D4.5 Report on completed FAIR data standard adoption and certifications of data repositories in the region

Author(s)	Tuomas J. Alaterä, Mari Kleemola, Henri Ala-Lahti (TUNI/FSD), Birger Jerlehag (GU/SND)
Status	Final
Version	1.0
Date	2022-08-31

Document identifier:	
Deliverable lead	Tampere University / Finnish Social Science Data Archive
Related work package	WP4
Author(s)	Tuomas J. Alaterä, Mari Kleemola, Henri Ala-Lahti (TUNI/FSD), Birger Jerlehag (GU/SND)
Contributor(s)	Mari Elisa Kuusniemi (HU), Trond Kvamme (Sikt), WP4 team
Due date	2022-08-31
Actual submission date	2022-08-31
Reviewed by	Siiri Fuchs (CSC), Anne Sofie Fink (DEIC)
Approved by	
Dissemination level	PU
Website	www.eosc-nordic.eu
Call	H2020-INFRAEOSC-2018-3
Project Number	857652
Start date of Project	2019-09-01
Duration	36 Months
License	Creative Commons CC-BY 4.0
Keywords	FAIR, CoreTrustSeal, data repository, certification, data standards, maturity evaluations, assessments, open data, project outputs

www.eosc-nordic.eu

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Abstract

The EOSC-Nordic project has pledged to implement FAIR in the Nordic and Baltic region. The tasks of EOSC-Nordic Work Package 4 were in particular "support for the introduction of FAIR information standards" and "support for the introduction of the FAIR certification system". This report summarises efforts of the WP4 task team to support selected communities to adopt a FAIR data standard. It also summarises the support process, visits the feedback gathered from the communities and discusses lessons learnt from communities that agreed to undertake a FAIRification effort or CoreTrustSeal certification self-assessment of their respective data repository.

The deliverable relies on other WP4 deliverables, which in detail describe how the supported repositories were selected and in which way FAIR maturity has been evaluated in detail. It first reports on the awareness raising and support activities done by the task team 2020-2022 and finishes with feedback from the community and lessons learnt. In the discussion section, the deliverable considers the sufficient levels of FAIR maturity, the applicability of certification, and needs for support and networks in the future.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Table of contents

www.eosc-nordic.eu

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Introduction

This report summarises efforts of the EOSC-Nordic Work Package 4 Task Team to support selected communities to adopt a FAIR data standard. It also summarises the process and lessons learnt from communities that agreed to undertake a CoreTrustSeal certification self-assessment of their respective data repository.

The task of the EOSC-Nordic project was to implement FAIR in the Nordic and Baltic region. The tasks of WP4 were in particular "support for the introduction of FAIR information standards" and "support for the introduction of the FAIR certification system". The implementation plan stated that this will be achieved by disseminating the benefits of FAIR and providing assessment-based recommendations on how to FAIRify data repositories and assist them in self-assessments for CoreTrustSeal certification. The task team's efforts have been focused on these recommendations and assistance directed to the selected research data repositories in the region. In addition, online events have been open to a larger audience across different disciplines and regions.

The availability of research data and an understanding of its nature are the basis on which demands for better reproducibility of research results, transparency of information and methods, and trust in the scientific process can be based on a data-oriented and computational research approach. Communicating this to both humans and machines is essential. This can be greatly facilitated by improving the repository's ability to produce and preserve FAIR (meta)data.

Broader outreach activities were described in task 4.1[1] (*D4.1 State of FAIR practices in the Nordics*) and the implementation plan in task 4.2[2] (*D4.2 Implementation Plan for the adoption of data standards and certification*).

Alongside the EOSC-Nordic project there were both the FAIRsFAIR[3] and SSHOC[4] EU-projects, which provided support for certification or FAIRification or both. In addition, European Research Infrastructures, like CESSDA, have their own internal trusted digital repository requirements and provide support for repositories. Therefore, some simultaneous activities were discipline-specific and some were regional. EOSC-Nordic WP4 has actively cooperated with other pan-European initiatives in the region. A clear view of the landscape of a trusted digital archive cannot be obtained from just one project. The need for future cooperation has been identified. However, it can be with confidence stated that improvements have taken place based on the support measures and more FAIR trusted digital repository services have emerged.

In EOSC-Nordic, a very practical and results-oriented approach was chosen for FAIR and certification support. This report focuses exclusively on the support process and lessons learnt, as well as a discussion of next steps.

[1] Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, Adil Hasan, Josefine Nordling, & Pauli Assinen. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). Zenodo. https://doi.org/10.5281/zenodo.4045402

[2] Henrik Jakobsen, Mari Kleemola, Andreas Jaunsen, Birger Jerlehag & Tuomas J. Alaterä. (2020). D4.2 Implementation Plan for the adoption of data standards and certification (internal project document).

[3] https://www.fairsfair.eu

[4] https://www.sshopencloud.eu

www.eosc-nordic.eu

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

FAIR Data Standards

The starting point of WP4 FAIR support was that metadata provisioning is the most efficient way to FAIRify data. This is noticeable in the FAIR principles too, as they place more emphasis on metadata than data. However, the FAIR Principles have not been written as a standard and they were not published with specific implementation guidelines. Actually, they are envisioned to have an extensive research community adoption element, which defines what and how is FAIR for that specific community. As a regional implementation project EOSC-Nordic included participants from several disciplines. Its aim was to use the Nordic and Baltic countries as a test-bed for FAIR adoption and certification. This highlighted the need for generic metadata that serves not only subject specialists but anyone doing cross-discipline research, non-experts, data harvesters or citizen scientists.

Therefore, to adopt FAIR data standards, the process of preparing metadata for the consumption of both machines and humans was seen as a crucial measure for success. To do this, an automated and machine-actionable FAIR evaluation process was designed.[5] These evaluation scores of almost a hundred Nordic and Baltic repositories are based on the operationalisation of the individual FAIR principles that measure findability, accessibility, interoperability and reusability at the digital object level. At the very minimum they provide a starting point against which to measure change.[6] The FAIR principles then form a standard and a machine-actionable metadata evaluation tool to test the adoption of the standard. In future, if automated FAIR evaluations become a standard practice, certain FAIR score levels could be seen as evidence of adopting a FAIR data standard.

The WP4 team produced FAIRicifation guidelines (Appendix 1) which were constantly referred to in webinars and repository specific support meetings. Following these recommendations was also essential in concluding the milestone *MS33: Completion of first FAIR data standards adoption in a repository*. The guidelines were intended to facilitate efforts of making a repository FAIRer by adopting FAIR data standards in practice.

While there has not been a single FAIRification path backed up by a certification standard or initiative, the support given focussed on repositories examining their FAIR scores and looking for improvements in one to one support calls. It was obvious from the start that different repositories have different capabilities and possibilities to update their machine readable metadata records. Also the variation from discipline to discipline was notable. Therefore, support suggestions were aimed at changes that are likely to increase the FAIRness of metadata and subsequently produce better results in FAIR evaluations. Equally important were observations done by the repositories themselves. During the early stages of support, many of these observations were elevated to the level of F-UJI Automated FAIR Data Assessment Tool[7] developers. Some proposed changes were taken into consideration and the team was able to convey information from the developers to the repositories.

Several FAIR assessment tools have emerged since EOSC Nordic started. The same metadata records, assessed by different tools, often get different FAIR scores since the tools interpret the FAIR principles differently. The EOSC Association Task Force on FAIR metrics and data quality[8] is trying to address these

[5] See Nordling, Josefine, Mihai, Hannah, Meerman, Bert, Alaterä, Tuomas, Kleemola, Mari, & Livenson, Ilja. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. Zenodo. https://doi.org/10.5281/zenodo.6880904

[6] See Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, Adil Hasan, Josefine Nordling, & Pauli Assinen. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). Zenodo. https://doi.org/10.5281/zenodo.4045402

[7] https://www.f-uji.net/

[8] https://www.eosc.eu/advisory-groups/fair-metrics-and-data-quality

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

ambiguities so we expect to see positive developments in this area in the near future. Feedback from the communities has helped to create better evaluation processes and a basis for what could become a FAIR data standard.

The team was able to deliver strong FAIRification support during the first half of the project. From late 2021 onwards due to changes in project staff, less FAIRfication support was available. Due to this, the team focused primarily in carrying out automated FAIR assessments and reporting about them, self-organised webinars and invited talks on FAIR recommendations in other relevant events. As there isn't any authoritative FAIR requirements implementation list, FAIR support was less structured and more adapted to either individual repository's needs or provided on a more generic level. FAIRification STEPs webinars reported in D4.3[9] were well received and provided lots of positive feedback. Same applies to Metadata for Machines (M4M) workshops where WP4 partnered with the GO FAIR Foundation in its ongoing series of open, community workshops.

