

Technical Repository Service Providers: Community Engagement #2



Collaborative Notes:

<https://docs.google.com/document/d/10fWEB-rz3x9W536uRrcmFNriGAQd8k1H/e/dit?usp=sharing&oid=104659450293150213734&rtpof=true&sd=true>

RDA Complex Citation WG:

<https://www.rd-alliance.org/groups/community-based-catalogue-requirements-trustworthy-technical-repository-service-providers/>

RDA 23rd Plenary Meeting (RDA P23) | Sustainable Science

12-14 November 2024, Costa Rica and Online

RDA 23rd Plenary Meeting

www.rd-alliance.org



Agenda

1. Welcome and session overview (WH, 5 min)
2. Brief review of WG and case statement for newcomers (AL, 15 min)
3. Report on phase 1 of WG activities and community consultation '0' (WH, 10 min)
 - a. 3.1 Consultation results from Plenary 22
 - b. 3.2 Draft inventory of desirable characteristics
4. Breakout sessions: reviewing desirable characteristics of repositories (30 min)
 - a. Group 1: Landscape scope review
 - b. Group 2: Scope, organisation, and applicability of characteristics - TRSP focus
 - c. Group 3: Details of landscape analysis - review
5. Feedback from breakouts (10 min)
6. Open discussion (10 min)
7. Planning and next steps (10 min)
 - a. Regular meetings
 - b. Task assignments



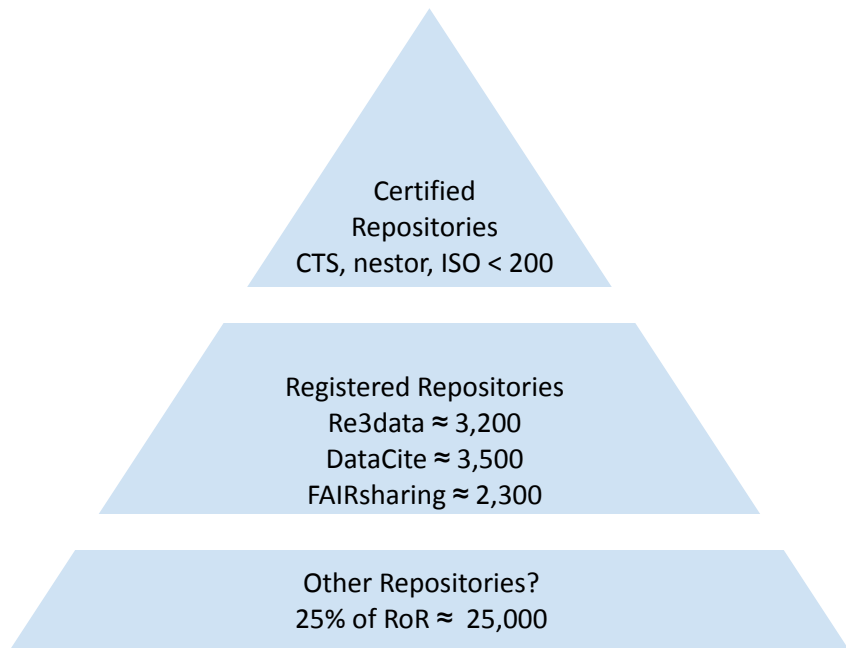
Case Statement, Work Programme

Allyson Lister, FAIRsharing

Demand and Supply: Trustworthy Repositories

Estimates for Research Data

Supply



Demand

Zenodo ≈ 320,000 deposits to date
DataCite ≈ 35,000,000 deposits to date
FigShare ≈ 2,000,000 deposits to date

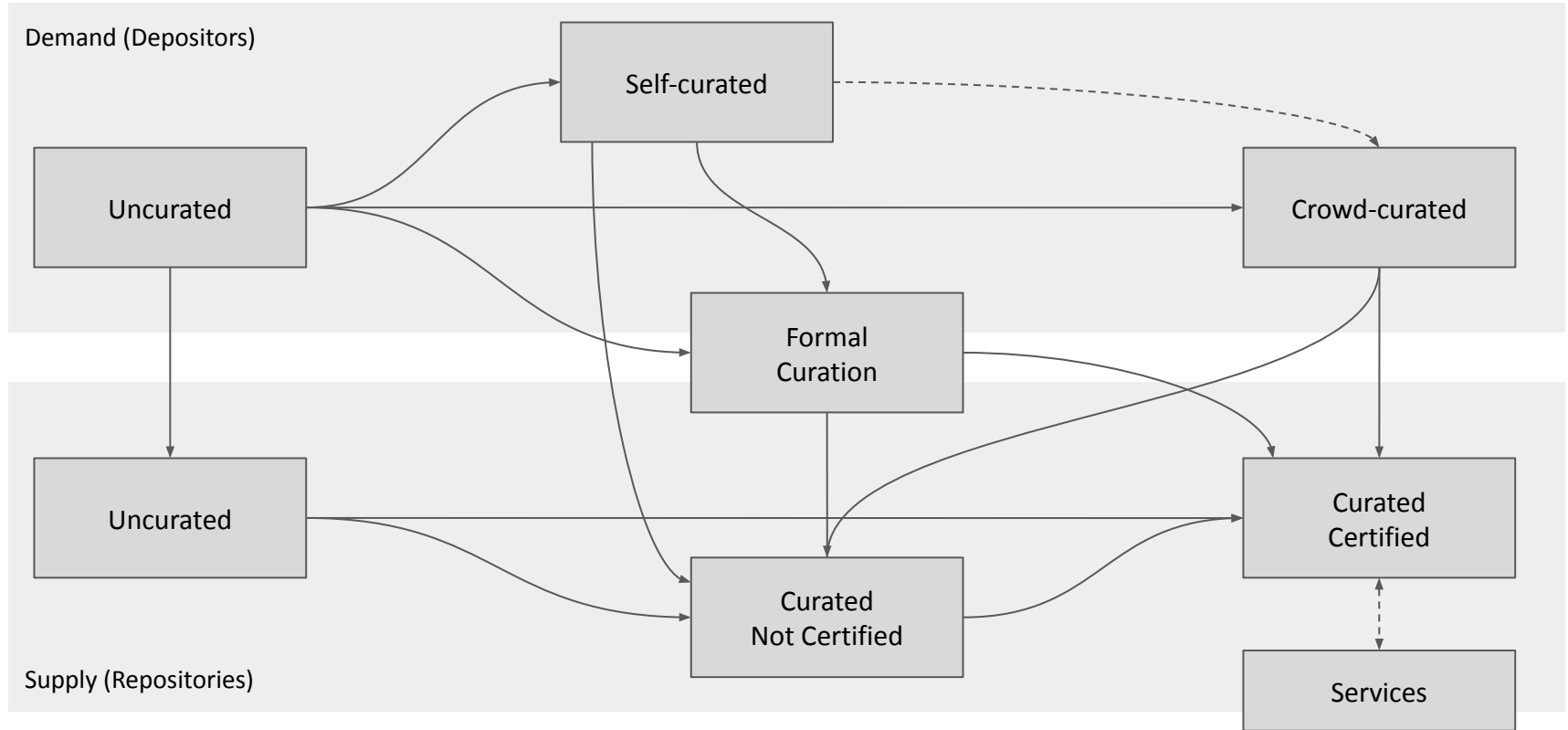
Peer-reviewed publications per annum
≈ 5,000,000

Back-of-the-envelope estimates

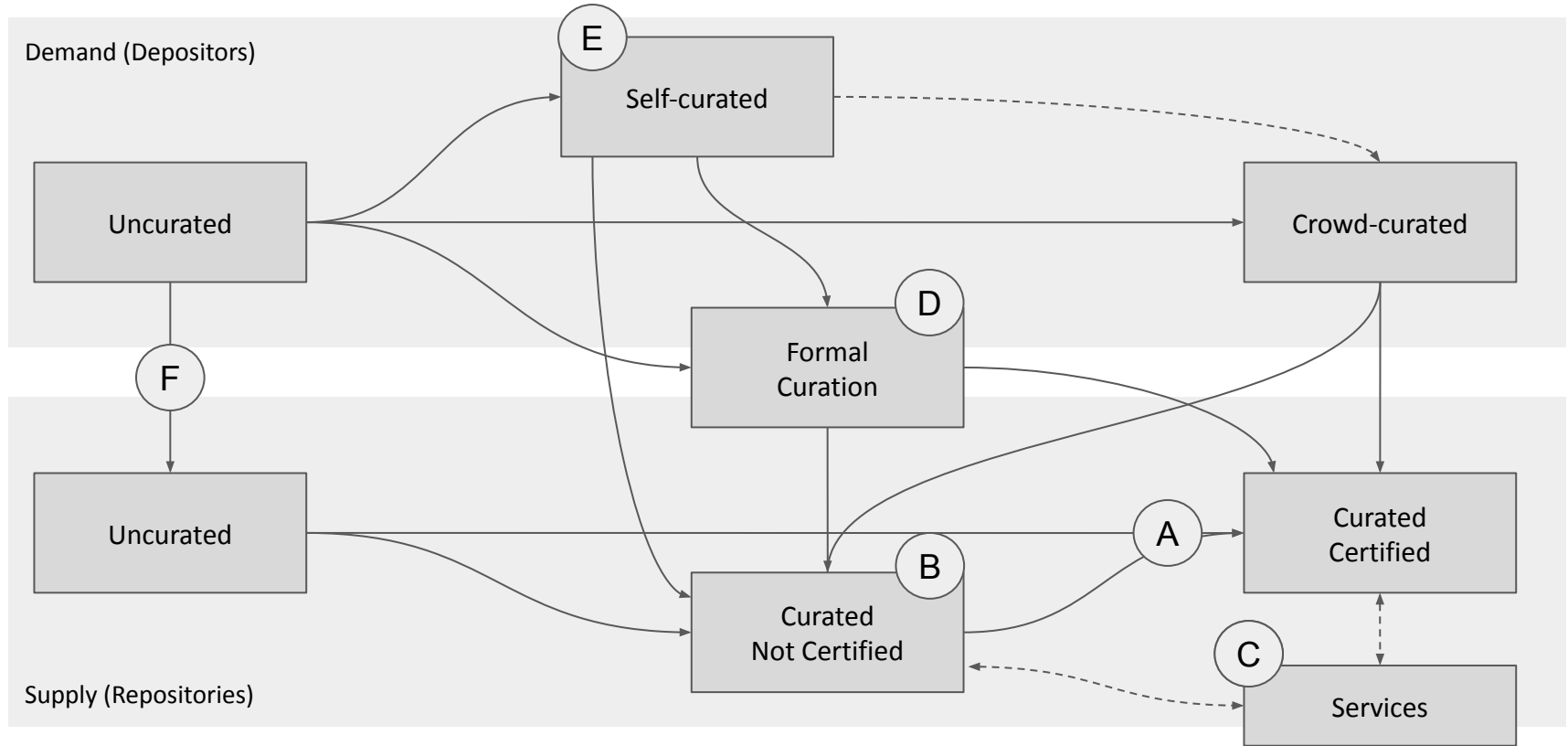
2.5 to 5m datasets per annum

Allocated to known repositories:
≈ **3 - 5 curated datasets** per day in **registered** repositories
≈ **30 - 55 curated datasets** per day in **certified** repositories

