LCRDM Netwerkdag

1 November 2022

RDA-NL community:
an interactive exploration of
the “10 things for curating reproducible and FAIR research” output

Marjan Grootveld, Kim Ferguson, Pascal Flohr, Ingrid Dillo
DANS

Session Structure

- Introduction to RDA and RDA.NL (10 minutes)
- Hands-on exploration and discussion of one of the recent RDA outputs (35 minutes)

Part of the outcome of the discussion has been added at the end of this slide deck.

Vision

Researchers and innovators openly share and **re-use** data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the **social and technical bridges** that enable open sharing and **re-use** of data.

THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

building the social and technical bridges that enable open sharing and re-use of data

53 FLAGSHIP OUTPUTS

including 8 ICT Technical Specifications

200+ ADOPTION CASES

across multiple disciplines, organisations & countries

86 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

32 Working Groups
53 Interest Groups
1 Community of Practice

12,559 INDIVIDUAL MEMBERS FROM 147 COUNTRIES

69% Academia & Research
14% Public Administration
11% Enterprise & Industry

58 ORGANISATIONAL MEMBERS
12 AFFILIATE MEMBERS

Who can join RDA?

Any individual or organisation, regardless of profession or discipline, with an interest in **reducing the barriers to the sharing and re-use of data** and who agrees to RDA's guiding principles of:

- Openness
- Consensus
- Inclusive
- Harmonization
- Community-driven
- Non-profit and technology-neutral

Individual Membership is free at https://www.rd-alliance.org/user/register

RDA Groups - Overview

Any RDA member may initiate or join a Working or Interest group. To become a member of the RDA, individuals should register with the RDA online community and affirm their support for the RDA Guiding Principles.

Many of our members propose Birds of a Feather Meetings at the bi-annual RDA Plenary Meetings to gauge interest and gather momentum and consensus for the creation of Working & Interest Groups.

Working Groups (WG)

Working Groups are short-term (18 months) and come together to develop and implement data infrastructure, which could be tools, policy, practices and products that are adopted and used by projects, organizations, and communities. Embedded within these groups are individuals who will use the infrastructure and help in making it broadly available to the public.

Interest Groups (IG)

Interest groups are open-ended in terms of longevity. They focus on solving a specific data sharing problem and identifying what kind of infrastructure needs to be built. Interest Groups can identify specific pieces of work and start up a Working Group to tackle those.

Read more about RDA Groups here

RDA Recommendations that Make Data Work

“Create - Adopt - Use”

- ✓ Adopted **code, policies, specifications, standards, or practices that enable data sharing**
- ✓ “Harvestable” efforts for which 12-18 months of work can eliminate a roadblock
- ✓ Efforts that have substantive applicability to groups within the data community but may not apply to all
- ✓ Efforts that can start today

59 flagship Recommendations & Outputs

More than 200 cases of Adoption in Different Domains, Organisations and Countries

