

Certification+FAIR support workshop for data repositories

The Hague, Data Archiving and Networked Services (DANS)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

...

Twitter: @FAIRsFAIR_EU





Welcome!



On 6/02, at @DANSKNAW

premises, #FAIRsFAIR is organising the first #Datarepositories workshop to support selected #repositories on achieving the @CoreTrustSeal certification. What procedures to follow? What are the #certification requirements? Know more fairsfair.eu/events/fairsfa...

CORPORATION CONTRACTOR OF



Workshop organizers









ほうくだい ちゅうくだい ちゅうくだい ちゅう





Workshop participants, representatives from:

<u>Apollo - United Kingdom</u>

DaSCH (Data and Service Center for the Humanities) - Switzerland

DASS-BiH (Data Archive for Social Sciences & The Humanities in Bosnia & Herzegovina) - Bosnia & Herzegovina

くくがく ちゅうくくがく ちゅうくうがく

DASSH - The Archive for Marine Species and Habitats Data - United Kingdom

ESRF Data Repository - France

IAGOS Data Center - France

ICOS Data Portal - Sweden

The Movebank Data Repository - Germany

SOCIB (Balearic Islands Coastal Ocean Observing and Forecasting System Data Repository) Spain

Tárki Data Archive - Hungary



What do you expect to take home from this 1st Certification Support workshop?

"Enough knowledge to put in place our internal **roadmap** for CoreTrustSeal accreditation."



"A clearer understanding of the **requirements**."

"A clear idea of what COreTrustSeal means by FAIR-compliant" "....get back ideas about how to improve or better implement the **fair principles** in our repository."

> "Also learning from others; what is their viewpoint? What are examples of bottlenecks/thresholds for those that did already an application?"

Icon made by Freepik from flaticon.com

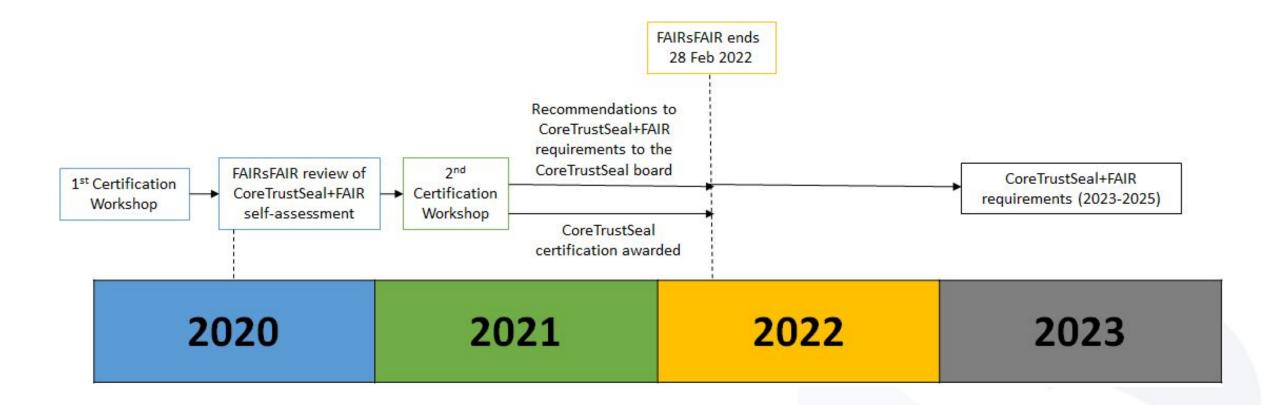


This workshop starts the journey towards...

- A support programme for CoreTrustSeal 2020-2022
- Sharing your repository practices; how they enable FAIR data
- Contributing to the Community review of CoreTrustSeal (2022): will take into account FAIR-elaboration of repository certification requirements (CoreTrustSeal+FAIR)
- peer-community support: helpful connections to other repositories with similar challenges
- Ready for CoreTrustSeal+FAIR 2023-2025!



CoreTrustSeal Meets FAIR Timeline



ほうくだい ちゅうくだい ちゅうくだいがく



Assessment work in FAIRsFAIR

- Evaluation and certification of repositories enabling FAIR data
- Evaluation and certification of FAIR data objects within a repository
- wider range of standard/assessment expectations, e.g. on services and software
- other initiatives that might be set for involvement in the EOSC

The **FAIRsFAIR-project** aims to supply practical solutions for the use of FAIR data principles

March 2019 – Feb 2022

22 partners

6 core partners DANS (coordinator), CSC, DCC, TrustIT, STFC, EUA

www.fairsfair.eu @FAIRsFAIR_EU



CoreTrustSeal in a NUTSHELL

Presenter: Mustapha Mokrane (DANS)



... that describe the characteristics of a Trustworthy Digital Repository.

Click the play button for a short video on CoreTrustSeal Trusted Digital Repositories and FAIR.

The video is licensed under a Creative Commons Attribution 4.0 International License.

The material is made by Elliott Walker, Yes Ditsen de Jong, Kent de Bruin and Niovi Dimitra Chatzipoufli as part of a cooperation project between the Master Digital Design of the Amsterdam University of Applied Sciences and DANS (Data Archiving and Networked Services).

October 2019.

and the second second



CoreTrustSeal Objective

Universal framework for the certification of data repositories at the core level to support long-term access to reusable data.

- Gives **data producers assurance** that data are preserved and remain reusable in the future;
- Provides **funding bodies confidence** that investments are maximized;
- Enables data consumers to choose the repositories where data are held;
- Supports repositories improve processes for efficient data archiving and distribution.



Core Requirements: history

Five criteria to determine whether digital data are sustainably archived:

- I. Data can be **found** on the web
- II. Data are **accessible**, within limits of IPR and personal data legislation
- III. Data are available in a **usable** format
- IV. Data are **reliable**
- V. Data can be **referred** to (with a persistent identifier)

L'Hours, H., Kleemola, M. and de Leeuw, L. (2019) **"CoreTrustSeal: From academic collaboration to sustainable services"**, IASSIST Quarterly, 43(1), pp. 1-17. <u>https://www.doi.org/10.29173/iq936</u>





Core Requirements: foundation

+

OAIS Reference Model

Built on the basis of the Open Archival Information System reference model

Designated Community

Preserving data accessibility and understandability for a 'designated community'

Communitydriven

+

sharing of expertise and effort for certification

Core Trustworthy Data Repositories Requirements

Reflect the characteristics of trustworthy repositories.



