



FAIRsFAIR

Fostering Fair Data Practices in Europe

Certification Support workshop

November 19, 2020, 12:30-13:30 CET
online

Ten FAIRsFAIR supported repositories & FAIRsFAIR (WP4) support team
Linus Čepinskas, Ilona von Stein, Frans Huigen (DANS), Hervé L'Hours (UKDA) and
Olivier Rouchon (CINES)



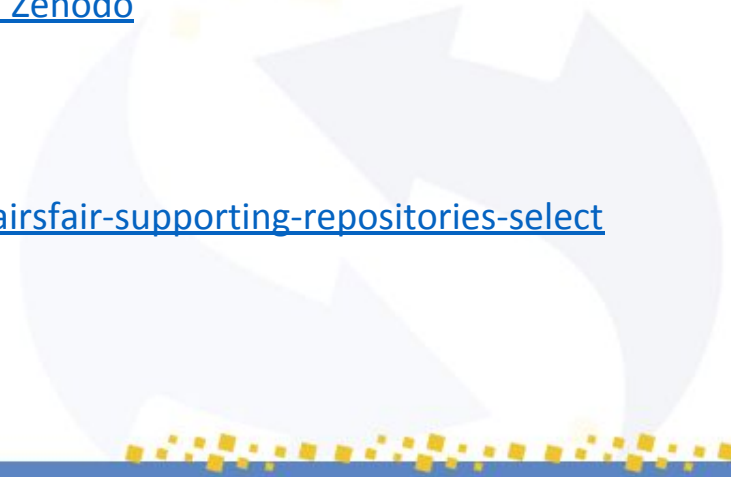
FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558



@FAIRsFAIR_EU

FAIRsFAIR Project Outputs (updated)

- Earlier meetings in the support programme:
 - workshop presentation 6 February ([Zenodo](#)) and workshop exercise material ([Zenodo](#)).
 - general support call for repositories 11 June ([Zenodo](#))
- FAIRsFAIR Deliverables and Milestones:
 - D4.1: Draft recommendations on FAIR dataset requirements ([Zenodo](#))
 - M4.2: Draft Maturity Model Based on Extensions and-or Additions to CoreTrustSeal Requirements ([Zenodo](#))
 - D3.3: Description of FAIRsFAIR's Transition Support Programme for Repositories ([Zenodo](#))
 - [All project outputs](#) and The FAIRsFAIR Community [on Zenodo](#)
- Blogpost
 - <https://www.fairsfair.eu/articles-publications/how-fairsfair-supporting-repositories-select-ed-coretrustseal-certification>



Meeting outline

- Analysis of your common issues towards repository certification by Linas Čepinskas and Frans Huigen (DANS)
- Highlighted topic: Continuity of Access and Long-term Preservation by Hervé L'Hours (UKDA)
- Highlighted topic: Workflows by Olivier Rouchon (CINES)
- Q&A by Ilona von Stein (DANS)
- Next steps by Ilona von Stein



Analysis of your common issues towards repository certification

- Prior to the survey, we took stock of 10 self-assessment reports.
- We identified at least 29 common issues with regard to the CoreTrustSeal requirements.



Initial analysis: results

R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
Context	Mission	Licences	Continuity plan	Compliance	Funding	Guidance & feedback	Integrity & authenticity	Metadata criteria	Processes & procedures	Long-term preservation	Expertise	Workflows	Data discovery	Data reuse	Operating systems	Infrastructures
Specific entity	Explicit mentioning of preservation	Compliance monitoring of applicants' self-assessed progress	Preservation plans & procedures	Processes and procedures	Staff, funding, governance	Interaction with experts and feedback integration	Data integrity & authenticity	Relevance of data to designated community	Processes in place	Documented approach	Procedures of data quality control	Archiving workflows & mngt	Search interfaces & discovery	Creation, curation and preservation of (meta) data	Relevant procedures and workflows	Risk analysis & controls
Designated community		Data usage licences		Types of personal information		Partnerships and groups involved		PID usage							Technical aspects of business continuity	
Choice between levels of curation								Data collection policy								
Relationships between repositories																