Celebrating 10 Years of the RDA

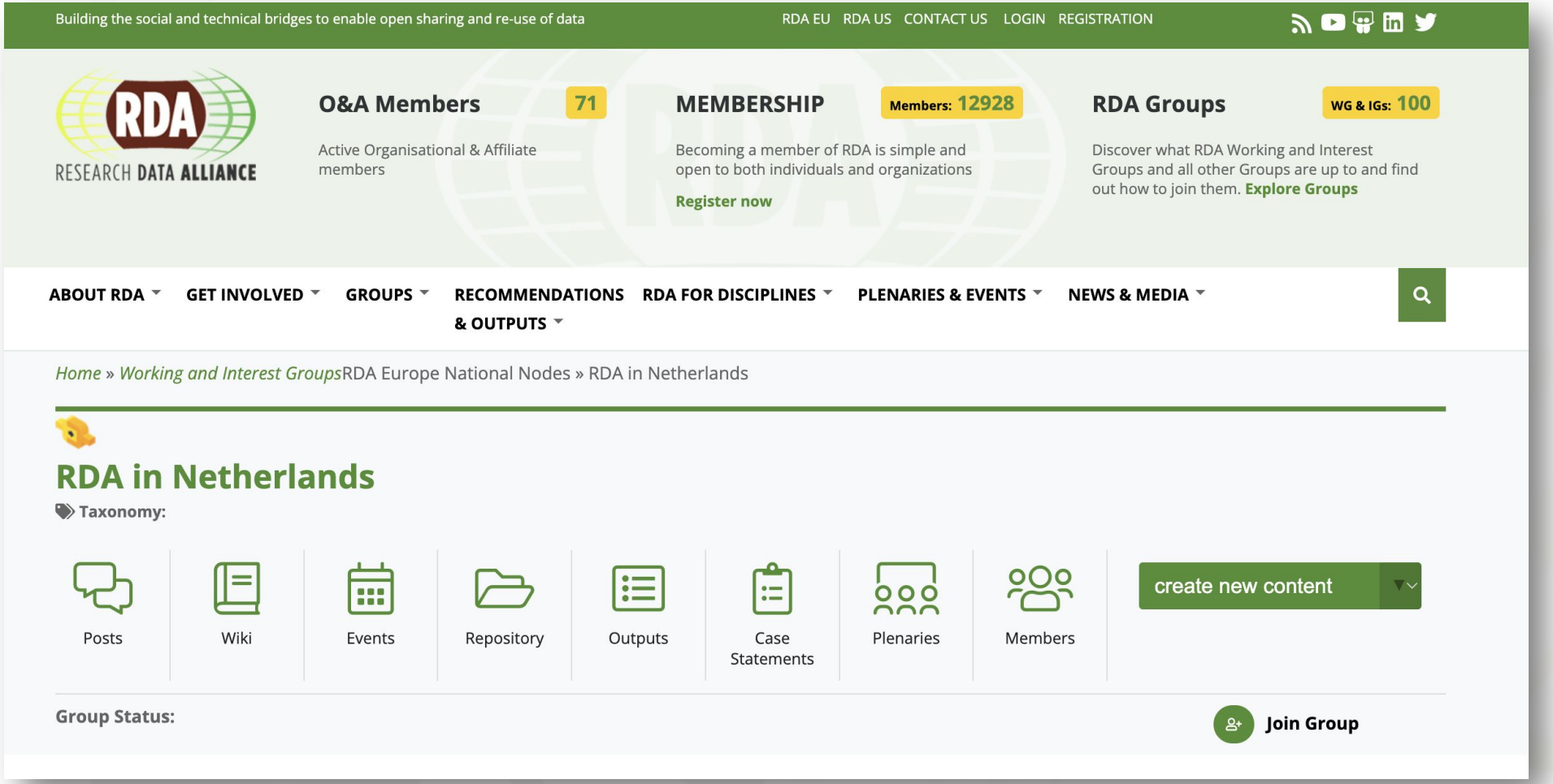
A Decade of Data 2013-2023

Research Data Alliance

Month 2023	Theme
February	FAIR data, software and hardware
March	A Decade of Data: The RDA's 20th Plenary meeting
April	Health and medical data
May	Metadata and technical infrastructure
June	Agriculture and environmental data
July	Research data policy
August	Disciplinary data
September	Sustainable development and responsible research
October	International Data Week - A Festival of Data
November	Research data management support and education

https://www.rd-alliance.org/plenaries-events/events/%E2%80%98-decade-data%E2%80%99-celebrating-10-years-research-data-alliance

RDA.NL Community



Contact

Questions, suggestions, ideas?
Let us know!

RDA Global

Email - enquiries@rd-alliance.org
Web - www.rd-alliance.org
Twitter - @resdatall
LinkedIn - www.linkedin.com/in/ResearchDataAlliance
Slideshare - http://www.slideshare.net/ResearchDataAlliance

RDA Europe

Twitter - @RDA_Europe

RDA US

Email – rdaus@rda-foundation.org
Twitter - @RDA_US

Data Archiving and Networked Services

DANS

RDA RESEARCH DATA ALLIANCE EUROPE

NL NATIONAL NODE

An interactive exploration of the "10 Things for Curating Reproducible and FAIR Research" output

Kim Ferguson (DANS) & Marjan Grootveld (DANS)

Materials co-prepared by Pascal Flohr (DANS)

November 1st, 2022 – LCRDM Networking Day – Utrecht

10.5281/zenodo.7267061

The plan for today

Time	Section
5 min	Introduction to the 10 Things
15-20 min	In groups: Thing 4 (Transparency) and Thing 6 (Access)
5 min	Exchange