Core Requirements: Stakeholders

Data producer	research integrityformatsdocumentation
Data consumers	terms of uselicencescitation
Data Repositories	 quality of storage organisational processes technical infrastructure along term availability



Current uptake: global





The 16 Requirements

Categories:

Background information (R0)

- 1. Organizational infrastructure (R1-6)
- 2. Digital object management (R7-14)
- 3. Technology and security (R15-16)

Applicant feedback

DOI: 10.5281/zenodo.667809



.



CoreTrustSeal Requirements

Background information

RO. Please provide context for your organization (repository type, designated community, level of curation...)

Organizational infrastructure

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

- **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.
- **R5.** The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.
- **R6.** The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

ふうぎょうか かくうぎょうかかくろ



CoreTrustSeal Requirements

Digital object management

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

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

R9. The repository applies documented processes and procedures in managing archival storage of the data.

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.

R12. Archiving takes place according to defined workflows from ingest to dissemination.

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.



CoreTrustSeal Requirements

Technology and security

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Applicant feedback



Compliance level

Applicant indicate a compliance level for each requirement:

- 0 Not applicable
- 1 The repository has not considered this yet
- 2 The repository has a theoretical concept
- 3 The repository is in the implementation phase
- 4 The guideline has been fully implemented in the repository

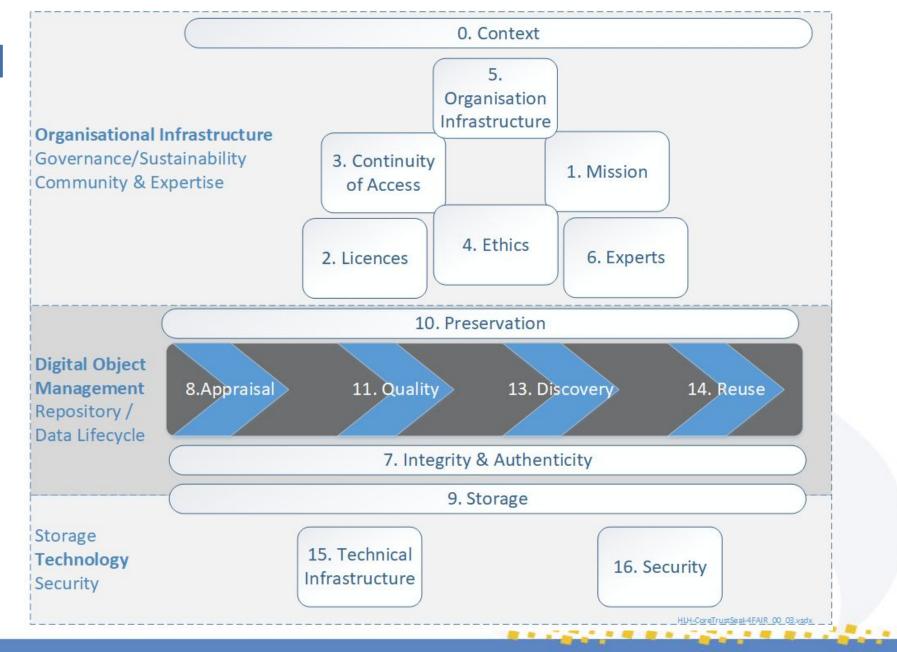


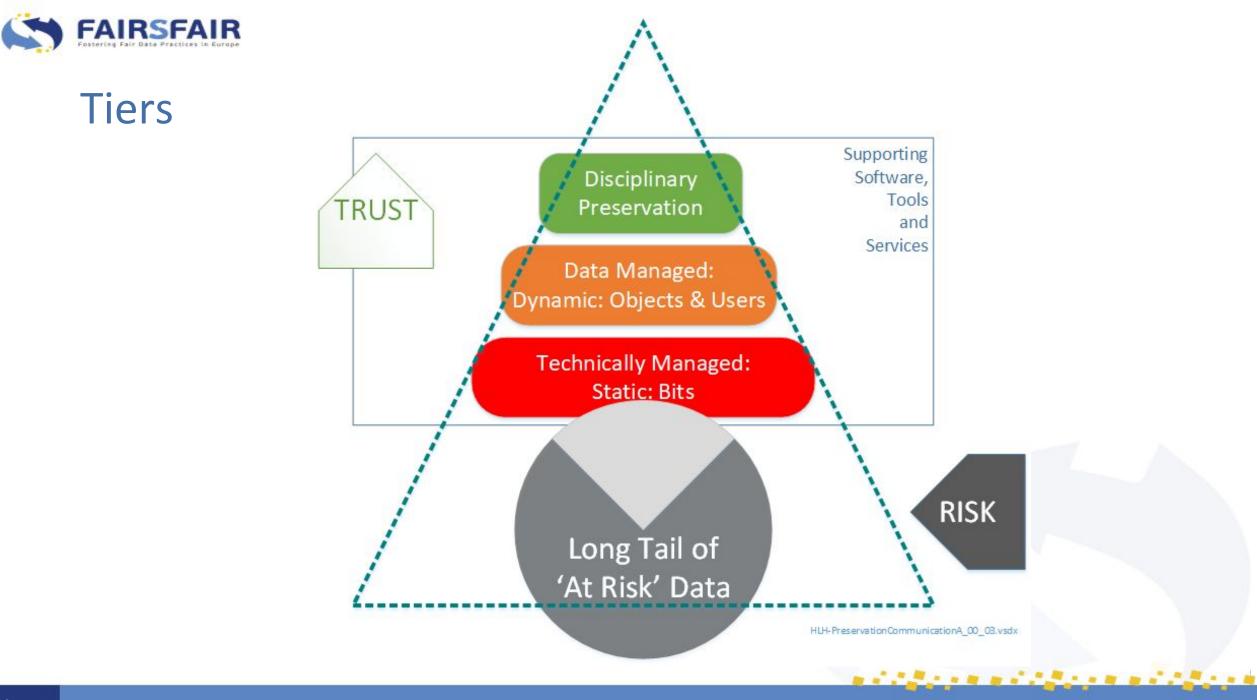
FAIR - CoreTrustSeal - Maturity

Presenters: Hervé L'Hours (UKDA) & Anusuriya Devaraju (PANGAEA)



