RDA 25th Plenary Meeting [part of IDW 2025]
Collaborative Notes

Group(s) name(s) / *session title*	BoF, Susanne Den Boer & Pedro Hernández Serrano - *A framework for assessing the value of research data*
Session link for more information	https://www.rd-alliance.org/members/susanne-den-boer/plenary-participation
Session programme time	Wednesday-Thursday, 15-16 October 23:00 – 00:30 UTC / Thursday, 16 October 09:00-10:30 AEST, Breakout 4
Presentation slides	*Value of Research Data_RDA_IDW2025_SdB-PH*

Session summary (to be filled by group chairs):
We will use the content in the table below to highlight your work to the RDA community in a report by the Technical Advisory Board (RDA P24 report example). **Please complete ALL fields below by *Friday, 31 October*, to be included in the report** and other communication. We also ask you to update your group webpage with the key takeaways from your post-plenary session.

1.	**Describe what your session is about in three sentences maximum:** This session addressed and discussed the value of research data as an important parameter for decision making in research projects. What different value types are there, what factors do we take into account to assess data value when collecting, storing, sharing, preserving data? We took starting point in some first work done on a data value framework by a national working group in Denmark as well as a quantitative analysis of (mainly) European research data policies that implicitly or explicitly define or mention data value.
2.	**Highlight a maximum of five key outcomes/actions/takeaways of your session:** *Example:* 1. *Data value is determined by the stakeholder and value is not static (it can change over time). It can depend on trends in research, or availability of new tools for analysis.* 2. *The session was successful in collecting 'data' (feedback, opinions) to further the work - see the detailed meeting notes below.* 3. *The session collected information on initiatives related to data value outside of RDA that should be further examined.* 4. *Follow up: the extensive meeting notes (see below) will be sent to the attendees for further feedback.* 5. *Great engagement and feedback by the attendees indicated that there is interest to explore this topic further. We will need to determine whether there would be*

	sufficient in-kind resources to establish an RDA interest group. We will follow up with attendees to identify whether this is the case.
3.	**What collaborations or synergies did you discover between your group and other RDA group(s) or external (outside of the RDA) organisations?** There seems to be no RDA group that specifically addresses data value to date. However, any group that addresses retention/archiving strategies, sample management/preservation, data management costs, etc. will likely consider data value as a factor. So our efforts (if formalised in a group) could support these discussions by developing a framework that helps in assessing whether data has value and to whom. As a complete newbie to RDA, I do not have a good overview of the RDA groups, but I could imagine this could be relevant for e.g. the RDA group 'Physical Samples and Collections in the Research Data Ecosystem IG or the group "Preservation Tools, Techniques, and Policies".

RDA
RESEARCH DATA ALLIANCE

Attendee Check-in

New to the RDA? Check out on where to start **here**
Get involved in the **RDA Community**
This meeting will take place in accordance with the **RDA Code of Conduct**

Please complete this table to indicate your attendance (add rows as needed):

Full Name	Affiliation	Location	Email
Susanne den Boer (PRESENTER)	University of Copenhagen	Copenhagen, Denmark	susanne.denboer@adm.ku.dk
Pedro Hernández Serrano (PRESENTER)	Maastricht University	Maastricht, the Netherlands	p.hernandezserrano@maastrichtuniversity.nl
Ryan McAllister	CSIRO	Brisbane	Ryan.McAllister@csiro.au
Aileen Day	PSDI	Southampton UK	a.e.day@soton.ac.uk
Katherine Rial	Helmholtz-Zentrum Berlin für Materialien und Energie	Germany	katherine.rial@helmholtz-berlin.de
Michelle Barker	ReSA	Australia	michelle@researchsoft.org
Kevin Ashley	Digital Curation Centre (DCC)	Scotland	director@dcc.ac.uk
Doug Schuster	US National Center for Atmospheric Research	USA	schuster@ucar.edu
Rolf Krahl	Helmholtz-Zentrum Berlin für Materialien und Energie (HZB)	Germany	rolf.krahl@helmholtz-berlin.de
Joseph Gum	Independent	USA	josephigum@gmail.com
Katie Hannan	CSIRO	Adelaide, Australia	Katie.Hannan@csiro.au
Judith Biernaux	FedOSC	Belgium	judith.biernaux@belnet.be
Richard Ferrers (RF)	ARDC	Australia	richard.ferrers@ardc.edu.au
Linlin Zhao	Deakin University	Australia	l.zhao@deakin.edu.au
Leo Chiloane	South African Environmental Observation Network	South Africa	pl.chiloane@saeon.nrf.ac.za
Su-Nee GOH	Nanyang Technological University	Singapore	sunee@ntu.edu.sg
Erin Gray	Geological Survey of Western Australia	Australia	erin.gray@dmpe.wa.gov.au

