

Assessing the FAIRness of Datasets in Trustworthy Digital Repositories: A Proposal



The First Big Set - Going up of Course!



STill Going.....



Made it!! looks steeper from the top



William Henry Playfair
1790-1857

Peter Doorn, Director DANS
Ingrid Dillo, Deputy Director DANS



@pkdoorn @dansknaw

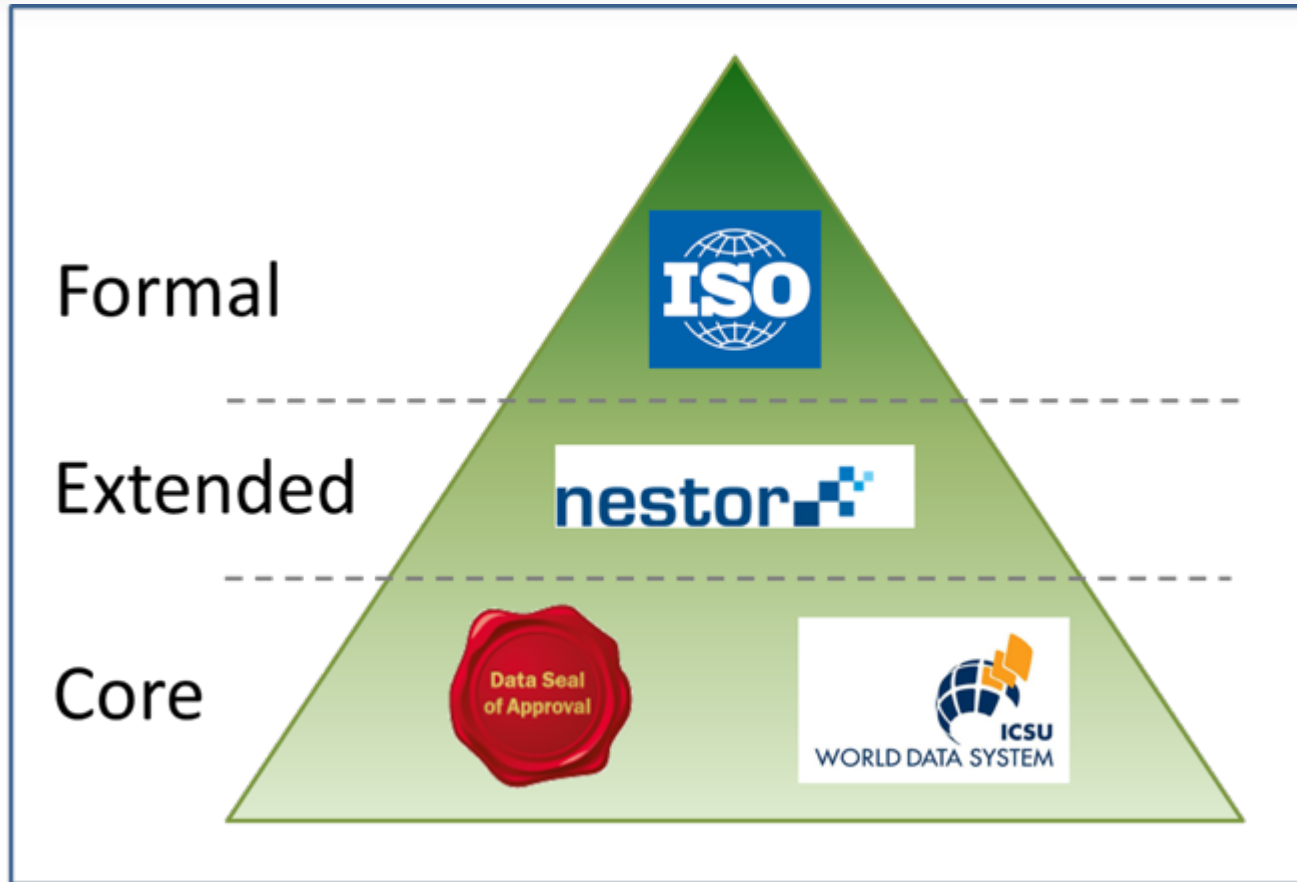


DANS and DSA



- 2005: DANS to promote and provide permanent access to digital research information
- Formulate quality guidelines for digital repositories including DANS
- 2006: 5 basic principles as basis for 16 DSA guidelines
- 2009: international DSA Board
- Almost 70 seals acquired around the globe, but with a focus on Europe

The certification landscape



DSA and WDS: look-a-likes

Communalities:

- Lightweight, community review

Complementarity:

- Geographical spread
- Disciplinary spread



Partnership



Goals:

- Realizing efficiencies
- Simplifying assessment options
- Stimulating more certifications
- Increasing impact on the community



Outcomes:

- Common catalogue of requirements for core repository assessment
- Common procedures for assessment
- Shared testbed for assessment

New common requirements

- Context (1)
- Organizational infrastructure (6)
- Digital object management (8)
- Technology (2)
- Additional information and applicant feedback (1)



Requirements dealing with data quality

R2. The repository maintains all applicable **licenses** covering data access and use and monitors compliance.