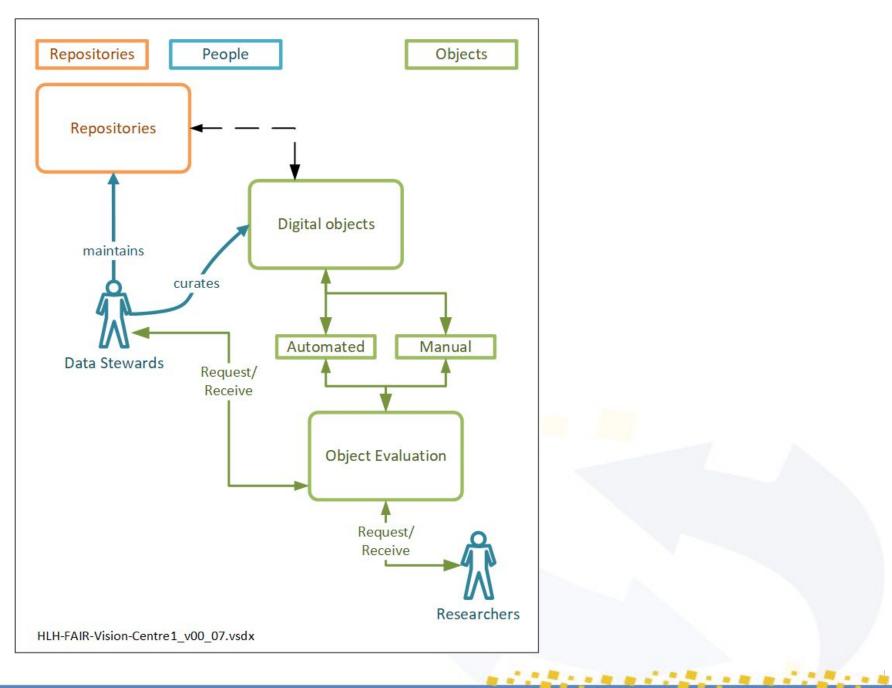
CoreTrustSeal





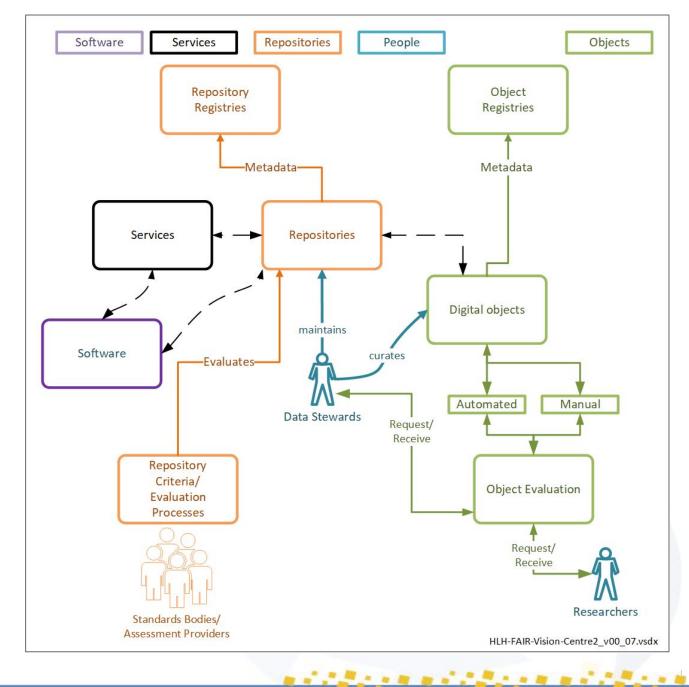


Initial Focus

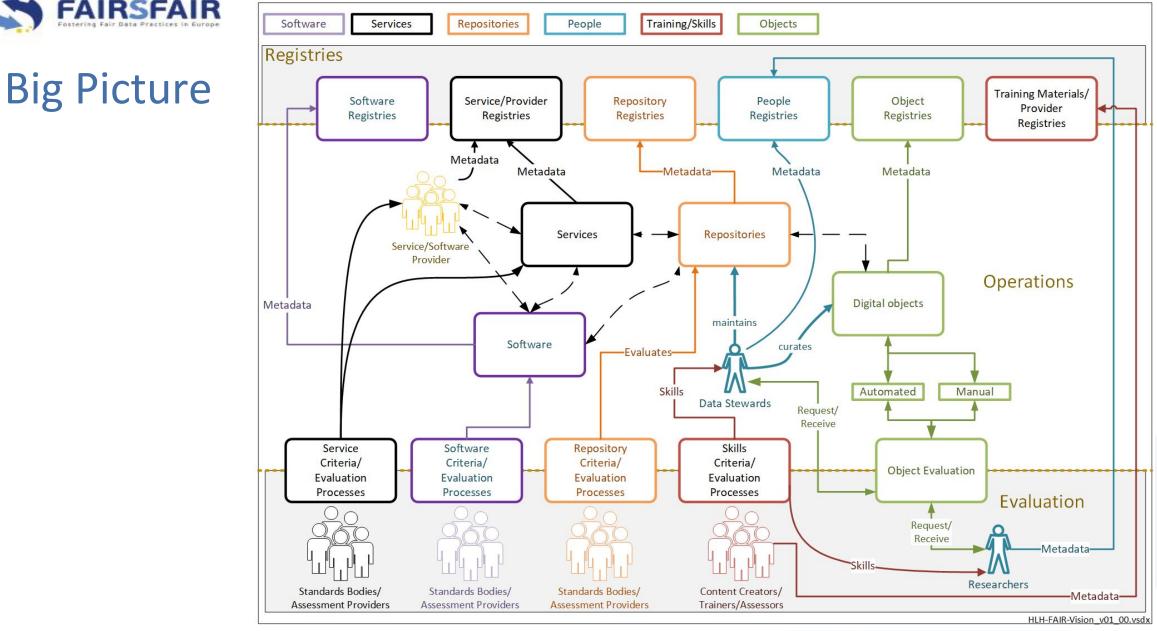




Wider Context







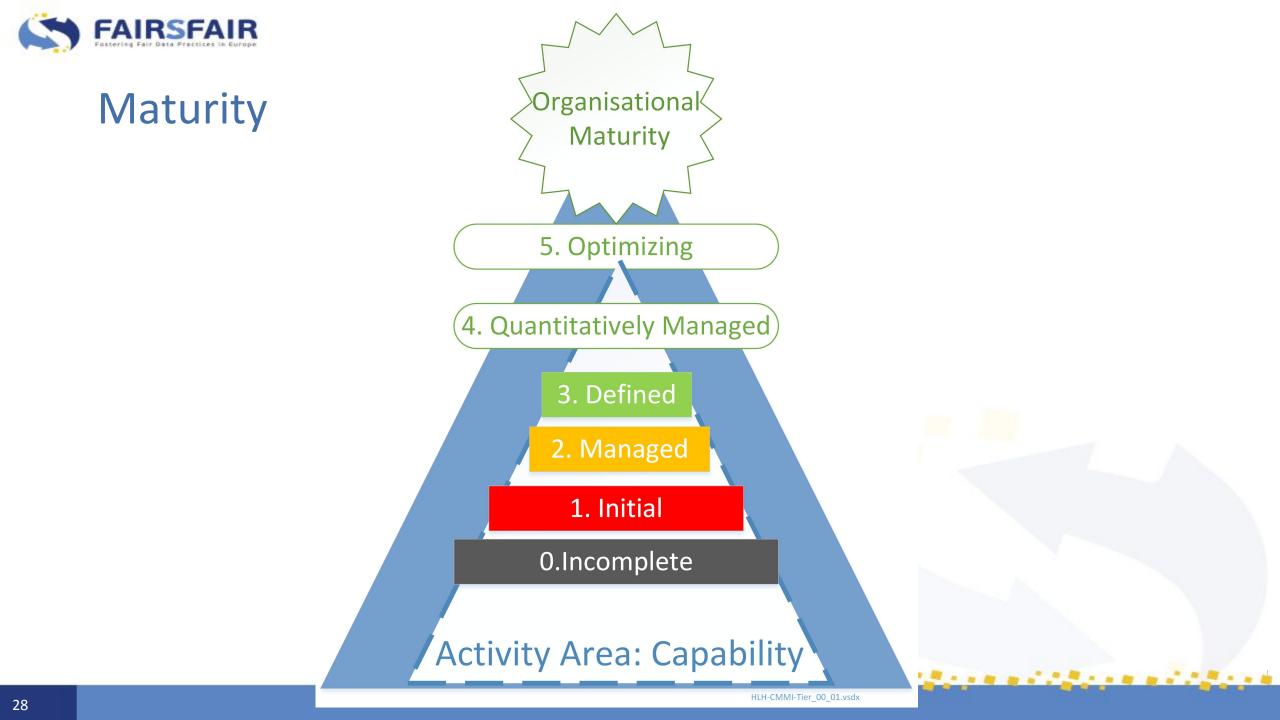
医马克勒氏试验 医马克勒氏试验 医马克勒氏试验

26



Objects Vs Enablers

- $\circ~$ F. A. I and R not news to repositories
- FAIR for Objects not fully defined
- FAIR enabling for repositories evolves in parallel
- Outcomes & Score: Pass/Fail vs Benchmark & Progress
- Maturity: Current State, Desired State, Roadmap
- Context is key





Maturity & FAIR

- Activity at 1, 2 or 3 (initial, managed, defined)
- Levels: 4 (quantitatively managed) and
 5 (optimising) support/depend on...
 - overall organisational maturity
- Aligned with repository practices that...
 enable FAIR Principles for objects.



Global FAIR vs FAIR in Context



ふうぎょうか かくうぎょうか かくうぎょう



CoreTrustSeal-FAIR mapping

F	R13 R15	 F1. (meta)data are assigned a globally unique and eternally persistent identifier. F2. data are described with rich metadata. F3. metadata specify the data identifier. R13. Data discovery and identification F4. (meta)data are registered or indexed in a searchable resource. R15. Technical infrastructure 	
A	R15 R16 R10	 A1 (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 the protocol is open, free, and universally implementable. R15. Technical infrastructure A1.2 the protocol allows for an authentication and authorization procedure, where necessary. R16. Security A2 metadata are accessible, even when the data are no longer available. R10. Preservation plan 	