Emma Crott	ARDC	Australia	Emma.crott@ardc.edu.au
Diana Roig-Sanz	ICREA	Spain	dsanzr@uoc.edu
Piy Ratana	Macquarie University	Australia	piy.ratana@mq.edu.au
Julie Shi	Scholars Portal	Canada	julie@scholarsportal.info
David Medyckyj-Scott	Bioeconomy Science Institute	New Zealand	medyckyj-scottd@landcareresearch.co.nz
Briget Lander	Bioeconomy Science Institute	New Zealand	Briget.lander@plantandfood.co.nz

Collaborative Session Notes

Mentimeter results:

1. At what type of organization are you employed?

University	7
Infrastructure provider	8
Funder	1
Library	0
Archive	0
Company	3
Government agency	0
Other	6

2. What is your role?

Researcher	5
Research leader	1
Administrative leader	2
Data steward / manager	9
IT officer	1
Librarian	4
IPR officer	0
Information security officer	0
Other	5

3. In which situations do you consider the value of research data as part of your role?

Grouped: in the planning phase of a project

- Choosing a research question
- During data management planning

Grouped: in relation to tools and infrastructure, incl. storage capacity

- When we run out of storage.
- When there's limited storage capacity
- Not yet but it would be good to decide what resources to onboard into our infrastructure
- System use for Usability assessment and improvement

- Development of decision support tools

Grouped: in relation to commercialisation
- Protection of potential commercial value according to University policies

Grouped: in relation to data publishing
- What data should be included in our repository in alignment with our core mission.
- Publishing
- Considering making data available

Grouped: in relation to preservation
- Determining what data is worth preserving
- Mid- and long-term storage decisions
- Developing a national framework for data management, thinking about data retention and disposal questions
- Do all versions of statistically measurements need to be preserved, or will the code suffice?
- When long term storage options need to be considered.

Grouped: in relation to decision making regarding data deletion / sample destruction
- When to discard data from the repository?
- When data clean ups are required.
- What may data that aren't irreplaceable measurements be deprecated?
- Data archaeology on orphan data - e.g. "what is this?"

Grouped: in relation to data reuse
- In data search: Normally when reusing datastes one would like to assess the most valuable ones
- Reuse value - maximise Ip value Storage costs Inc retention Unique record of environment

Grouped: in relation to specific research disciplines and their data types
- Clinical trials data
- Longitudinal data and samples with information on infectious disease exposures and immune response.
- Massive set of genomics data
- Data on armed group activity and civilian responses in conflict zones

Grouped: in relation to specific dataset 'parameters'
- What variables provide the most long term scientific value?
- Assessing if we're storing sufficient metadata
- What fidelity of measurements needs to be preserved?
- Whether the data has proper metadata and if they are labeled for the current tasks.

Grouped: in relation to communication or guidance about good data management practice
- Promoting open science to non-practising researchers

- Explaining to funders, university leaders etc why they should care about data
- Provide guidance to repository managers on curation

Grouped: in relation to value being a core principle for the organisation

- I work on FAIRification of biomedical data and samples, including pro and retrospective FAIR. Understanding what to FAIRify + what samples to preserve = key challenge because of the investment needed.
- It's fundamental to what we do - my organisation helps others maximise the value of their research data

4. Please, name some examples of data sets of HIGH value

Grouped: references to specific data collections

- ECMWF Reanalysis Version 5
- ECMWF model output/weather forecast
- New South Wales Human Services Data Set
- PLIDA
- National soils database
- National database of biological specimens
- National Landcover database
- ACMI
- WebCSD from CCDC https://www.ccdc.cam.ac.uk/
- Ausstage & Auslit

Grouped: examples referring to specific research disciplines
Health sciences:

- iEEG data that was collected continuously for 3 years labeled by clinicians
- Human genome

Climate / earth sciences:

- Underpinning climate and weather data
- In situ earth observations
- Irreproducible earth system observations.