R3. The repository has a **continuity plan** to ensure ongoing access to and preservation of its holdings.

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with **disciplinary and ethical norms**.

R7. The repository guarantees the **integrity and authenticity** of the data.

Requirements dealing with data quality

R8. The repository accepts data and metadata based on **defined criteria to ensure relevance and understandability** for data users.

R10. The repository assumes responsibility for **long-term preservation** and manages this function in a planned and documented way.

R11. The repository has appropriate expertise to address **technical data and metadata quality** and ensures that sufficient information is available for end users to make quality-related evaluations.

R13. The repository enables users to **discover the data** and **refer to them in a persistent way** through proper citation.

R14. The repository enables reuse of the data over time, ensuring that **appropriate metadata** are available to support the understanding and use of the data.

New requirements are out now!

RESEARCH DATA SHARING WITHOUT
BARRIERS

RDA EU

RDA US

CONTACT US

MY PROFILE

LOGOUT

SUPPORT



ABOUT RDA

GET INVOLVED

GROUPS

RECOMMENDATIONS &
OUTPUTS

RDA FOR DISCIPLINES

PLENARIES

EVENTS

NEWS & MEDIA

WDS and DSA Announce Unified Requirements for Core Certification of Trustworthy Data Repositories developed through the RDA DSA-WDS partnership Working Group

Home » News & Articles » WDS And DSA Announce Unified Requirements For Core Certification Of Trustworthy Data Repositories Developed Through The RDA DSA-WDS Partnership Working Group

<http://www.datasealofapproval.org/en/news-and-events/news/2016/11/25/wds-and-dsa-announce-uni-ed-requirements-core-cert/>

<https://www.icsu-wds.org/news/news-archive/wds-dsa-unified-requirements-for-core-certification-of-trustworthy-data-repositories>

Data Archiving and Networked Services



Resemblance DSA – FAIR principles

DSA Principles (for data repositories)	FAIR Principles (for data sets)
data can be found on the internet	Findable
data are accessible	Accessible
data are in a usable format	Interoperable
data are reliable	Reusable
data can be referred to	(citable)

The resemblance is not perfect:

- usable format (DSA) is an aspect of interoperability (FAIR)
- FAIR explicitly addresses machine readability
- etc.

A certified TDR already offers a baseline data quality level

Combine and operationalize: DSA & FAIR

- Growing demand for quality criteria for research datasets and a way to assess their fitness for use
- Combine the principles of core repository certification and FAIR
- Use the principles as quality criteria:
 - Core certification – digital repositories
 - FAIR – research data (sets)
- Operationalize the principles as an instrument to assess FAIRness of existing datasets in certified TDRs





Badges for assessing aspects of data quality and “openness”





These badges do not define good practice, they certify that a particular practice was followed.



 **BRONZE:** data is openly licensed, available with no restrictions, accessible and legally reusable.

 **SILVER:** satisfies the Bronze requirements, the data is documented in a machine readable format, reliable and offers ongoing support from the publisher via a dedicated communication channel.

 **GOLD:** satisfies the Silver requirements, is published in an open standard machine readable format, has guaranteed regular updates, offers greater support, documentation, and includes a machine readable rights statement.

 **PLATINUM:** satisfies the Gold requirements, has machine readable provenance documentation, uses unique identifiers in the data, the publisher has a communications team offering support. This is an exceptional example of an information infrastructure.

- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ make it available in a non-proprietary open format (e.g., CSV as well as of Excel)³
- ★★★★ use URIs to denote things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵

[5-star deployment scheme](#) for Open Data

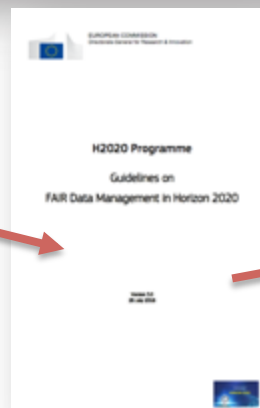
Sources: Open data institute (UK), Centre for open science (US), Tim-Berners Lee

Different implementations of FAIR

Creation



Requirements for new data creation



RISK ALERT



Assessment



Establishing the profile for existing data

Transformation



Transformation tools to make data FAIR (Go-FAIR initiative)

FAIR badge scheme



2 User Reviews
1 Archivist Assessment
24 Downloads

- First Badge System based on the FAIR principles: proxy for data quality assessment
- Operationalise the original principles to ensure no interactions among dimensions to ease scoring
- Consider Reusability as the resultant of the other three:
 - the average FAIRness as an indicator of data quality
 - $(F+A+I)/3=R$
- Manual and automatic scoring

Findable (defined by metadata (PID included) and documentation)

1. No PID nor metadata/documentation
2. PID without or with insufficient metadata
3. Sufficient/limited metadata without PID
4. PID with sufficient metadata
5. Extensive metadata and rich additional documentation available



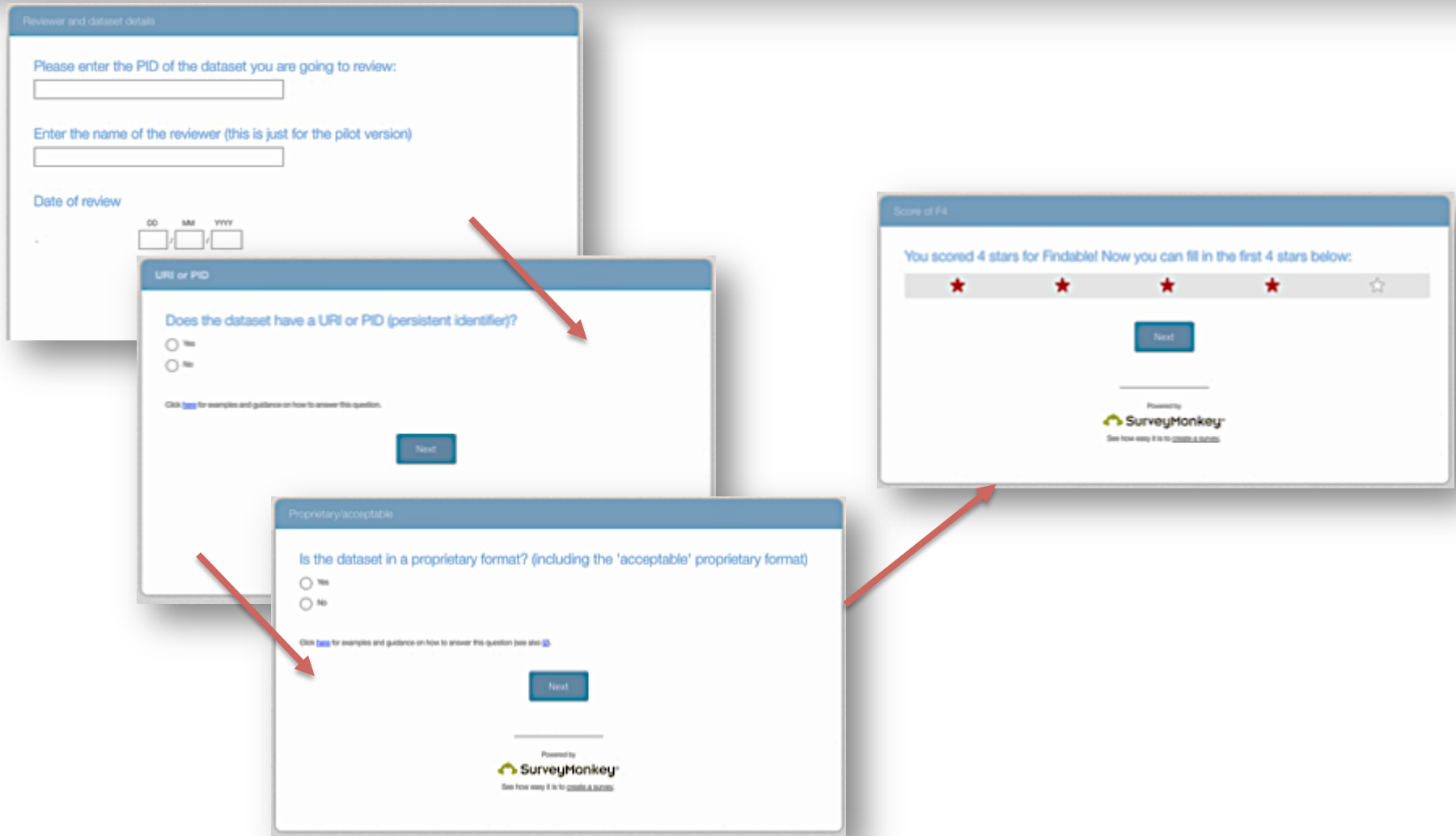
Accessible (defined by presence of user license)

