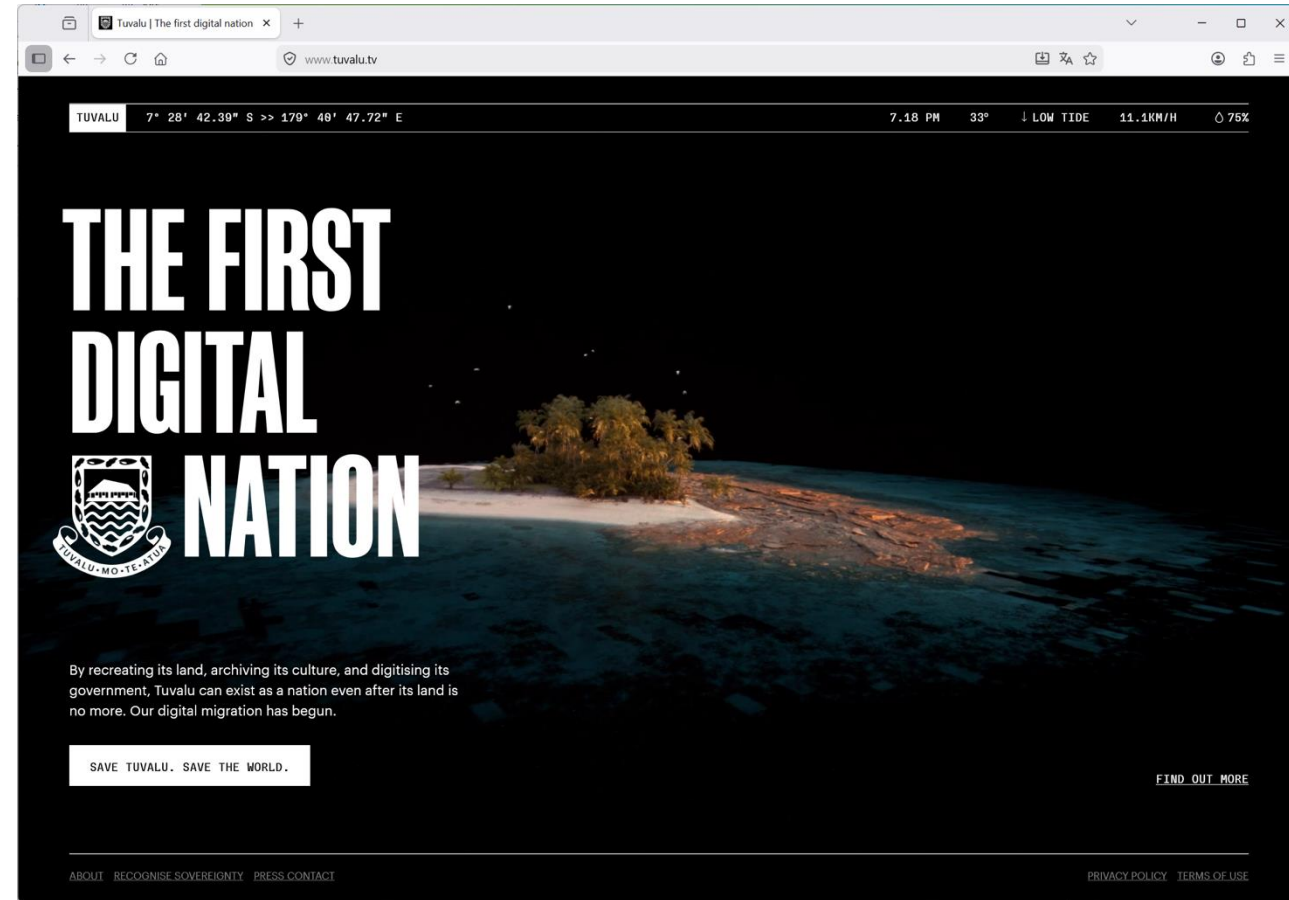
Curation, formats and guidelines

Valentijn Gilissen | Data Processing Team Leader /
Preservation Officer / Data Steward / Senior Datamanager

11 June 2026

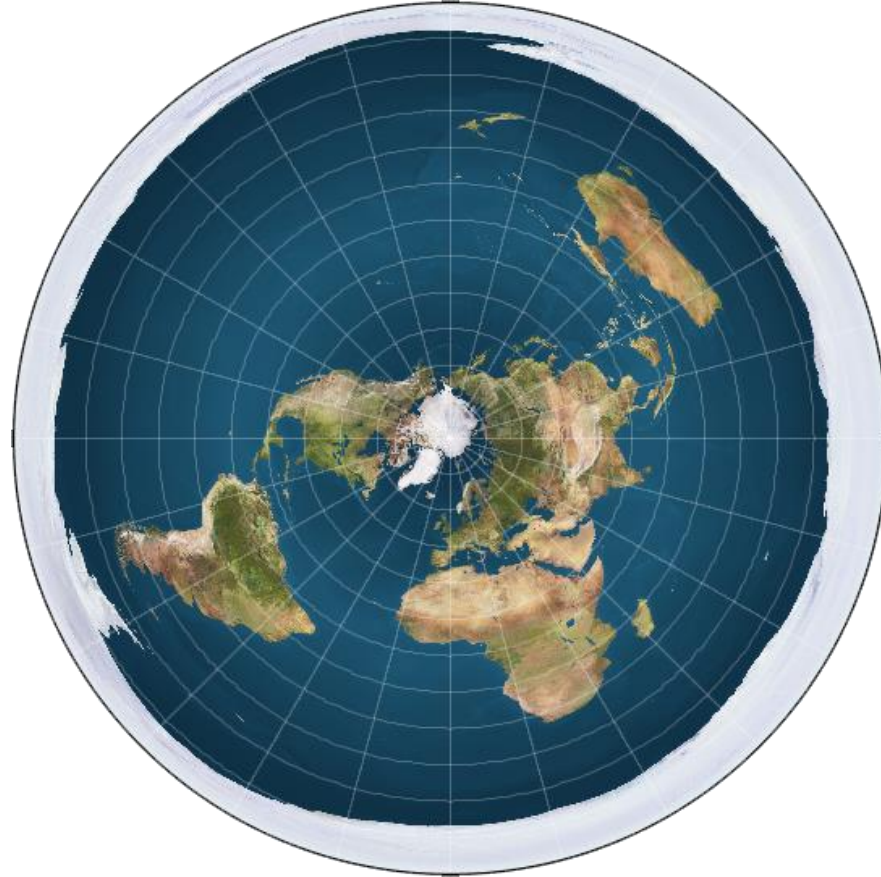


What makes our world last?



Left: Simon Kofe presenting the keynote 'Digital Nation' at iPRES 2025, Wellington NZ, 4/11/2025
Right: <https://www.tuvalu.tv/>

World of... science?!?



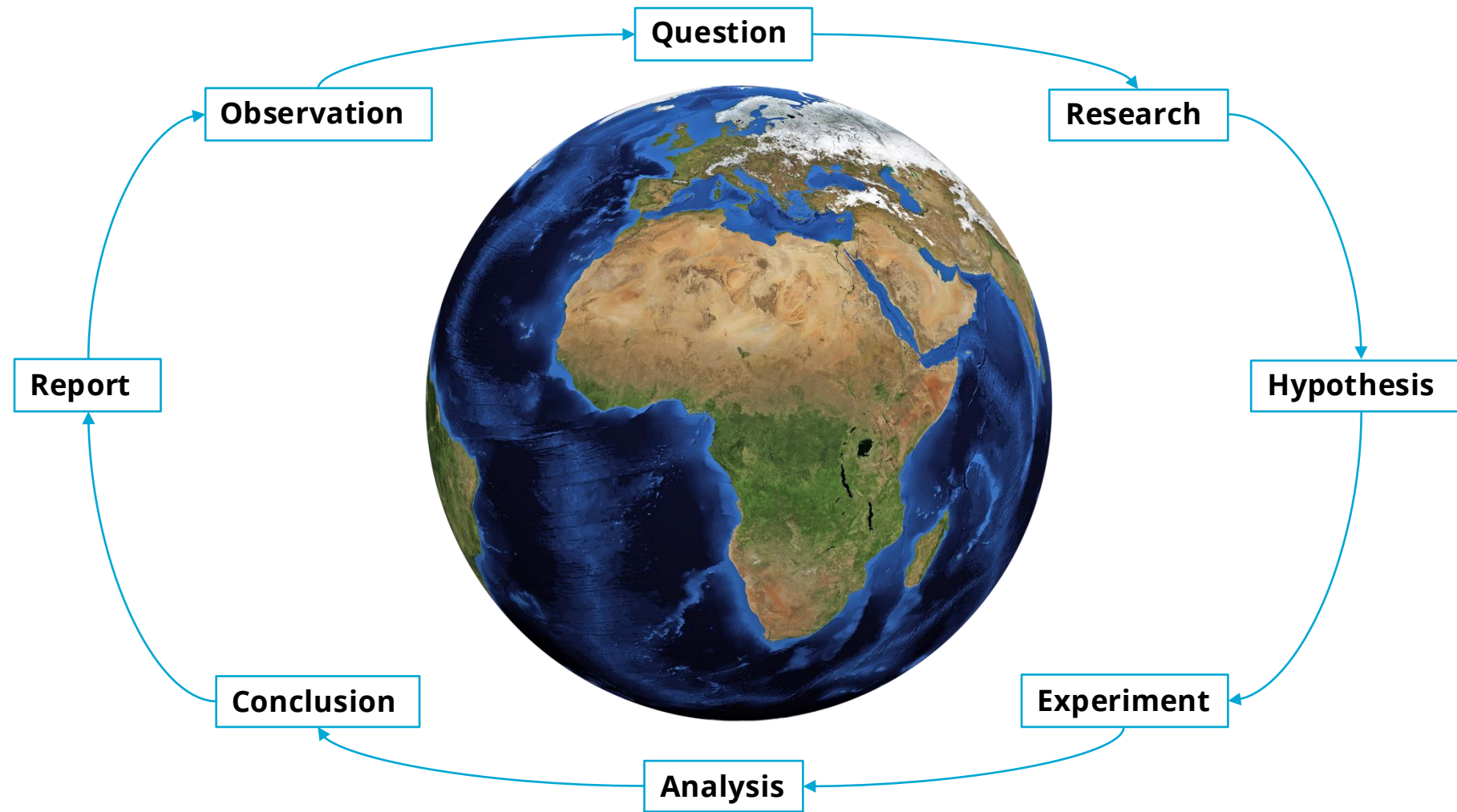
Trekky0623 at English Wikipedia ("I made this map myself"), Public domain, via Wikimedia Commons

World of science!