Social sciences:

- Historical census data

Grouped: examples referring to specific data types or characteristics
Data underpinning results:

- Dataset linked from the journal paper's data availability statement
- Publication data

'Old' data, or data spanning long periods:

- Heritage data (very old journals, books, records...)
- Long period data (planetary phenomena that occur once every century)
- Historical topographic maps

Other:

- Massive sets of genomics data
- Survey results

- Data in reference data bases
- The raw data from the national census
- Unrepeatable observations - weather etc
- Anything someone wants to use
- Outputs with errors where you've realised you made a mistake and you have to redo something.

5. Please, name some examples of data sets of LOW value

Grouped: in relation to how data are presented

- Non-FAIR datasets
- Little metadata
- Low quality data with no provenance or metadata
- Data in proprietary formats.
- Very coarsened data (e.g. data that have been anonymised and can no longer be used to for clinical inference)
- Data lacking metadata.

Grouped: data where trustworthiness is questionable

- Untrustworthy data
- Data of poor quality (big error margins, ...)

Grouped: in relation to the information the data carry

- Data lacking novelty
- Raw data of a measurement that did not work
- Simulation output parameters that only support very narrow use cases and are expensive to store.
- Datasets that don’t inform any action or decisions

Grouped: data types

- Some modelling outputs
- easily replicable data (?)
- Data from replication studies(?)
- Simulation outputs that can easily be recreated by preserving the code
- Early covid data
- Model outputs

Grouped: other

- there should not be difference between value of the data. all data have an intrinsic value the change on context
- Community need
- Menti survey responses 😂

6. What is it that makes data sets have value?

Grouped: in relation to how data are presented

- Rich metadata
- Well organised
- Rich metadata, linkage
- well structured
- Making it FAIR
- metadata for context
- Metadata, provenance data, including ethical provenance, broad scope of permissions, FAIR data

Grouped: in relation to resources needed to (re)produce or economic value

- Very expensive to collect and reproduce
- Unrepeatable work
- The cost to get/create the data
- Unrepeatable
- The value the markets set 🤑
- rarity of data

Grouped: in relation to perceived importance or trustworthiness

- Relevance
- Verifiable
- Data that has informed significant decisions (even if the data is intrinsically flawed)
- data that underpins publications
- Community benchmark dataset.
- Usefulness
- Its match with current important research questions (climate change, ...)
- Data that underlines a publication
- Comprehensiveness ; accuracy (and 12 others identified by wang and strong)
- Are derived from a coordinated set of experiences.
- National scale

Grouped: in relation to reusability and reuse

- Reuse
- High reuse potential for large number of communities
- Their difficulty to be replicated/recaptured
- Used in a "Highly Influential Scientific Assessment"
- Are used in many scientific research projects
- lots of accesses/use/reuse
- reuse of the data
- A use and usability

Grouped: as determined by stakeholders

- Value to whom? Retired researchers may not value their data at all. While the archive may think it’s high value.

- High number of citations
- Domain expert opinion
- Research impact area does a case study on it and says it has impact
- Community need/use
- A user
- Labeled by domain experts