Survey

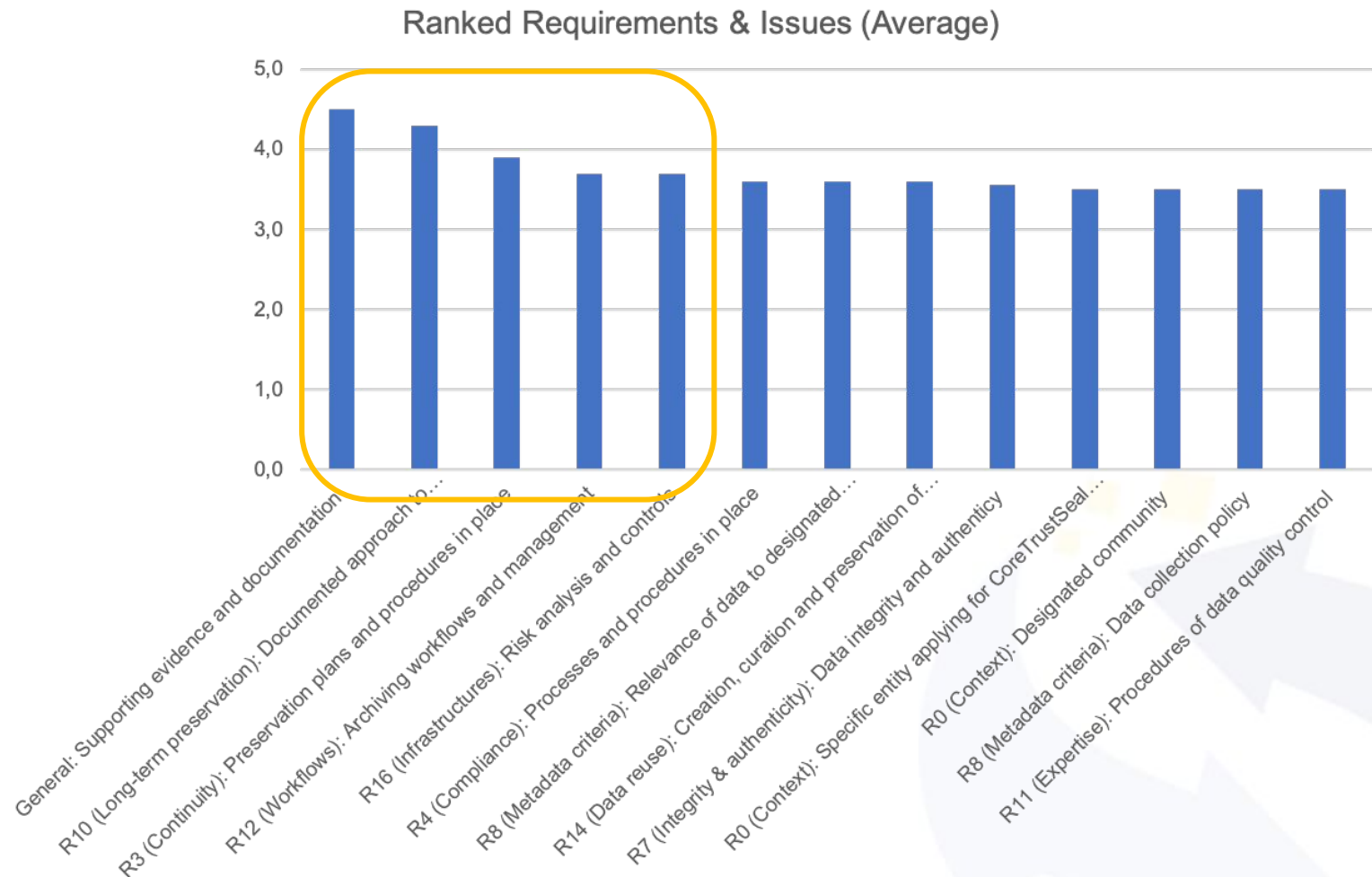
- In recent survey, we asked repositories to rank these 29 common issues in order of their importance.
- The objective was to address these issues separately and provide practical guidance on how to meet the CoreTrustSeal requirements.