Simple Process Model: State of Curation



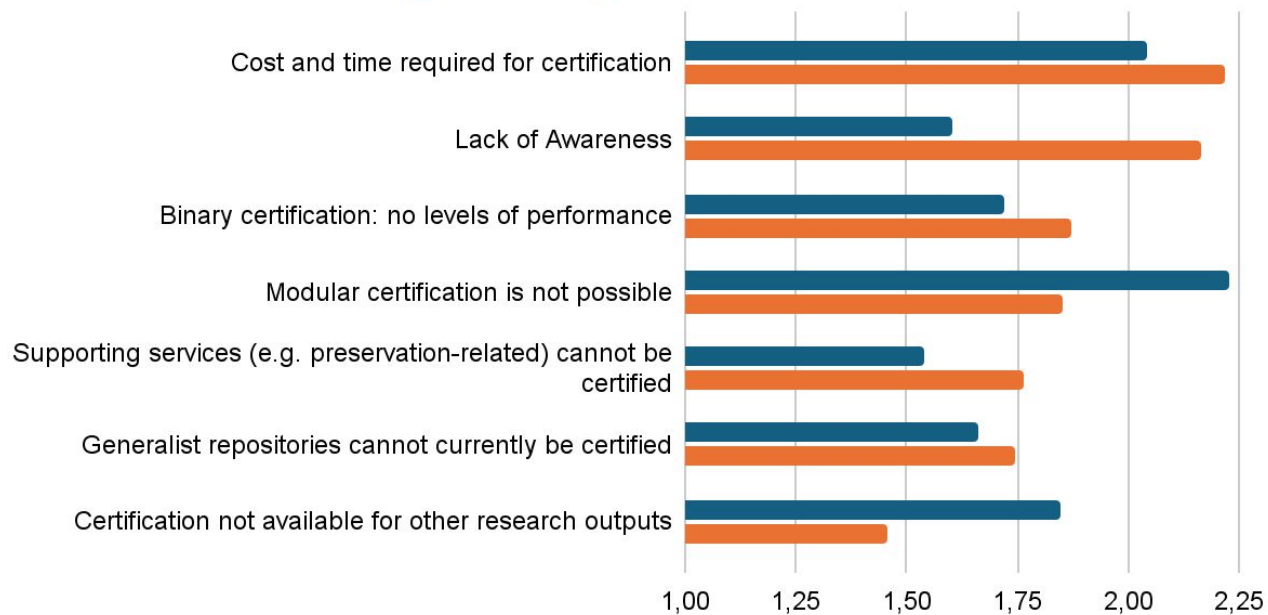
Simple Process Model: State of Curation



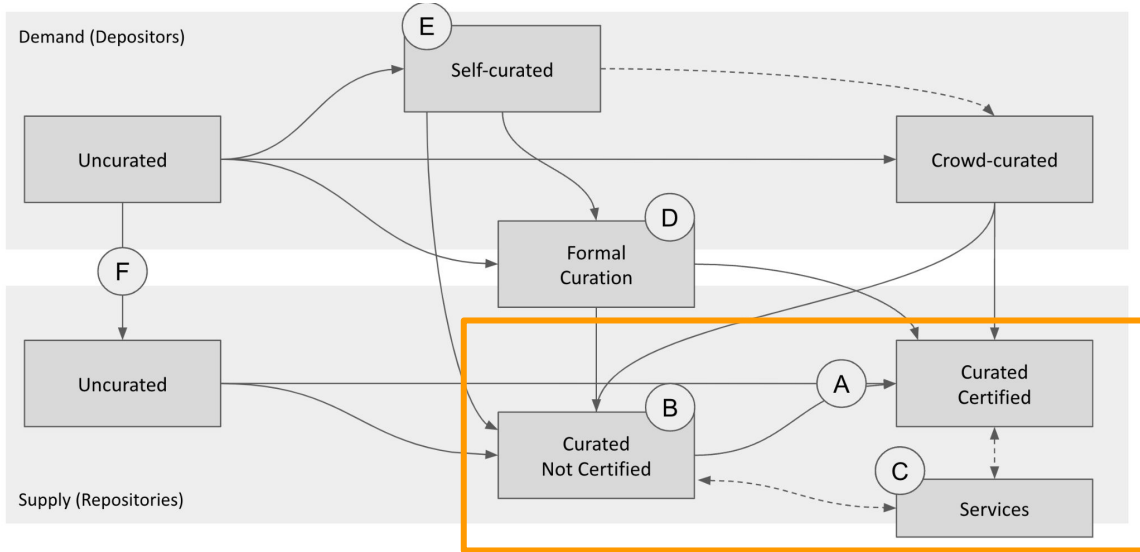
Relative Importance of Certification Issues

Ranking of Issues

■ Deviation ■ Weighted Average



Bottlenecks and Issues



Workgroup Focus

- A. Rate of certification
 - a. Too costly and complex
 - b. Bar too high
 - c. Certification capacity
- B. Scope of certification
 - a. Types of outputs
 - b. Granularity of certification
 - c. Features and performance
- C. Inclusion of service providers (TRSPs)

Not in Scope

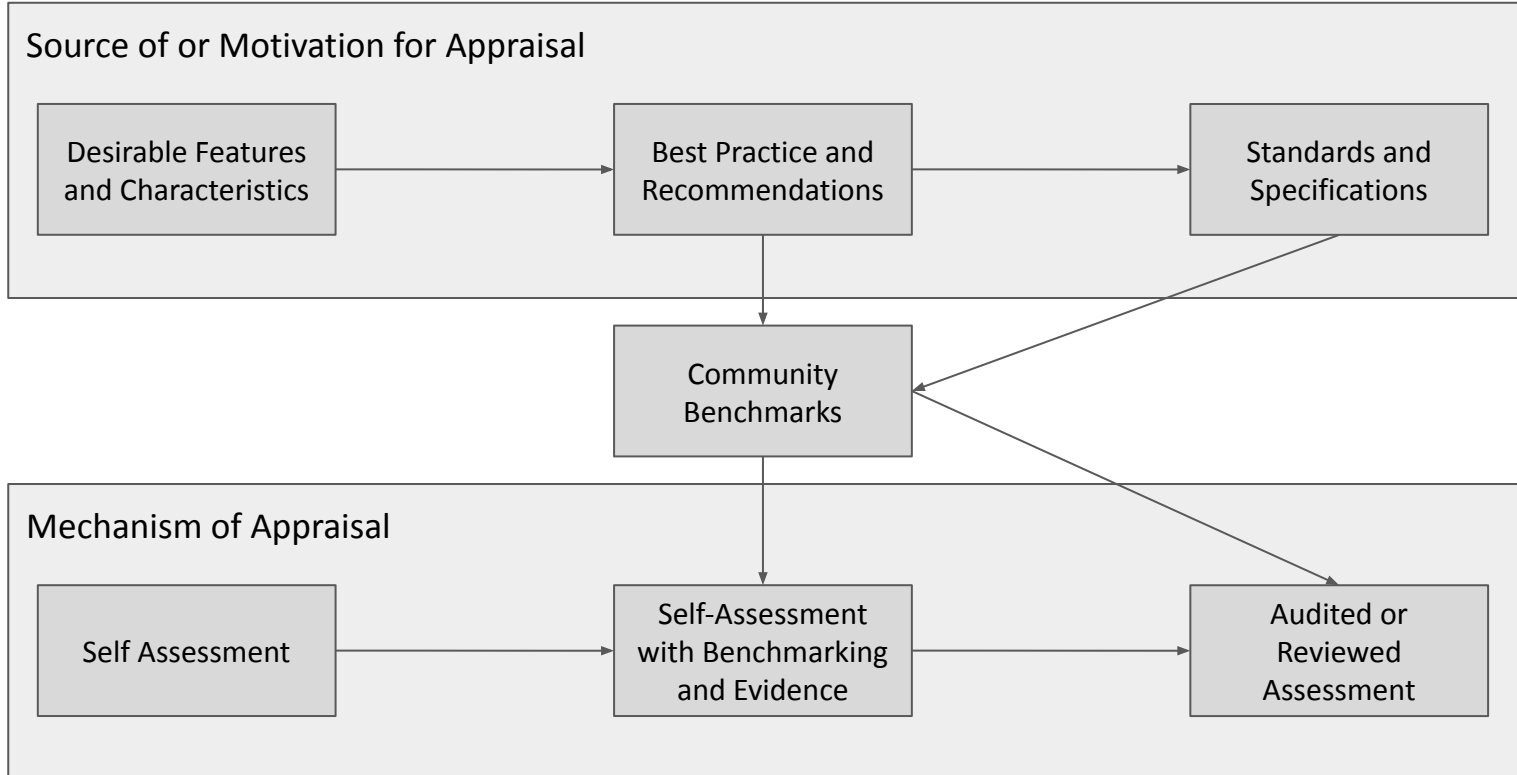
- D. Rate of Curation
- E. Loci and Mechanisms of Curation
- F. Uncurated flows

Working Group Intentions and Outcomes

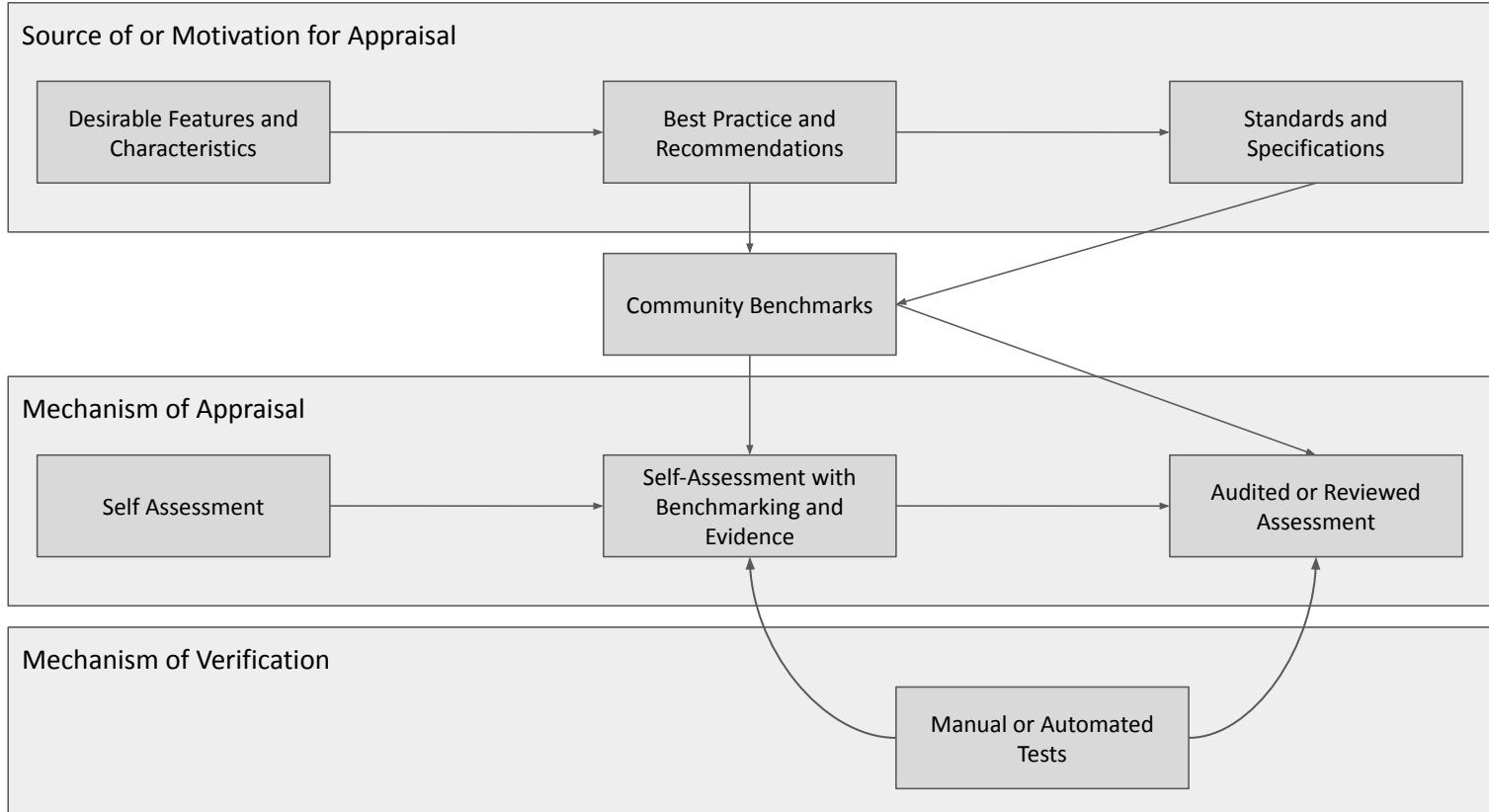
- Establish a set of **modular repository features and characteristics**, confirming and assessing the needs and expectations of the community in respect of the trustworthy repository ecosystem.
- Determine which of these features can be delegated to **data repository service providers**.
- Define and describe **alternatives to formal certification** that improves the appraisal and selection of
 - repositories that are not formally certified,
 - services contributing to the features expected by the community.
- Contribute to an improvement to the number and scale of successful certifications.
- Assist with extending the scope of certification beyond data repositories
- Community makes better informed decisions about the use of infrastructure while the above is under way.