Grouped: other

- See value literature
- everything we can think of

General feedback to the concept of data value

- What is valuable now isn't necessarily valuable in 6 weeks - however it may be valuable in 100 years again. *Time* is an important factor.
- National archives of Australia retention policies - some things should be kept very long….*how often should we re-appraise* data value?
- Value of data also depends on the *scientific discipline*.
- Value of data for preservation purposes - *how much does it cost to preserve*? How much work is involved in preserving the data? Value needs to be weighed against institutional capacity
- Example of the current "assessment" provided for NCAR's GDEX repository
    - Non observing system based datasets (e.g. Model output datasets typically larger than 10 TBs) may be purged from the GDEX after 5 years if:
        - Community usage metrics don't support continued maintenance of the dataset.
        - The dataset has been superseded by an updated version and it is determined that the copy of the old version is no longer of significant research value.
    - The GDEX will attempt to contact the original submitter to provide advance notice in the event that a decision has been made to purge a dataset. This will allow the author to find an alternative repository or pay for continued archival in the GDEX if needed.
- *Is data depreciated over time*? Does its value decrease, or could it become more valuable in the future?
- Key questions: What is value? When is value determined?
- Data can hold significant scientific value, but assessing it *depends on context and purpose*.
- *How often* should data value appraisals be conducted?
- It is *hard to underestimate the importance of data*. Once data are lost, their value is lost as well.
- The *concept of data value can play a role in institutional insurance discussions*: is research data included in the insurance cover? How and who values it? Institutional accountants? Since unlike other assets, data can't be replaced like for like and because of the challenges of estimating the value, data might be dropped from a policy/cover. If it is included, how is depreciation handled? The purpose of insurance

is to provide financial protection against potential losses from unexpected events; it's risk transfer. If data is not included, the fact there is a risk around data is lost and thus data is not included in enterprise risk assessments nor is is treated as an asset that needs to be managed (investment) unlike a building, library collection etc.

- The importance of data is leading to increased attention from some governments, who are trying to understand and value a research institution's data holdings to assess its performance and potential for commercialisation (*the data has value because it underpins a technology product*) - contribution to the economy.
- Ultimately, how much are institutions willing to invest to maintain, preserve, and secure valuable data? What do policies say and what do the institutions really do, what monetary value are they willing to spend on the management of data, that will give an idea of the value that institutions attribute to datasets
- Socializing the reality that data will be lost
- Comment/thought - we keep too much data and far too little metadata
- See also difference between value and values. See also Accounting definition of value - Statement of Accounting Concepts (s82-90, Framework; Recognition of assets). Value is now part of the definition of What is Marketing? Economics has a rich history of considering the nature of value. Value is a key part of the Innovation literature (see PhD above; a value theory of innovation; ie Innovation is new things that create value). New things can also destroy value.
- Feedback from the "What about model data?" Determining Best Practices for Preservation and Replicability project might be relevant to this discussion. It includes a rubric to inform researchers of what (Earth System Science) model output needs to be preserved and shared to "communicate knowledge". In many cases project participants determined that a combination of selected model outputs and the full suite of code used to generate those outputs is sufficient to meet the "knowledge communication" goal.
    - https://modeldatarcn.github.io/
    - Schuster, D. C., M. S. Mayernik, G. L. Mullendore, and J. W. Marquis, 2023: What about Model Data? Best Practices for Preservation and Replicability. Bull. Amer. Meteor. Soc., 104, E2053–E2064, https://doi.org/10.1175/BAMS-D-22-0252.1.
    - The rubric evaluates the following concepts:
        - CommunityCommitment
            - Is it anticipated that your simulation workflow outputs will have broad community impact and downstream reuse?
        - Research Workflow Accessibility
            - Would it be reasonable to expect others in your academic discipline to rerun your full simulation workflow? ●
        - DataAccessibility
            - Would it be reasonable to expect others to access and use simulation workflow outputs?
        - Research Feature Replicability
            - Are physical features generated by a simulation replicable?
        - Cost

        - Is it more costly to re-run a full simulation workflow or preserve model output products in a trusted repository?
    - There a value dimensions which equate to value types. In my research I found 12 dimensions of value.

Feedback specifically to the analysis of RDM policies

- Value can be objective or subjective. Where subjective then this assumes a beneficiary or value actor or value target. Where objective, then you can say X is valuable (so assumes value is socially agreed).
- In UK, many policies are derived from a policy template (e.g. produced by DCC) and this is why the policies in UK are very similar.
- Many institutional policies are based on templates. There is a question regarding what universities claim to do, and what they actually do in practice. It would be interesting to review concrete activities universities undertake that demonstrate that they value research data.
- Having a DM policy in itself implies that an organisation values data.
- RDM policy may need to be looked at alongside other policies as well e.g. digital preservation, retention.