10.5281/zenodo.7267061

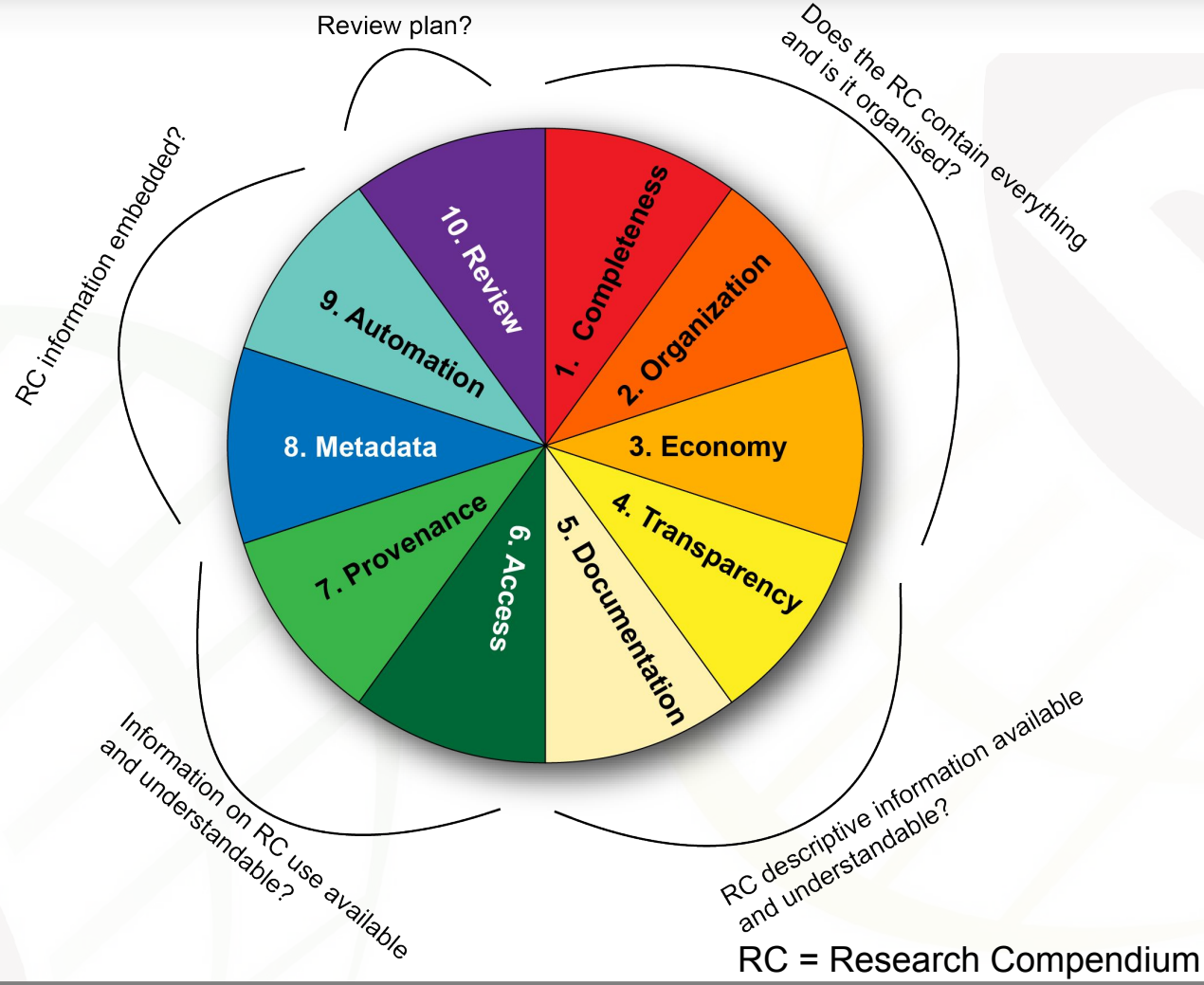
10 Things for Curating Reproducible and FAIR Research

[RDA CURE-FAIR Working Group]

How to curate FAIRly? A main challenge is reproducibility.