	R15 R11	 I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles. R15. Technical infrastructure (Business Information? Object Model?) I3. (meta)data include qualified references to other (meta)data. R11. Data quality 	
R	R11 R2 R7 R15	 R1. meta(data) have a plurality of accurate and relevant attributes. R11. Data quality R1.1. (meta)data are released with a clear and accessible data usage license. R2. Licenses R1.2. (meta)data are associated with their provenance. R7. Data integrity and authenticity R1.3. (meta)data meet domain-relevant community standards. R15. Technical infrastructure 	

. .



Findable

- F1. (meta)data: globally unique and eternally **persistent identifier**.
 - R13. Data discovery and identification
- F2. data are described with **rich metadata**.
 - R13. Data discovery and identification
- F3. metadata specify the data identifier.
 - R13. Data discovery and identification
- F4. (meta)data are **registered** or **indexed** in a **searchable** resource.
 - R15. Technical infrastructure



Accessible

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - R15. Technical infrastructure
- A1.1 the **protocol** is **open**, free, and universally implementable.
 - R15. Technical infrastructure
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

- R16. Security
- A2 metadata are accessible, even when the data are no longer available.
 - R10. Preservation plan



Interoperable

- I1. (meta)data use a formal, accessible, shared, and broadly applicable
 language for knowledge representation.
 - R15. Technical infrastructure
- I2. (meta)data use **vocabularies that follow FAIR principles**.
 - R15. Technical infrastructure (Business Information? Object Model?)
- I3. (meta)data include qualified **references to other (meta)data**.
 - R11. Data quality