Survey results: overall ranking

Requirements	Average
General: Supporting evidence and documentation	4,5
R10 (Long-term preservation): Documented approach to long-term preservation	4,3
R3 (Continuity): Preservation plans and procedures in place	3,9
R12 (Workflows): Archiving workflows and management	3,7
R16 (Infrastructures): Risk analysis and controls	3,7
R4 (Compliance): Processes and procedures in place	3,6
R8 (Metadata criteria): Relevance of data to designated community	3,6
R14 (Data reuse): Creation, curation and preservation of (meta)data	3,6
R7 (Integrity & authenticity): Data integrity and authenticy	3,6
R0 (Context): Specific entity applying for CoreTrustSeal certification	3,5
R0 (Context): Designated community	3,5
R8 (Metadata criteria): Data collection policy	3,5
R11 (Expertise): Procedures of data quality control	3,5
R15 (Operating systems): Relevant procedures and workflows	3,4
General: Compliance level of your self-assessed progress	3,4
R9 (Process & procedures): Processes and procedures in place	3,4
R8 (Metadata criteria): PID usage	3,3
R1 (Mission): Explicit mentioning of preservation	3,2
R5 (Funding): Number of staff, funding scheme, organisational structure and governance	3,2
R0 (Context): Scope of curation	3,1
R0 (Context): Choice between levels of curation	3,1
R6 (Guidance & feedback): Interaction with experts and feedback integration	3,1
R15 (Operating systems): Technical aspects of business continuity	3,1
R2 (Licences): Compliance monitoring of applicants' self-assessed progress	2,8
R13 (Data discovery): Search interfaces and discovery	2,8
R2 (Licences): Data licences	2,7
R0 (Context): Relationships between repositories	2,6
R4 (Compliance): Types of personal information and size of the subset of data	2,5
R6 (Guidance & feedback): Partnerships and groups involved	2,5

Survey results: top issue areas



R03. Continuity of Access

R03 Continuity is about the *organisation's* ability to:

- Function in all normal circumstances
- Respond to abnormal circumstances
- Consider the Worst Case Scenario(s)

Less about the tech (R15) though good to provide a statement that aligns well with R15. Technology.

Provides assurance that the organisation can persist and survive to deliver other services with minimum risk to objects.

R03. Continuity of Access



R03. Continuity of Access

“ If there is no formal, written agreement between the repository and [a successor] organization, then the Compliance Level can be at a maximum of 3 only”

NB: 3 is a ‘pass’

- Some organisations have succession built into their model
- Others will never achieve a formal plan until forced
- Demonstrate that you understand the issues and are ready to face the challenge if it happens.

R10. Preservation (what it's not!)

R03 Continuity asks:

“The level of responsibility undertaken for data holdings, including any guaranteed preservation periods”

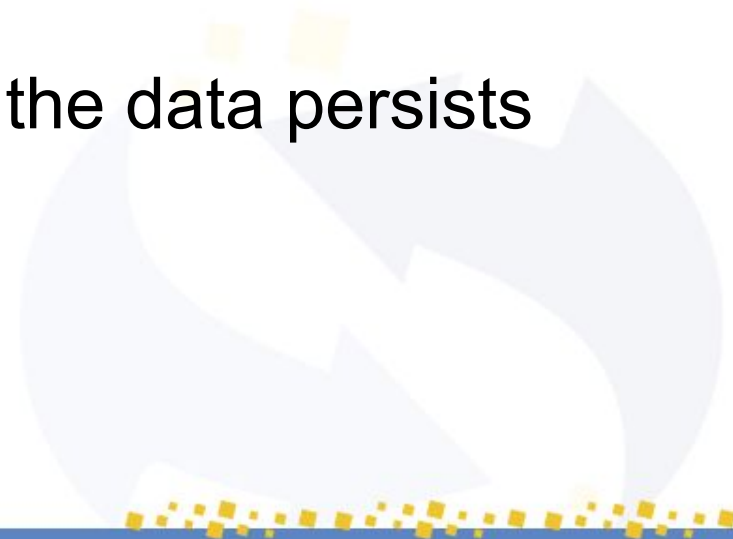
- The focus is less on “Preservation”
- More on “Level of responsibility” as *context for organisational resilience* (sole custodian, shared responsibility)

The data-related actions that this responsibility implies are covered under R10. Preservation