10 Things: Standard-based guidelines for the curation of reproducible and FAIR data and code.

DOI: 10.15497/RDA00074

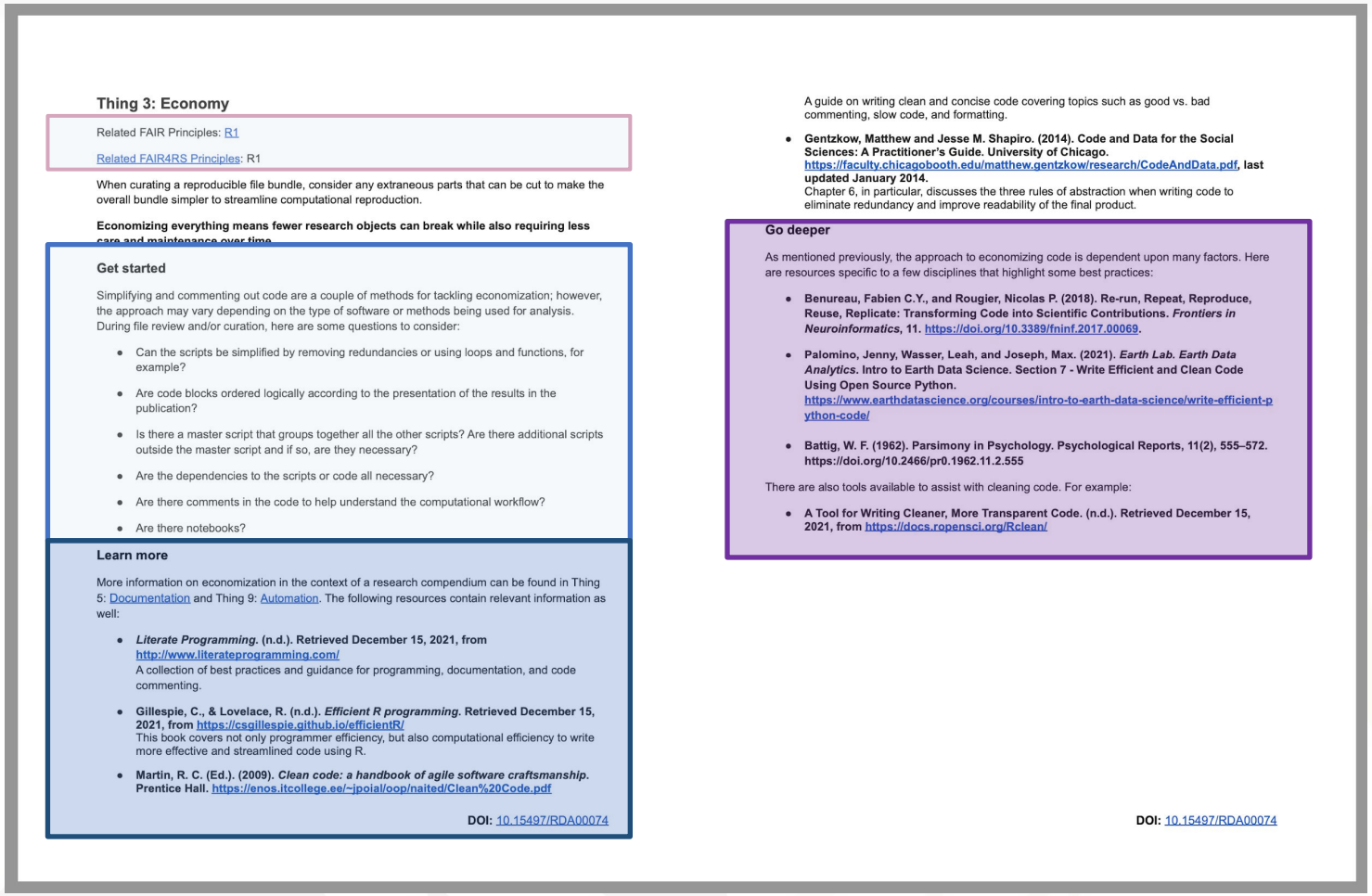


10 Things for Curating Reproducible and FAIR Research cont.

FAIR(4RS) principles

Get started

Learn more



Thing 3: Economy

Related FAIR Principles: R1

Related FAIR4RS Principles: R1

When curating a reproducible file bundle, consider any extraneous parts that can be cut to make the overall bundle simpler to streamline computational reproduction.

Economizing everything means fewer research objects can break while also requiring less care and maintenance over time.