Reusable

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R11. Data quality
- R1.1. (meta)data are released with a clear and accessible **data usage license**.
 - R2. Licenses
- R1.2. (meta)data are associated with their **provenance**.
 - R7. Data integrity and authenticity
- R1.3. (meta)data meet **domain-relevant community standards**.
 - R15. Technical infrastructure



FAIR Dependencies and Action Points

Dependencies

- **R1. Mission/Scope**
 - Trustworthy Digital Repository Mission Enables FAIR

• R3. Continuity of access

Organisational Continuity reduces Rrisk to FAIR Data

\circ 10. Preservation plan

Preservation ensures FAIR over time



FAIR Dependencies and Action Points

Action

- **R8. Appraisal**
 - FAIRness Evaluated at Appraisal

• **R11. Data quality**

■ FAIRness Ensured through Data Quality curation measures

• **R14. Data reuse**

■ FAIRness assured for, and communicated at point of re-use



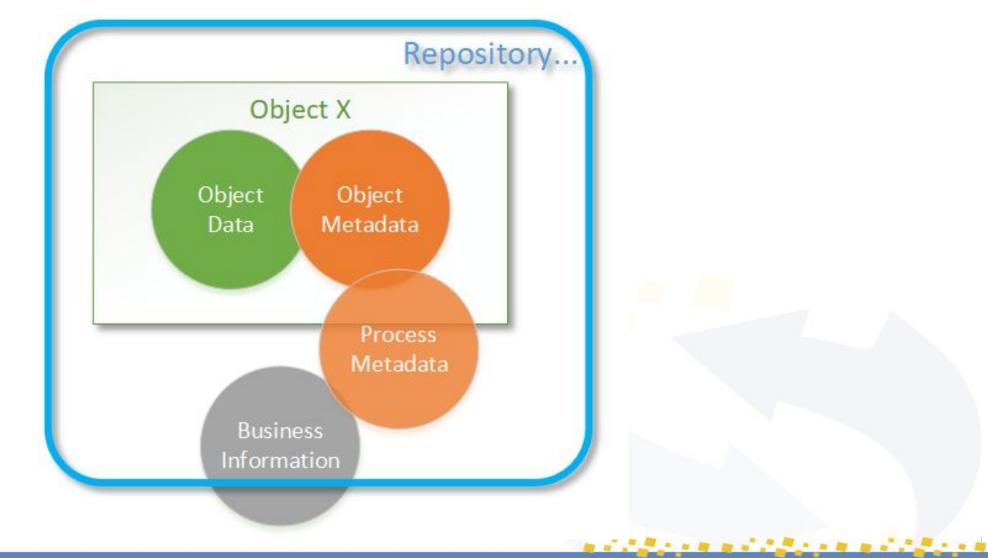
What have we (or FAIR) missed?

- R4. Confidentiality/Ethics
- **R5. Organizational infrastructure**
- **R6. Expert guidance**
- **R12. Workflows**





Repositories, Objects & Evidence





Enabling FAIR

- Score and Outcomes which:
 - Support Progress
 - Acknowledge Complex collections
- Interactions between Enabling and Object Evaluation
 - Evolving



FAIR Assessment of Data Objects in Trustworthy Digital Repository

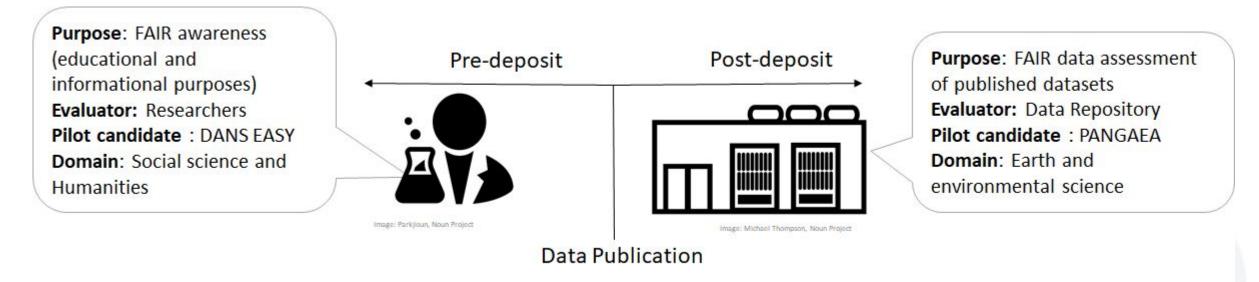
- Data objects are individual research datasets that are measured, collected, or created for scientific analysis.
- Trustworthy digital repository refers to a certified data repository, e.g., CoreTrustSeal.





Use case-based Iterative Assessment Approach

- Expected outcomes:
 - **Requirements** (including **metrics**) for assessing the FAIRness of data objects.
 - Tools to assess data objects in the context of two selected use-cases.



Pilot testing will cover 3 repositories from other domains

1 . M. . .



satifyd.dans.knaw.nl

Tools Implementation

* 🗛 😡 🌒 🗋

Data Archiving and Networked Services

SATIFYD

Self-Assessment Tool to Improve the FAIRness of Your Dataset

Welcome to SATIFYD: the DANS Self-Assessment Tool to Improve the FAIRness of Your Dataset. This tool will show you how FAIR (Findable, Accessible, Interoperable, Reusable) your dataset is and will provide you with tips to score (even) higher on FAIRness. Ideally, you use this tool prior to the deposit in EASY.

The 12 questions touch upon the FAIR data principles but do not strictly follow them. While answering the questions, the score per letter will be displayed underneath each letter. The more 'blue' the letters get, the more FAIR your dataset is. An overall score is provided at the end of the page.

Some questions are posed more than once (e.g. on metadata and data standards or usage licences), because the topics are relevant in more than one letter.

Want to know more? Please click here -

If you have any questions, please let us know by sending an e-mail 🖂



* Wilkinson, Mark D., Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, et al. 2019. 'Evaluating FAIR Maturity through a Scalable, Automated, Community-Governed Framework'. Scientific Data 6 (1): 1–12. <u>https://doi.org/10.1038/s41597-019-0184-5</u>

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.