“Broaden the Umbrella of Trust”

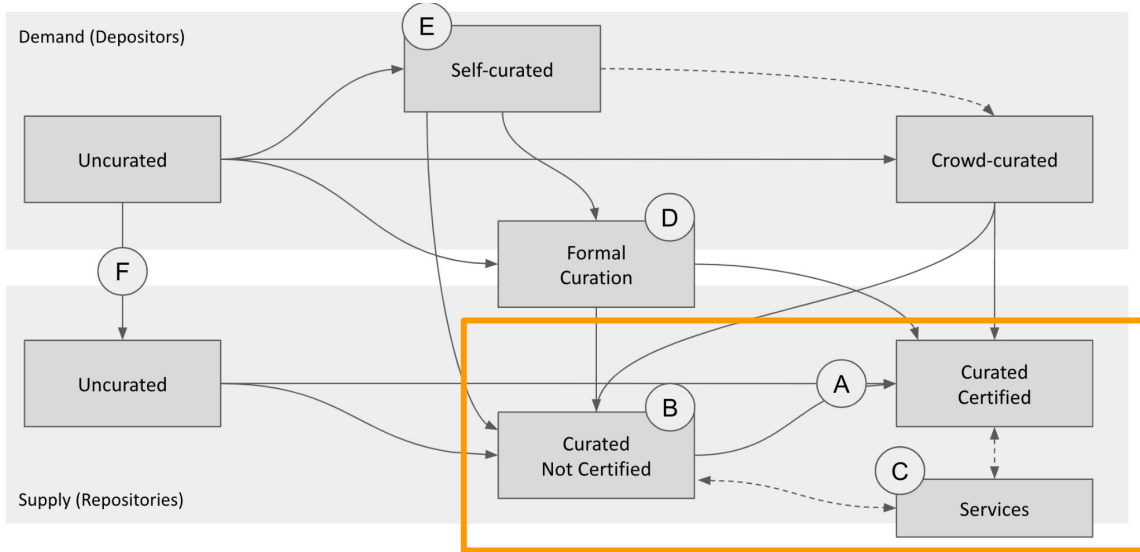
Degrees of Assurance in Performance Appraisal



Degrees of Assurance in Performance Appraisal



Value Proposition

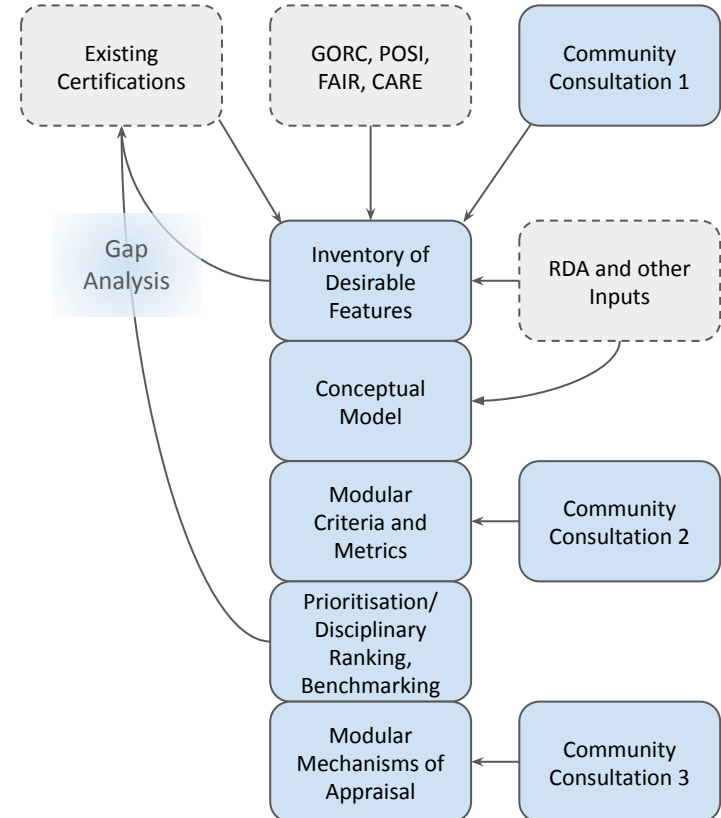


1. Efficiency of certification
 - a. Component services can be certified once
 - b. Modular, decentralised certification
 - c. Improved rate of appraisal
2. Appropriate criteria that can be made applicable to non-data repositories can accelerate the movement towards assessment and future certification of such repositories;
3. The WG will work towards community consensus on selecting the most appropriate repository or service for a given context, which will promote the investment and (re)use of the highest-performance service that meets user needs within that context.

Broad community engagement | Input to EOSC FIDELIS, EDEN | Input for Certification Authorities

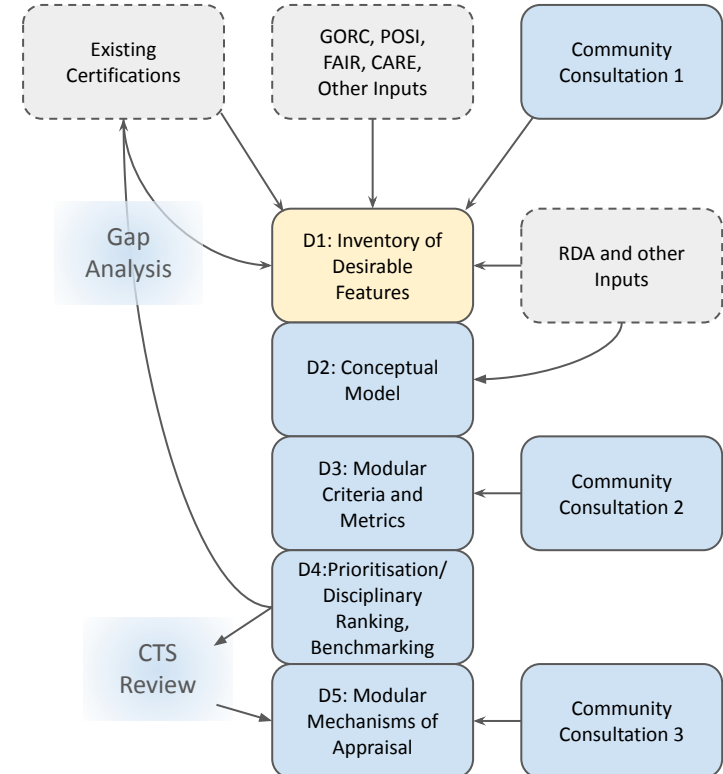
Working Group Outputs and Process

- **“Inventory of Expectations”**
 - Through community consultation and a scoping review, compile an inventory of performance expectations for repositories and related services, and create a mapping of this inventory to the portfolio of existing criteria applied by the certification authorities.
 - Based on the results of the community consultation and scoping review, develop a conceptual model of the actors and service providers in the ecosystem.
- **“Prioritisation”**
 - Use the generated mapping developed earlier to verify importance of individual features to the community
 - Identify gaps between expectations and current certification options,
 - Determine the criteria whereby an appropriate repository or service will be selected based on the context of the end user (scientific discipline, institutional affiliation, country, data formats, etc.).
- **“Appraisal and Verification”**
 - Determine which of the selected criteria should apply to generalist repositories, to services, other actors, and what level of performance can be expected of different actors in the ecosystem.
 - Define the ways on which level of performance can be verified, including but not limited to formal certification.
 - Determine, through stakeholder consultation, priorities for criteria implementation and identify metadata that can be associated with the implementation of each criterion to facilitate modular, decentralised assessment and certification.



Amended Work Plan

Project Timeline	Task	RDA Plenaries	Other Important Events
M-3 - M0	D.1	Plenary 22 - Virtual	Community Consultation 1
M1 - M3			Endorsement of WG
		Plenary 23 - Costa Rica	Community Consultation 2
M4 - M7	D.2		EOSC FIDELIS, EDEN Kick-Off
	D.3		Community Consultation 2a
			Plenary 24 - Virtual
M8 - M11	D.4		Finalise Gap Analysis
		IDW, Australia	
		D.5	CTS review: Gap Analysis and TRSP Criteria
Community Consultation 3			
Outputs for Community Review			
M16 - M18		Plenary 25	WG Conclusion