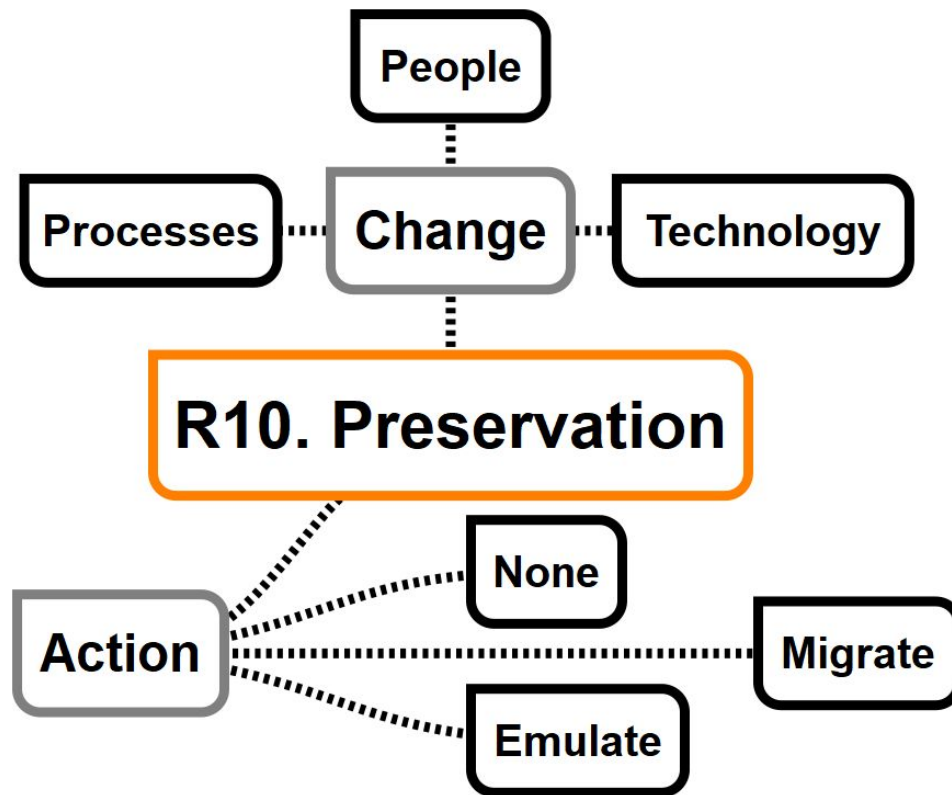
R10 Preservation (What it's Not)

- R03 is not about Preservation
 - But it does provide assurance that the organisation can persist and survive to deliver other services with minimum risk to objects.
 - Including Preservation

The organisation persists... so the data persists



R10 Preservation (What it is)



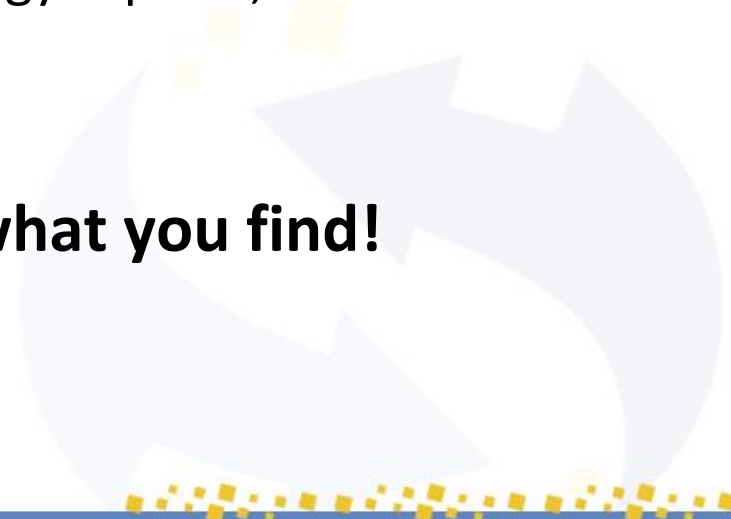
R10 Preservation: Understand the need for change