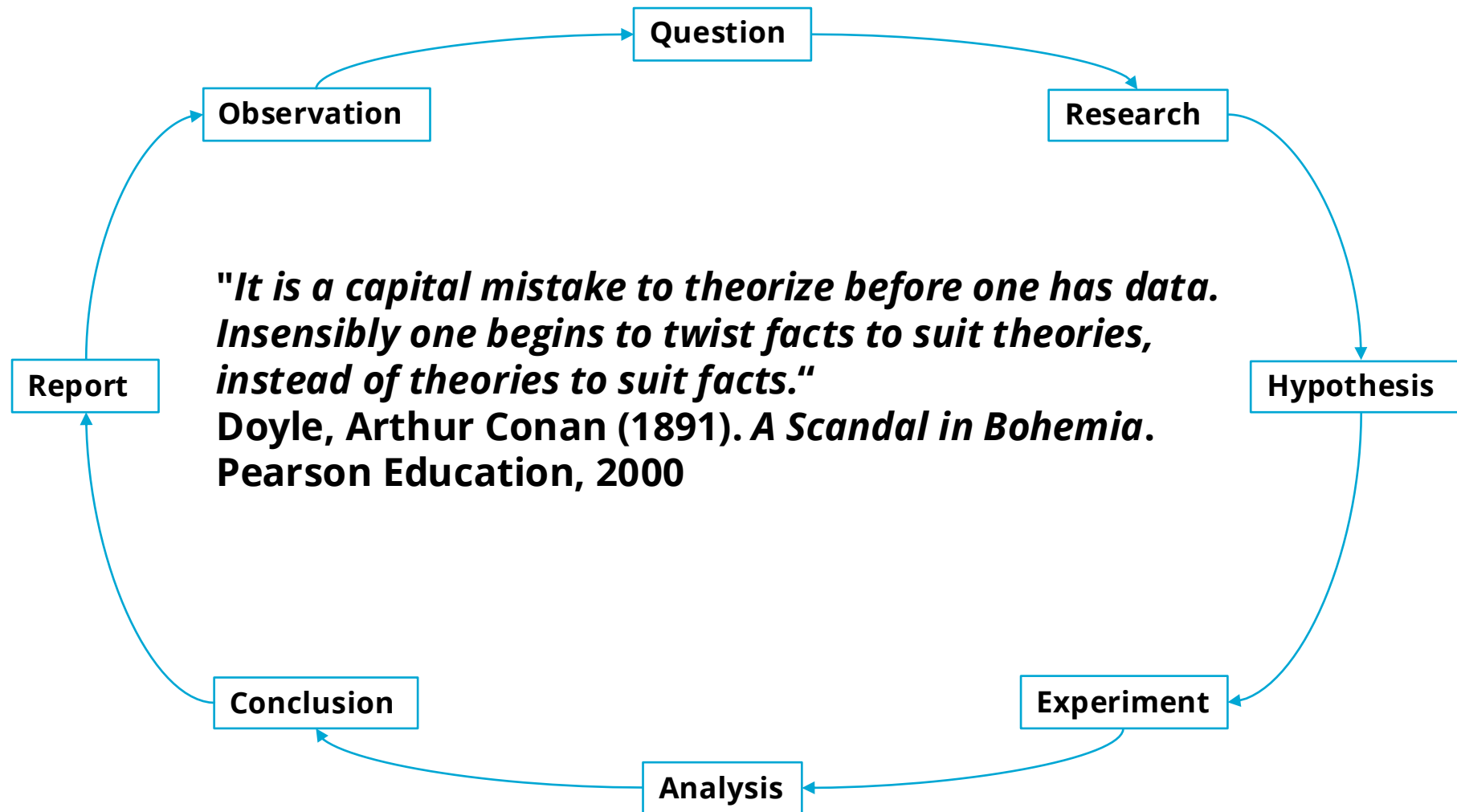


Image by qimono via Pixabay

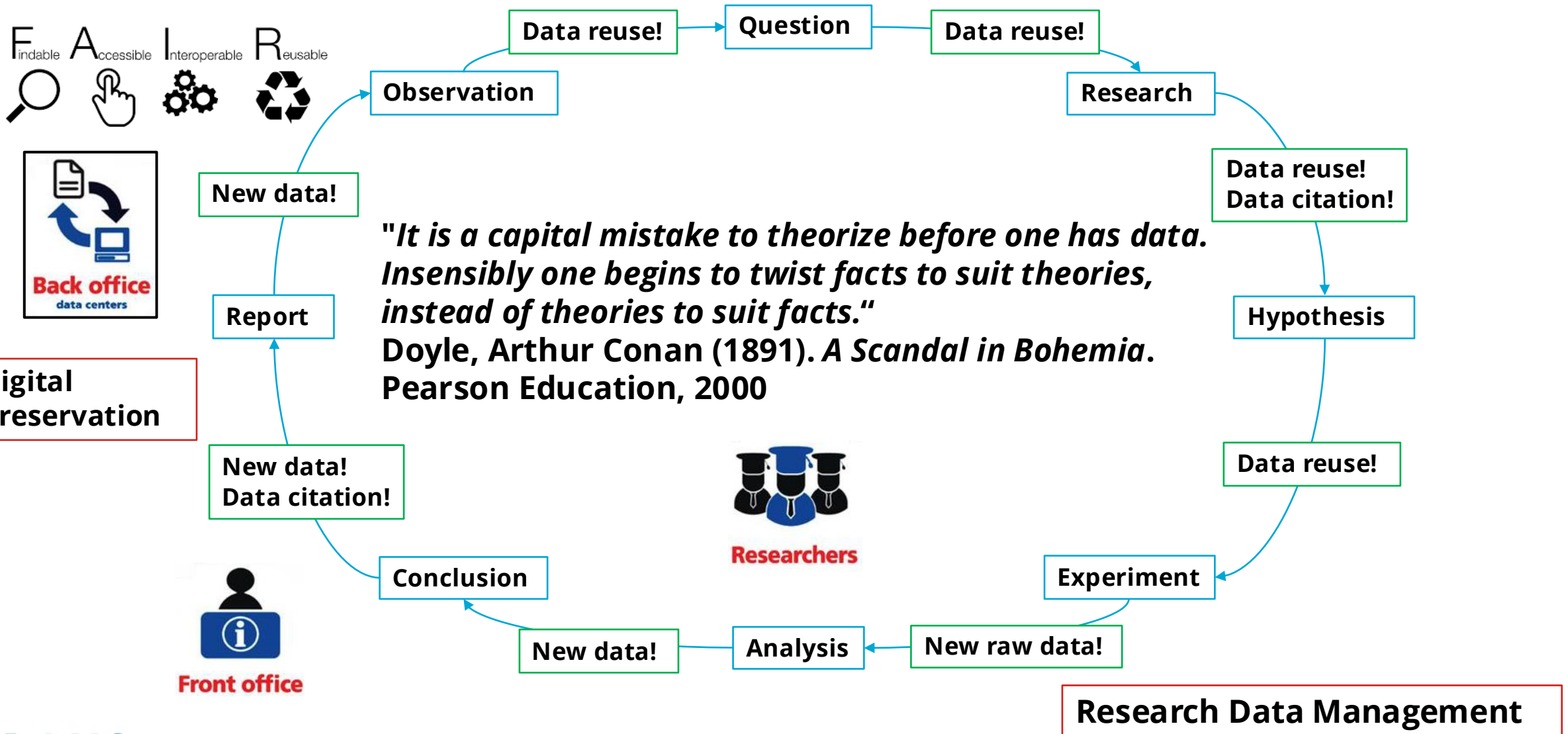
The Scientific Method



The Scientific Method (and Sherlock Holmes)



The Scientific Method (and Sherlock Holmes)



DANS Data Stations



DANS Data Station Social Sciences and Humanities

This Data Station allows you to deposit and search for data within the field of SSH

657 Downloads

Search this dataverse... Advanced Search + Add Data

1 to 10 of 7,297 Results

- Strikes in the Netherlands, 1810-1995
Jun 5, 2023
J.H.A. van der Velden, 2000, "Strikes in the Netherlands, 1810-1995", <https://doi.org/10.17026/dans-zd4-mv9c>, DANS Data Station Social Sciences and Humanities, V3
The dataset contains data on all known strikes in the Netherlands during the period 1810-1998. Dataset also includes data on other incidents of labour unrest.
- Research project "That's What Friends are For"
Jun 5, 2023
Y.H.M. van den Berg, 2021, "Research project "That's What Friends are For", <https://doi.org/10.17026/dans-zfz-4znr>, DANS Data Station Social Sciences and Humanities, V2
Our relationships with our parents and our friends are considered the most essential and rewarding relationships in life. Especially in emerging adulthood, individuals spend a large and increasing amount of their time with others outside of the family. However, studies examining...
- The Effectiveness of the MLU in Amending the Bilateral Tax Treaty Network: (On) the Measure of Multilateral Success
Jan 31, 2023
D.M. Broekhuijsen, 2023, "The Effectiveness of the MLU in Amending the Bilateral Tax Treaty Network: (On) the Measure of Multilateral Success", <https://doi.org/10.17026/dans-2cv-g4jt>, DANS Data Station Social Sciences and Humanities, V2
Database related to article. In the article, the authors discuss the implications of their study on the effectiveness of the Multilateral Instrument (MLI) in terms of states that are parties to the MLI, covered tax treaties, and bilateral tax treaties whose application has been...
- Omvang, achtergrond en kosten aanvullend en particulier onderwijs
Jan 20, 2023
K van der Ven, 2023, "Omvang, achtergrond en kosten aanvullend en particulier onderwijs", <https://doi.org/10.17026/dans-zqr-92hp>, DANS Data Station Social Sciences and Humanities, V2
Verkenning naar de omvang, motieven en kosten van aanvullend en particulier onderwijs. Deze bestanden bevatten de respons van enquêtes in het kader van het onderzoek naar de omvang, kosten en motieven van aanvullend en particulier onderwijs. Resultaten van de enquêtes zijn terug...
- Transcriptions of Subscriptions in the Manuscript Witnesses included in the ECM of Mark
Jan 5, 2023
C.T. Elmelund, 2023, "Transcriptions of Subscriptions in the Manuscript Witnesses included in the ECM of Mark", <https://doi.org/10.17026/dans-zdf-b9sk>, DANS Data Station Social Sciences and Humanities, V1
Full new transcriptions of the subscriptions to Mark's Gospel in the manuscripts included in the ECM (Editio Critica Maior) of Mark with corrections to both the ECM apparatus and the transcriptions of the NT.VMR (<https://nvmr.uni-muenster.de/>) database. The transcriptions present...
- e-RUPI and Economics
Jun 21, 2023

DANS Data Station Social Sciences and Humanities

This Data Station allows you to deposit and search for data within the field of SSH

DANS Data Station Social Sciences and Humanities >

Interview data on "Mobility- and behaviour-based early-warning system after the first wave of COVID-19"

Embargued Version 2.0



F.A. Metz, 2023, "Interview data on "Mobility- and behaviour-based early-warning system after the first wave of COVID-19", <https://doi.org/10.17026/dans-z3c-54eq>, DANS Data Station Social Sciences and Humanities, V2

Cite Dataset

Learn about Data Citation Standards.