Community Consultation 1

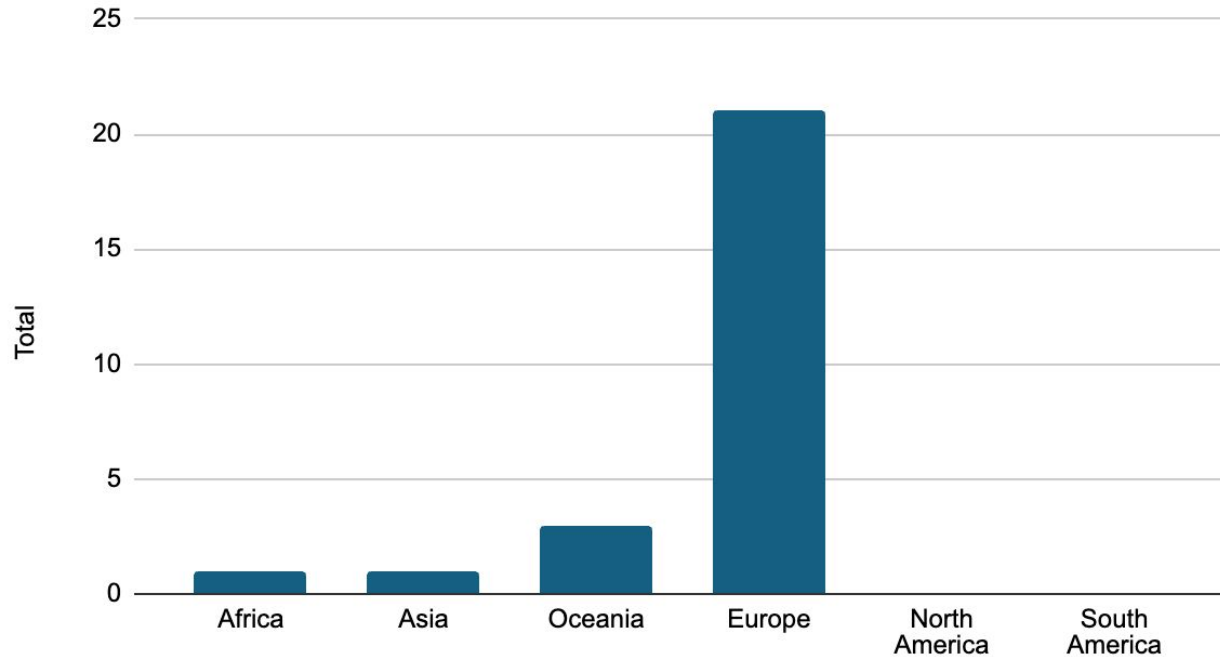
Feedback

Virtual Plenary 22

Wim Hugo, DANS

Location of Participants

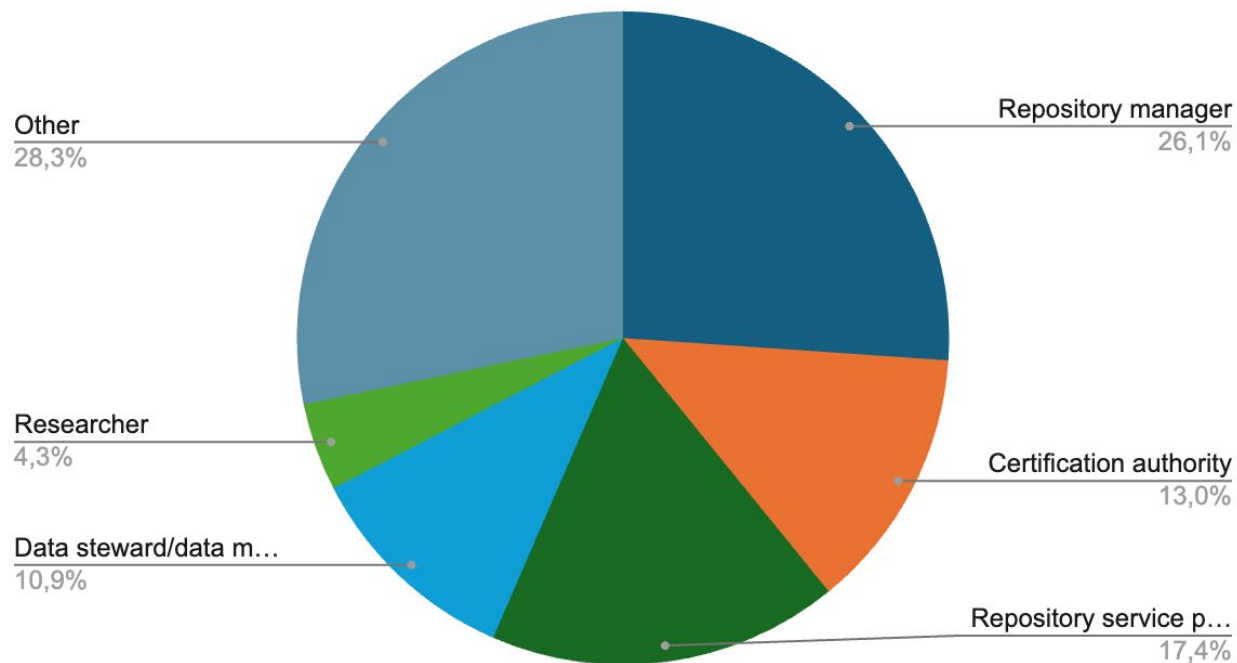
Location of Participants



Participant Role



Total

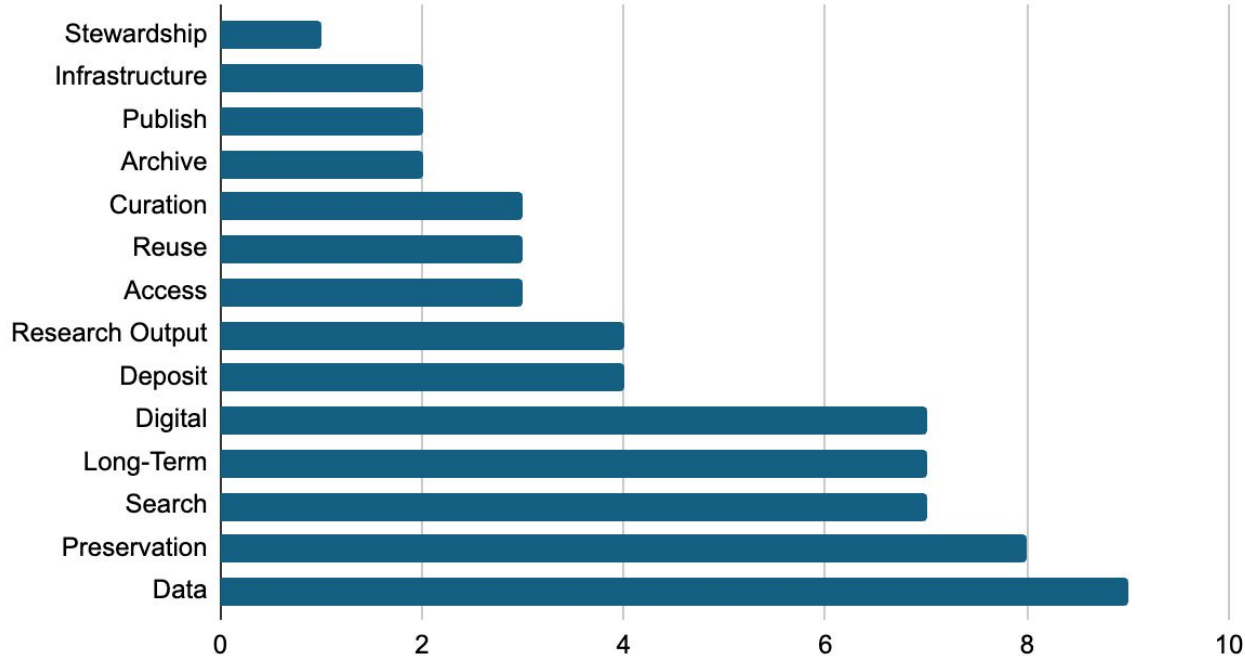


Repository: Definition

- A service providing stable and citable copies of research outputs for the long term
- place to store digital objects
- A database for storing any digital research objects... Not just data! That allows deposition
- Institution or organization providing long-term preservation of digital research objects
- an infrastructure for long term storage of data for the purposes of retention and reuse
- A service to science opening and enriching
- Entity providing long term storage and curation
- a service to facilitate data opening and sharing, with curation and standard
- A place for long term curation
- One could have an intermediary level 'on the way to certification'
- By it's activities, functions, levels of care and levels of responsibility.
- Facility the enables data reuse
- Platform for sharing and preserving digital resources.
- Archive for digital research outputs that provide quality assurance and preservation services
- A place for digital assets with a plan to keep them accessible and alive.
- Deposit service for research outputs
- Infrastructure to store, preserve and publish output
- publishes and preserves data
- Discovery and dissemination of research outputs
- Place to steward and secure services for scholarly/info. assets
- Platform for depositing and accessing digital data. It guarantees the preservation and accessibility of data sets over the long term.
- Repository contains technical platform, processes for preservation and management for data reuse.
- Archive and search/ discovery service
- Something whose main (but not sole) functions are the deposition and long-term storage of data.
- Service that enables access and preservation of quality data