In addition to the recommendations in the Appendix 1, certain generic recommendations were tested and deemed successful. These were

- promoting the use of schema.org and other generic, widely used metadata standards embedded or linked to in a machine actionable linked data format (eg. JSON-LD, RDF)
- signposting generic metadata as typed links in HTTP headers as recommended by signposting.org[10]
- expressing data licences in machine actionable format, for example when using standard Creative Commons licences.

Above all adopting any FAIR data standards begins with exposing your repository metadata to machines. This means there has to be a landing page which is findable on the web. Therefore, a consistent use of persistent identifiers is one of the very first steps to take.

(FAIR) Certification of Trustworthy Digital Repositories

The EOSC Nordic project has strived to increase the maturity of the Nordic and Baltic research data repositories in order to ensure that valuable data holdings become FAIR and, when needed, remain FAIR in the long run. *The Turning FAIR into Reality Report* (2018)[11] proposed a model "denoting the minimal components needed to offer an ecosystem that enables the creation, curation, and reuse of FAIR Digital Objects in an effective and sustainable way". Data repositories are a key element in this FAIR ecosystem since they provide the organisational context for FAIR Digital Objects. Of critical importance are trustworthy digital repositories (TDRs) that provide long term digital preservation for a designated community of users. A TDR ensures technical continuity through file format migration or emulation, and maintains data and metadata so that it remains understandable to their designated community of users in the long term[12].

[9] Nordling, Josefine, Mihai, Hannah, Meerman, Bert, Alaterä, Tuomas, Kleemola, Mari, & Livenson, Ilja. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. https://doi.org/10.5281/zenodo.6880904

[10] FAIR Signposting Profile: https://signposting.org/FAIR/ [19.8.2022]

[11] European Commission (2018). Turning FAIR into reality - Final report and action plan from the European Commission expert group on FAIR data https://doi.org/10.2777/1524

[12] L'Hours, Hervé, Kleemola, Mari, von Stein, Ilona, van Horik, René, Herterich, Patricia, Davidson, Joy, Rouchon, Olivier, Mokrane, Mustapha, & Huber, Robert. (2022). FAIR + Time: Preservation for a Designated Community (02.00). Zenodo. https://doi.org/10.5281/zenodo.5797776

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

The CoreTrustSeal[13] is the basic level certification for TDRs. It contains 16 Requirements that cover organisational infrastructure, digital object management and technology (Diagram 1). For each Requirement, the repository needs to provide a self-assessment statement and supporting evidence[14]. This offers a flexible, human-driven assessment and evaluation framework for heterogeneous data repositories. Additional factors in opting for CoreTrustSeal are that it is community-driven, the documentation is openly available, and there were already over one hundred CoreTrustSeal certified repositories when EOSC Nordic began. The EOSC Nordic WP4 support team also had a wide range of expertise in CoreTrustSeal Requirements, and there were support materials available from other projects and organisations[15].

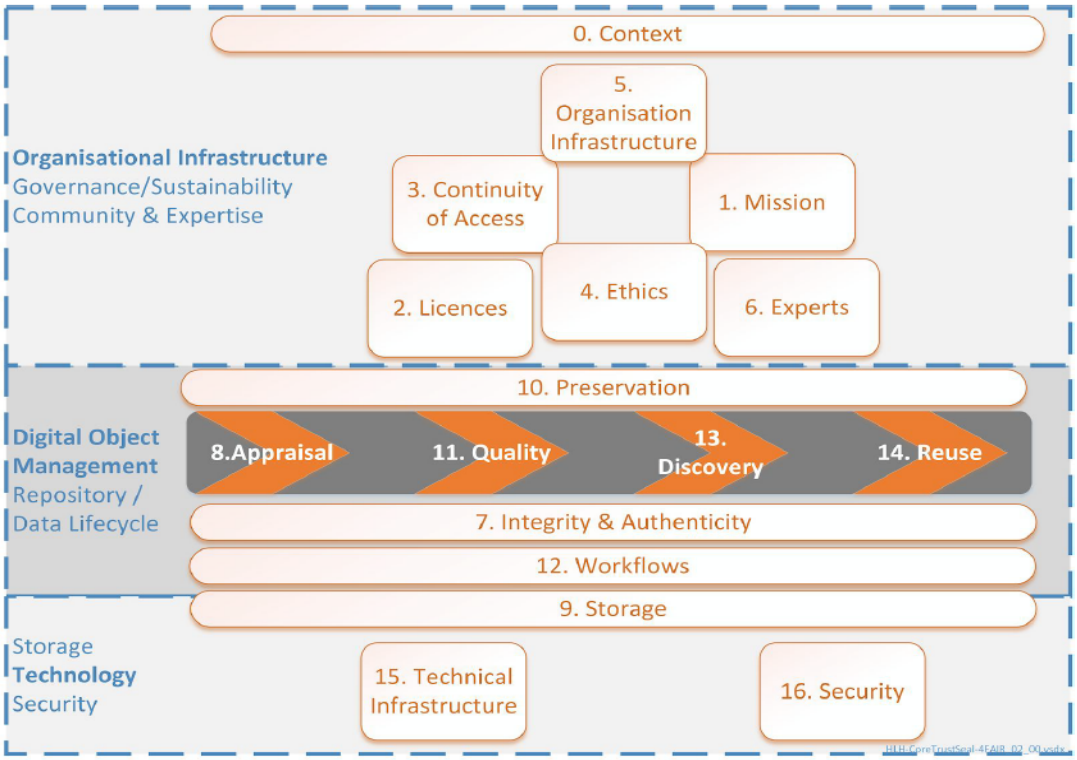

Diagram 1. CoreTrustSeal Requirements in Brief.[16] *Applies to the 2020-2022 requirements.*

When it comes to FAIR, there are a lot of activities around FAIR indicators, metrics and tests but, to date, no formal standard, process or governance structure in place to certify FAIR digital objects or assess the FAIRness of data repositories. The FAIRsFAIR project has used the term 'FAIR enabling' for the steps taken by repositories to ensure digital objects become and remain FAIR. The project developed a model, called *CoreTrustSeal+FAIRenabling CapMat* for repositories to self-assess their current levels of capability and to

[13] https://www.coretrustseal.org/

[14] CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022 (v02.00-2020-2022). Zenodo. https://doi.org/10.5281/zenodo.3632533

[15] These include e.g. the FAIRsFAIR and SSHOC projects that had similar repository support programmes, and organisations like CESSDA ERIC and CLARIN ERIC that require CoreTrustSeal certification from their members.

[16] L'Hours, Hervé, von Stein, Ilona, Huigen, Frans, Devaraju, Anusuriya, Mokrane, Mustapha, Davidson, Joy, de Vries, Jerry, Herterich, Patricia, Cepinskas, Linas & Huber, Robert. (2020). D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures (1.0). Zenodo. https://doi.org/10.5281/zenodo.5360937

www.eosc-nordic.eu

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

plan for increased maturity.[17] The FAIR Working Group of the EOSC Executive Board considers the CoreTrustSeal as the right level for research data repositories and recommends that the CoreTrustSeal+FAIR approach proposed by the FAIRsFAIR project should be extensively tested[18].

The EOSC Nordic activities have focussed on three aspects that cover the current TDR and FAIR certification landscape (see details in Table 1):

- Repository support for FAIRification of metadata records.
- Repository support for self-assessment using the CoreTrustSeal Requirements.
- Testing of the CoreTrustSeal+FAIR approach proposed by the FAIRsFAIR project.

Support Activities for FAIRification and Certification

Support for the adoption of a FAIR data standard and completing the CoreTrustSeal self-assessment was a joint task, i.e. those seeking help in both were supported by the same WP4 task group members in online meetings. Sometimes the calls were themed as either FAIR or certification meetings, or a separate FAIRification meeting was organised, in which case the FAIR evaluation task leader provided special support on the evaluation results.

Repositories were encouraged to participate in the support process to the extent relevant to them. In the initial meetings with each repository, a goal was set for the support process and the first tasks were agreed on. The leading idea was that the repository would select those support options that help them to identify the areas of improvement.

Table 1. Support levels and modes provided by the EOSC-Nordic WP4.

Support level	Modes of support	Target group
Raising awareness of good repository practices based on standards	Webinars, EOSC-Nordic knowledge hub, blog posts and other openly available materials	All repositories
FAIRification of metadata records	FAIRification workshops	All repositories
Self-assessment against selected CoreTrustSeal requirements	Review and feedback on self-assessment texts on selected requirements (typically 3-5), invitation-only workshops to share experiences	Selected repositories
CoreTrustSeal	Review and feedback on self-assessment	Selected repositories

[17] Hervé L'Hours, Maaike Verburg, Jerry de Vries, Linas Cepinskas, Ilona von Stein, Robert Huber, Joy Davidson, Patricia Herterich, & Benjamin Mathers. (2022). Report on a maturity model towards FAIR data in FAIR repositories (D4.6) (V2.0). Zenodo. https://doi.org/10.5281/zenodo.6699520

[18] European Commission, Directorate-General for Research and Innovation, Jones, S., Aronsen, J., Beyan, O., et al., Recommendations on certifying services required to enable FAIR within EOSC, Genova, F.(editor), Publications Office, 2021, https://data.europa.eu/doi/10.2777/127253

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

self-assessment with the intention of a formal CoreTrustSeal application submission	texts of all 16 requirements, invitation-only workshops to share experiences	
CoreTrustSeal + FAIR approach	Peer support for testing of CoreTrustSeal + FAIR approach as proposed by FAIRsFAIR	1-2 repositories that are already CoreTrustSeal certified

Raising awareness of good repository practices based on standards was a continuous activity throughout the project. This included FAIRification activities such as the five open-to-all FAIRification Steps webinars in task 4.1.3 and the closed support recipient webinars in task 4.2.2. These events continued as planned until February 2022.