Get started

Simplifying and commenting out code are a couple of methods for tackling economization; however, the approach may vary depending on the type of software or methods being used for analysis. During file review and/or curation, here are some questions to consider:

- Can the scripts be simplified by removing redundancies or using loops and functions, for example?
- Are code blocks ordered logically according to the presentation of the results in the publication?
- Is there a master script that groups together all the other scripts? Are there additional scripts outside the master script and if so, are they necessary?
- Are the dependencies to the scripts or code all necessary?
- Are there comments in the code to help understand the computational workflow?
- Are there notebooks?

Learn more

More information on economization in the context of a research compendium can be found in Thing 5: Documentation and Thing 9: Automation. The following resources contain relevant information as well:

- ***Literate Programming*. (n.d.). Retrieved December 15, 2021, from http://www.literateprogramming.com/**
  A collection of best practices and guidance for programming, documentation, and code commenting.
- **Gillespie, C., & Lovelace, R. (n.d.). *Efficient R programming*. Retrieved December 15, 2021, from https://csgillespie.github.io/efficientR/**
  This book covers not only programmer efficiency, but also computational efficiency to write more effective and streamlined code using R.
- **Martin, R. C. (Ed.). (2009). *Clean code: a handbook of agile software craftsmanship*. Prentice Hall. https://enos.itcollege.ee/~jpoial/oop/naited/Clean%20Code.pdf**
  A guide on writing clean and concise code covering topics such as good vs. bad commenting, slow code, and formatting.
- **Gentzkow, Matthew and Jesse M. Shapiro. (2014). Code and Data for the Social Sciences: A Practitioner's Guide. University of Chicago. https://faculty.chicagobooth.edu/matthew.gentzkow/research/CodeAndData.pdf, last updated January 2014.**
  Chapter 6, in particular, discusses the three rules of abstraction when writing code to eliminate redundancy and improve readability of the final product.

Go deeper

As mentioned previously, the approach to economizing code is dependent upon many factors. Here are resources specific to a few disciplines that highlight some best practices:

- **Benureau, Fabien C.Y., and Rougier, Nicolas P. (2018). Re-run, Repeat, Reproduce, Reuse, Replicate: Transforming Code into Scientific Contributions. *Frontiers in Neuroinformatics*, 11. https://doi.org/10.3389/fninf.2017.00069.**
- **Palomino, Jenny, Wasser, Leah, and Joseph, Max. (2021). *Earth Lab. Earth Data Analytics*. Intro to Earth Data Science. Section 7 - Write Efficient and Clean Code Using Open Source Python. https://www.earthdatascience.org/courses/intro-to-earth-data-science/write-efficient-python-code/**
- **Battig, W. F. (1962). Parsimony in Psychology. Psychological Reports, 11(2), 555–572. https://doi.org/10.2466/pr0.1962.11.2.555**

There are also tools available to assist with cleaning code. For example:

- **A Tool for Writing Cleaner, More Transparent Code. (n.d.). Retrieved December 15, 2021, from https://docs.ropensci.org/Rclean/**

DOI: 10.15497/RDA00074

Go deeper

DOI: 10.15497/RDA00074

10.5281/zenodo.7267061

Research compendium

[The Turing Way community handbook]: :

"A research compendium is a collection of all digital parts of a research project including data, code, texts (protocols, reports, questionnaires, metadata)."

So that the results can be reproduced.

Illustration by Scriberia with The Turing Way community. CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807

Two of the Things

Thing 4: Transparency

From the 10 Things:

"Transparency requires that details about how data were captured and transformed, and how they were analyzed to produce published results, are included in the research compendium."

Thing 6: Access

From the 10 Things:

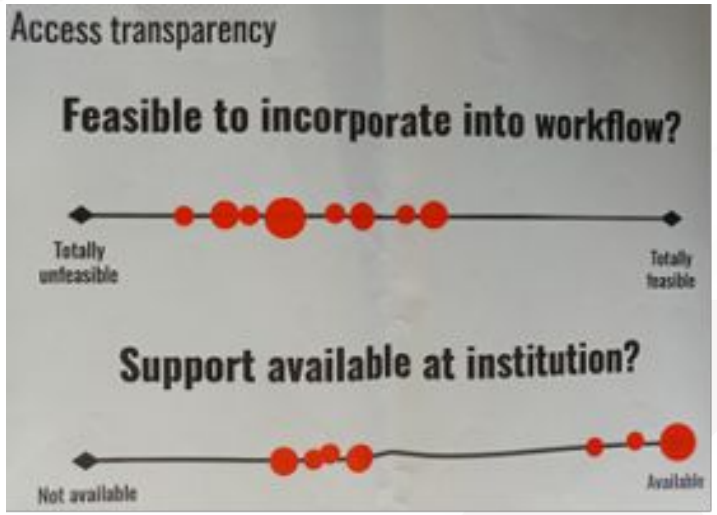
"Because the research compendium serves as the evidence-base for published findings, enabling others to access it with minimal physical, legal, and technical barriers is essential for supporting and promoting … reproducibility."