1. Metadata nor data are accessible
2. Metadata are accessible but data is not accessible (no clear terms of reuse in license)
3. User restrictions apply (i.e. privacy, commercial interests, embargo period)
4. Public access (after registration)
5. Open access unrestricted

Interoperable (defined by data format)

1. Proprietary (privately owned), non-open format data
2. Proprietary format, accepted by Certified Trustworthy Data Repository
3. Non-proprietary, open format = 'preferred format'
4. As well as in the preferred format, data is standardised using a standard vocabulary format (for the research field to which the data pertain)
5. Data additionally linked to other data to provide context

Creating a FAIR data assessment tool



Reviewer and dataset details

Please enter the PID of the dataset you are going to review:

Enter the name of the reviewer (this is just for the pilot version)

Date of review

DD MM YYYY

URI or PID

Does the dataset have a URI or PID (persistent identifier)?

☐ Yes

☐ No

[Click here](#) for examples and guidance on how to answer this question.

Next

Proprietary/acceptable

Is the dataset in a proprietary format? (including the 'acceptable' proprietary format)

☐ Yes

☐ No

[Click here](#) for examples and guidance on how to answer this question (see also [5](#)).

Next

Score of F4

You scored 4 stars for Findable! Now you can fill in the first 4 stars below:

★ ★ ★ ★ ☆

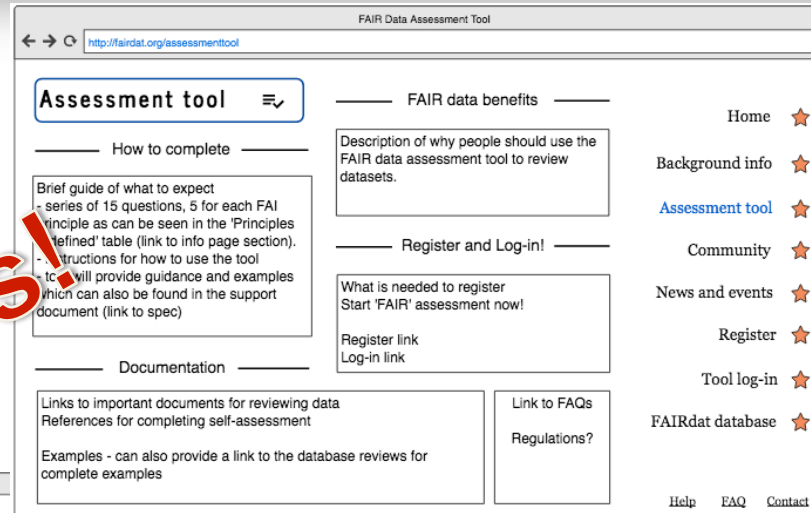
Next

Powered by **SurveyMonkey**

See how easy it is to create a survey

Website FAIRDAT

Neutral, Independent
Analogous to DSA website



FAIR Data Assessment Tool

Assessment tool

How to complete

Brief guide of what to expect

- series of 15 questions, 5 for each FAIR principle as can be seen in the 'Principles defined' table (link to info page section).
- instructions for how to use the tool
- tool will provide guidance and examples which can also be found in the support document (link to spec)

Documentation

Links to important documents for reviewing data

References for completing self-assessment

Examples - can also provide a link to the database reviews for complete examples

FAIR data benefits

Description of why people should use the FAIR data assessment tool to review datasets.

Register and Log-in!

What is needed to register

Start 'FAIR' assessment now!

Register link

Log-in link

Link to FAQs

Regulations?