Support activities for s*elf-assessment against selected CoreTrustSeal requirements* became mostly the first step, as most repositories decided to upgrade their goals from selected requirements to all requirements with the intention of submitting a formal CoreTrustSeal application either during the project or shortly thereafter. *CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission* was scheduled to last until the end of 2021 (M28). However, in the fall of 2021, it became clear that some repositories would need support beyond this point, as progress was steady but gradual. Considering the reallocation of some PMs, the task team was able to extend the general support to April 2022 and continue certification support with the remaining repositories until the end of the whole EOSC-Nordic project, which was later extended to November 2022.

Some of the supported repositories opted to focus exclusively on *FAIRification of metadata records*. Therefore, the support of the WP4 task team was aimed at helping to make metadata and/or data objects more FAIR, and did not include any certification support. We expected to support up to five individual repositories with their efforts to adopt more FAIR practices. This turned out to be the case, although support given measured in working hours was less than expected.

Same is true for those repositories which did not request any detailed FAIRification support. They received general information via the FAIRification webinars, but no detailed one to one support was provided beyond reporting their FAIR scores.

Testing of *CoreTrustSeal + FAIR approach* was to take place in August 2021 to January 2022 by 1 to 2 repositories already CoreTrustCertified. As there were none in the pool of supported recipients, this task was carried out by the project partners themselves, Finnish Social Science Data Archive in December 2021 and Sikt in August/September 2022.

The WP4 task team expected that by the end of 2021 the possible outcomes would be that repositories are encouraged to:

1. submit a formal CoreTrustSeal application
2. continue working on self-assessment statements after the EOSC-Nordic support process or
3. not submit a formal CoreTrustSeal application.

As described above, the deadline was practically postponed until the end of the project. The third recommendation was thought to occur if the repository is not eligible for the CoreTrustSeal certificate. This was the case with one repository, despite their commitment and hard work to develop their practices. At the time of writing, it seems likely that almost all repositories will be prompted to apply.

Since the issuing of the CoreTrustSeal certificate can take 6-8 months after submission, it is assumed that any certificate will only be awarded after the EOSC-Nordic project is completed. The general measure of

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

success is not the number of research data repositories that formally apply for CoreTrustSeal, but the data repositories that committed to improving their data management practices and processes. This goal has been achieved excellently. At least 8 repositories have done this, and this does not include those which did not choose to receive certification support. The task team estimated that it will be able to support 3–5 data repositories to the level where they are ready to submit a formal application. Three have already been advised to submit an application, and up to three more will follow. For more detailed results, see section "Outcomes of the repository specific support".

Raising awareness and events

TDR certification as well as evaluation and assessment of FAIRness of digital objects or data repositories are all still relatively new and evolving topics. The CoreTrustSeal Requirements were launched in 2017[19], the FAIR principles published in 2016[20] and RDA's FAIR Data Maturity Model emerged in 2020[21]. Thus one of the goals of the project was raising awareness of good repository practices based on standards. This was achieved by organising events which all, due to COVID-19, took place virtually.

All repositories in the EOSC-Nordic sample were invited to the kick-off webinar in April 2020, to the FAIRification webinar series and other general events. In addition, we organised events restricted to the supported repositories. All events have been reported in articles published on the EOSC Nordic website and materials are available in the EOSC Nordic Knowledge Hub[22].

- Workshop 'FAIRification of Nordic+Baltic data repositories' (April 22, 2020)
- FAIRification of repositories webinar series
    - STEP1: Global Unique Identifiers for Datasets (Nov 26, 2020)
    - STEP2: FAIR principle F3 – Metadata includes the identifier of the data it describes (Feb 3, 2021)
    - STEP3: Generic metadata standards (Apr 29, 2021)
    - STEP4: Domain specific metadata (Oct 7, 2021)
    - STEP5: Value and Limitations of (FAIR) Automated Evaluators (Feb 8, 2022)
- "From Self-Assessment to Certification with FAIR Results" Certification workshop (Jun 3, 2021)
- Workshop for participating communities in adoption of FAIR data standards - focus on domain specific metadata - M4M event (Dec 7-8, 2021)
- Workshop network of FAIR-enabling trustworthy digital repositories (TDRs) (Jan 13, 2022, Co-organised with FAIRsFAIR and SSHOC projects)
- Workshop for stakeholders for the evaluation of FAIR incentives, May 9, 2022
- Satellite event Becoming a FAIR-enabling trustworthy digital repository, Jun 1, 2022, in connection with NeIC2022 conference

Each event had typically 2-4 invited presentations followed by time for discussion, which made it possible to share the knowledge, experiences and best practices. The topics were chosen based on the feedback and questions from the repositories. All in all, the events had more than 600 participants in total so there is great interest in FAIR and certification among the repositories and other stakeholders.

[19] L'Hours, H., Kleemola, M., & de Leeuw, L. (2019). CoreTrustSeal: From academic collaboration to sustainable services. IASSIST Quarterly, 43(1), 1–17. https://doi.org/10.29173/iq936

[20] Wilkinson, M.; Dumontier, M.; Aalbersberg, I. et al. 2016, The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 https://doi.org/10.1038/sdata.2016.18

[21] FAIR Data Maturity Model Working Group. (2020, June 25). FAIR Data Maturity Model. Specification and Guidelines (Version 1.0). http://doi.org/10.15497/rda00050

[22] https://eosc-nordic.eu/knowledge-hub/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Repository specific support

Initially, almost 100 Nordic and Baltic data repositories were identified. This is the sample which was included in the automated FAIR evaluation test from the beginning or the project[23]. During the project, the task team picked out a total of 23 potential repositories which could benefit from FAIRification or certification support. All these repositories were invited to a kick-off webinar in August 2020 where available support levels were introduced. One repository declined the invitation as they were already in the final stages of finishing their CoreTrustSeal application.

After the webinar, the repositories were approached with a questionnaire which served as an expression of interest for support and provided the task team with basic information of the FAIR and certification status of each applicant and their ability to commit to the work ahead. Initially 8 repositories from five countries were selected to participate in the support process. In April 2021, the task team opened more spaces in the support programme. Additional five repositories, from the pool of 11 identified new potential candidates, expressed their interest. The decision to allow more repositories into the support programme was made based on availability of resources and because some repositories not initially approached were now determined as suitable candidates. This process resulted in a total of 13 repositories participating in the support programme. The majority of the repositories were specialist, domain or subject-based repositories or institutional repositories, in the fields of environmental sciences or social science and humanities.

Table 2. Supported repositories and their types, coverage and goals for the support process.

Name of the repository (* indicates a later addition into the support process)	**Repository type, coverage**	**Support level provided by the WP4 team**	**Repository's goal for the support process**
AIDA Data Hub, Sweden	Domain specific repository, special coverage medical imaging AI	Raising awareness of good repository practices based on standards	Raising awareness
Bolin Centre Database, Sweden	Domain specific repository, special coverage climate research	FAIRification of metadata records	Fairification
DTU*, Denmark	Institutional repository, general coverage	CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission	CoreTrustSeal application
DataDOI, Estonia	Institutional repository, general coverage	CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission;	CoreTrustSeal application and FAIRification

[23] See Nordling, Josefine, Mihai, Hannah, Meerman, Bert, Alaterä, Tuomas, Kleemola, Mari, & Livenson, Ilja. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. Zenodo. https://doi.org/10.5281/zenodo.6880904

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

		FAIRification of metadata records	
FinBIF, Finland	Domain specific repository, special coverage biology and biodiversity	CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission; FAIRification of metadata records	CoreTrustSeal application and FAIRification
Icelandic Social Science Data Service (DATICE), Iceland	Domain specific repository, special coverage SSH	CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission; FAIRification of metadata records	CoreTrustSeal application
Lithuanian Data Archive for Social Sciences and Humanities (LiDA), Lithuania	Domain specific repository, special coverage SSH	CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission	CoreTrustSeal application and FAIRification
Metabolic Atlas, Sweden	Research project repository, coverage metabolism	Raising awareness of good repository practices based on standards	Raising awareness
Musiikkiarkisto (Music Archive Finland)*, Finland	Archive, coverage musicology and cultural studies	Self-assessment against selected CoreTrustSeal requirements	CoreTrustSeal self-assessment or application
NIRD Research Data Archive, Norway	Institutional repository, general coverage	Raising awareness of good repository practices based on standards; self-assessment against selected CoreTrustSeal requirements	Self-assessment against selected CoreTrustSeal requirements
SU*, Sweden	Institutional repository, general coverage	Self-assessment against selected CoreTrustSeal requirements; FAIRification of metadata records	FAIRification, and CoreTrustSeal self-assessment
USN*, Norway	Institutional repository, general coverage	Raising awareness of good repository practices based on standards	Raising awareness

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

qsarDB*, Estonia	Domain specific repository, special coverage oceanography	CoreTrustSeal self-assessment with the intention of a formal CoreTrustSeal application submission; FAIRification of metadata records	CoreTrustSeal application and FAIRification

It should be noted that it is relatively challenging even for a team of FAIR and repository certification experts to flawlessly identify which actors are in fact “data repositories”. Changes in the repository field are also quite rapid. In WP4 our definition was quite broad, as we saw fit most organisations which hosted and provided metadata and digital research data for research purposes. We deliberately excluded repositories which already had a certification or provided only statistical or administrative data, or which did not have a clear research connection to the Nordic and Baltic region. Still, some initially selected repositories in the sample were not in scope.