Run evaluation

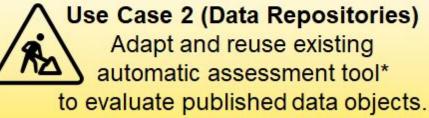
Globally Unique Identifier (GUID) to be Evaluated: This is the GUID for the metadata (1) of your data record

Hints: We currently recognize INCHI Keys, DOIs, Handles and URLs. Contact the developer to add new GUID types to this list - for example, we plan to support LSIDs soon, and others.

If you have a DOI for your data deposit, a DOI returns metadata to our system, so please use your DOI unless you have specifically created a better metadata record. For DOIs, enter the DOI without the 'doi:' prefix (e.g. just '10.1002/abd.123')

Select a collection to evaluate the resource







Break out working session

Presenter: Frans Huigen (DANS)



Break out working session

Blue, Yellow, Red, and Green groups, working on questions concerning designated CoreTrustSeal requirements

	Repository representatives	Requirements (every group does R0)	FsF partner representative
Blue	Apollo, DaSCH, ICOS	1 - 5 - 9 - 12	UKDA
Yellow	DASSH, Movebank, SOCIB	2 - 6 - 10 - 14	DANS + PANGAEA
Red	ESRF, IAGOS Tárki	3 - 8 - 11 - 15	CINES
Green	ESRF, DaSCH, DASS-BiH	4 - 7 - 13 - 16	DANS + DCC

- Is the requirement clear and understandable for you? If not, why?
- How will you meet the requirement?
 Will that be difficult?
- What actions have already been taken on this requirement? (do not worry if there are none!)
- What would your repository need to do more, to comply with this requirement?
- What stakeholders or experts would you need to comply with this requirement?



CoreTrustSeal, process and procedures

Presenters: Olivier Rouchon (CINES)



Two steps certification process

• Self assessment based on 16 Requirements

- Written responses
- URLs of documented public evidence
- Compliance level
- Peer review by two expert and independent reviewers under the responsibility of the CoreTrustSeal Standards and Certification Board

- Iterative feedback and revision process
- Successful applications are made publicly available
- Administrative fee of 1,000 euro
- Certification valid 3 years



Example

XIV. Data reuse

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.

Compliance Level

Response

Guidance:

Repositories must ensure that data can be understood and used effectively into the future despite changes in technology. This Requirement evaluates the measures taken to ensure that data are reusable.

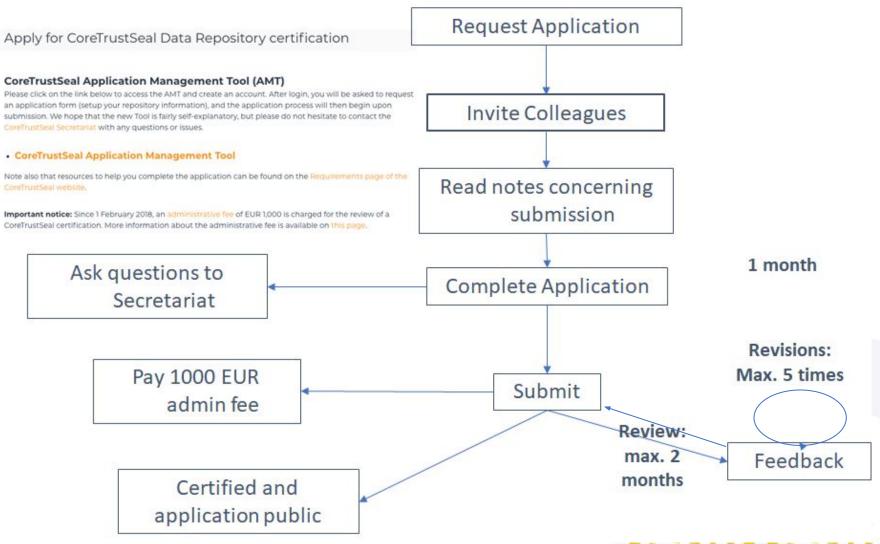
For this Requirement, responses should include evidence related to the following questions:

- Which metadata are required by the repository when the data are provided (e.g., Dublin Core or content-oriented metadata)?
- Are data provided in formats used by the Designated Community? Which formats?
- Are measures taken to account for the possible evolution of formats?
- Are plans related to future migrations in place?
- How does the repository ensure understandability of the data?

Reuse is dependent on the applicable licenses covered in R2 (Licenses).



Application process



. .



Resources / Help

- <u>Requirements 2020-2022</u>
- <u>Requirements with Extended</u>
 <u>Guidance 2.0</u>
- Successful applications
- <u>Webinar</u>

CORE	Home	About Y	Certification ~	Certified Repositories •	Apply 🗸	Contact	
Search markers						created	i on 🗸
CoreTrustSeal certification 2017-2019	Link to o success		oad oplication			0	00
PORTULAN CLARIN https://portulanclarin.net CoreTrustSeal certification 2017-2019						5	00
Departamento de Informática, Faculdade de Ciencias da Unive	rsidade de Lisboa,	Edifício C6, C	ampo Grande, 1749-	016 Lisboa, Portugal			
C ZBW Digital Long-Term Archive https://zbw.eu/ CoreTrustSeal certification 2017-2019						5	00

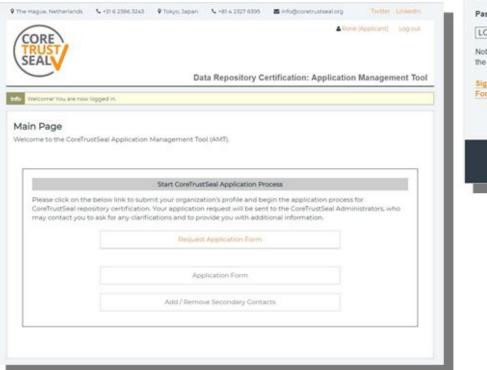
https://www.coretrustseal.org/why-certification/search-repositories/

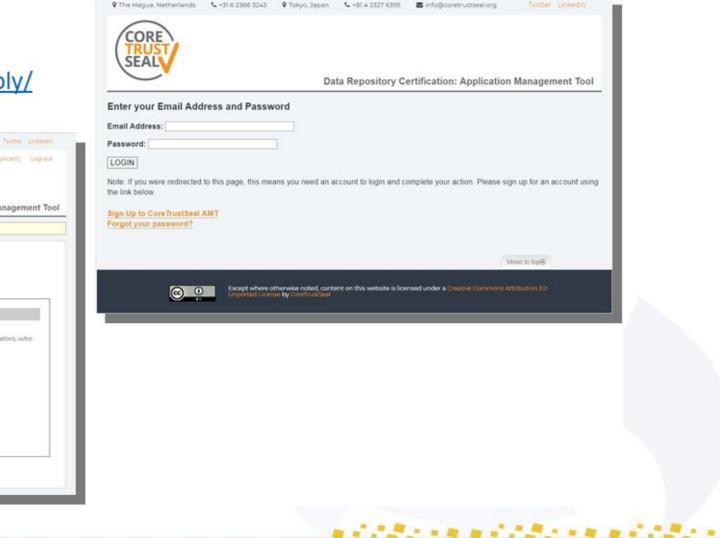
. . .