Research data value types

TYPE	DEFINITION BY THE DANISH WORKING GROUP	EXAMPLES or COMMENTS
Cultural / societal value	The impact that research data have on our cultural heritage. *What is the importance of the research data for our understanding of society now and in the future?*	Value changes over time with new information.(RF). Eg artwork, indigenous data
Scientific value	The contribution that the data make to support or further develop conceptual frameworks within a research discipline. *What is the importance of the data in advancing the field of research?*	Value is social, to a scientific discipline. Value stories eg gravity, big bang.
Market value	The economic significance of data, e.g. the direct financial gains to the project or organisation or the indirect impact on future value generation in society. *Can the data be commercialised by the organisation?*	Value is financial (Cost, able to sell). Ability to sell an asset, cost to maintain an asset.
(Re)production value	The resources it requires to create the data or recreate the data when they are lost. *Can data be recreated at all?*	Value is time (RF) time-saving, convenience, cost to reproduce.
Compliance value	The impact of non-compliance to ethical, legal or contractual obligations. Positively correlated with value. *What are the consequences for the data providing organisation, human participants, cultural heritage, populations, animals, ecosystems?*	Value is regulatory, about power and duty.(RF) ANd time to comply with rules.

Additional types added by session attendees:		
Sentimentality		Value is emotional (RF)
Cost	The resources required to support long-term curation and access to a dataset	What is the financial and environmental cost to provide long-term preservation and access to a dataset (e.g. E-infrastructure, Human resources, pollution, etc…)?
Reputation	Institution reputation as well as the researchers reputation	

Feedback to value types:

- Value of novelty, value of learning, value of function, value is need, value is beauty (Aesthetic), value is reliability, simplicity.
- Can the data be regenerated/reproduced is a critical question along with the effort/specialization required to regenerate a dataset
    - Preserving the documented code/workflow used to produce a digital data is cheaper for “synthetic datasets”, and may be more valuable than the resultant data outputs in the long term.
    - Principle: Irreproducible measurements/observations (e.g. earth systems observations) need to be the top priority for long-term preservation (as long as possible)
- What about the negative value on data,
    - Another example of intrinsically negative value turning into positive value: retaining malware, etc. to study cybersecurity, computer science, etc.
- Value assigned by stakeholders: mandated data deposit for publicly funded research, data availability statements from journal publishers. Preservation may be implied if repositories are defined as places that “archive” data
- If data is derived from a larger source (e.g. corpus of text), is it sufficient to describe the methods and not retain their derived datasets for long periods

Intrinsic factors that influence value

FACTOR	**DEFINITION**	**EXAMPLE**
Originality, uniqueness	The novelty of the data	
Reliability	The degree to which data are collected via methods that are quality-controlled, preferably peer-reviewed and assessed to be appropriate to address the research questions.	
Representativeness	The degree to which the data cover all relevant factors (variables, cases, time periods etc.) needed to address the research question	

Organisation and standardisation of data	The degree to which data are organized in a clear, logical, and systematic manner and widely accepted file formats are used	
Standardisation and richness of metadata	The richness of embedded metadata and the degree of adherence to metadata standards, vocabularies, ontologies.	
Additional factors added by session attendees:		
Sensitivity	The security risks to retaining the data	

Extrinsic factors that influence value

FACTOR	DEFINITION BY DANISH WORKING GROUP	EXAMPLE or COMMENTS
Relevance	The degree to which data are aligned with current needs (scientific, societal, political, …)	
Demand	The number of stakeholders for whom the data has value	
Exclusivity	The abundance of data, when the data are in high-demand	
Innovation, technological compatibility	The degree to which data can be exploited using available technologies	
Accessibility	The ease of access to the data	
Reputation of the source	The degree to which there is trust in the producer of the data	
Additions by the session attendees:		
Ethics		Health data from vulnerable populations that was obtained without individuals' consent
Cost		Need to include "Cost" for long term data curation as a value factor

Feedback to value factors:

- Intrinsic /extrinsic value is discussed in literature of philosophy (See Stanford Encyclopaedia of Philosophy) See also ibid. Value Theory
- Marx wrote about a Labor Theory of Value, that the value of X is how many work hours took to create X. Value also relates to risk and return eg failure of fridge.