At the kickoff meeting, different levels of support available were introduced and two members of the task team were assigned to lead each repository-specific support process. In the first meeting with the repository representatives, preliminary goals were set. The task team had envisioned a standard approach that most repositories could follow, but this was always adapted to the repository’s needs at the first meeting. Over time, there was fairly little change in the goals initially set.

From this point on, each individual support process took its own path and schedule. The schedule was mostly based on the repository’s ability to dedicate resources to the work, and in small scale, dependent on the availability of task team members. One virtual meeting lasted approximately an hour, but often included also a pre-meeting commentary, especially regarding filling in the CoreTrustSeal self-assessment template.

Outcomes of the repository specific support

Even though a few support processes are still ongoing, we can report with a high level of certainty on the expected outcomes.

Status at the end of August 2022 (out of 13 participants):

- 4 support processes are ongoing
- 9 support processes have been finished.

The four ongoing support processes are in their final stages and all focus on the certification support.

Status of the 9 finished support processes:

- 3 repositories are ready or have submitted their CoreTrustSeal application
- 1 repository successfully finished the self-assessment but was deemed not eligible for certification at the moment
- 2 received support for FAIRification only and achieved the goal of raising their FAIR score
- 2 repositories withdrew early from further involvement besides webinars, mostly due to time constraints
- 1 repository withdrew after the initial meeting because it was evident that they were not in scope for CoreTrustSeal certification.

With these figures in mind it can be said that the support processes have been very successful. In practice, none of the repositories dropped out once they committed to work on improving their FAIR score or carry out the CoreTrustSeal self-assessment.

EOSC-Nordic project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857652

At the end of August 2022, three supported repositories have either submitted or are ready to submit their CoreTrustSeal applications. All three had set submitting an application as a goal. Currently, the support process is ongoing with four repositories. At least three are aiming to submit their applications before the end of the project, and the WP4 support team sees that they should be ready to do so. If all are granted with a certificate, this would raise the number of CoreTrustSeal certified data repositories in the Nordic and Baltic region from ten to sixteen.

The support process is open until the end of the project, but this in practice means seeing that the remaining certification support processes are wrapped up before the end of October 2022 when CoreTrustSeal ceases to approve new submissions based on the current set of certification criteria. An updated version of the requirement will be introduced in January 2023[24].

One repository made very thorough work with all CoreTrustSeal certification requirements but in the end reached a conclusion that the current mission and charge of the organisation does not make them fully in scope of the CoreTrustSeal.

When comparing repository goals to end results, we can state that there were not any actual drop-outs. Three repositories started with a modest goal aiming to learn more and later during the process define what was desired. One repository's organisational structure changed remarkably during the process and it seemed to make going through the self-assessment undesirable during the project's lifespan. However, they were able to return as a support recipient in late spring 2022, due to the extension of the project. Two of the later additions to the support process were repositories which very efficiently worked though the self-assessment and FAIRification recommendations, and were among the first ones to finish the support process.

Since no formal FAIR data certification scheme emerged during the project, goals related to FAIRification were measured with a) awareness raising and b) improvement in F-UJI scores. Two repositories focused only on improving the FAIRness. They did so in an efficient manner and implemented the changes in their metadata. Recommendations (Appendix 1) provided by the WP4 team were used to guide this work. Diagram 2 shows the FAIRscore development of those repositories which received dedicated FAIRification support. Clear improvement overtime is seen. Changes in the F-UJI tool version explain the even or in some cases negative development from April to August 2021.

24 https://www.coretrustseal.org/why-certification/meeting-community-needs/trustworthy-data-repository-requirements-review-2023-2025/ [23.8.2022]

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

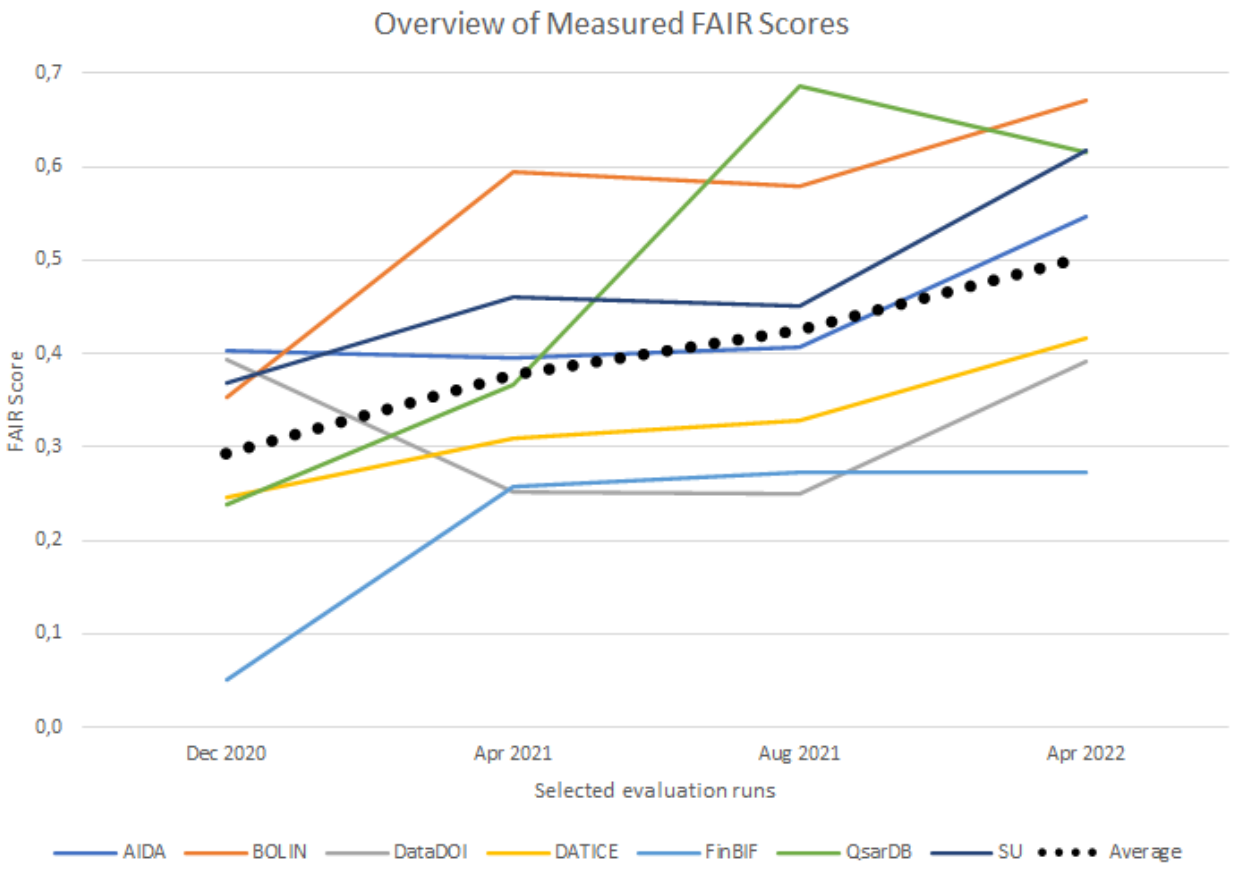

Diagram 2. FAIR scores of repositories which received dedicated FAIRification support during the WP4 support programme.

In F-UJI, FAIR scores run from 0 to 100 (in the picture scaled to 0-1). Getting to 100 is generally speaking extremely difficult or outright impossible and requires extensive optimisation for FAIR and also includes efforts to make actual data objects, not only metadata, more FAIR. Roughly speaking, scores above 50 are already above moderate and closer to 70 very good. These scores cannot be reached without providing machine-actionable metadata about the data object, machine-actionable licences, persistent identifiers, use of some controlled vocabularies, and without providing some information of relations between data and data creators and other related entities.