Application management tool

Apply via: https://www.coretrustseal.org/apply/







Application management tool

Data	Repository Certification: Application Management Tool	
nization Profile nion Name Required		Data Repository Certification: Application Manageme
Required		Main Page
uired		
Required		A Repository
•		Application Review Progress
Required		Oraft Submitted Under Review Feedback Available Approved
itute Required		
Idata Entry		Organization Profile
ess https://www.re3data.org/ to find the DOI of your entry, or	control on activities were reportioned if deale and exist in	
tative Name Required	r suchini an entry su your republicity it is does not exist.	CoreTrustSeal Application Form The online form for your repository to certify against the Core Trustworthy Data Repositories Requirements can be
tarove Name Required		accessed below. The form does not need to be filled out on a single occasion, and you may save your draft, request simple feedback (assistance) from the CoreTrustSeal Admin, or submit the completed form for review.
quired		Application Status: Draft
		Piperse note: The form is editable ONLV when it has status 'Draft' or 'Feedback Available'.
•		Hease note: The form is equable UNLY when it has status that, or Heedback Available,
cription of Repository		Application Form
		Contributors Completing the CoreTrustSeal application will very likely require discussion with other data management and
2		technical experts in your organization. To share the application with your colleagues, please add them as contributo using the link below. It is recommended to include at least one person as an additional contact for your organization
Cancel		Rorie (Applicant)/rorie.edmunds+uset@icsu-wdsorg)
		House (Hother and Londer on Linuary Lase (Bir 20, March 6)
	Linue to top@	Add / Bernove Secondary Contact
Except where otherwise noted, content of	on this website is licensed under a Creative Commons Attribution 1.0	
CO Content of Unported Laboratory Construction		

かんかがく かかん かがく かかか かんかがく かい



Recap application process

- Self-assessment based on requirements
- Online tool available
- Extended guidance (document and webinar) available
- Evidence provided:
 - Bullet lists of topics/questions for discussion and inclusion are neither exhaustive nor prescriptive.
 Highly preferred that responses are written in full prose.
 - URLs to evidence strongly encouraged
 - Maturity ratings strongly encouraged
 - Responses in English
 - Evidence documents in other languages need a short summary in English
- Review of the self assessment by two reviewers under the responsibility of the CoreTrustSeal Board

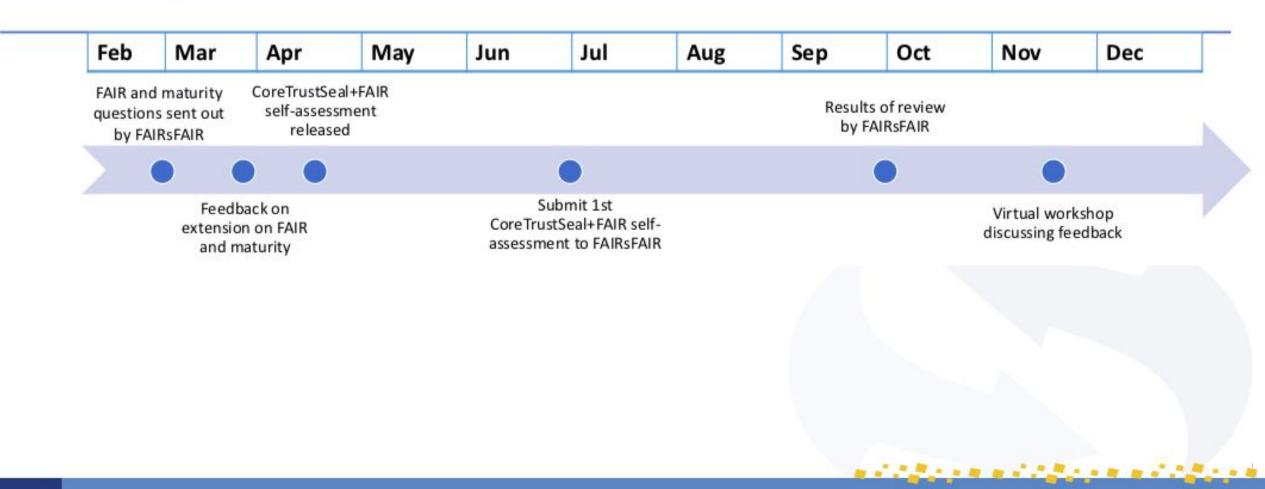


Needs and expectations certification support programme

Presenter: Ilona von Stein (DANS)



Roadmap – 2020





How might we best support you?

• Support Routes

• Support Approaches

• Monitoring progress

Get your laptop or mobile and go to: www.menti.com



Mentimeter Results

How would you rank the usefulness of these possible Support Routes?



Mentimeter

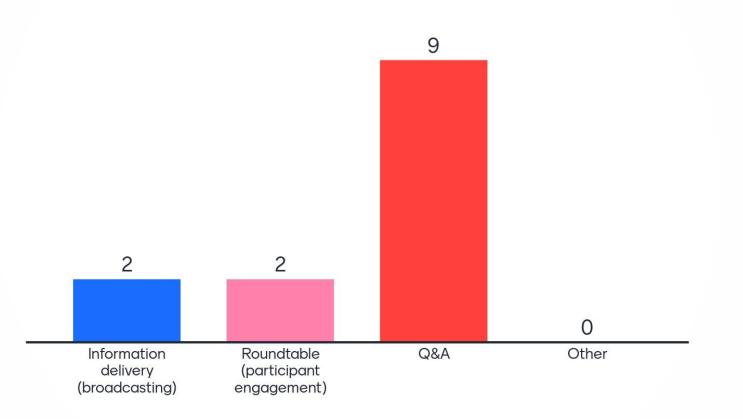
▲ 12

......