Useful resources on research data value

Type (project, report, group, …)	Title (if known)	Why relevant?	Link or contact person
Guidance	Appraisal Guidance for the Preservation of Research Data	Assessing value	https://zenodo.org/records/5942236
Report	Retention and Disposal of Research Data: from current to best practices.		https://doi.org/10.5281/zenodo.10076891
Report	Business Intelligence and Reporting of Research Data		https://doi.org/10.5281/zenodo.10076883
Publication	What about Model Data? Best Practices for Preservation and Replicability	Rubric to assess value -what artifacts to preserve and share to communicate knowledge	https://doi.org/10.1175/BAMS-D-22-0252.1
Tool	Repository Crisis Scorecards	Scorecard to assess risk factors/broad view of potential costs for recovery or increased resiliency	https://doi.org/10.5281/zenodo.15208172 or josephigum@gmail.com
Review	The Little Book of Value: How Innovation Creates Value for Consumers - Volume 1 (Richard Ferrers)		https://doi.org/10.6084/m9.figshare.7376879
PhD thesis	A consumer *value* theory of innovation in 3G mobile phones: a grounded theory approach (Richard Ferrers)		https://figshare.com/articles/dataset/A_consumer_value_theory_of_innovation_in_3G_mobile_phones_a_grounded_theory_approach_/680002
Poster	Towards a Value Theory of Innovation (Richard Ferrers)		https://figshare.com/articles/dataset/Towards_a_Value_Theory_of_Innovation/1399227
Blog	Value Management: Innovation 2.0. Exploring:conceiving Value Management1 from my phd thesis2 on innovation. (Richard Ferrers)		https://valman.blogspot.com/
Conference	Imperial College Nov 19-20, 2025 - see Customer Value Foundation		8th Global Conference on Creating Value
Webpage	What if Data were not Forever?	Australian work on what data to keep	Rob Cook, Rhys Francis

Article	Research Data Culture Conversation - A Macro View of Retained Australian Academic Research Data	Includes a first approximation to likely sector wide data retention cost.	Rhys Francis
Guidance	DCC/ANDS guidance on data appraisal: How to Appraise and Select Research Data for Curation | DCC	Has been adapted into a number of discipline-specific and centre-specific appraisal guidelines. Value assessment is implicit to appraisal (and this builds on archival practice that long pre-dates work on research data.	DCC
Collected works	Lots of work by Beagrie & Houghton on valuing research data facilities	these have been used to justify the expenditure on (e.g.) a data archive by making an assessment of the value added by that archive, which again implicitly involves valuing the underlying data asset.	Beagrie & Houghton
Collected works	Wang & Strong's work on data quality	Way of looking at value since once of its uses is to estimate what someone would pay in order to have access to data with a particular set of qualities	Wang & Strong

Have you worked on initiatives related to research data value and can we contact you for a brief online interview? If so, leave your details here:

Name	**Affiliation & country**	**Email**	**Short description of the initiative**
Kevin Ashley	DCC	director@dcc.ac.uk	DCC/ANDS guide on appraisal of research data; UK national archives work on selecting government data for permanent preservation; multiple initiatives on assigning value to research data repositories and their services, and more…
Katie Hannan	CSIRO	katie.hannan@csiro.au	Worked on a flowchart for retention and

			disposal that took into consideration storage and publication options, CSIRO's records disposal authority from the National Archives of Australia, the National Archives of Australia's General Records Authority and Normal Administrative Practice.
Doug Schuster	NCAR	schuster@ucar.edu	https://modeldatarcn.github.io/
Richard Ferrers	ARDC Australia	richard.ferrers@ardc.edu.au	Research Data as an asset
Joseph Gum	ESIP/AGU	josephigum@gmail.com	AGU Impactful Datasets, Initiative to have people nominate datasets important to them
Mark Rehbein	IMOS Australia	mark.rehbein@utas.edu.au	Method to evaluate value of datasets for prioritisation and 'co-investment' purposes