Repository: Definition Analysis

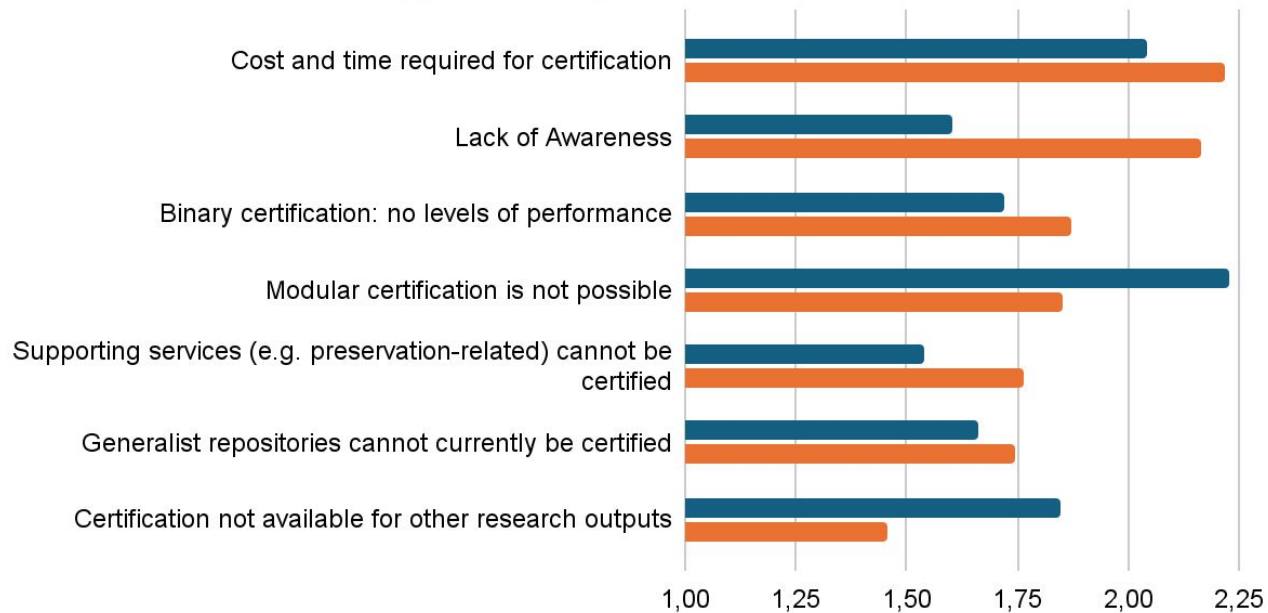
Major characteristics



Relative Importance of Issues

Ranking of Issues

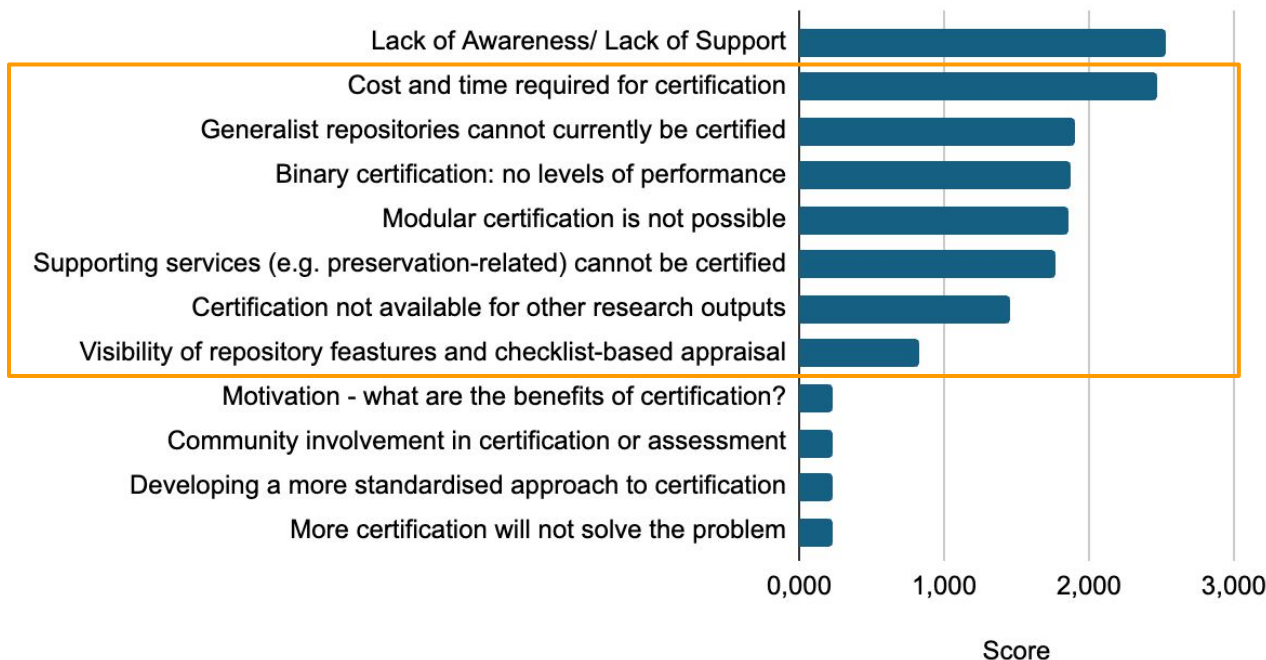
■ Deviation ■ Weighted Average



Consolidated Issues



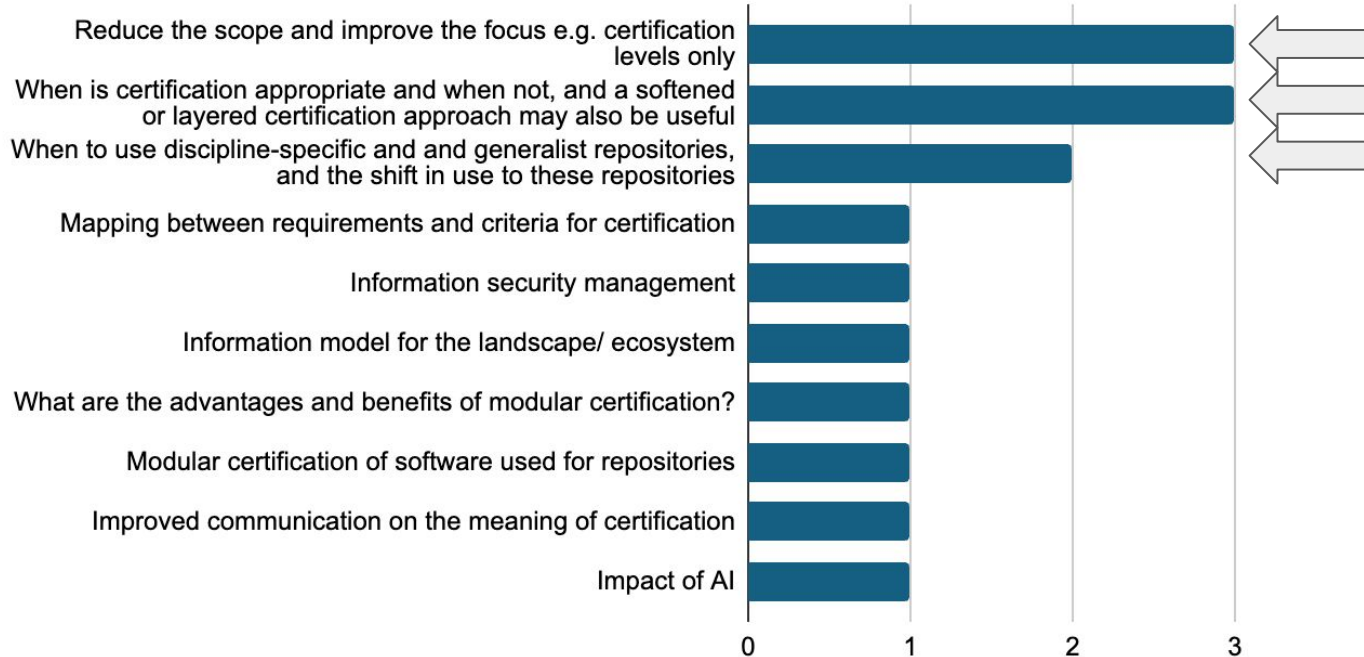
Consolidated Issues



Scope of Work



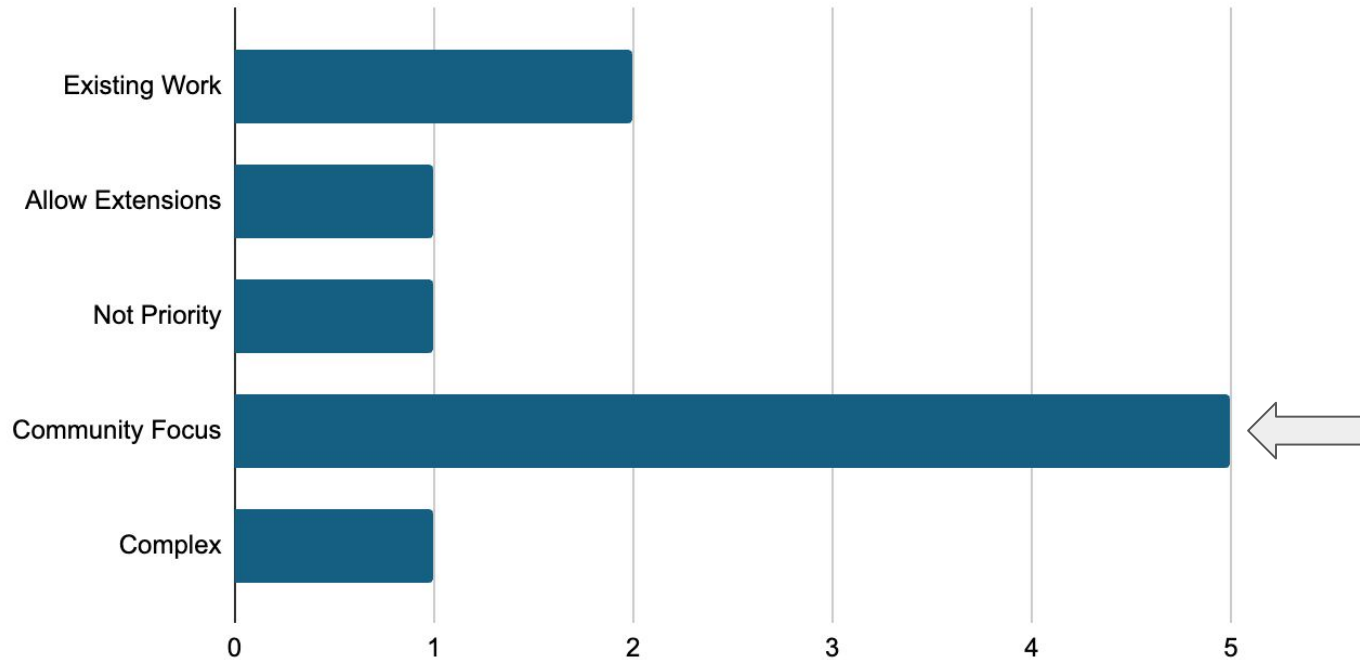
Scope of Work in WG



Compiling an Inventory



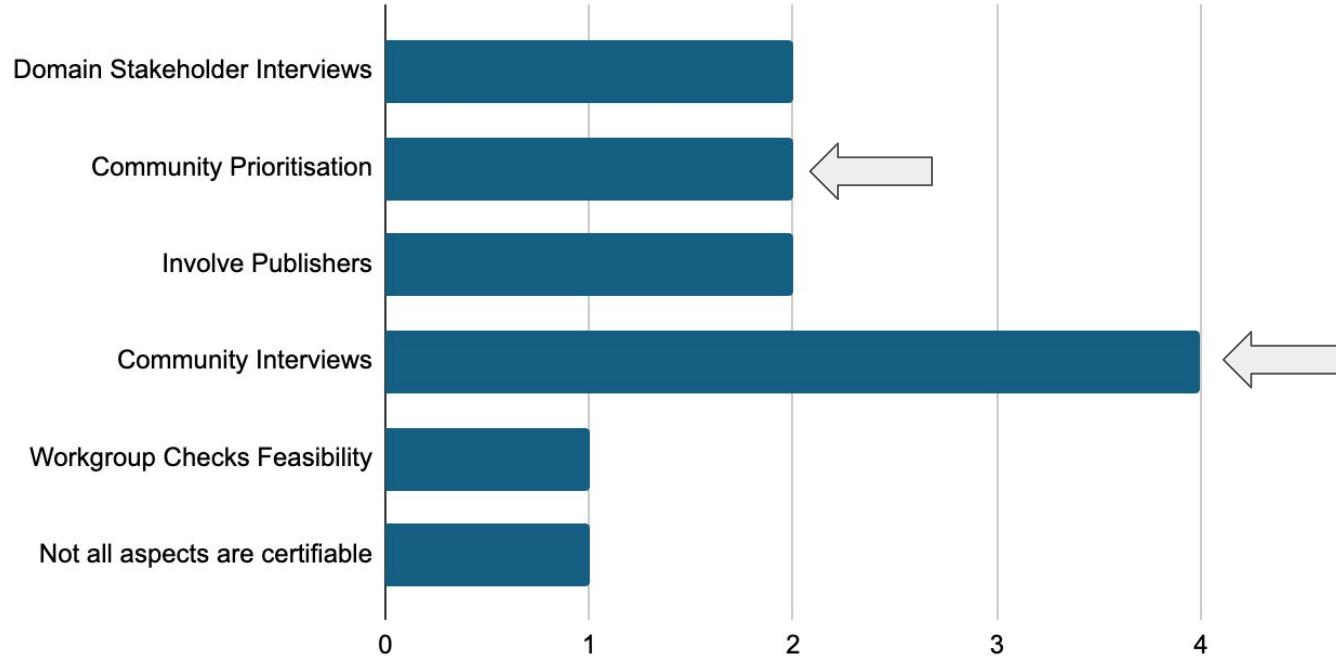
Compiling and Inventory of Characteristics