The task team had more than 80 online meetings or other one to one communication events with the supported repositories. In addition, comments were provided on the self-assessment sheets (CoreTrustSeal certification) or dealt with via email. Biweekly team meetings were held almost throughout the whole project. In these meetings, the progress of the task in general was reviewed but also experiences shared regarding the progress of the individual support processes. In practice, both tasks 4.2.2 (FAIRification) and 4.2.3 (certification) were advanced under the task 4.2.3 meetings. In total, 33 virtual team meetings were held. Normally, a meeting lasted up to an hour.

Testing the CoreTrustSeal+FAIR approach

Since there is no formal FAIR certification mechanism in place, we chose to follow the FAIR Working Group recommendation to test the *CoreTrustSeal+FAIRenabling CapMat* model proposed by the FAIRsFAIR project. The model maps and aligns the CoreTrustSeal Requirements with the FAIR Data Principles to support repository self-assessment of FAIR enabling capability. In short, the mappings align the repository

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

characteristics necessary to achieve TDR status with those that demonstrate a TDR is enabling FAIR (meta)data. The capability maturity (CapMat) approach is designed to support self-assessments and the model focuses on the provision of supporting evidence.[25]

Using the first version of the *CoreTrustSeal+FAIRenabling CapMat* model, we assessed the maturity of the Finnish Social Science Data Archive (FSD) in December 2021. The assessment was completed as an internal review and feedback was provided to the FAIRsFAIR team for further development of the model. FSD can be considered a mature research data repository: it has the CoreTrustSeal certification[26] and during the EOSC Nordic project it has improved the FAIR score of its metadata significantly[27]. But even for a mature repository, the *CoreTrustSeal+FAIRenabling CapMat* assessment proved helpful in identifying gaps and for setting targets for progress. For example, FSD needs to address issues related to de-archiving practices and there is room to improve the interoperability of (meta)data. Areas where FSD reaches a high level of capability include, for example, organisational structure, confidentiality/ethics, preservation plan and data discovery.

Another EOSC Nordic partner, Sikt[28], is planning to carry out a *CoreTrustSeal+FAIRenabling CapMat* assessment during the last two months of the project.

Review of identified risks and realisation of the risks

In the implementation plan, several support process related risks were identified. In the following table those risks are described and the realisation and impact shortly evaluated.

Table 3. Risk and mitigation actions and observed realisation.

Description of risk	Expected mitigation action	Outcome and impact
Covid-19 related risks	Online events instead of face-to-face, as well as rescheduling events and activities. The project remains flexible to adapt to community availability.	All F2F events were replanned as virtual events. This had a minor effect on the commitment and a somewhat negative effect on hands-on peer support, but on the other hand increased the total number of participants.
Changes in organisation of WP4 project team	This is a minor risk as we have a number of individuals with certification competence in the WP4 project team and it is unlikely that several of them will leave the project in the remainder of the process.	In the end, changes in the organisation were a greater risk than expected. Because of various reasons, several project partners and individuals withdrew during the course of the project. Not all were replaced

[25] Hervé L'Hours, Maaike Verburg, Jerry de Vries, Linas Cepinskas, Ilona von Stein, Robert Huber, Joy Davidson, Patricia Herterich, & Benjamin Mathers. (2022). Report on a maturity model towards FAIR data in FAIR repositories (D4.6) (V2.0). Zenodo. https://doi.org/10.5281/zenodo.6699520

[26] https://www.coretrustseal.org/wp-content/uploads/2020/11/Finnish-Social-Science-Data-Archive.pdf

[27] Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, Adil Hasan, Josefine Nordling, & Pauli Assinen. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). Zenodo. https://doi.org/10.5281/zenodo.4045402

[28] https://sikt.no/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

		partly due to the lack of individuals available. Redistribution of PMs was a noticeable task. Some support activities, especially regarding repository specific FAIR support, had to be scaled down.
Repository's wish to withdraw from the process	Provide support materials, encouragement, keep the repository in the loop.	Some withdrew, but very early in the process and with no major effect on the support action. Early withdrawals saved WP4 resources and allowed participation of additional repositories.
Lack of commitments from a community / repository	Provide various levels of complexity / effort to suit differences in ambition.	Commitment is often connected to the resources available and we experienced some related to those repositories short on time of statt. Encouragement and extended support timelines were needed.

Landscape in the Nordic and Baltic region

To better understand the Nordic and Baltic research data repository landscape, the task team conducted desk research to examine the features of the data repositories in our sample. The desk research was done by accessing the publicly available information on the repository websites between October and November 2021. After removing duplicates, non-repositories and repositories that could not be examined because information on them was not publicly accessible, 86 repositories were included in the desk research.

What is a research data repository?

The term 'data repository' is very general and can be used to refer to e.g. general data storage services, data archives or libraries. Nearly a fourth (20) of the repositories in our sample were hosted in Norway. Finland was represented by 13 repositories and Sweden and Denmark were represented by 11 each. Six repositories from Iceland and six from Latvia were included in the sample, while Estonia had seven repositories. Lithuania was the least represented with only two repositories. In addition, 10 repositories were hosted in multiple countries or could not be placed in a single country.

A significant majority of the repositories in the sample, 79%, were specialist repositories focusing on specific fields or disciplines, while the rest were generalist without a specialisation. The repositories were also categorised into types according to the CoreTrustSeal typology[29].

[29] The CoreTrustSeal typology is: Domain or subject-based repository; Institutional repository; National repository system, including governmental; Publication repository; Library/Museum/Archives; Research project repository; Other.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

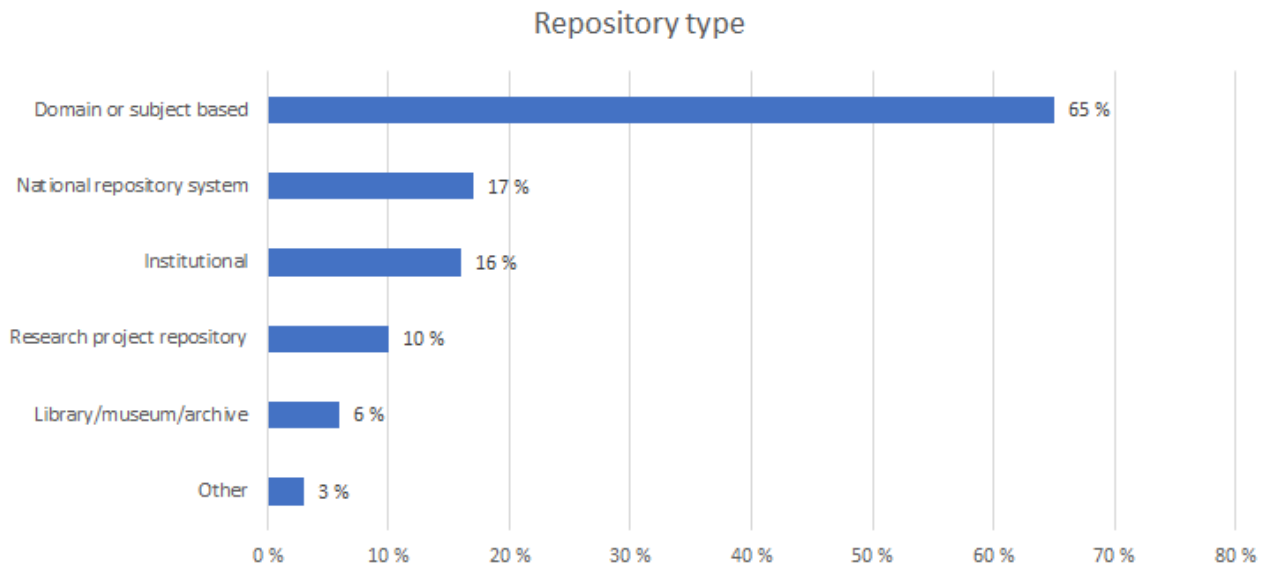

Figure 1. Shares of repository types in the sample following the CoreTrustSeal typology (N=86). A repository could be assigned more than one type.

The specialist repositories in the sample represented various fields and disciplines. Table 4 presents the broad fields and more specific disciplines under them, which were assigned to the repositories during desk research based on the information available on the repository websites. Repositories in the field of natural and life sciences were most represented, as over a half of the repositories in the list focused on disciplines under it. Almost a fourth of the repositories fell under social sciences and humanities, while linguistics comprised 16% of the repositories and other fields 6%.

Table 4. Categorisation of specialisms amongst specialist repositories (N=68).