Change in response to the needs of the Designated Community

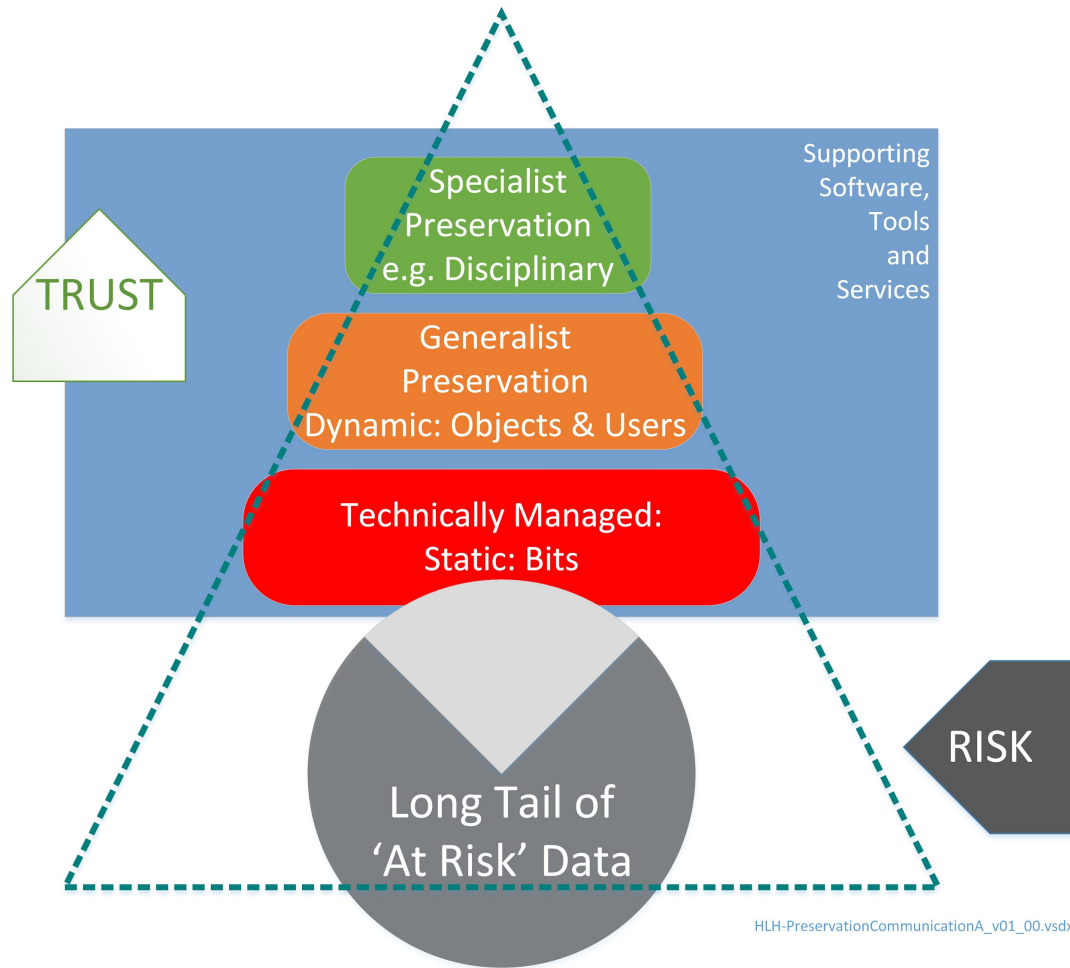
- **People:** Disciplinary communities have more specific needs
- **Processes:** Research focus and practice changes over time
- **Technology:** general technical environment specific
designated community technical environment

Tech watch can cover some of the technology aspects, for the rest...

Ask the community, act on what you find!



R10 Preservation: Foundations & Layers

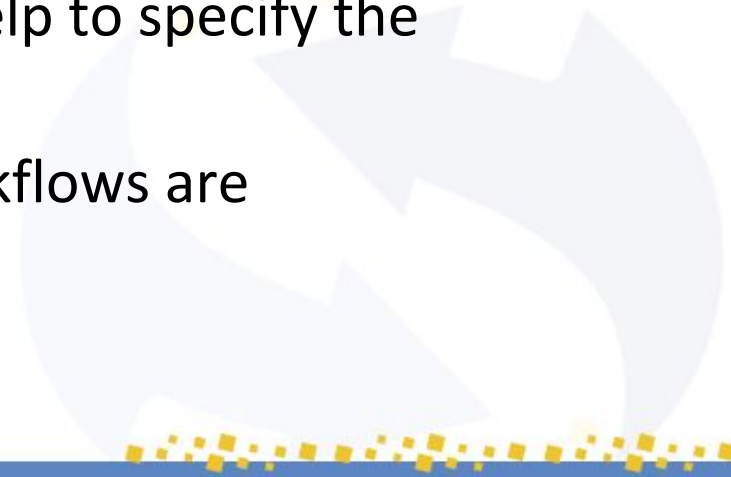


R12. Workflows

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

To ensure the consistency of practices across datasets and services and to avoid ad hoc actions, workflows should be defined according to the repository's activities and clearly documented. Provisions for managed change should be in place. The OAIS reference model can help to specify the workflow functions of a repository.