Access Dataset

Contact Owner

Share

Dataset Metrics

0 Downloads

Description

During the COVID-19 pandemic infectious disease models have increasingly guided policy, but also fueled a discussion about the role of models in policymaking. Little is known about how policymakers and public administrations deal with model uncertainties and related ambiguities. In contrast to the widely stated evidence-policy gap, the article argues that evidence-based policymaking is a collaborative effort shared among different roles in the policy advisory process. Whether projections guide policy depends on trustful relationships among those producing scientific knowledge and those translating it to political action. We conducted in-depth interviews with employees working along the policy advisory process in the case of Dutch regional public entities of disaster- and crisis response. Results reveal task-sharing among highly specialized roles. Knowledge producers tend to use model results and are more concerned with questions of uncertainties than those focusing on policy implications. Trustful relationships with advisors impact whether models are considered by policymakers.

Date: 2022-09-21

Subject

Social Sciences

Keyword

COVID-19, evidence-based policymaking, policy advisory process, crisis management, mathematical modelling, uncertainty, Dutch Safety Regions

License/Data Use Agreement



Files Metadata Terms Versions

Search this dataset...

Filter by
File Type: All Access: All

Sort

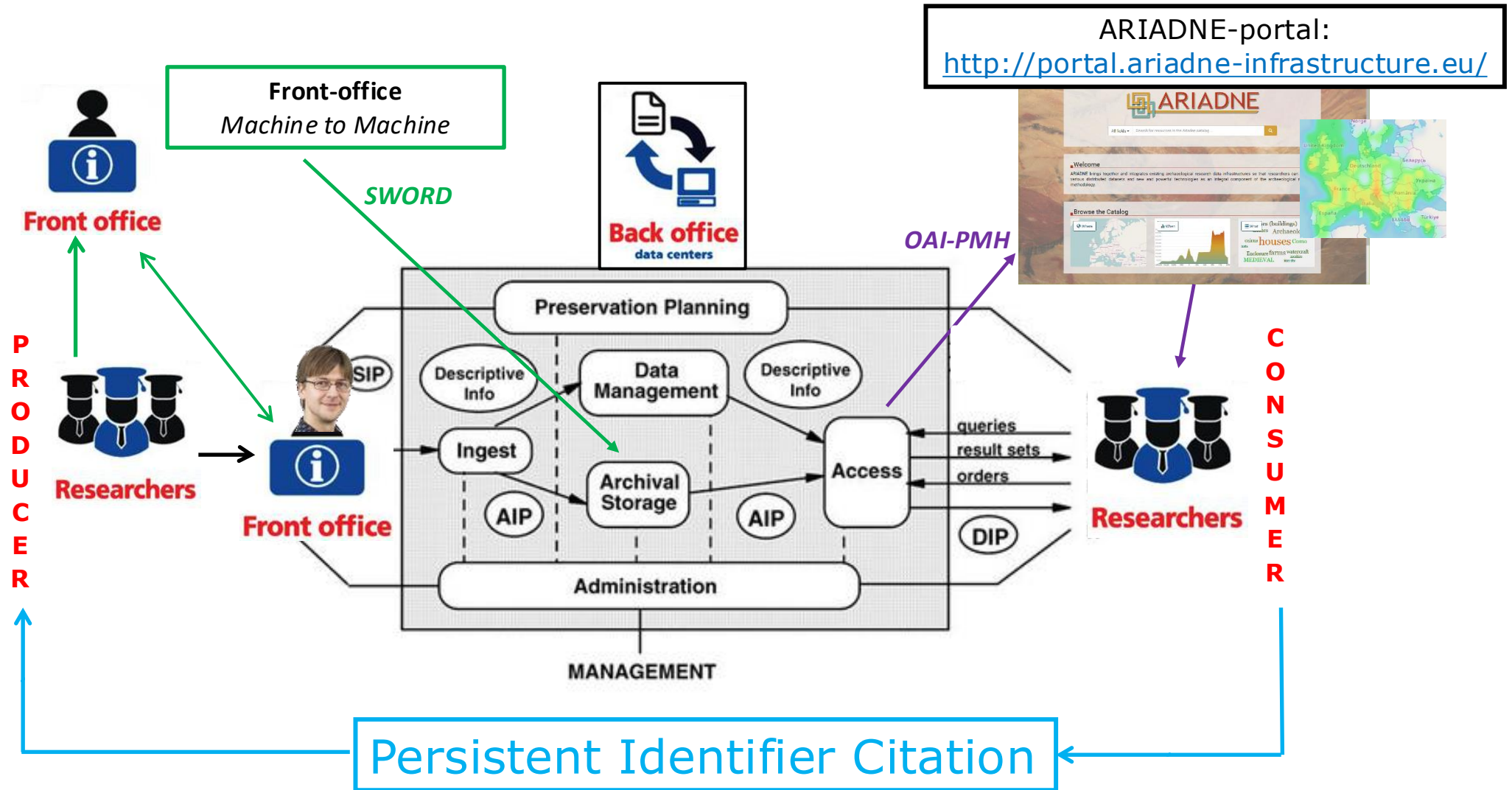
1 to 4 of 4 Files

Download

<https://ssh.datastations.nl/>; <https://archaeology.datastations.nl/>;
<https://lifesciences.datastations.nl/>; <https://phys-techsciences.datastations.nl/>



The magic(?) box of the repository



Dataset curation

2023-2025



Levels of Curation

Select all relevant types from:

A. Content distributed as deposited

B. Basic curation - e.g. brief checking, addition of basic metadata or documentation

C. Enhanced curation - e.g. conversion to new formats during ingest, enhancement of documentation and metadata

D. Data-level curation - as in C above, but with additional editing of deposited data

DANS

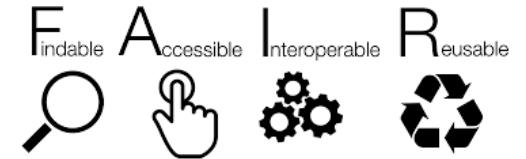
- Inventorise deposit

Basic curation

- Check if the dataset is valid and complete
- Check metadata, edit/modify where necessary
- Check for privacy sensitive data
- Publish dataset

Enhanced curation

- Migrate files to Preferred Formats
- Modify directory structure
- Make a new version



Conform *Data Stations Policy*; *DANS Selection Policy*; *Data Processing Manual*

Dataset curation

2026-2028



Levels of Curation and Preservation

Select all relevant types from:

Z. Level Zero. Content distributed as deposited.

D. Deposit Compliance with defined criteria, e.g. data formats, metadata elements, and compliance with legal and ethical norms

C. Initial Curation to meet defined criteria. This may include, e.g., the correction or enhancement of metadata and/or data content, or the creation of dissemination formats

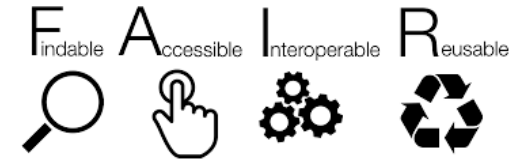
A. Active preservation. The repository takes long-term responsibility for ensuring that the data and metadata can be understood and rendered as required by the designated community for reuse.

DANS

- Inventorise deposit

Initial Curation

- Check if the dataset is valid and complete
- Check metadata, edit/modify where necessary
- Check for privacy sensitive data
- Publish dataset

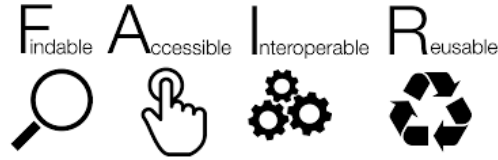


Active Preservation

- Migrate files to Preferred Formats
- Modify directory structure
- Make a new version

Conform *Data Stations Policy*; *DANS Selection Policy*; *Data Processing Manual*

Dataset curation



- **Validation** - DANS evaluates deposits on suitability for the Designated Community for the Data Station. DANS reserves the right to reject deposits submitted for evaluation if such datasets do not meet the criteria for acceptance.
- **Submission Package** - DANS always retains a copy of the data exactly as submitted by the depositor.
- **Initial Version** - The submitted data and curated metadata become the first version of a submitted dataset and form the basis for any preservation copies made of the dataset.
- **Enhancements** - DANS may make changes to the metadata and add conversion files to align them with community expectations, increase *FAIR* quality, and improve the long-term sustainability of the dataset. Changes to the data, and any changes to the metadata after the initial published version, will result in a new version of the dataset.
- **Publication** - If the dataset meets all criteria and after completion of the curatorial review, DANS will publish the dataset on behalf of the depositor and provide it with a *persistent identifier* (PID).

Conform *Data Stations Policy*; *DANS Selection Policy*; *Data Processing Manual*

Open Access metadata

Trustworthy repository

A

Standardised exchange protocols

Codebooks

Embargo when needed

Open Access when possible

Clear folder structure

F

Data citation

Access and licence information

Licence

Clear names for files and folders

Persistent Author Identifiers

F
Findable

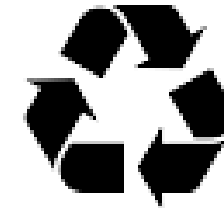
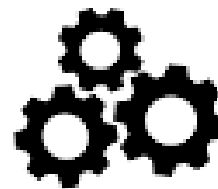
A
Accessible

I
Interoperable

R
Reusable

Authenticity

Metadata in scheme



Provenance

R

Preferred formats

Version control

Data standards

Workflows

Support metadata mapping

Common file formats

Vocabularies

I

APIs

Metadata

Title

Subtitle
Alternative Title
Alternative URL
Other Identifier
Author (+ROR; ORCID; ...)
Point of Contact
Description
Subject
Keyword
Notes
Language
Producer
Production Date
Production Location

Contributor
Funding Information
Distributor
Distribution date
Depositor
Time Period
Date of Collection
Data Type
Series
Software
Data Sources

SSH

Keyword Getty AAT
Keyword ELSST
Topic Classification CESSDA
Universe
Frequency
Sampling Procedure
Characteristics of Data Collection Situation
Actions to Minimize Losses
Weighting
Response Rate

Rights Holder
Personal Data in Dataset?
Language Of Metadata

Archaeology

Archis ID
Report (ABR)
Methods of Recovery (ABR)
Subject (ABR)
Artefact (ABR)
Temporal (ABR)
Keyword Getty AAT

Audience

Collection
Relation or Related Material

Temporal Coverage
Spatial Point
Spatial Box
Spatial Coverage

Access conditions & licenses

- [CC0-1.0](#) is a licence with which the copyright of the dataset is given up, putting the dataset in the public domain. A user may copy, change or distribute the dataset without notification. Formal attribution is not required, but this is of course still common practice in science.
- With [CC-BY-4.0](#) the copyright of the dataset remains with the rightsholder, but the user can share and change the dataset. This licence requires attribution and the user has to state if they have changed the dataset.
- With [CC-BY-SA-4.0](#) the same rules apply as for CC-BY-4.0, but in addition the user has to distribute the dataset they have reused under the same licence as the original (CC BY-SA 4.0).
- With [CC-BY-NC-4.0](#) the same rules apply as for CC-BY-4.0, but the dataset may not be used for commercial purposes.
- With [CC-BY-ND-4.0](#) the same rules apply as for CC-BY-4.0, but the user is not allowed to distribute the dataset if they have changed or extended the dataset.
- The licences [CC-BY-NC-ND-4.0](#) and [CC-BY-NC-SA-4.0](#) are a combination of the aforementioned licences.