Prioritisation Process

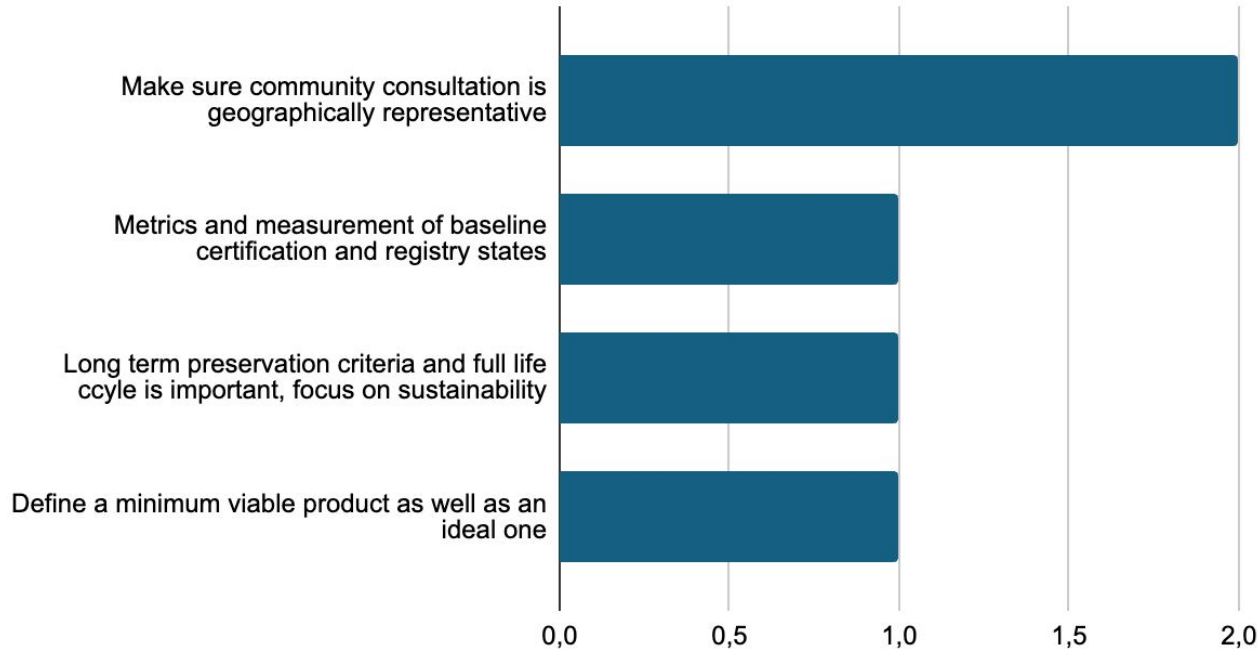


Process for Prioritisation of Criteria



Suggested work plan improvements

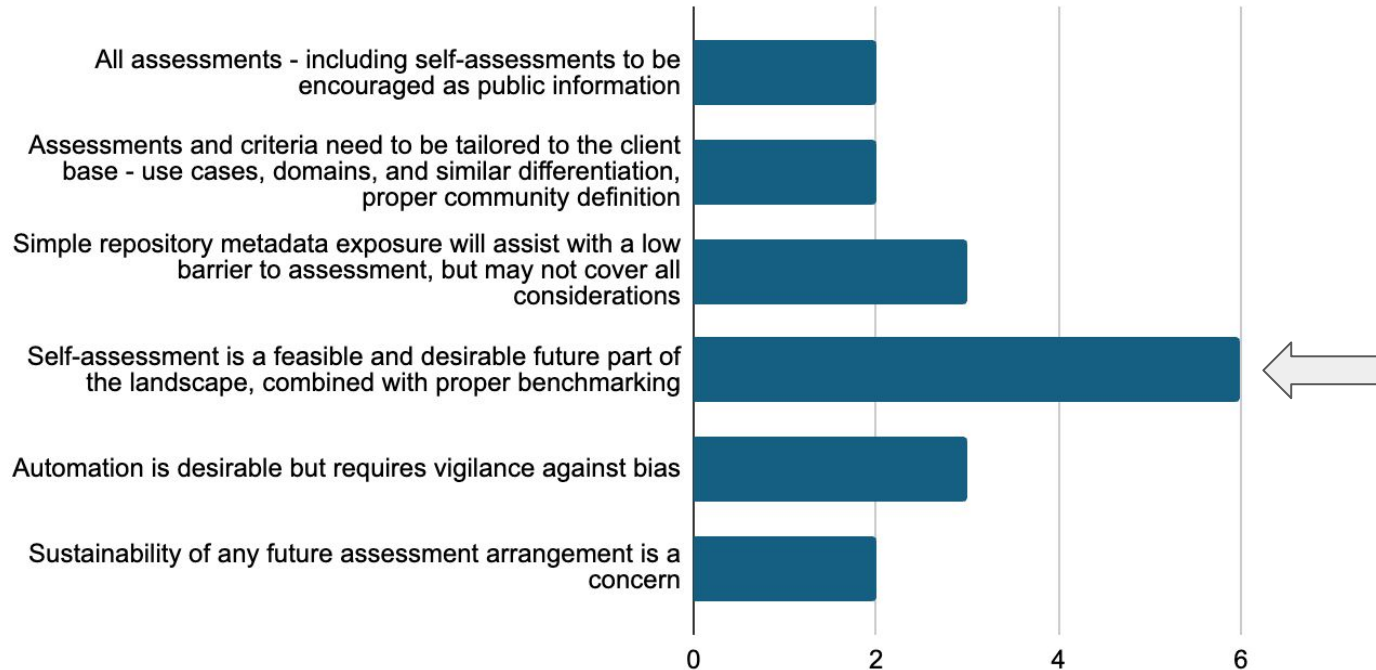
Work Plan Suggestions



Approaches to Appraisal and Certification

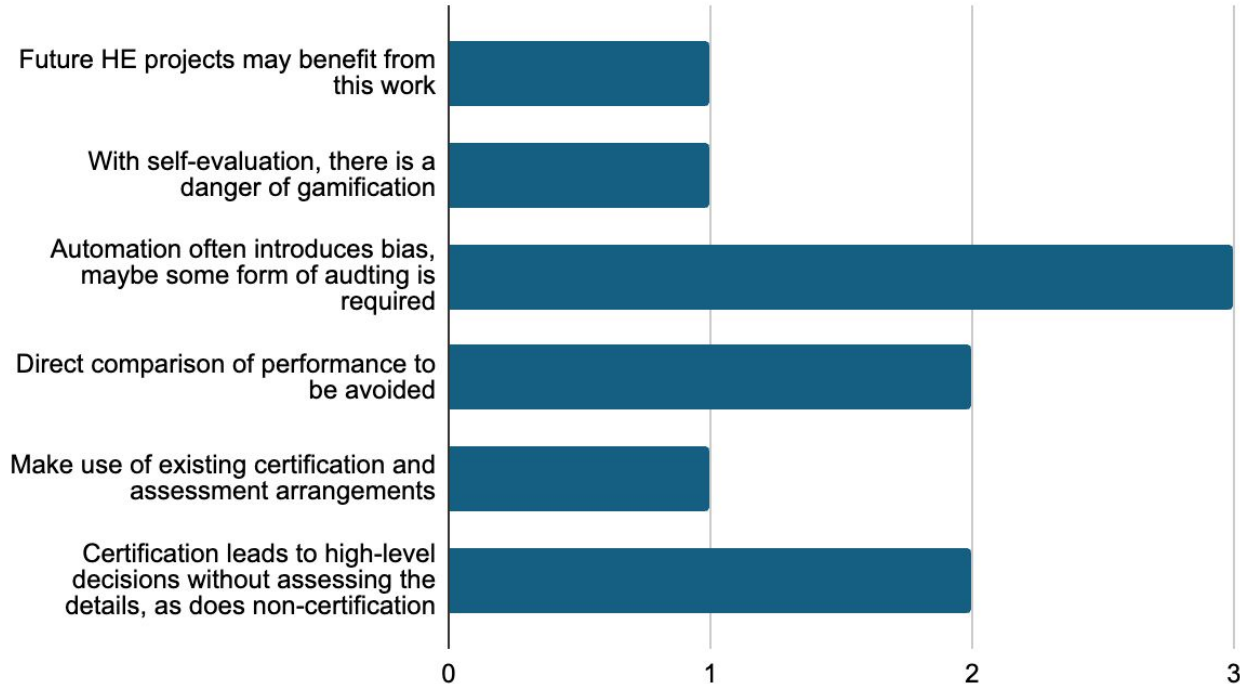


Approaches to Appraisal and Certification



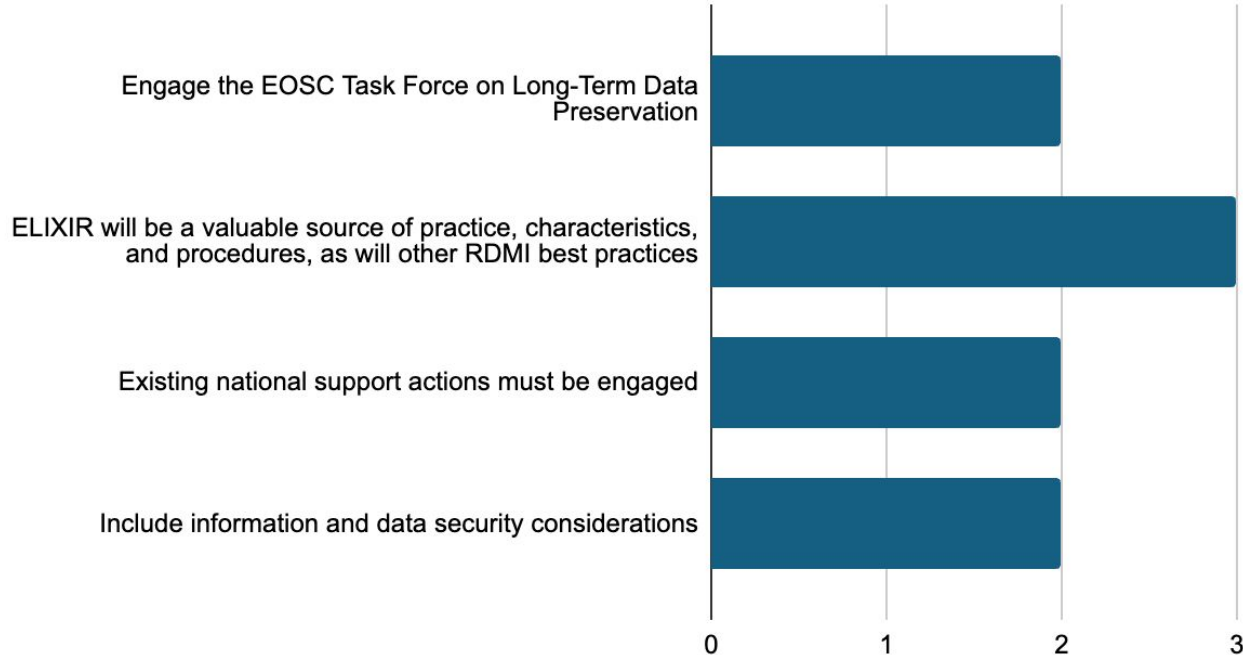
Other Benefits and Side Effects

Other benefits and side effects



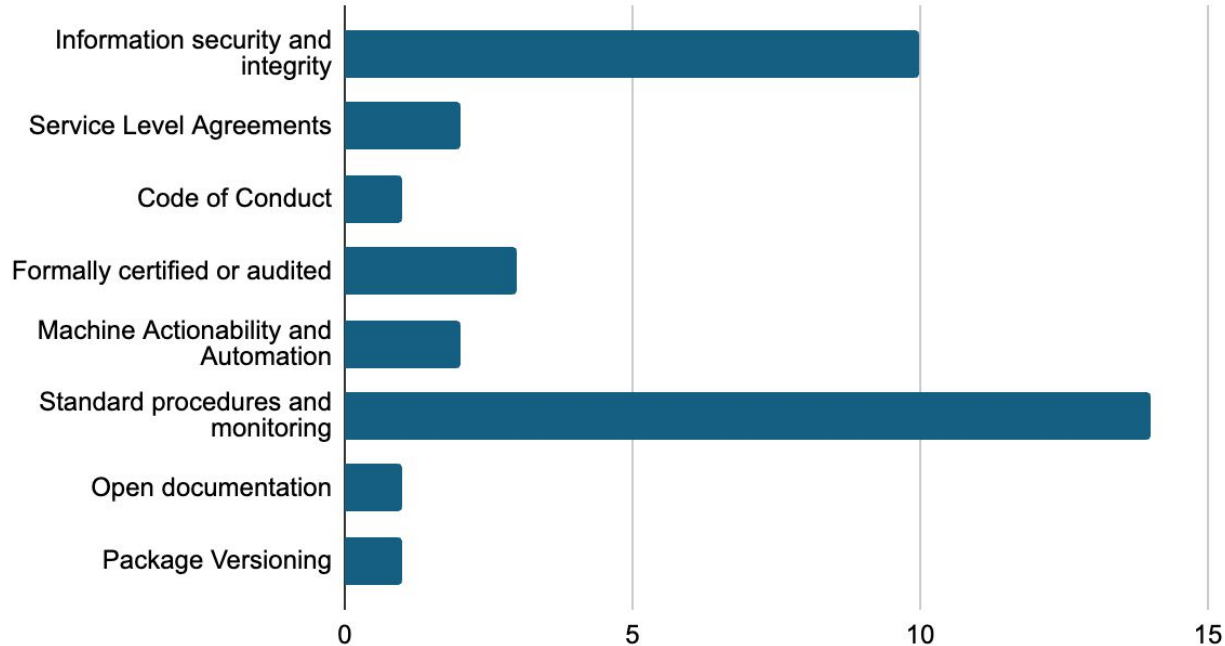
Important additional inputs identified by participants

Important additional inputs



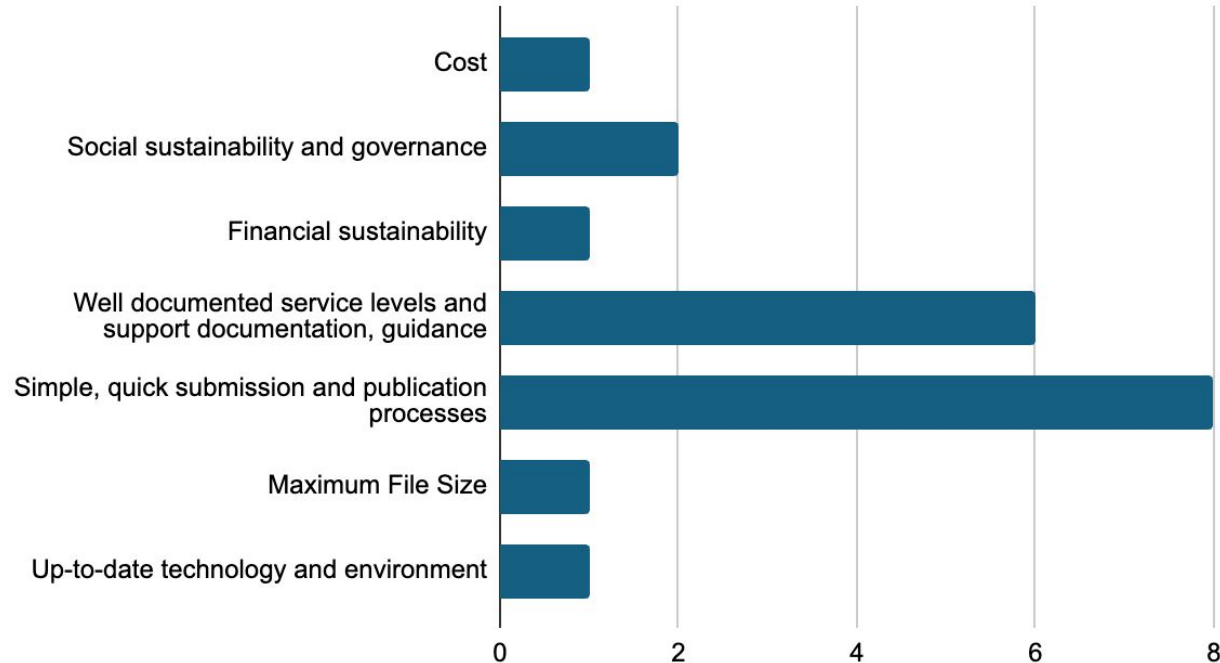
Important additional service provider criteria

Service Provider Criteria



Other criteria identified by participants

Other Criteria



Desirable Features and Characteristics

Draft Inventory

Wim Hugo, DANS

Normative and Informative

Desirable characteristics of repositories and services can be normative or informative.

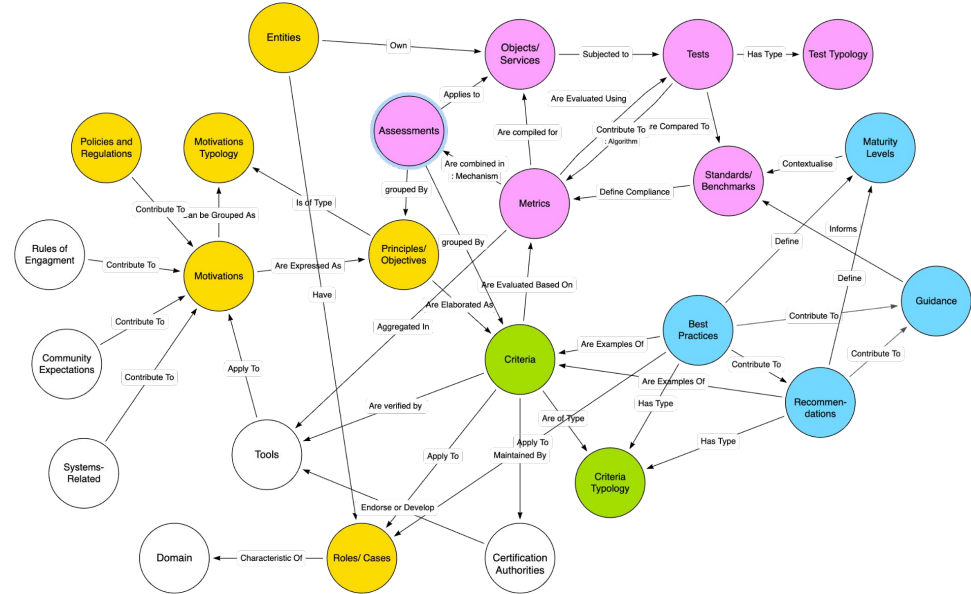
Normative characteristics are used in assessments and certification, and can be specified and/ or standardised in respect of criteria, metrics, tests, benchmarks and performance levels, and assessment rubrics.

Informative characteristics are used to select an appropriate repository or service based on end user needs, and will be appraised differently in each case. Standardising these criteria makes it possible to objectively select services and to potentially rank them.

Conceptual Model: Assessment and Appraisal

Considerable effort has gone into a conceptual model for assessment developed by the FAIRCORE4EOSC Project, and currently being validated in the OSTRAILS, QUANTUM, and EOSC Beyond Projects.