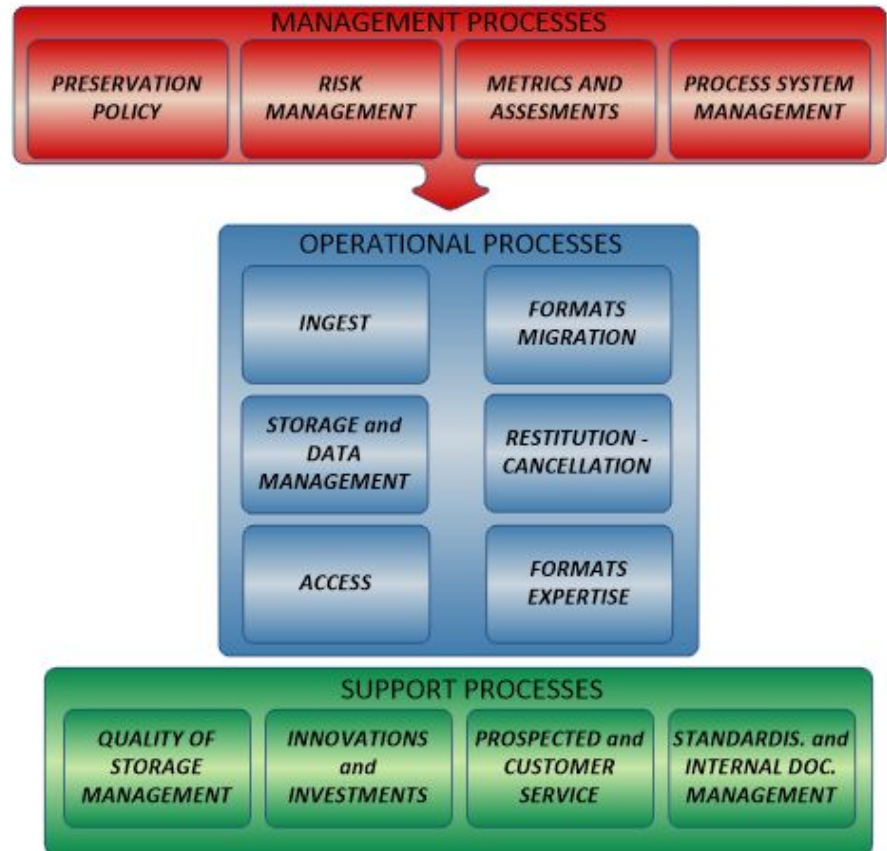
This Requirement confirms that all workflows are documented



R12. Workflows

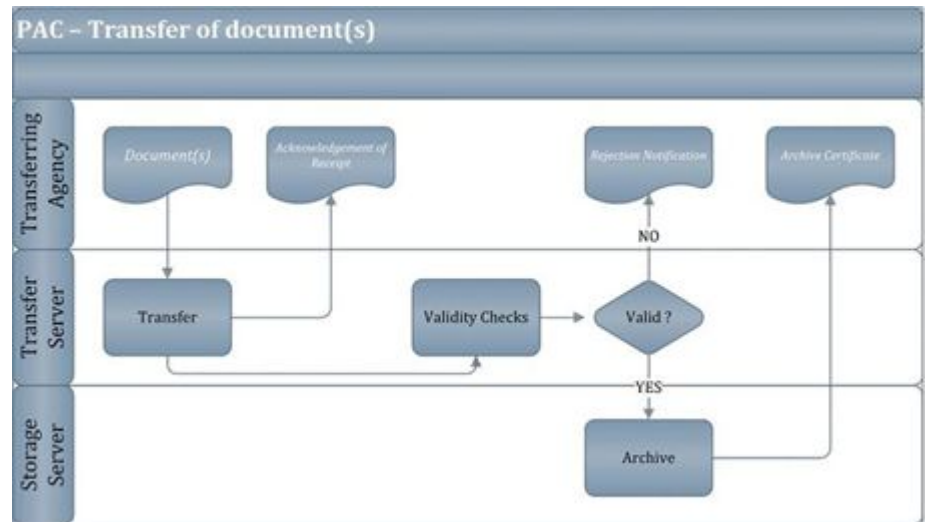
For this Requirement, responses should include evidence related to the following:

- Workflows/business process descriptions.
- Change management of workflows.