Access conditions & licenses

pseudonymisation/anonymisation

The User will act in accordance with the Netherlands Code of Conduct for Research Integrity, the GDPR and other applicable laws and regulations.

The User will always cite the dataset in the research results they publish, in whatever form, when it has been used in the research.

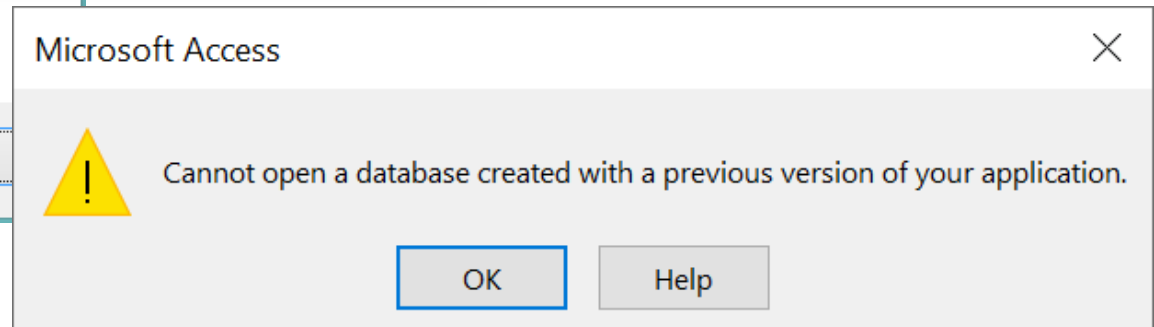
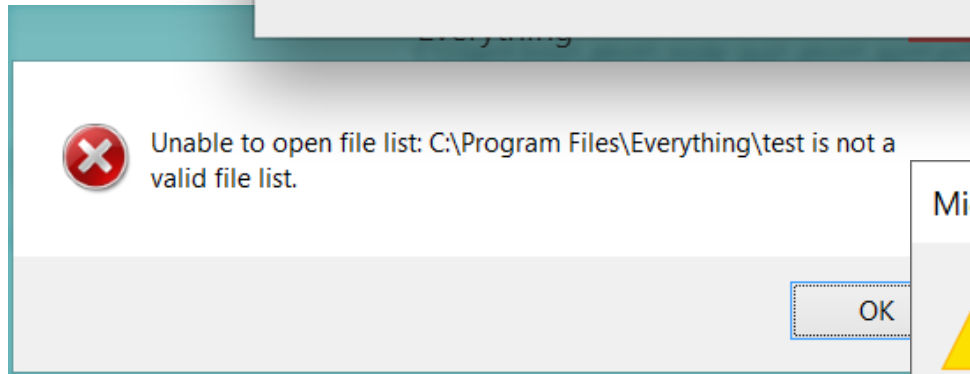
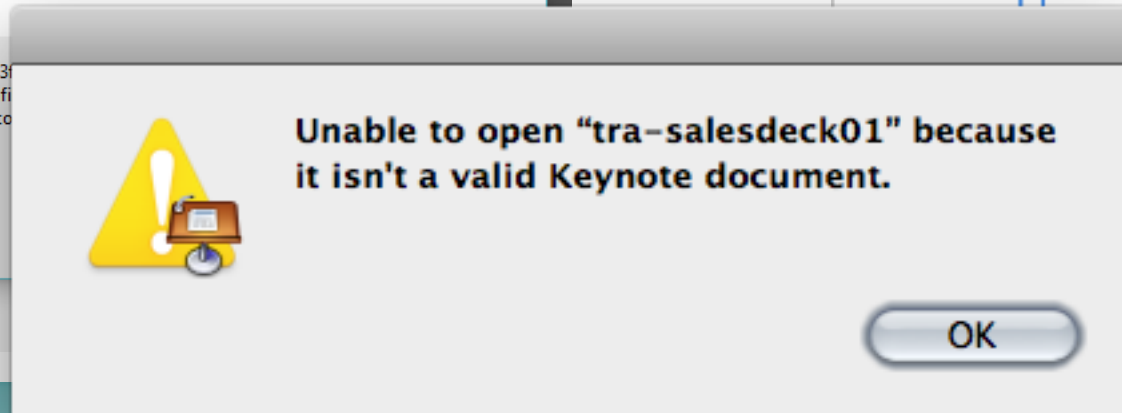
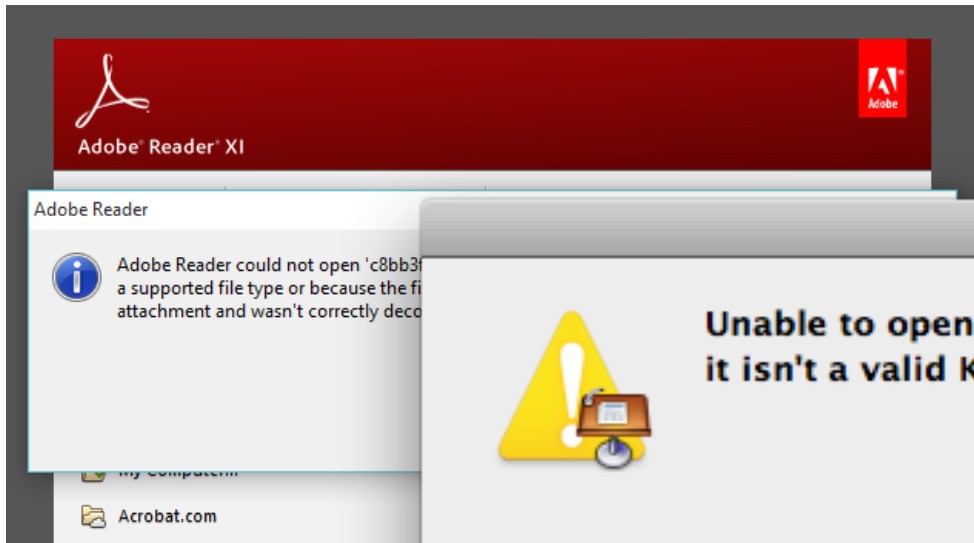
Embargo

For distribution or disclosure of the entire dataset or of substantial parts thereof, the User must first request permission from the holder of the rights to the dataset.

The User will always be responsible for the processing of personal data made available within the meaning of the GDPR and any other relevant privacy legislation, as well as for complying with any conditions set by the depositor.

Restricted access

Preferred Formats



Preferred Formats



=> TGA; RAW; CDR



=> PCX; BMP; PSD



=> JPG; TIF; PNG

Preferred Formats

UK Data Service Learning Hub page titled "Recommended formats". The page includes a search bar and navigation links for "Find data", "Deposit data", "Learning Hub", "Training and events", "About", "News", "Impact", "Help", and "Contact". The breadcrumb trail is: Home > Learning Hub > Research data management > Formatting data > Recommended formats.

Recommended formats

File formats recommended

The table contains guidance on file formats that you may need to convert your data files into. We welcome queries from researchers about the suitability of your file formats for the

Library of Congress page titled "Recommended Formats Statement". The page includes a search bar and navigation links for "ASK A LIBRARIAN", "DIGITAL COLLECTIONS", and "LIBRARY CATALOGS". The breadcrumb trail is: The Library of Congress > Preservation > Resources > Recommended Formats Statement.

Archaeology Data Service (ADS) page titled "File formats". The page includes a search bar and navigation links for "Search data", "Deposit data", "Help & guidance", "News & events", "Blog", and "About". The breadcrumb trail is: Help & guidance > Guides to Good Practice > File formats.

File formats

Keaton Niven, Archaeology Data Service / Digital Antiquity: Guides to Good Practice

National Archives website page titled "Lijst voorkeursformaten". The page includes a search bar and navigation links for "Home", "Onderzoeken", "Beleven", and "Archiveren". The breadcrumb trail is: Home > Archiveren > Voorkeursformaten Overheid.

Informatiesoort	Extensie	Versie en/of profiel	Formaat naam	Status	PUID	Opmerkingen
Afbeelding			Toggle Image	Voorkeur	fmt/353	
				Voorkeur	fmt/11 fmt/12 fmt/13	
				Acceptabel	fmt/42 fmt/43 fmt/44	
				Acceptabel	x-fmt/392 fmt/151	
				Acceptabel	fmt/3 fmt/4	
				Voorkeur	fmt/91 fmt/92 fmt/413	

Norm Voorkeursformaten

- Duurzame bestandsformaten
- Selectiecriteria voor bestandsformaten
- Voorkeursformaten Overheid
- Lijst voorkeursformaten
- Onderbouwing voorkeursformaten
- Voorkeursformaten in de praktijk
- Begrippenlijst
- Toelichting compressie
- Veelgestelde vragen
- Wijzigingenbeheer Norm Voorkeursformaten

printversie

Swedish National Data Service (SND) page titled "Choosing a file format". The page includes a search bar and navigation links for "Find Data", "Manage Data", "Describe & Share Data", "News & Events", "About Us", and "Contact". The breadcrumb trail is: Manage Data > Guides > Choosing a file format.

Choosing a file format

Manage Data

- Plan
- Organise

Preferred Formats

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

- are frequently used
- have open specification
- are independent of specific software, developers or vendors

In practice, it is not always possible to use formats which meet all the ideal characteristics of a preferred format.

<https://dans.knaw.nl/en/file-formats>

The screenshot shows the DANS website page for 'Preferred Formats'. The page includes a search bar, navigation links for 'Home', 'Data Stations', 'Data expertise', 'News', 'Agenda', and 'Ask your question'. The main content explains that preferred formats are those with long-term guarantees in terms of usability, accessibility, and sustainability. It lists criteria for preferred formats: frequently used, open specifications, and independence from specific software or vendors. A table summarizes preferred and non-preferred formats across various categories.

Type	Preferred format(s)	Non-preferred format(s)
Text documents	<ul style="list-style-type: none">• PDF/A (.pdf)• ODT (.odt)	<ul style="list-style-type: none">• Microsoft Word (.doc)• Office Open XML (.docx)• Rich Text File (.rtf)• PDF other than PDF/A (.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)• Markdown (.md)
Programming languages	<ul style="list-style-type: none">• MATLAB• NetCDF• Text-Fabric	
Spreadsheets	<ul style="list-style-type: none">• ODS (.ods)• CSV (.csv)	<ul style="list-style-type: none">• Microsoft Excel (.xls)• Office Open XML Workbook (.xlsx)• PDF/A (.pdf)
	<ul style="list-style-type: none">• SQL (.sql)	<ul style="list-style-type: none">• Microsoft Access (.mdb, .accdb)

Preferred Formats

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

- are frequently used
- have open specifications
- are independent of commercial developments

In practice, formats with the following characteristics are preferred:

<https://dans.knaw.nl/en/file-formats>

The screenshot shows the DANS website with a table titled "Preferred formats are file formats of which DANS – based on international agreements – is confident that they will offer the best long-term guarantees in terms of sustainability and accessibility." The table is divided into two columns: "Preferred Format" and "Non-preferred Format".