The model defines the relationship between principles, criteria, metrics, tests, benchmarks, and assessments (or appraisals)

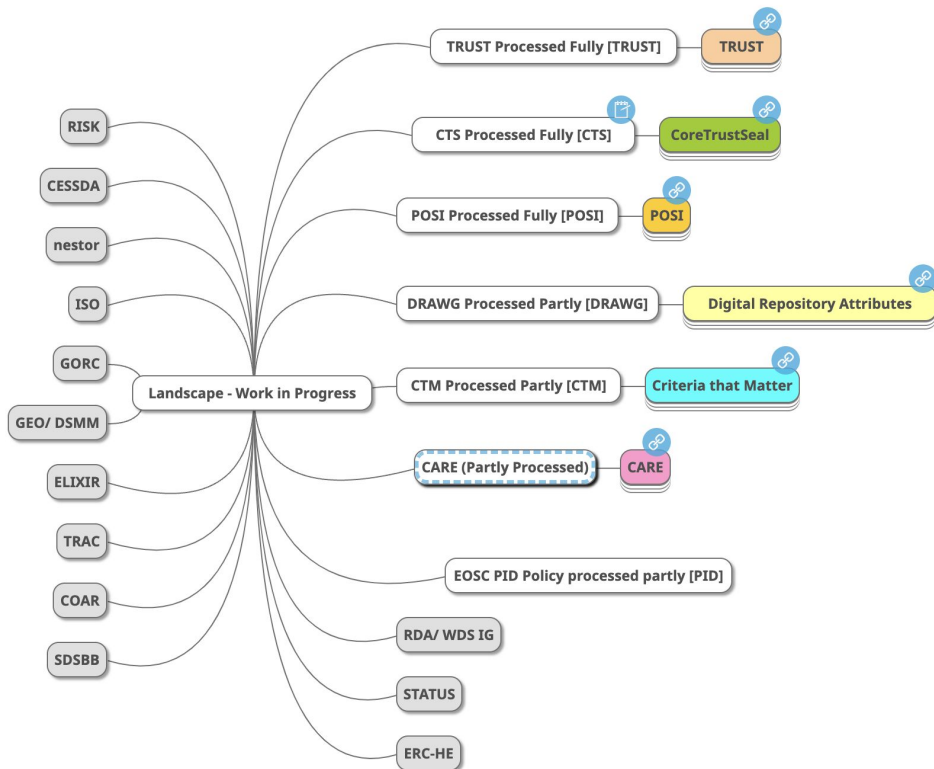


Selection

When an individual, project, institution, or community selects an appropriate repository or service:

1. They can mix and match normative and informative criteria
2. They can define their own metrics, tests, and benchmarks

Landscape Analysis - Scope



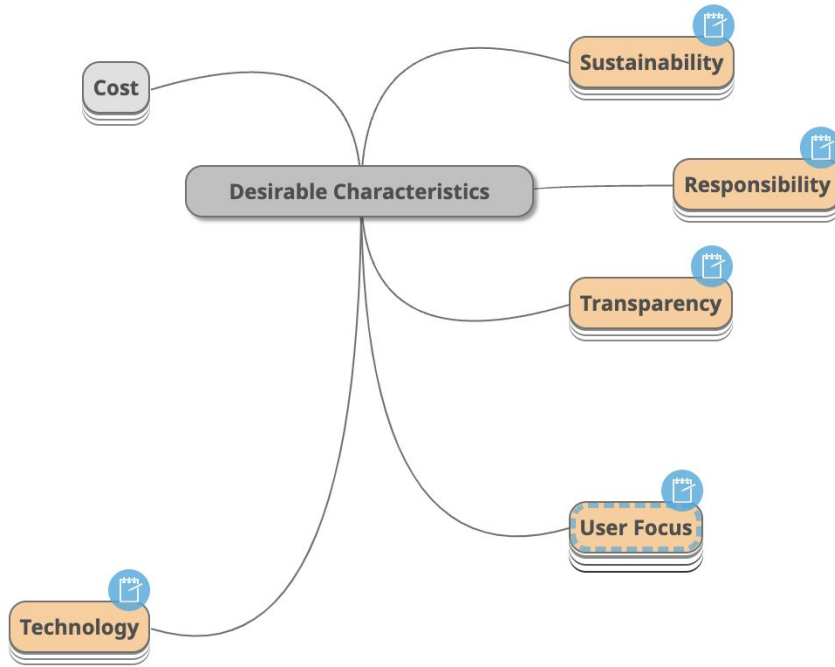
#	Label	Source of Information	Ref
1	TRUST	Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7	[1]
2	CTS	CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012	[2]
3	RISK	Frank, R.D. (2024), "Constructing risk in trustworthy digital repositories", Journal of Documentation, Vol. 80 No. 6, pp. 1508-1527. https://doi.org/10.1108/JD-08-2023-0157	[3]
4	POSI	Bilder G, Lin J, Neylon C (2020), The Principles of Open Scholarly Infrastructure, retrieved 08-11-2024, https://doi.org/10.24343/C34W2H	[4]
5	CESSDA	CESSDA, CESSDA Trust and Landscape Report 2024, https://docs.google.com/document/d/131oqNWGwZGICdvSOCKLgUFOoP1bbiFFmpANW4alH7E/edit?usp=sharing	[5]
6	nestor	nestor criteria : Catalogue of Criteria for Trusted Digital Repositories, Version 2, 2009, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2010030806 http://nbn-resolving.de/urn:nbn:de:0008-2010030806 https://d-nb.info/1189191830/34	[6]
7	ISO	International Organization for Standardization. Audit and certification of trustworthy digital repositories. 2012. https://www.iso.org/standard/56510.html	[7]
8	GORC	Woodford, C., Treloar, A., Leggott, M., Payne, K., Jones, S., Lopez Albacete, J., Madalli, D., Genova, F., Dharmawardena, K., Chibhira, N., Åkerström, W. N., Macneil, R., Nurnberger, A., Pfeifferberger, H., Tanifuji, M., Zhang, Q., Jones, N., Sesink, L., Wood-Charlson, E., & RDA GORC International Model WG. (2023). The Global Open Research Commons International Model, Version 1 (1.0). Zenodo. https://doi.org/10.15497/RDA00099	[8]
9	CARE	Carroll, S.R., Garba, I., Figueroa-Rodriguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020) 'The CARE Principles for Indigenous Data Governance', Data Science Journal, 19(1), p. 43. Available at: https://doi.org/10.5334/dsj-2020-043	[9]

[Landscape Analysis Table](#)

[Landscape Analysis Map](#)

(needs [MindMup](#) to edit)

Landscape Analysis - Inventory

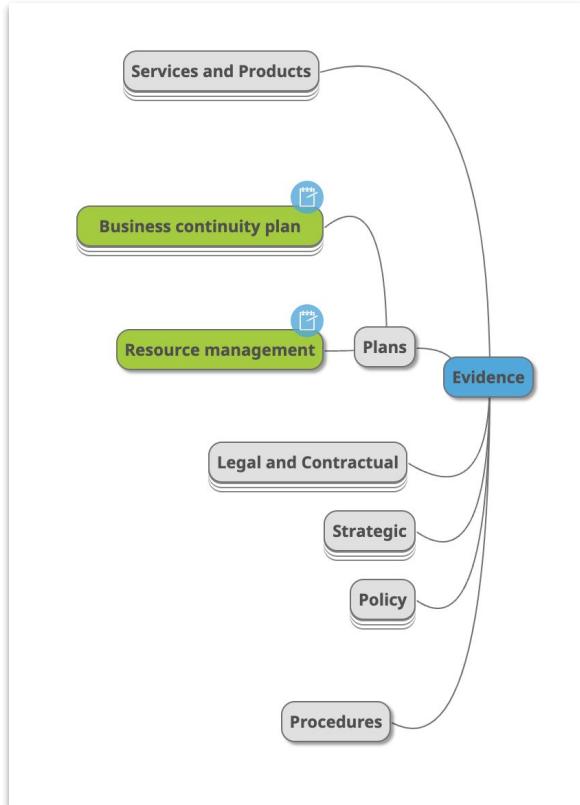


Characteristic or Feature	Description	Sources	Assessm	
			Certified Repos/ Services	Self- As with t
Sustainability				
Social Sustainability	Aspects such as open governance, community involvement and engagement, planning and continuity arrangements, forum and working group participation and creation. Frequent user surveys, publication of roadmaps.			
Governance		CTS, POSI, TRUST, CARE	Yes	
Expert Review		CTS	Yes	
Planning/ Continuity		CTS, POSI, TRUST	Yes	
Community Engagement		CTS, POSI, TRUST, CARE	Yes	
Financial Sustainability	Demonstrable longer-term funding and income sources. Structural income is adequate.	POSI, CTS, TRUST	Yes	
Technical Sustainability	Open availability of the technical means to reestablish the service elsewhere	POSI, CTS, TRUST	Partly	
Responsibility				
Schema	Establishing appropriate metadata schema, as well as assisting to identify applicable file formats and schema			
Metadata and Data		CTS, TRUST	Yes	
Applicable and appropriate		CTS, TRUST	Yes	
Stewardship				
Documentation	Ensuring that deposit documentation is adequate, meets well-defined quality standards or guidelines, and is technically sound (schema validation, file formats,...)	CTS, TRUST	Yes	
Quality control		CTS, TRUST, CTM	Yes	
Technical validation		CTS, TRUST, CTM	Yes	
Long-term persistence	Format and bit-level preservation	CTS, TRUST, CARE	Yes	
Transfer	Allow transfer of stewardship	CTS, TRUST	Yes	
Confidentiality and Ethics	Determine if ethical norms have been adhered to, if consent has been granted by subjects, confidentiality respected, and similar concerns	CTS, TRUST, CTM	Yes	
Confidentiality		CTS, TRUST	Yes	
Ethics		CTS, TRUST, CARE	Yes	
Integrity	Validation of provenance and authenticity, protection and recording of provenance, verification of depositors/ content editors	CTS, TRUST	Yes	
Authenticity - determine and protect		CTS, CTM	Yes	
Identify depositors		CTS	Yes	
Rights Management	Aspects such as protection of the rights of the	CTS, TRUST, CTM	Partly	

[Desirable Characteristics Table](#)

[Landscape Analysis Map](#)

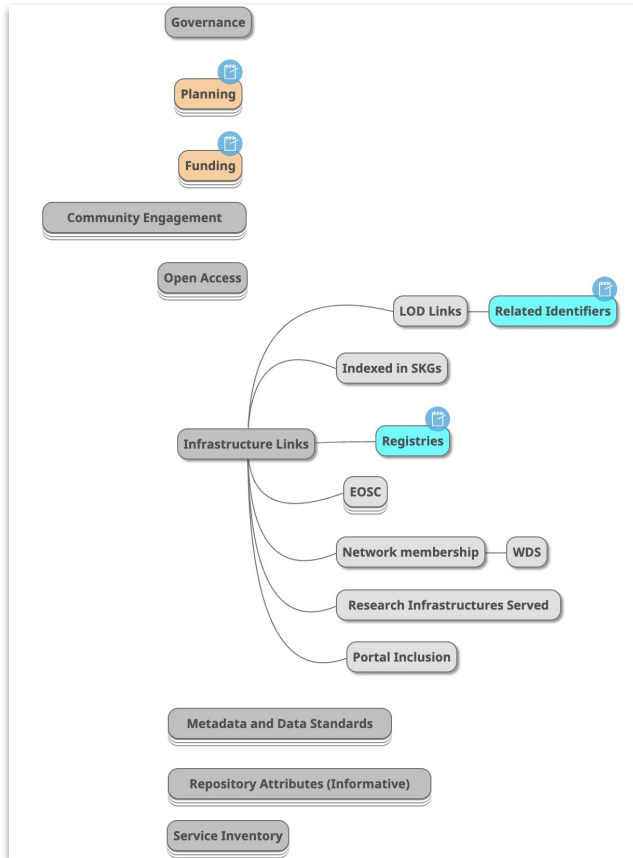
Landscape Analysis - Evidence



- Evidence is linked to or is included in tests
- No significant work done except recording it when encountered
- Not a formal WG output

[Landscape Analysis Map](#)

Landscape Analysis - Benchmarks and Implementations



- This focuses on ‘how’ and not ‘why’ and ‘what’
- No significant work done except recording it when encountered
- Not a formal WG output