R12. Workflows

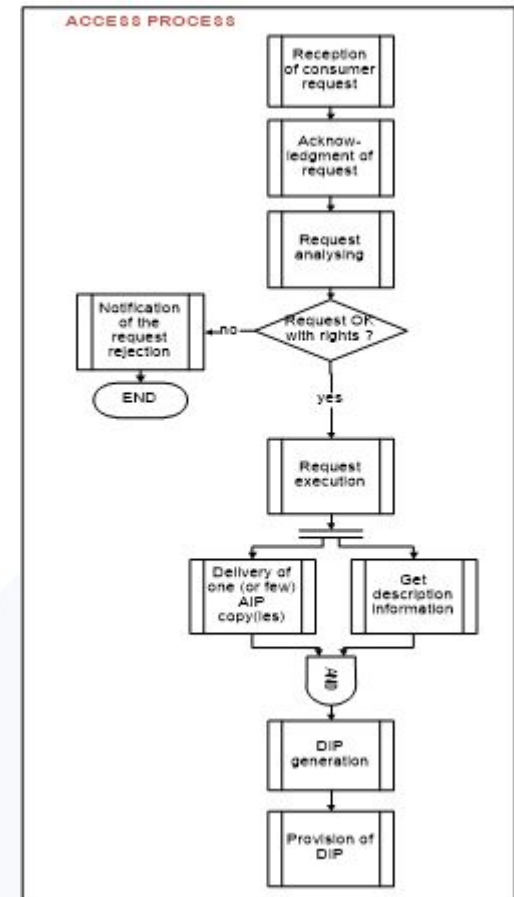
- Qualitative and quantitative checking of outputs
- Clear communication to depositors and users about handling of data.



- The types of data managed and any impact on workflow.
- Decision handling within the workflows (e.g., archival data transformation).

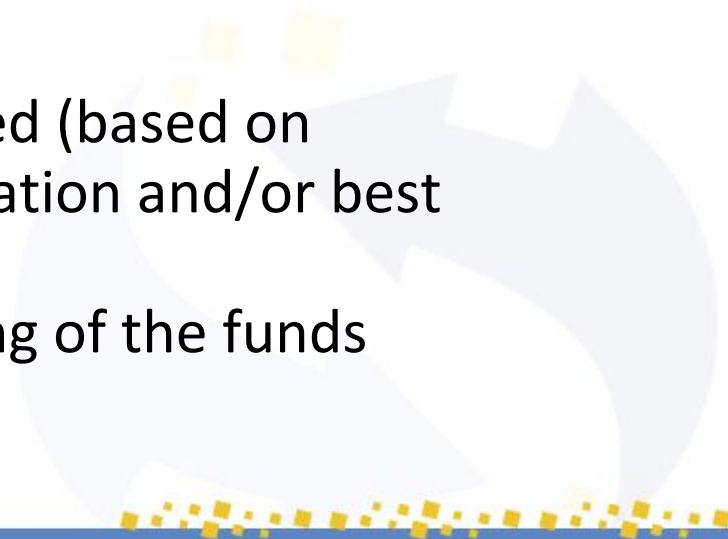
R12. Workflows

- Levels of security and impact on workflows (guarding privacy of subjects, etc.).
- Example:
<https://www.cines.fr/en/long-term-preservation/archiving-solutions/pac/archive-workflow/>



Next steps for the selected repositories

- Feb 2021: final versions of the self-assessments to be submitted to CoreTrustSeal
 - depending on your progress and readiness
- FAIRsFAIR support team at their disposal for further help
- Upon request, further topical, in-depth discussion sessions
- Autumn/winter 2021:
 - other half of the funding provided (based on successful CoreTrustSeal certification and/or best efforts)
 - lightweight reporting on spending of the funds



Nexts steps for FAIRsFAIR

- keep supporting you on your journey towards certification
- keep providing support material, also for a wider repository audience
 - we actively use your experiences for this
- The RDA FAIR data maturity model WG released [final core assessment object metrics](#).
 - In 2021 we will use your FAIR-questions as input to our recommendations on how to integrate FAIR data properties into certification repository requirements

FAIRsFAIR repository support programme upcoming events



<https://conference.codata.org/FAIRconvergence2020/>



<https://www.agu.org/Fall-Meeting>