Preferred Format	Non-preferred Format
<ul style="list-style-type: none">• WaveFront Object (.obj)• Polygon file format (.ply)• X3D (.x3d)• glTF 2.0 (.gltf, .glb)• COLLADA (.dae)• LASer (.las, .laz)• IFC (.ifc)	<ul style="list-style-type: none">• Autodesk FBX (.fbx)• Blender (.blend)• glTF 1.0 (.gltf, .glb)• 3D PDF (.pdf)• Google Draco (.drc)• Artec (.a3d)• Agisoft Metashape (.psx & .psz)• STL (.stl)• VRML (.wrl, .wrz, .vrm)

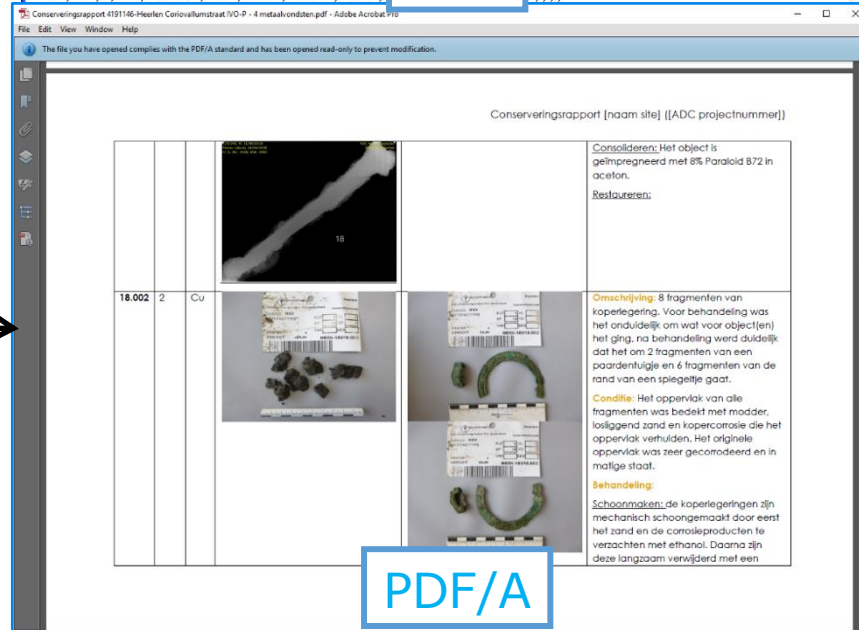
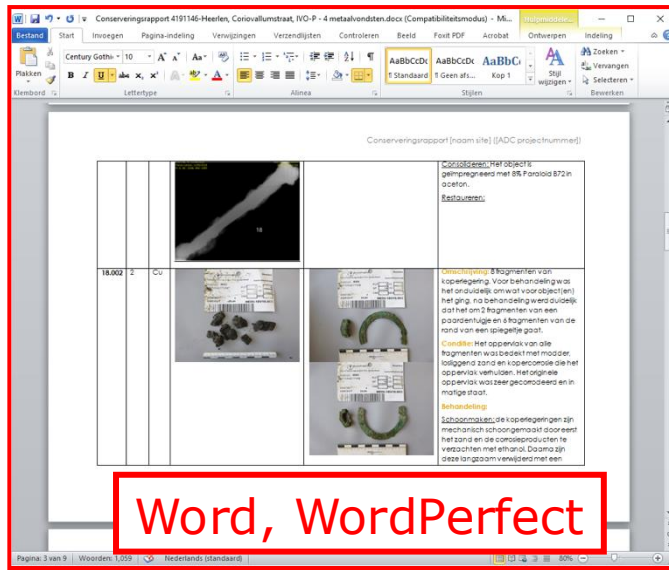
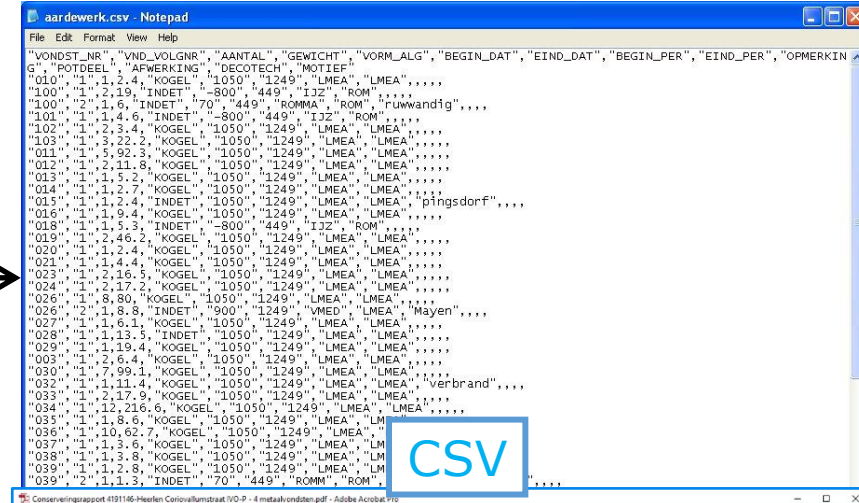
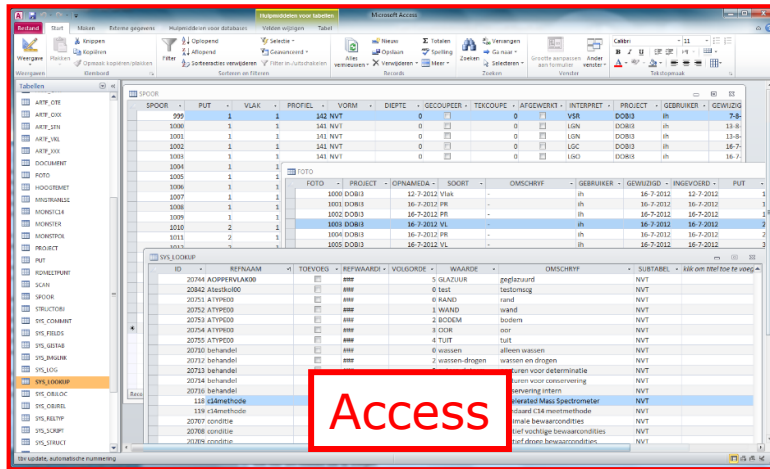
Below the table, there are sections for "Spreadsheets" and "Databases".

Spreadsheets	Databases
<ul style="list-style-type: none">• ODS (.ods)• CSV (.csv)	<ul style="list-style-type: none">• Microsoft Excel (.xls)• Office Open XML Workbook (.xlsx)• PDF/A (.pdf)
	<ul style="list-style-type: none">• SQL (.sql)• Microsoft Access (.mdb, .accdb)

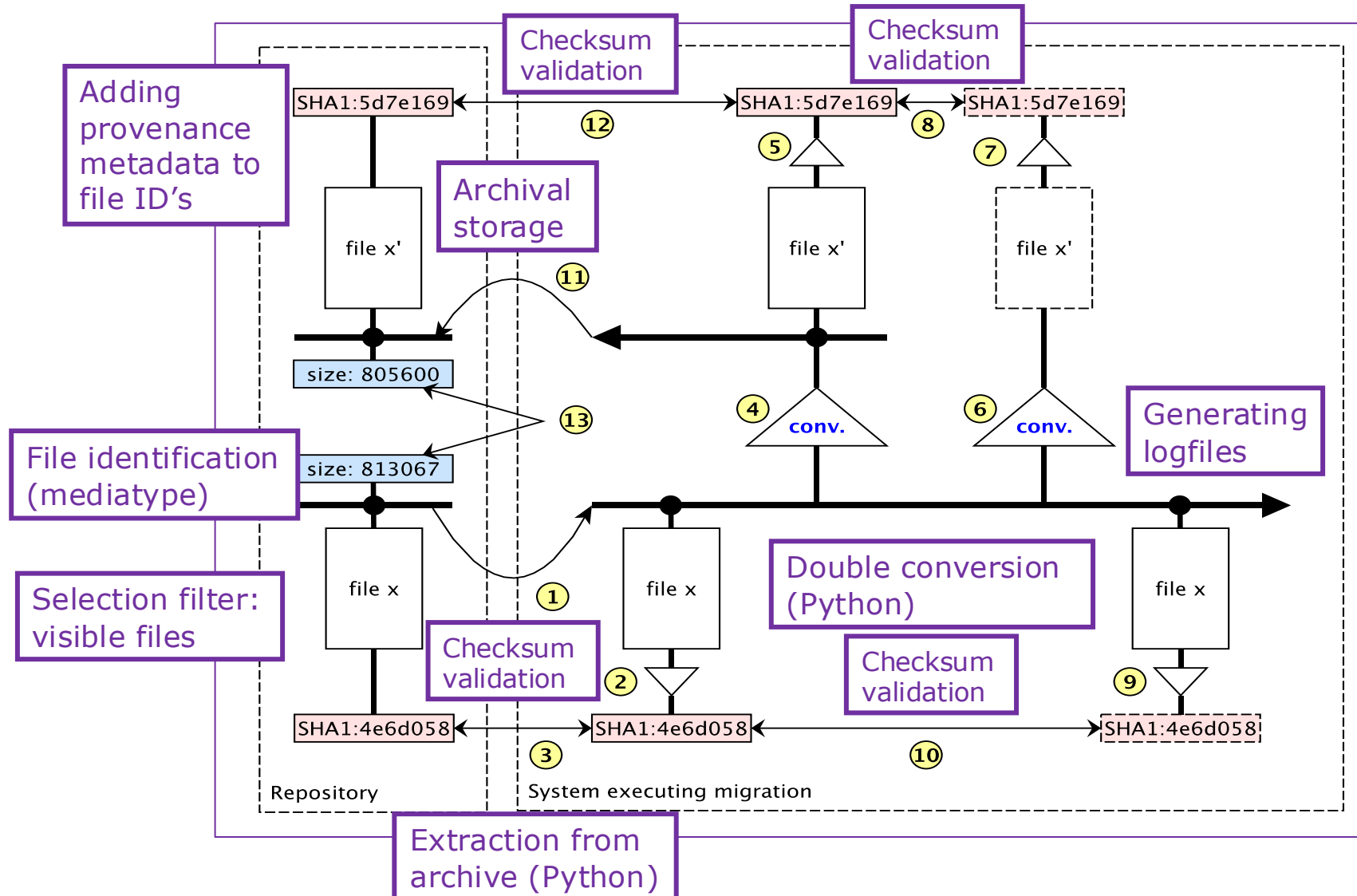
Preferred Formats

- Formats can still be in danger of falling into disuse and becoming unsupported. Long-term preservation should aim to make use of robust file formats
- If it is not possible to obtain 'preferred formats', non-preferred formats should still be accepted, although less long-term guarantees can be given
- When formats are converted/migrated, the original data should be preserved as well
- Certain non-preferred formats can be made available as current usage formats
- A policy with 'non-acceptable formats' risks receiving incomplete or lower-quality data / A policy with 'acceptable formats' risks a lack of effort towards sustainability
- Emulation may offer helpful solutions, but relying on emulation may simply shift preservation concerns from formats to software
- File format strategies need to be informed by expert input from data users and communities