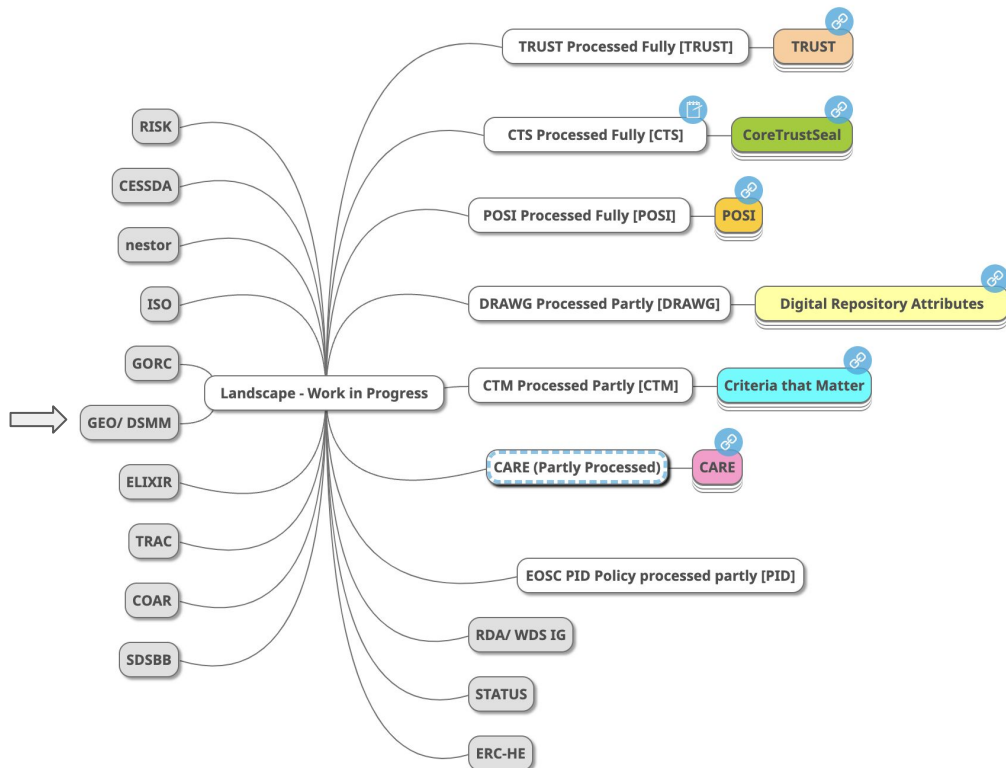
[Landscape Analysis Map](#)

Breakout Sessions

Breakout Groups: Scope of Discussion

- Group 1 (Online): Are we missing any significant ***sources*** of repository or service features and characteristics?
 - [Spreadsheet Tab 1](#) (Landscape) and [detailed slide](#)
 - Add new items at the bottom or report back
- Group 2: Are we missing obvious ***features or characteristics***? Should we organise them differently?
 - [Spreadsheet Tab 2](#) (Characteristics) and details ([Group 2a](#), [Group 2b](#))
 - Start from the top down - main groupings, next level, etc. Group 2a is more important.
 - Add new items at the bottom or report back
- Group 3:
 - Look at the scope and content of what we want to confirm - [detailed slide](#)
 - [Spreadsheet Tab 2](#) (Characteristics)
 - Record group consensus

Breakout Group 1: Online

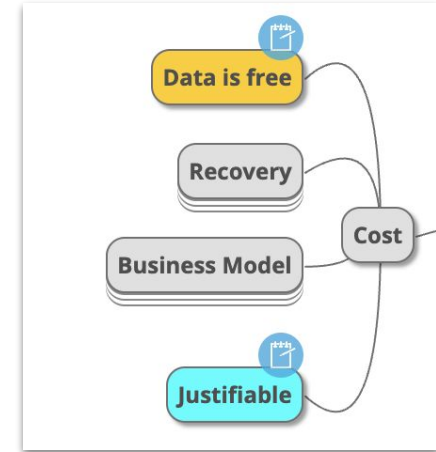
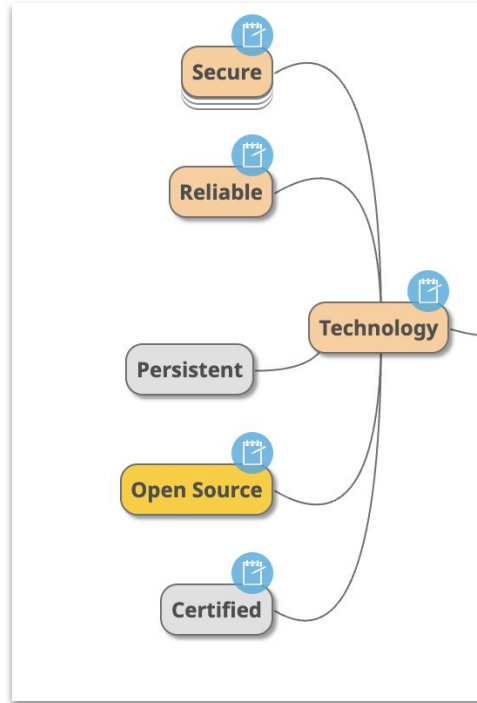
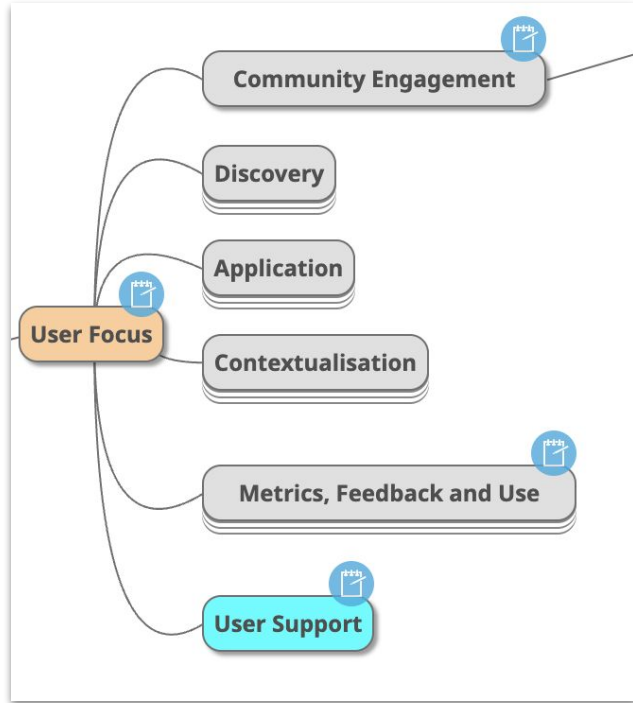


#	Label	Source of Information	Ref
1	TRUST	Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7	[1]
2	CTS	CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012	[2]
3	RISK	Frank, R.D. (2024), "Constructing risk in trustworthy digital repositories", Journal of Documentation, Vol. 80 No. 6, pp. 1508-1527. https://doi.org/10.1108/JD-08-2023-0157	[3]
4	POSI	Bilder G, Lin J, Neylon C (2020), The Principles of Open Scholarly Infrastructure, retrieved 08-11-2024, https://doi.org/10.24343/C34W2H	[4]
5	CESSDA	CESSDA, CESSDA Trust and Landscape Report 2024, https://docs.google.com/document/d/131oqNWGwZGICdvSOCKLgUFOoP1bbiFFfmpANW4alH7E/edit?usp=sharing	[5]
6	nestor	nestor criteria : Catalogue of Criteria for Trusted Digital Repositories, Version 2, 2009, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2010030806 http://nbn-resolving.de/urn:nbn:de:0008-2010030806 https://d-nb.info/1189191830/34	[6]
7	ISO	International Organization for Standardization. Audit and certification of trustworthy digital repositories. 2012. https://www.iso.org/standard/56510.html	[7]
8	GORC	Woodford, C., Treloar, A., Leggott, M., Payne, K., Jones, S., Lopez Albacete, J., Madalli, D., Genova, F., Dharmawardena, K., Chibhira, N., Åkerström, W. N., Macneil, R., Nurnberger, A., Pfeiffenberger, H., Tanifuji, M., Zhang, Q., Jones, N., Sesink, L., Wood-Charlson, E., & RDA GORC International Model WG. (2023). The Global Open Research Commons International Model, Version 1 (1.0). Zenodo. https://doi.org/10.15497/RDA00099	[8]
9	CARE	Carroll, S.R., Garba, I., Figueroa-Rodriguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020) 'The CARE Principles for Indigenous Data Governance', Data Science Journal, 19(1), p. 43. Available at: https://doi.org/10.5334/dsj-2020-043	[9]

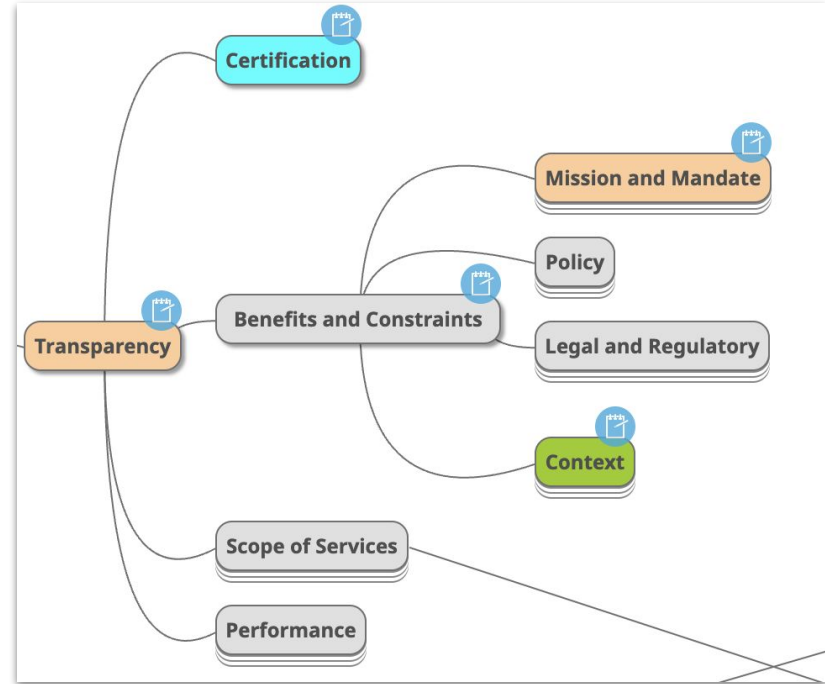
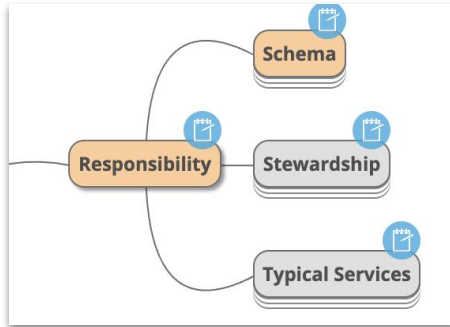
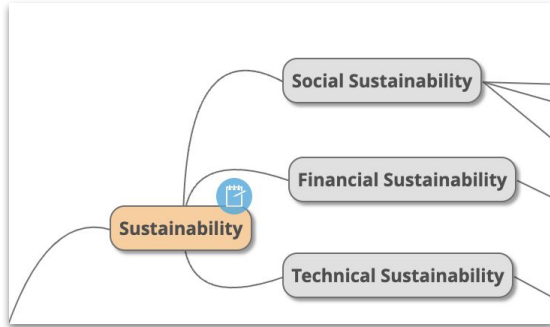
[Landscape Analysis Table](#)

[Landscape Analysis Map](#)

Breakout Group 2a: Content and its Organisation



Breakout Group 2b: Content and its Organisation



Breakout Group 3: Scope of Our Proposed Analysis

Assessment Type		
Certified Repos/ Services	Self- Assessment with Evidence	Self- Assessment

Yes ✓

Partly

No

Maybe

Research Output			
Datasets	Articles and Reports	Software	Semantic Artefacts

Mandatory

Optional

Informative

No

Applicability	
Stakeholder	TRSPs

All

End User

Depositor

Publisher

Funder

Other

Yes

Partly

No

Maybe

Breakout Sessions - Feedback

Open Discussion

Next

Regular Workgroup Meetings

Starting 11 December (15h00 UTC)

- Meeting structures and ways of working
- Additional co-chair(s) welcome



Link to WG
page
Please join to
access
meeting info!

Task Groups are proposed for the first phase of work to end January 2025 :

- Task 1 - Inventory of Desirable Characteristics
- Task 2 - Characteristics Applicable to TRSPs
- Task 3 - Characteristics Applicable to Other Research Outputs

Please volunteer to join the tasks, and we encourage members to be task coordinators

2 task meetings/ work sessions by end January