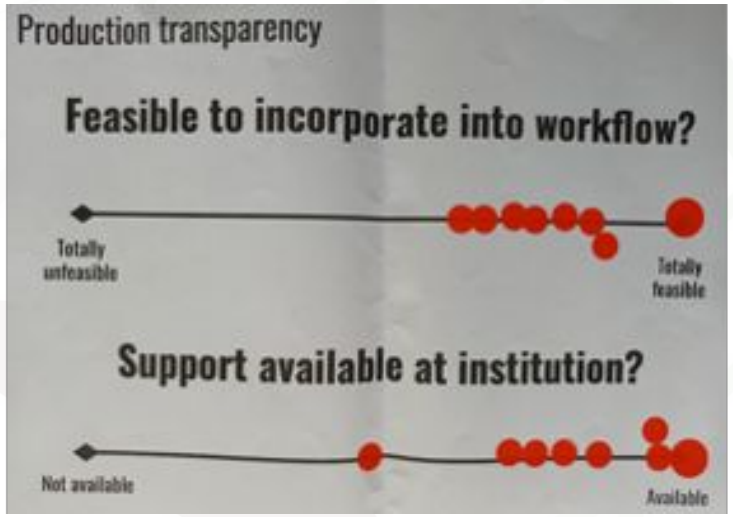
10.5281/zenodo.7267061

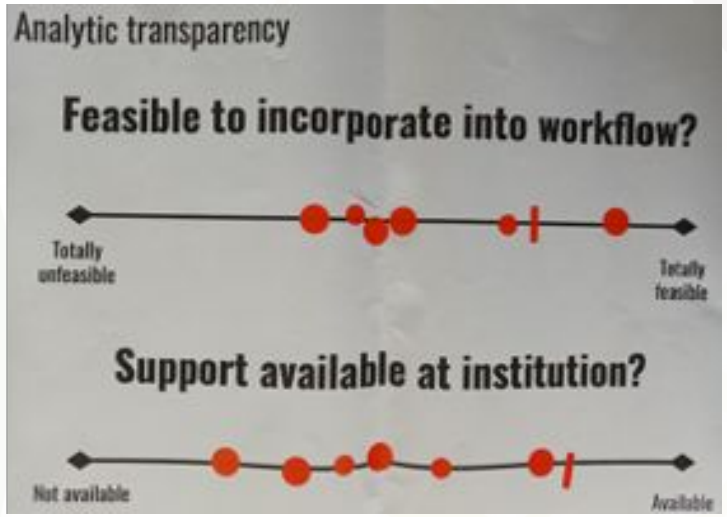
Transparency – input from participants

Three types of transparency, with several guiding questions (see https://doi.org/10.15497/RDA00074):

- Access transparency through information about location, origin, access protocols of compendium
- Production transparency through details about the study design and research methods
- Analytic transparency: disclosure of analytical steps taken