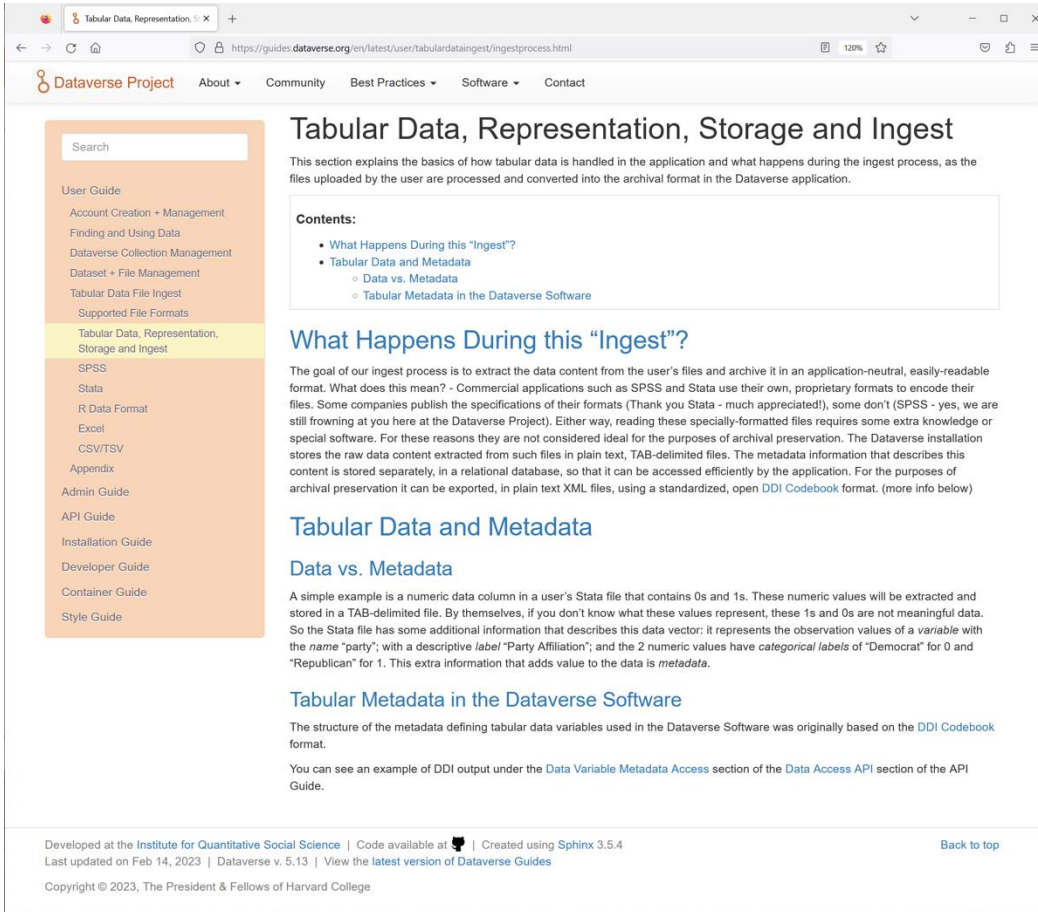
Enhanced Curation



Enhanced Curation



Tabular Ingest



The screenshot shows a web browser window displaying the Dataverse Project website. The page title is "Tabular Data, Representation, Storage and Ingest". The main content area includes a search bar, a navigation menu on the left, and a main text area. The navigation menu lists various guides, with "Tabular Data, Representation, Storage and Ingest" highlighted. The main text area contains a section titled "What Happens During this 'Ingest'?" and another titled "Tabular Data and Metadata".

Tabular Data, Representation, Storage and Ingest

This section explains the basics of how tabular data is handled in the application and what happens during the ingest process, as the files uploaded by the user are processed and converted into the archival format in the Dataverse application.

Contents:

- What Happens During this "Ingest"?
- Tabular Data and Metadata
 - Data vs. Metadata
 - Tabular Metadata in the Dataverse Software

What Happens During this "Ingest"?

The goal of our ingest process is to extract the data content from the user's files and archive it in an application-neutral, easily-readable format. What does this mean? - Commercial applications such as SPSS and Stata use their own, proprietary formats to encode their files. Some companies publish the specifications of their formats (Thank you Stata - much appreciated!), some don't (SPSS - yes, we are still frowning at you here at the Dataverse Project). Either way, reading these specially-formatted files requires some extra knowledge or special software. For these reasons they are not considered ideal for the purposes of archival preservation. The Dataverse installation stores the raw data content extracted from such files in plain text, TAB-delimited files. The metadata information that describes this content is stored separately, in a relational database, so that it can be accessed efficiently by the application. For the purposes of archival preservation it can be exported, in plain text XML files, using a standardized, open [DDI Codebook](#) format. (more info below)

Tabular Data and Metadata

Data vs. Metadata

A simple example is a numeric data column in a user's Stata file that contains 0s and 1s. These numeric values will be extracted and stored in a TAB-delimited file. By themselves, if you don't know what these values represent, these 1s and 0s are not meaningful data. So the Stata file has some additional information that describes this data vector: it represents the observation values of a *variable* with the *name* "party"; with a descriptive *label* "Party Affiliation"; and the 2 numeric values have *categorical labels* of "Democrat" for 0 and "Republican" for 1. This extra information that adds value to the data is *metadata*.

Tabular Metadata in the Dataverse Software

The structure of the metadata defining tabular data variables used in the Dataverse Software was originally based on the [DDI Codebook](#) format.

You can see an example of DDI output under the [Data Variable Metadata Access](#) section of the [Data Access API](#) section of the [API Guide](#).

Developed at the [Institute for Quantitative Social Science](#) | Code available at [GitHub](#) | Created using Sphinx 3.5.4
Last updated on Feb 14, 2023 | Dataverse v. 5.13 | [View the latest version of Dataverse Guides](#) [Back to top](#)
Copyright © 2023, The President & Fellows of Harvard College

- *The goal of our ingest process is to extract the data content from the user's files and archive it in an application-neutral, easily-readable format.*
- *The Dataverse installation stores the raw data content extracted from such files in plain text, TAB-delimited files. The metadata information that describes this content is stored separately, in a relational database, so that it can be accessed efficiently by the application. For the purposes of archival preservation it can be exported, in plain text XML files, using a standardized, open DDI Codebook format.*

Dataset versioning

DANS Data Station Social Sciences and Humanities

This Data Station allows you to deposit and search for data within the field of SSH

Publish Dataset

⚠ Are you sure you want to republish this dataset?

Due to the nature of changes to the current draft this will be a major release (2.0)

License/Data Use Agreement



CC-BY-NC-ND-4.0

License Description

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Continue

Cancel

and behaviour-based early-warning COVID-19"

ir-based early-warning system after the first wave
NS Data Station Social Sciences and Humanities,

Access Dataset ▾

Contact Owner

Share

Dataset Metrics ⓘ

0 Downloads ⓘ

sease models have increasingly guided policy, but
odels in policymaking. Little is known about how
al with model uncertainties and related ambiguities.
licy gap, the article argues that evidence-based
among different roles in the policy advisory process.
n trustful relationships among those producing
to political action. We conducted in-depth interviews
visory process in the case of Dutch regional public
ults reveal task-sharing among highly specialized
del results and are more concerned with questions of

uncertainties than those focusing on policy implications. Trustful relationships with advisors
impact whether models are considered by policymakers.

Date: 2022-09-21

Files

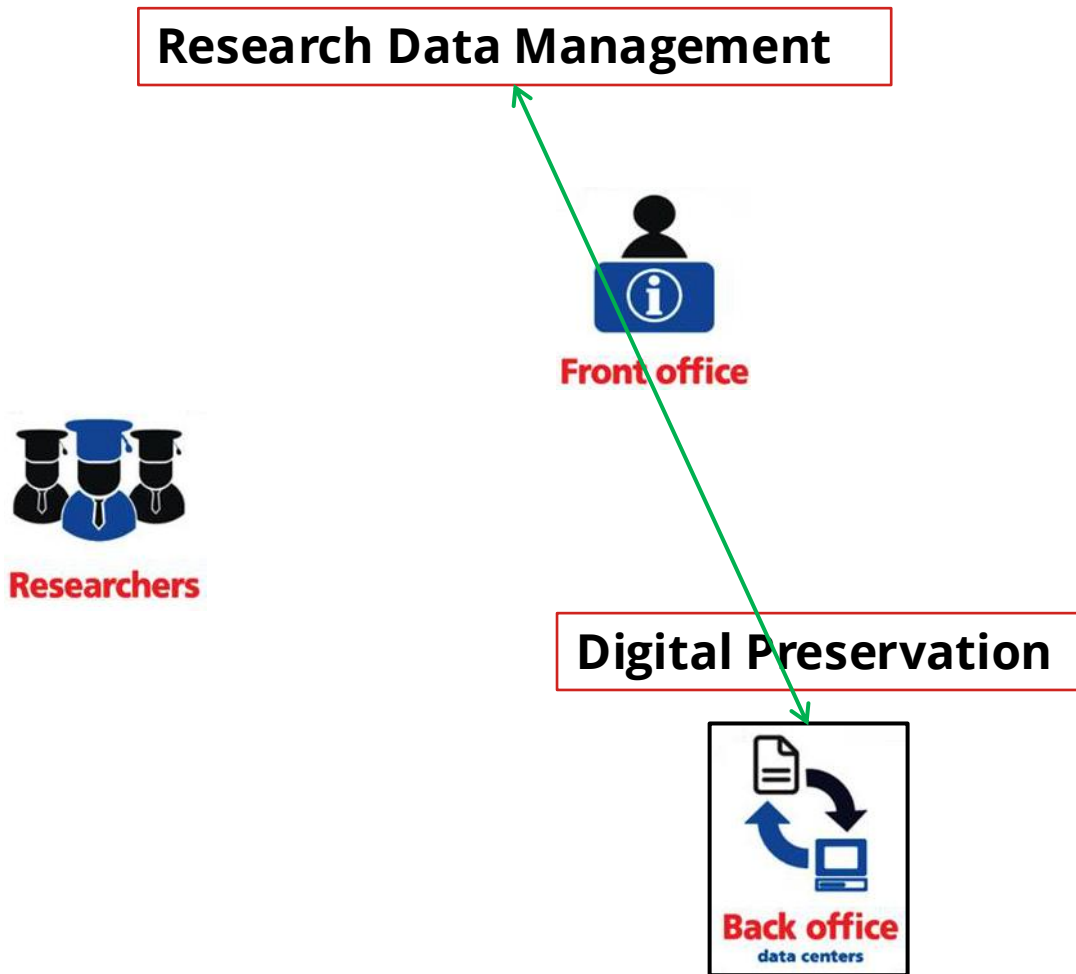
Metadata

Terms

Versions

Dataset Version	Summary	Contributors	Published on
2.0	Citation Metadata: DV PID Version (Changed); Bag ID (Changed); Files (Added: 1; Changed File Metadata: 1); View Details	V.L. Gilissen	2022-06-07
1.0	This is the first published version.	Chris Baars, Chris Baars, EASY Migration	2022-05-27