Field of science	Areas of specialism included in the field
Social sciences and humanities (24%)	Social sciences broadly defined, elections, political science, immigration, folklore studies, literary research, history
Linguistics (16%)	Computational linguistics, language technology
Natural and life sciences (54%)	Biology, bioscience, ecology, genomics, palaeontology, oceanography, marine science, geophysics, geology, geography, climate research
Other (6%)	Psychology, health science, medicine

We were also interested in what information about the repository itself was readily available to the users and other stakeholders (Table 5). We looked for information essential for understanding the services and for reusing the data: mission statement, description of designated community, data preservation approach, model citations, use of PIDs, and terms and conditions.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Table 5. Shares of the repositories with mission statement, description of designated community, long-term preservation, model citation, terms and conditions and persistent identifier. (N=86).

Mission statement	Description of designated community	Long-term preservation	Model citation	Terms and conditions	Persistent identifier (PID)
72%	52%	48%	54%	77%	60%

Nearly three out of four repositories had a mission statement and about half of them had a description of their designated community. Almost half of the repositories explicitly mentioned that they were responsible for long-term preservation. A little over a half of the repositories provided a model citation for their data. Over three fourths had terms and conditions for data use available on their website. Their contents and extent varied between repositories.

Common repository registries like re3data[30] and FAIRsharing[31] help users to discover appropriate repositories and provide overview of existing services. As many as 72% of the repositories in our sample had a record on re3data, while only 22% had a record on FAIRsharing. There was notable overlap with having a record on both registries, as all of the repositories with a FAIRsharing record also had a re3data record. It appears that re3data is the dominant registry for repository information in the current repository practice.

Repository certification status and FAIR maturity in the Nordics and Baltics

In the desk research, we also looked for information about repository certifications. Only 17 (20%) of the repositories in our sample mentioned a certification on their website. The most common certification was CoreTrustSeal, with 71% of the certified repositories having been awarded it. Other mentioned certifications included CLARIN certificate[32], ISO 27001[33] and WDS[34]. As illustrated in Figure 2, Nordics and Baltics have a relatively low number of CoreTrustSeal certified repositories compared to western Europe but a clearly higher number compared to southern Europe.

[30] https://www.re3data.org/
[31] https://fairsharing.org/
[32] There was notable overlap between CoreTrustSeal and CLARIN, as over half of the CLARIN-certified repositories also had the CoreTrustSeal. This is to be expected because becoming a CLARIN centre requires at least initiating the CoreTrustSeal application process. See: https://www.clarin.eu/content/assessment-procedure
[33] ISO/IEC 27001 Information security management. https://www.iso.org/isoiec-27001-information-security.html
[34] World Data System (WDS) https://www.worlddatasystem.org/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Figure 2. CoreTrustSeal certifications in Europe in August 2022. *At the end of August 2022, there are ten certified repositories in the Nordic and Baltic region.*[35]

There is no data available about the FAIR maturity of European research data repositories. The EOSC-Nordic project has tracked the uptake of FAIR in the Nordic and Baltic region by a repeated mass FAIR score testing of metadata records from the repositories in the sample. The results are reported in *D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR*[36]. The average FAIR score in April 2022 was 24.3%. The FAIR scores of most repositories did not change notably during the project but some repositories were able to increase their FAIR score significantly.

[35] Screenshots from the CoreTrustSeal website: https://www.coretrustseal.org/why-certification/certified-repositories/ [30.8.2022]. Currently certified are: Denmark: CLARIN Centre at the University of Copenhagen, International Council for the Exploration of the Sea (ICES); Estonia: CELR META-SHARE; Finland: Finnish Social Science Data Archive (FSD), Language Bank of Finland; Norway: CLARINO Bergen Centre, DataverseNO, Norwegian Marine Data Centre (NMD), NSD's Research Data Archive; Sweden: Swedish National Data Service (SND).

[36] Nordling, Josefine, Mihai, Hannah, Meerman, Bert, Alaterä, Tuomas, Kleemola, Mari, & Livenson, Ilja. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. https://doi.org/10.5281/zenodo.6880904

www.eosc-nordic.eu

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Collaboration and networking

There is a wide array of ongoing activities implementing the Turning FAIR into Reality[37] recommendations. The CoreTrustSeal has been adopted widely globally and across domains[38]. The FAIRsFAIR Synchronisation Force report[39] provides an overview of projects involved in building a FAIR ecosystem for the European Open Science Cloud (EOSC). The EU-funded FAIRsFAIR and SSHOC projects had certification support tasks similar to EOSC-Nordic, all of which had an initial reference point of the CESSDA Trust approach[40].

SSHOC targeted repositories in the SSH domain and provided CoreTrustCertification support but did not encompass FAIR evaluations or support[41]. There was certain overlap between EOSC-Nordic and SSHOC both in project staff and targeted repositories and therefore a dialogue and collaboration between the projects was opened early on and was proven fruitful. FAIRsFAIR selected ten data repositories through an open call for their CoreTrustSeal certification support program and twelve repositories for their FAIRification and interoperability improvement program. The repositories in the certification support programme had an explicit goal to submit a CoreTrustSeal application and FAIRsFAIR provided them also financial support[42].

Discussions among the support teams from these three projects made it clear that research data repositories are seeing a growing number of actors and partners included across the data lifecycle and outsourcing of activities is common. One result of this is that key concepts for TDRs such as Designated Community and active preservation are not clear for all stakeholders. With this in mind the projects collaborated and wrote a working paper *FAIR + Time: Preservation for a Designated Community*[43] to fill this knowledge gap.

Together these three projects took the initiative of organising a workshop to explore the idea of a Network of TDRs in January 2022, with a goal of bringing together the various stakeholders and brainstorm ways to sustain the efforts to empower trustworthy and FAIR-enabling repositories in the EOSC ecosystem[44]. The meeting resulted in writing a working paper (unpublished at time of writing) that will serve as input for the EOSC Task Force on Long Term Digital Preservation[45]. Such a network could provide a coordinated approach that is required to advocate TDRs' role in the ecosystem and to ensure that the importance of TDRs is taken into consideration in policy-making and funding programmes. In many cases, also in the Nordic and Baltic

[37] European Commission (2018). Turning FAIR into reality - Final report and action plan from the European Commission expert group on FAIR data https://doi.org/10.2777/1524

[38] https://www.rd-alliance.org/rda-coretrustseal-adoption-story-across-domains-and-regions [23.8.2022]

[39] Davidson, Joy, Dillo, Ingrid, Grootveld, Marjan, Hodson, Simon, & Pittonet Gaiarin, Sara. (2021). D5.6 Report 3 of the Synchronisation Force (V1.0). Zenodo. https://doi.org/10.5281/zenodo.5595863

[40] L'Hours, Hervé, van Horik, René, Kleemola, Mari, Recker, Jonas, Štebe, Janez, & Jerlehag, Birger. (2020). CESSDA Trust Group: Overview of Support Approaches (v02.00). Zenodo. https://doi.org/10.5281/zenodo.4020399

[41] Mari Kleemola, Henri Ala-Lahti, Tuomas Alaterà, Hervé L'Hours, Benjamin Jacob Mathers, Daan Broeder, René van Horik, Birger Jeriehag, Emiliano Degl'Innocenti, Maurizio Sanesi, & Niko Koski. (2022). D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories. Zenodo. https://doi.org/10.5281/zenodo.6530203

[42] Maaike Verburg, Ilona von Stein, Linas Cepinskas, Hervé L'Hours, Patricia Herterich, Joy Davidson, Kevin Ashley, Olivier Rouchon, Andrea Greco, Serenella Muradore Gallas, & Sara Pittonet Gaiarin. (2021). D4.3 Report on the certification support and guidance for repositories and reviewers (V1.0). Zenodo. https://doi.org/10.5281/zenodo.6656437

[43] L'Hours, Hervé, Kleemola, Mari, von Stein, Ilona, van Horik, René, Herterich, Patricia, Davidson, Joy, Rouchon, Olivier, Mokrane, Mustapha, & Huber, Robert. (2022). FAIR + Time: Preservation for a Designated Community (02.00). Zenodo. https://doi.org/10.5281/zenodo.5797776

[44] Ilona von Stein, Ingrid Dillo, Christian Cuciniello, Linas Cepinskas, Maaike Verburg, Hervé L'Hours, Tuomas Alaterä, & Henri Ala-Lahti. (2022, January 14). Towards a network of European FAIR-enabling Trustworthy Digital Repositories. Zenodo. https://doi.org/10.5281/zenodo.5849658

[45] https://www.eosc.eu/advisory-groups/long-term-data-preservation

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

region, funding for research data repositories is project-based and therefore finite, not allowing for real long-term sustainability.

The EOSC-Nordic project period coincided with the CoreTrustSeal Requirements Review[46] that took place in spring 2022 so the task team was able to provide feedback based on the experiences from the support process and dialogue with repositories.

Feedback from the communities and lessons learnt

The feedback received by the WP4 support programme has been very positive. Due to the fairly small number of supported repositories and not having responses to the feedback survey from all of them (N=7), it is not justified to generalise the results in detail. In the feedback form, all questions were on a scale from one to seven. The median response varied almost always between five and seven. All respondents felt that the WP4 support process was helpful, fulfilled their needs very well and was provided on an adequate level (Figure 3).