Home

Background info

Assessment tool

Community

News and events

Register

Tool log-in

FAIRDat database

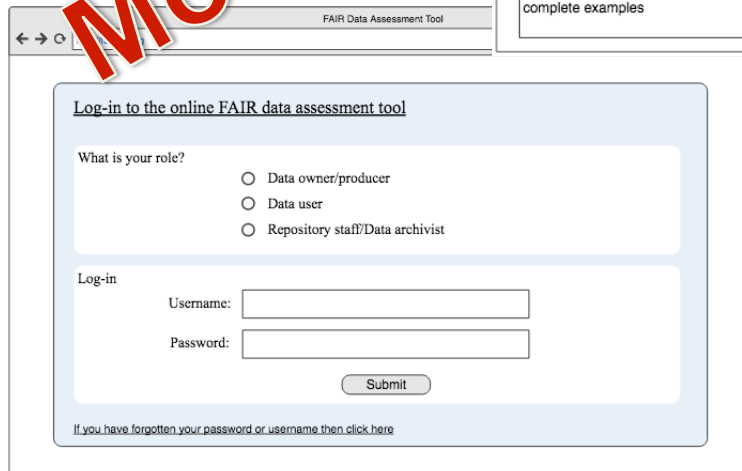
Help

FAQ

Contact

To contain FAIR data assessments from any repository or website, linking to the location of the data set via (persistent) identifier

The repository can show the resultant badge, linking back to the FAIRDAT website



FAIR Data Assessment Tool

Log-in to the online FAIR data assessment tool

What is your role?

☐ Data owner/producer

☐ Data user

☐ Repository staff/Data archivist

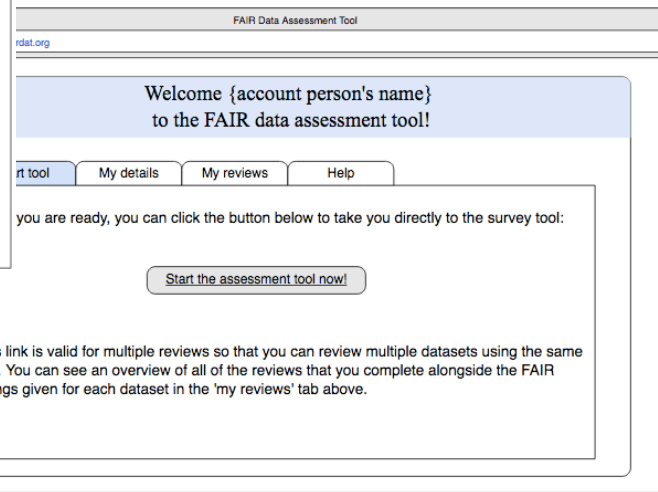
Log-in

Username:

Password:

Submit

If you have forgotten your password or username then click here



FAIR Data Assessment Tool

rdat.org

Welcome {account person's name} to the FAIR data assessment tool!

rt tool

My details

My reviews

Help

you are ready, you can click the button below to take you directly to the survey tool:

Start the assessment tool now!

This link is valid for multiple reviews so that you can review multiple datasets using the same tool. You can see an overview of all of the reviews that you complete alongside the FAIR ratings given for each dataset in the 'my reviews' tab above.



2 User Reviews
1 Archivist
Assessment
24 Downloads

Display FAIR badges in any repository (Zenodo, Dataverse, Mendeley Data, figshare, B2SAFE, ...)

The image displays three mockups of research data repositories, each featuring FAIR (Findable, Accessible, Interoperable, Reusable) badges. A large red diagonal watermark reading "Mockups!" is overlaid across the center of the image.

Zenodo Mockup: The top left shows the Zenodo interface with a search bar and navigation links. Below, a "Recent uploads" section features a dataset titled "FIGURE 5 in Molecular and bioacoustic differentiation of Boophis occidentalis with description of a new treefrog from north-western Madagascar". The dataset includes a FAIR badge and a "View" button. A "Sep 12: Major update" banner is also visible.

Dataverse Mockup: The top right shows the Dataverse interface with a search bar and navigation links. Below, a "Share, publish, and archive your data" section features a FAIR badge and a "View" button. A "Sep 12: Major update" banner is also visible.

DANS Mockup: The bottom left shows the DANS interface with a search bar and navigation links. Below, a "32,530 RESULTS IN PUBLISHED DATASETS" section features a FAIR badge and a "View" button. A "Sep 12: Major update" banner is also visible.

Dataverse Search Results Mockup: The bottom right shows a search results page from Dataverse. It displays a list of datasets, including "Internet Banking Espousal in Bangladesh: A Probing Study" and "Archival Data for Consider the Redirect: A Missing Dimension of Wikipedia Research". Each result includes a FAIR badge and a "View" button.

Thank you for listening!



"Tell us what
you think!"



peter.doorn@dans.knaw.nl

ingrid.dillo@dans.knaw.nl

www.dans.knaw.nl

<http://www.dtls.nl/go-fair/>

<https://eudat.eu/events/webinar/fair-data-in-trustworthy-data-repositories-webinar>