Research Data Management vs Digital Preservation



Hickman, Jonathan (w), Steve Epting (p, i), Rick Magyar (i), Frank D'Armata (c), Joe Caramagna (l), Tom Brevoort (e), Lauren Sankovitch (e); Incursion from *Memento Mori*, New Avengers Vol 3 #1 (March 2013), Marvel Comics

Research Data Management

Personal data – pseudonymisation/anonymisation

Selecting a repository



Front office

File names – anonymisation

Discipline-specific requirements

Selection of relevant data files

Selection of relevant datasets

Data provenance

Choosing a licence for reuse

Personal data – awareness of rules and regulations

Organising data in folders

Digital Preservation



Researchers

Choosing file formats

Providing dataset metadata

File formats policy

Restricting file access

File format conversions

Data authenticity

Documenting data: codebooks/instructions

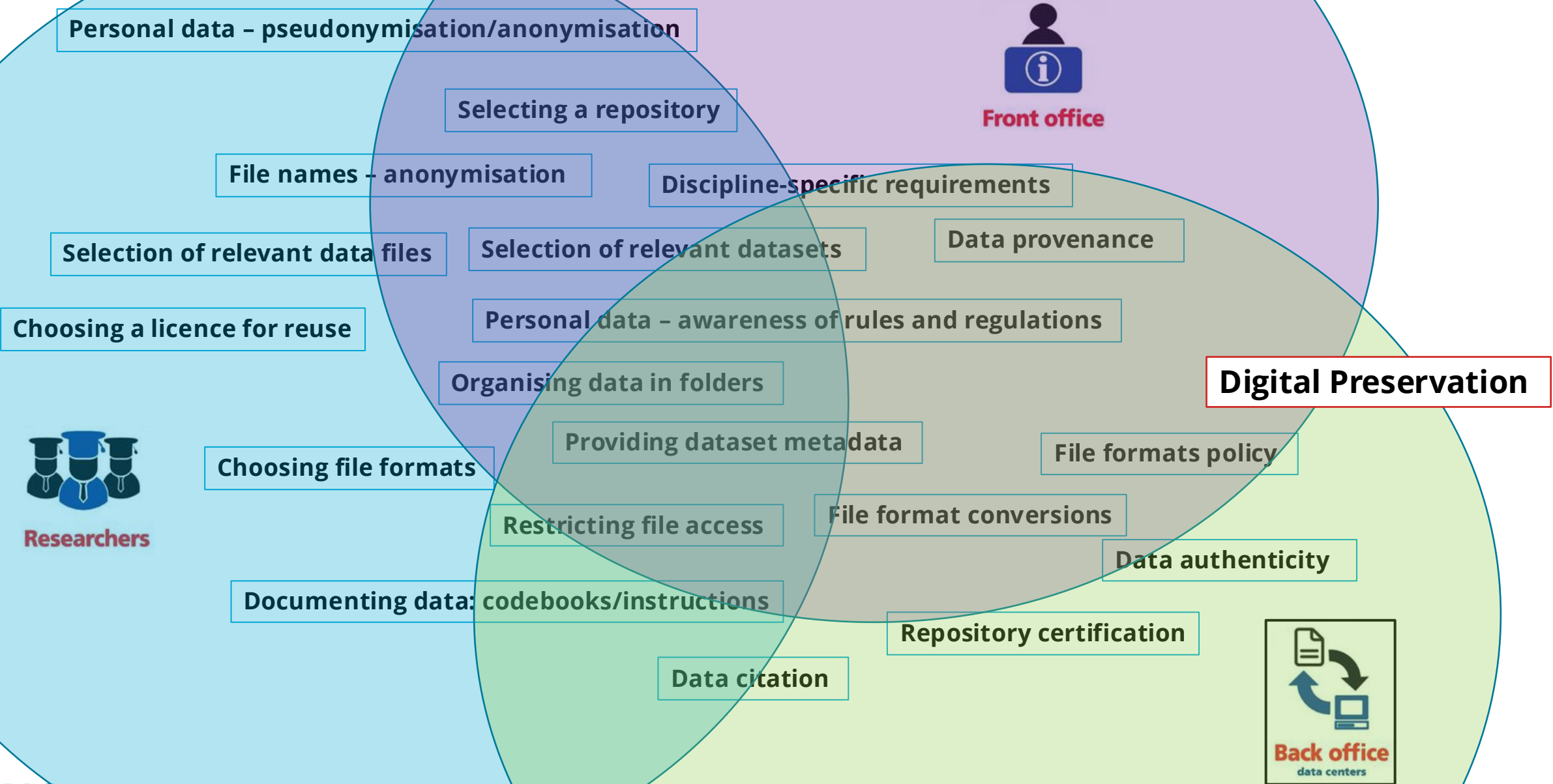
Repository certification

Data citation



Back office
data centers

Research Data Management



Digital Preservation

Initiatives

European **O**pen **S**cience **C**loud

Enhancing **D**igital preservation strategies at **E**uropean and **N**ational level

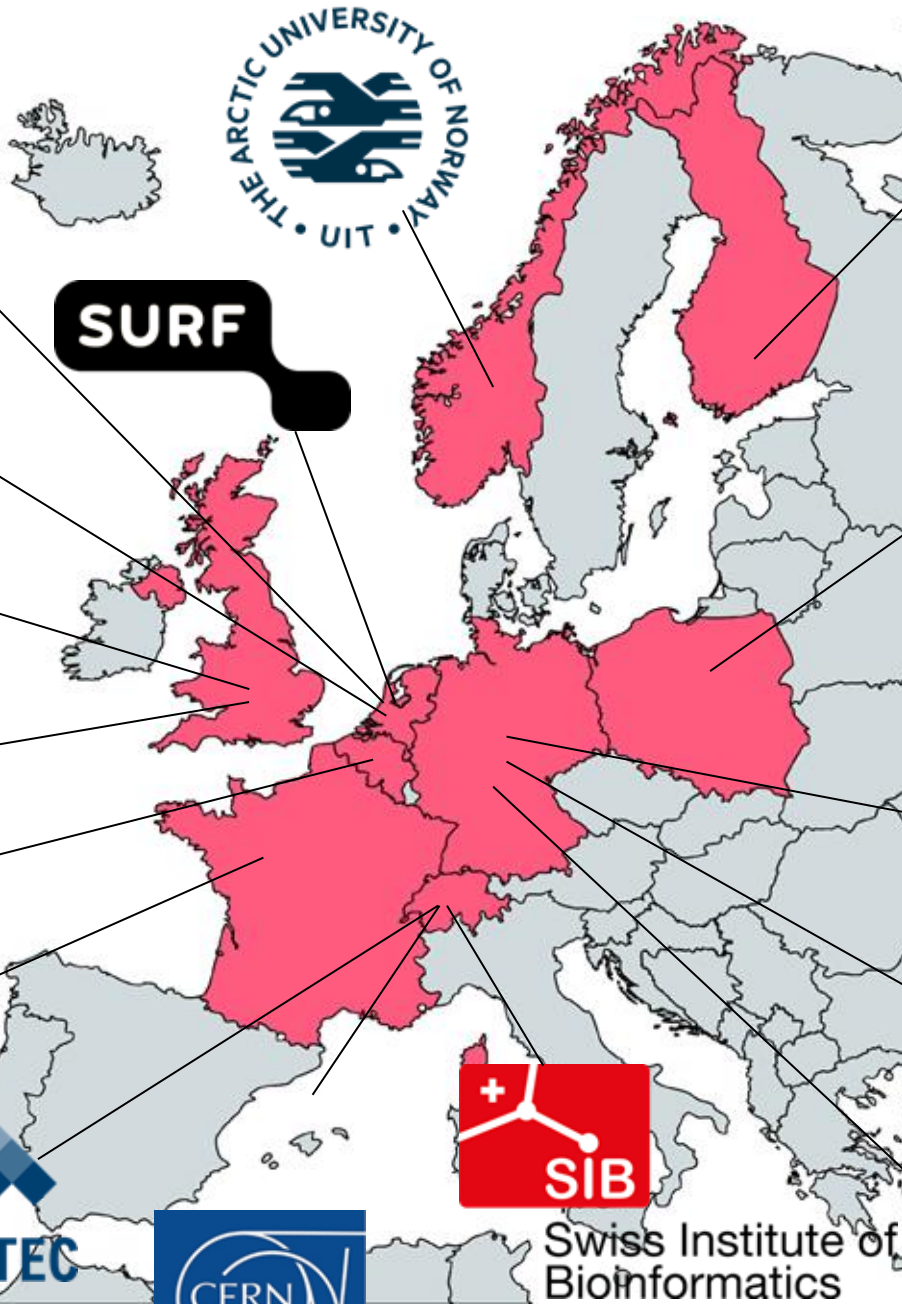


The EOSC EDEN project is funded by the European Union under the Horizon Europe framework programme and aims to support and promote digital preservation practices and standards at the European and national levels.

This support will be complemented by the development of user-centric tools, services, and standards designed to facilitate the creation of a distributed European infrastructure for digital data preservation, retention, and access.

These tools will be developed following extensive stakeholder mapping encompassing individuals, groups, and organisations with an interest in or influence on, the project's development, implementation, and outcomes.