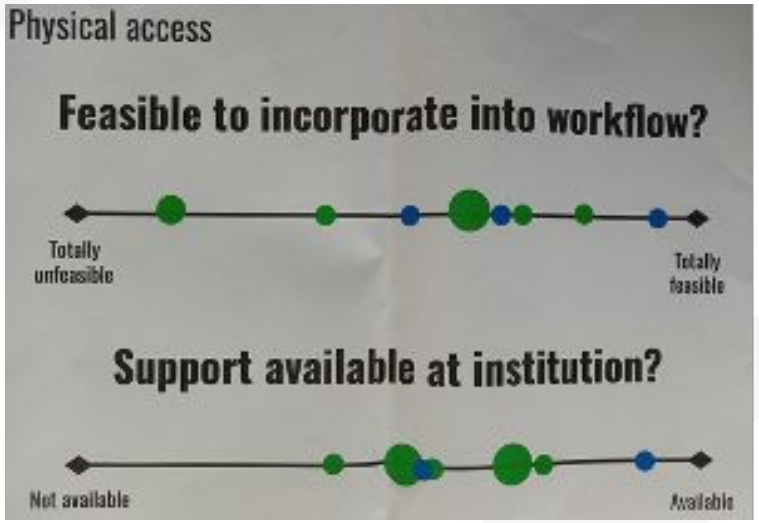


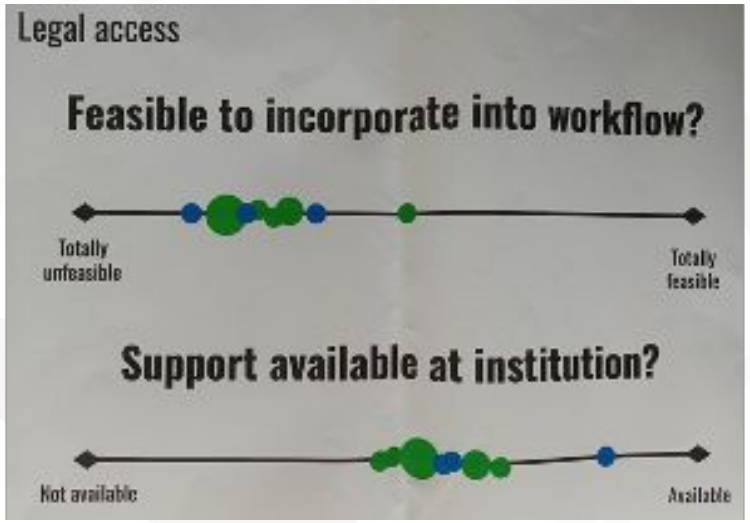
Size of the dots has no meaning.

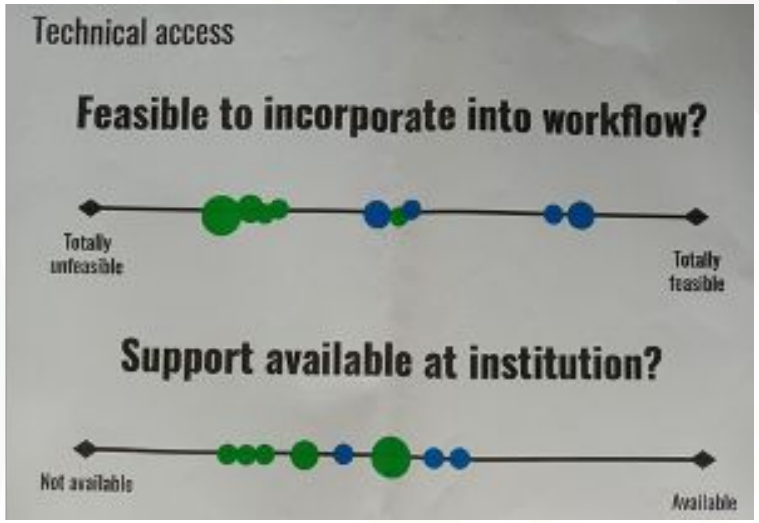
Access – input from participants

Three types of access, with several guiding questions (see https://doi.org/10.15497/RDA00074):

- Physical access to the “research compendium”, in a trustworthy repository
- Legal access, via licences etc.
- Technical access: technology required for reproducing the results should be reasonably accessible







Size and colour of the dots have no meaning.

Data Archiving and Networked Services

DANS

RDA RESEARCH DATA ALLIANCE EUROPE

NL NATIONAL NODE

These slides can be found at Zenodo: https://doi.org/10.5281/zenodo.7261034

kim.ferguson@dans.knaw.nl | marjan.grootveld@dans.knaw.nl | ingrid.dillo@dans.knaw.nl

Twitter: @DANS_knaw_nwo

dans.knaw.nl/en

Further reading

Arguillas et al. 2022. 10 Things for Curating Reproducible and FAIR Research, Zenodo, https://doi.org/10.15497/RDA00074

23 Things Revisited, https://www.lcrdm.nl/23things

The Turing Way community handbook, https://the-turing-way.netlify.app