Mentimeter Results

Which Support Approach do you like best?



å 13

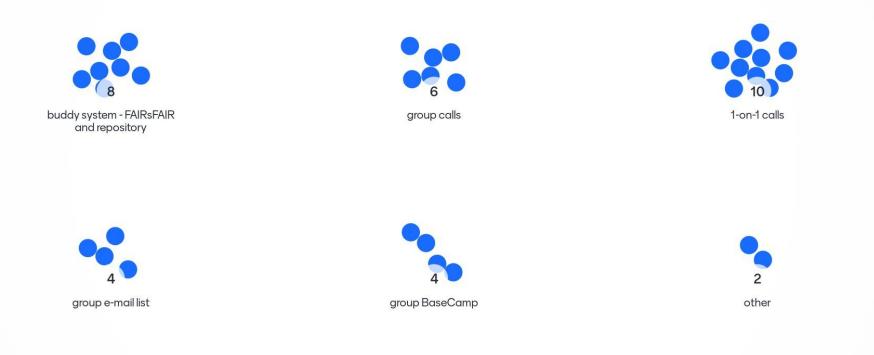
......

Mentimeter



Mentimeter Results

What is a good way to monitor progress? Please indicate your top 3.



.....

Mentimeter



Getting started!

Presenters: Patricia Herterich (DCC) and Anusuriya Devaraju (PANGAEA)

This exercise is available at <u>http://doi.org/10.5281/zenodo.3741693</u>



Stakeholder mind map for your institution

- GOAL: identify all the people in your institution that you will need to involve in your certification application and FAIR enabling process
- one person can hold several roles
- depending on your institutional structure, not all example roles will be applicable



Example Repository Roles

During the support process the following roles may be useful in preparing your application:

- Senior champion a Director or Senior manager
- Curation practitioner with experience of data management workflows
- Technical contact with knowledge of applications, systems and services
- Application lead co-ordinates the application
- Executive sponsor will sign off the assessment for submission



Roadmap

- GOAL: identify key activities and tasks you need to carry out to understand CoreTrustSeal+FAIR in your institutions, put your self-assessment together, highlight dependencies and identify risks
- Inspiration via the RDA CoreTrustSeal adoption story across domains and regions:

https://www.rd-alliance.org/rda-coretrustseal-adoption-story-across-domains-a nd-regions

• Risk matrix:

	Negligible	Marginal	Critical	Catastrophic
Certain	High	High	Extreme	Extreme
Likely	Moderate	High	High	Extreme
Possible	Low	Moderate	High	Extreme
Unlikely	Low	Low	Moderate	Extreme
Rare	Low	Low	Moderate	High



Break out groups

	Repository representatives	FsF partner representative
Blue	Apollo, ICOS, DaSCH	UKDA
Yellow	DASSH, Movebank, SOCIB	DANS + PANGAEA
Red	IAGOS, Tárki	CINES
Green	ESRF, DASS-BiH	DANS + DCC

Plenary feedback in 1 minute madness: Let us know your key risk / most urgent action / most difficult stakeholder to get on board / other!



Wrap-up in postcards

- Think about two takeaways that you want to bring home from this workshop
- Write them on the postcards provided
- Address the postcard to yourself, make sure the postcard will reach your desk
- We will post it for you in a couple of weeks!

Presenter: Ilona von Stein (DANS)



Image from https://www.wikihow.com/Write-a-Postcard



Did this day meet your expectations in terms of learning?

Mentimeter





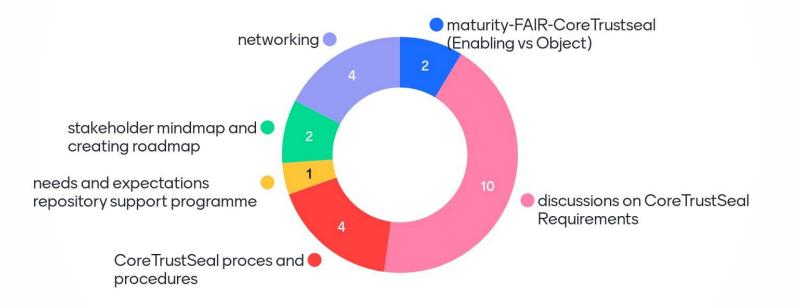
0

No, it did not meet my expectations

.....



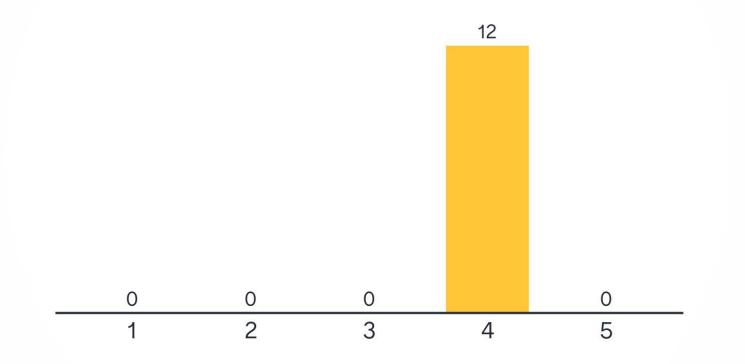
Which parts of the day were most valuable for you? max. 2 answers



Mentimeter



On a scale from one to five what would be your overall mark for the quality of the workshop? (1 equals very poor and 5 equals excellent)



Mentimeter

▲ 12



After today, which keywords (max. 3) are important for you in the light of FAIR-enabling repository certification?



衬 Mentimeter



If at some time we will have a 2nd F2F workshop, which topics would you like to have covered?

8 7 5 5 4 FAIR data Assessing the Developing Addressing Changes to FAIRness of the related to evidence specific CoreTrustSeal repository data objects statements requirements requirements

Mentimeter

▲ 13



Thank you very much!