Initiatives



16 partners from nine countries with expertise in:

- digital preservation
- data quality
- curation
- FAIR data management and services
- hosting of repositories and archives

Initiatives



Core Preservation Process (CPP) = a specific action that every Trustworthy Digital Archive (TDA) should undertake (either directly or through an associated service)

CPP-001 Checksum Generation & Recording

CPP-002 Checksum Validation

CPP-003 Integrity Checking

CPP-004 Data Corruption Management

CPP-005 Identifier Management

CPP-006 AIP Batch Export

CPP-007 Virus Scanning

CPP-008 File Format Identification

CPP-009 Metadata Extraction

CPP-010 File Format Validation

CPP-011 Replication

CPP-012 Risk Mitigation

CPP-013 Object Management Reporting

CPP-014 File Migration

CPP-015 Emulation & Rendering Tools

CPP-016 Metadata Ingest & Management

CPP-017 Disposal

CPP-018 Community Watch

CPP-019 Data Quality Assessment

CPP-020 Rights Management

CPP-021 AIP Versioning

CPP-022 Significant Properties Definition

CPP-023 Risk Definition and Extraction

CPP-024 Enabling Discovery

CPP-025 Enabling Access

CPP-026 File Normalisation

CPP-027 File Repair

CPP-028 Creation of Derivatives

CPP-029 Ingest

CPP-30 Refreshment

Initiatives



Enhancing Digital Preservation Strategies at European and National Level



	eden-fidelis.eu
	linkedin.com/company/eosc-eden
	@eosc-eden.bsky.social
	@EOSC-EDEN

#EOSCEDEN



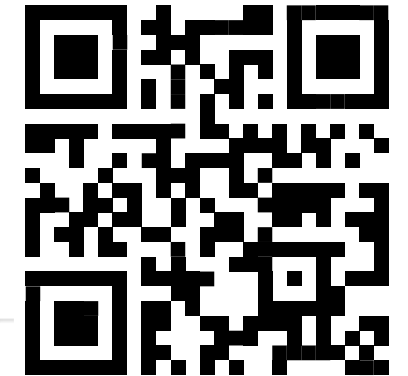
github.com/EOSC-EDEN



[EOSC EDEN Zenodo Community](#)

M1.1 Report on Identification of Core Preservation Processes

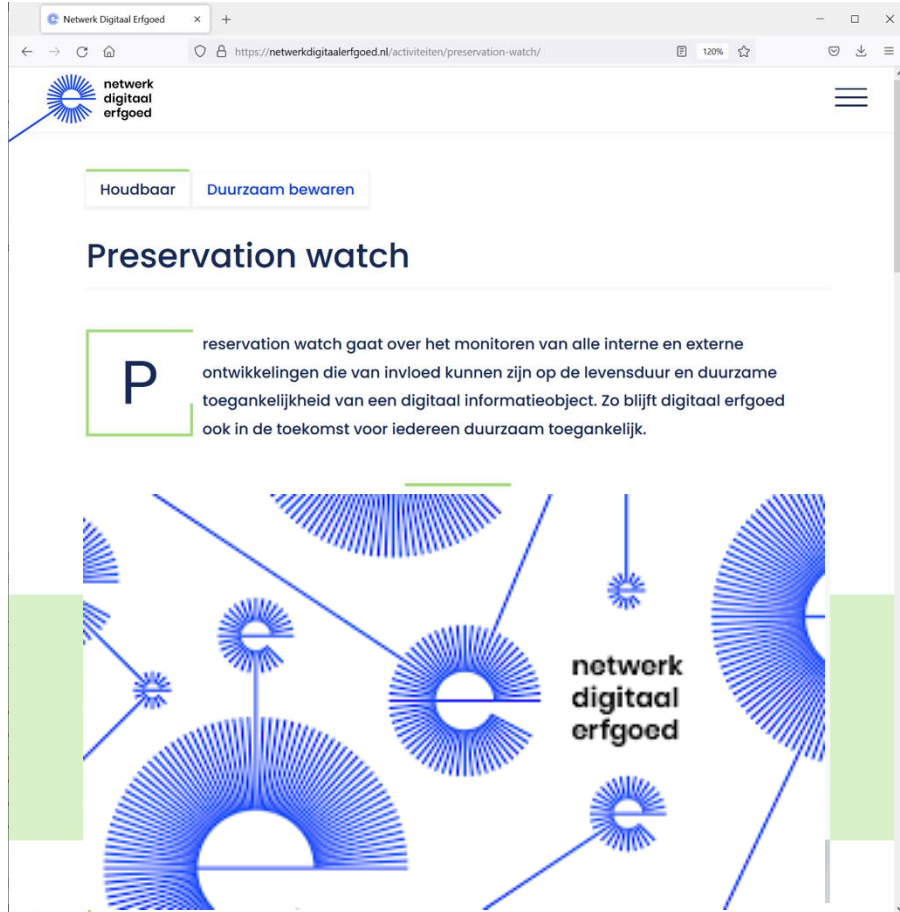
EOSC EDEN T1.2 ; Lindlar, Micky (Work package leader)¹ ;
Caron, Bertrand (Project leader)¹ ; Benauer, Maria (Project member)¹ ;
Kylander, Johan (Project member)² ; Dekeyser, Kris (Project member)³ ;
Addis, Matthew (Project member)⁴ ; Levlín, Mattias (Project member)² ;
Laukkanen, Mikko (Project member)² ; Lehtonen, Juha (Project member)² ;
Burger, Felix (Project member)¹ ; Koho, Tiina (Project member)² ;
Schwab, Franziska (Other)¹ ; Molloy, Laura (Other)⁵ ;
Zhang, Fen (Other)³ 



<https://doi.org/10.5281/zenodo.16992451>

edu link:
<https://edu.nl/4ecty>

Initiatives



Dutch Digital Heritage Network

KNAW Humanities Cluster (includes DANS)

National Library of the Netherlands

National Archive

Netherlands Institute for Sound and Vision

Het Nieuwe Instituut

Netherlands Cultural Heritage Agency

Preservation Watch

Netherlands Institute for Sound and Vision

DANS

EYE Filmmuseum

National Library of the Netherlands

LIMA

National Archive

RHC Eindhoven

Amsterdam City Archives

The Utrecht Archives

Initiatives

www.wegwijzervoorkeursformaten.nl

Stappenplan

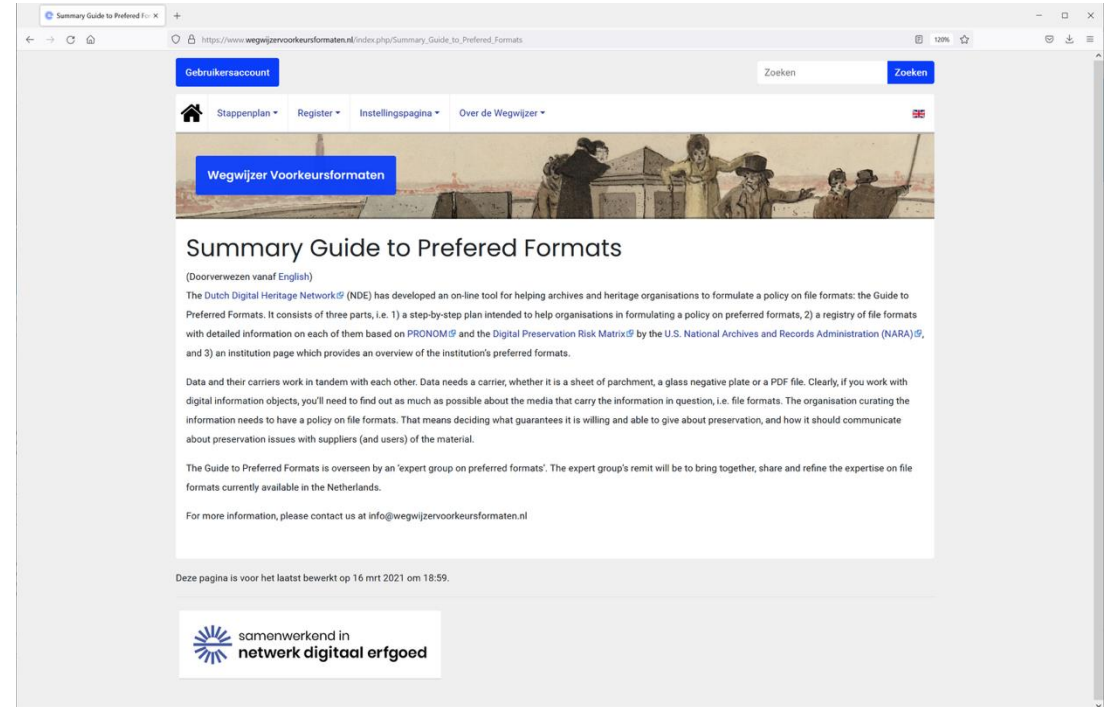
Interactive **step-by-step module** to help institutions create their own file formats policies through comparisons with other policies, via available information and weighing importance

Register

Register of file formats with information derived from international sources: PRONOM; Wikidata; Library of Congress; NARA; COPTR

Instellingspagina

Institution pages with overviews derived from the step-by-step plans and the register



Initiatives

<https://coptr.digipres.org>

Community Owned digital Preservation Tool Registry (COPTR)

COPTR helps practitioners find tools needed for [long term digital preservation](#) tasks. It describes 608 tools and 36 workflows.

- Find a digital preservation tool using the **Tools Grid** (or by stage, function, content type or file format). Or watch this video to [learn more](#).
- View digital preservation workflows
- Add a tool, or find out more about [how you can get involved](#)
- Add a workflow
- About COPTR, including how to [Contact Us](#), how COPTR is structured and how to access COPTR data via the API.

NEW FOR JULY 2021: COPTR has been re-launched with new functionality. Find out all about it in this [video presentation](#).

COPTR Partners

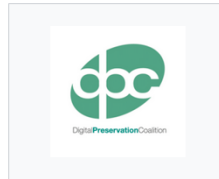
COPTR was created and launched with the support of the [Aligning National Approaches to Digital Preservation](#) initiative and has been populated and maintained by members of the COPTR partner organisations:



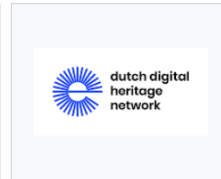
The Digital Curation Centre (DCC) [↗](#)



The Digital Curation Exchange (DCE) [↗](#)



The Digital Preservation Coalition (DPC) [↗](#)



Dutch Digital Heritage Network (DDHN) [↗](#)



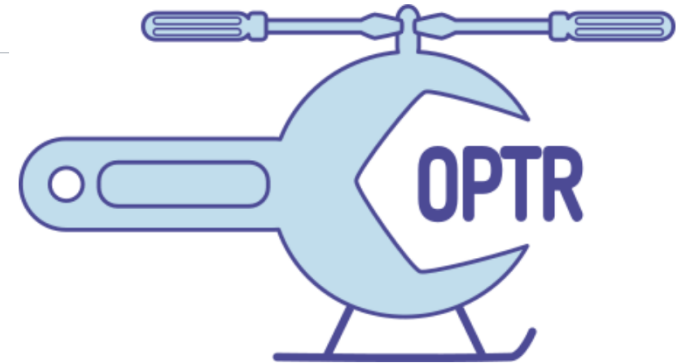
National Digital Stewardship Alliance (NDSA) [↗](#)



Network of expertise in long-term storage of digital resources in Germany (NESTOR) [↗](#)



The Open Preservation Foundation (OPF) [↗](#)



COPTR Home
Community Owned Workflows
Recent changes

Find tools

Tools grid
By lifecycle stage
By function
By content type
By file format
All tools

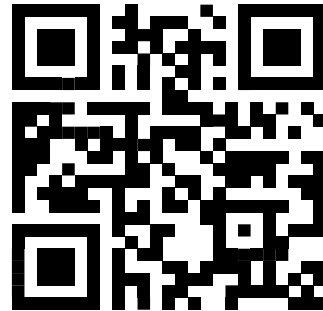
Help

About COPTR
Video guides to using COPTR
Mediawiki help
Wikitext cheat sheet
COPTR data structures

Tools

What links here
Related changes
Special pages
Printable version
Permanent link
Page information
Cite this page
Browse properties

Thank you for your attention



DANS Open Day
Open data, open science



DANS

Anna van Saksenlaan 51 | 2593 HW The Hague | The Netherlands | +31 (0)88 003 46 66
National centre of expertise and repository for research data | An institute of the KNAW and NWO

✉ [DataLink](#) [in LinkedIn](#) [🦋 BlueSky](#) [🐙 Mastodon](#) [🌐 www.dans.knaw.nl](#)