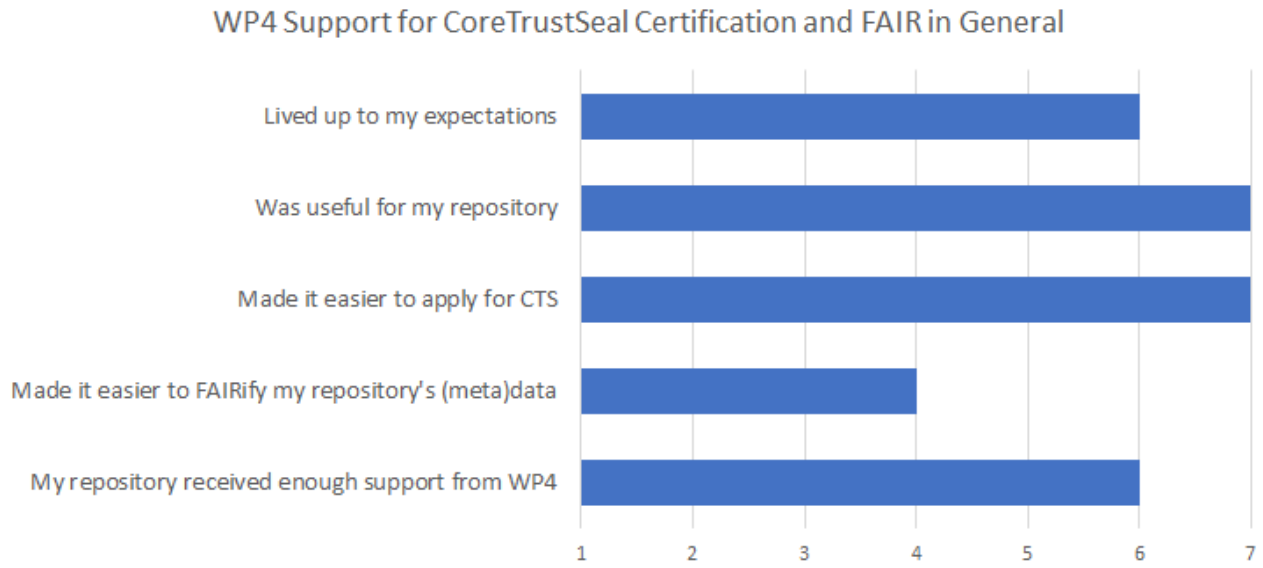

Figure 3. Views on support provided by the WP4 task team.

According to feedback, repositories felt that certification was more important than FAIRification (Figure 4). However, the difference was not great, and some emphasised FAIRification strongly. Repositories that received only FAIR support participated, on average, less in one-to-one support meetings. In most cases, they were involved in joint webinars and held 1-2 discussions with the WP4 team early during the project. It seems (for more information, see aforementioned D4.3) that often after these meetings, repositories made the necessary changes to their metadata and no longer needed support from the task team. The CoreTrustSeal certification process, however, required more support meetings and a cumulative approach.

46 https://www.coretrustseal.org/why-certification/meeting-community-needs/trustworthy-data-repository-requirements-review-2023-2025/ [23.8.2022]

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

It is noteworthy that the lowest grades in feedback apply to FAIR:

- Support process made it easier to FAIRify my repository's (meta) data: 4
- My repository has the tools needed to provide FAIR metadata: 5
- My repository has the tools needed to provide FAIR metadata: 5

This is probably explained by the fact that most repositories focussed on certification but in addition the implementation of FAIR is not yet a well-established activity and the desirable implementations are very discipline- and organisation-specific. Regarding CoreTrustSeal very precise and highly skilled guidance was available. In addition, the ability of the WP4 team to provide support specifically for detailed FAIR issues decreased slightly during the project due to changes in participating organisations and individuals. A fair bit of FAIR knowledge had to be regained during the project. For the last 12 months, FAIR support was primarily offered at webinars, though there were some discussions at support meetings.

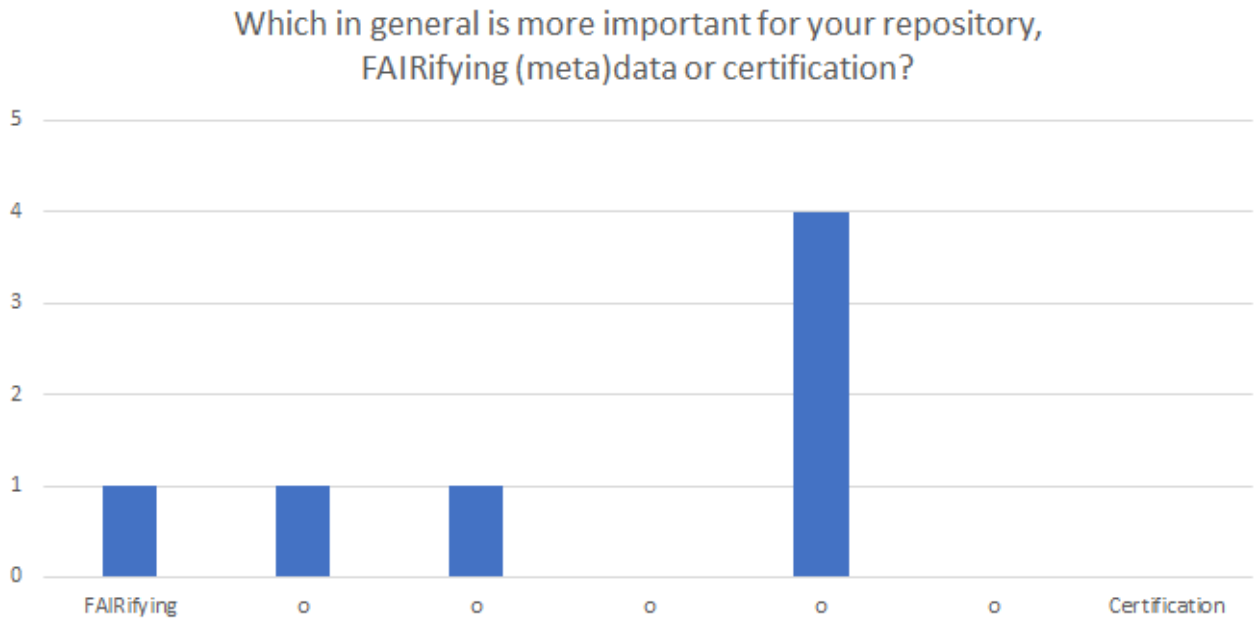

Figure 4. Repositories' views on the importance of FAIR or certification. *The question was asked on a seven-point scale, where the extremes were either only FAIRification or certification.*

As we know from the individual support processes, most of the repositories focused on certification and considered it important. It is still good to note that FAIRification was of primary importance to some, unlike certification.

All respondents agree that support for both certification and FAIRification will continue to be needed. Six replied "definitely yes" and one "probably yes". This outcome is consistent with both the FAIRsFAIR and SSHOC projects. It is therefore worth considering ways in which research infrastructures provide support to their members or how to build interdisciplinary support networks where peer support and best practice can be shared. This also applies to the various technology platform service providers, who should be able to provide adequate means to provide, for example, machine-actionable metadata.

In particular, all respondents emphasise that the WP4 support process for both certification and FAIR certification helped them identify areas where they need to improve their current practices (Figures 5 and 6). This was also evident in many one-to-one support discussions. When there is a clear "list of requirements" or principles for which guiding practices have been developed, they become more concrete

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

and easier to apply to existing workflows. This supports the idea that the CoreTrustSeal criteria may be beneficial in promoting trustworthy digital preservation, even if there is no intent to apply for the certificate or the repository is not fully within the scope of CoreTrustSeal, and therefore unable to apply for it. Similarly, providing general guidance on how FAIR principles apply to metadata requirements and good data management or data curation will facilitate the development of technical and operational solutions that in turn increase FAIR capability.

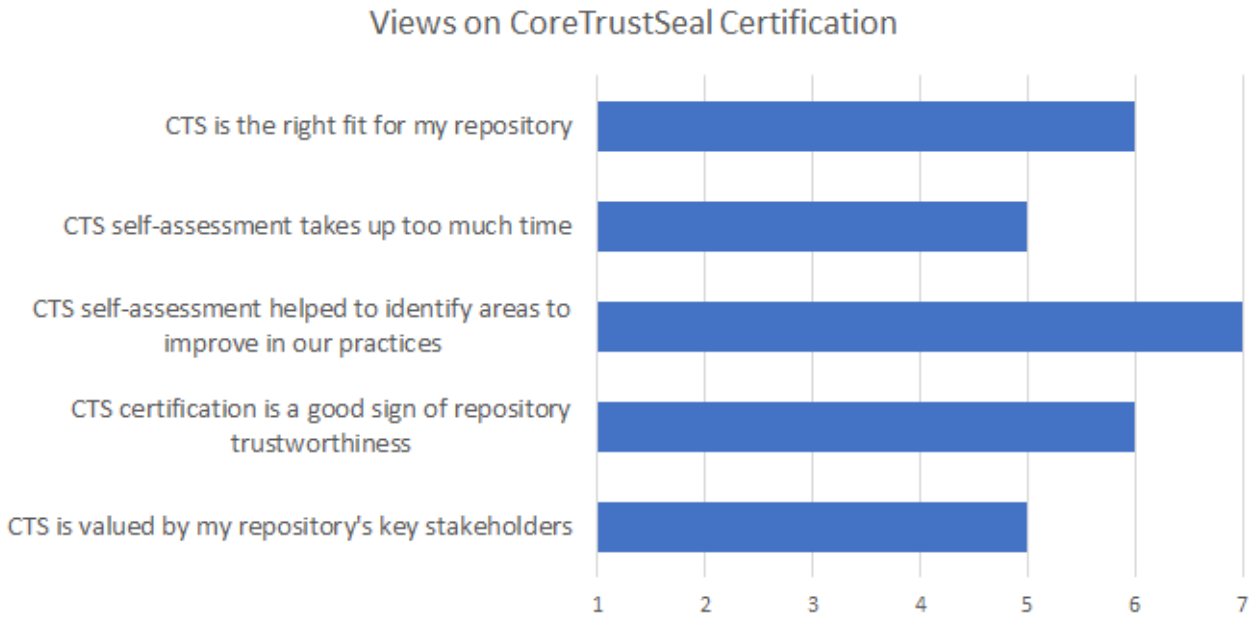

Figure 5. Repositories' views on CoreTrustSeal certification.

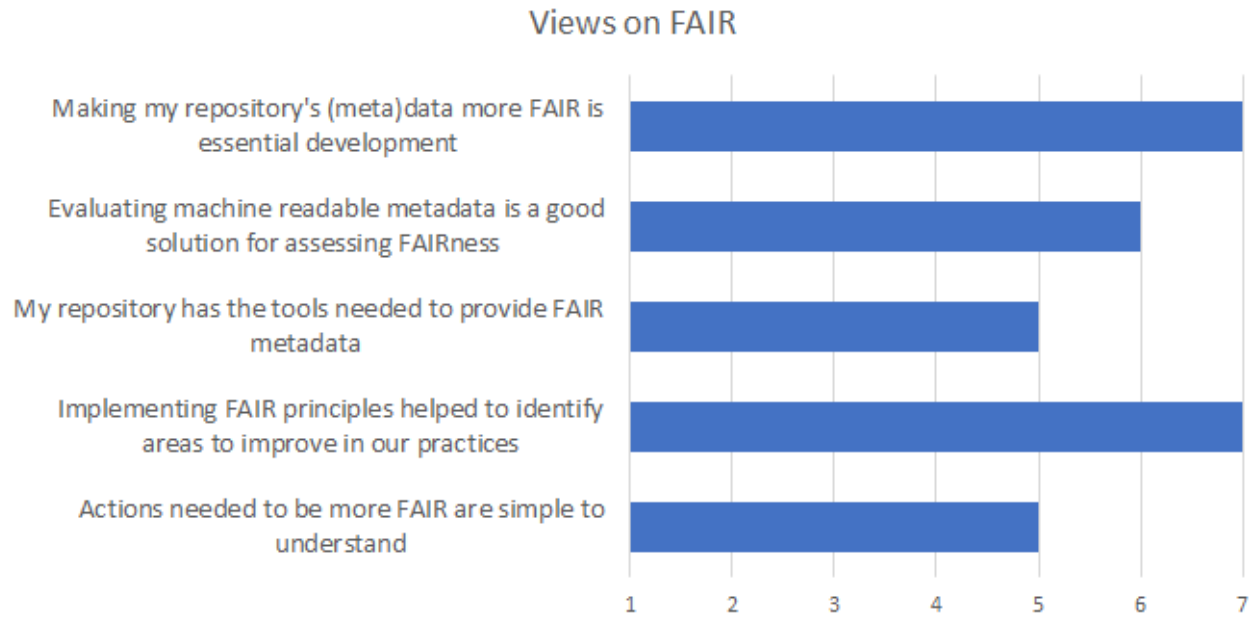

Figure 6. Repositories' views on FAIR.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Based on the feedback, it is also clear that both activities require considerable attention from the repository. CoreTrustSeal self-assessment was considered to take a significant amount of time and resources, primarily to produce sufficient documentation. Also, the still ongoing support processes are those where time was not earlier available or the production of documentation or making systems compliant with the requirements required considerable effort. In terms of FAIR capability, this was reflected in the fact that several archives found that they did not have ideally sufficient tools to provide FAIR-compliant metadata. This further emphasises the role of platform service providers and the fact that FAIR requirements must also be able to be met at source, not only by relying on evaluation results achieved, for example, by harvesting metadata into a common data catalogue.

In the future, some attention should be paid on finding out if CoreTrustSeal is valued by repository key stakeholders. Although the feedback here was mostly positive, there were a few comments during the discussions in which repositories stated that their funders or host organisations did not consider the actual certificate particularly important, even though the repository saw the self-assessment valuable to improve their practises. On the other hand, some organisations have set CoreTrustSeal as a mandatory certificate.

Supported repositories seem to value individual support modes above all else (Figure 7). This is hardly surprising, as direct repository-specific feedback is likely to provide the most help with challenges that the repository has. Regarding online webinars, the separate feedback from these events has been very positive, but those may not be as helpful compared to tailored, repository specific support.

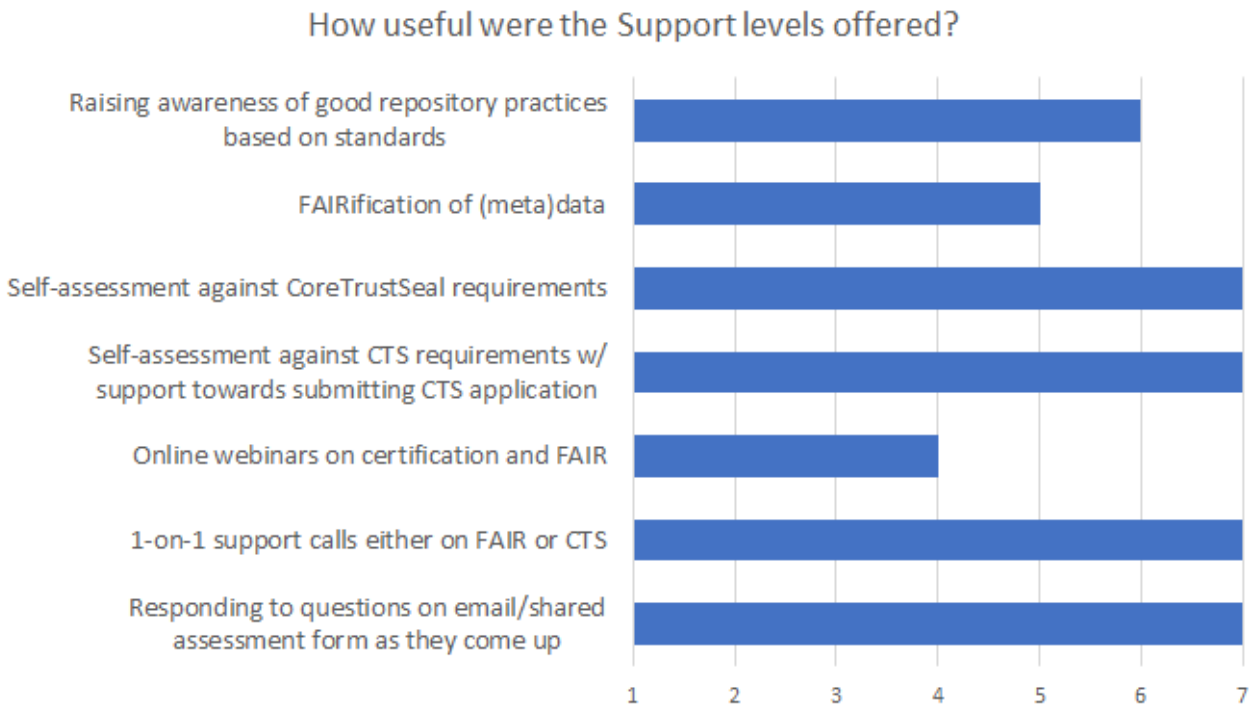

Figure 7. Usefulness of the WP4 support levels provided. *These seven support levels were communicated to the recipients and they were all taken advantage of, but to varying degrees.*

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Discussion

Two general lessons have been learnt during the EOSC-Nordic support programme. Firstly, the starting point matters and secondly, support and guidance is necessary, but not sufficient: the repositories need to have adequate resources to implement the required improvements.

The effort needed to make metadata more FAIR or to get ready to seek certification depends a lot on the starting point. Regarding metadata, if basic steps like persistent identifiers, rich generic and discipline-specific metadata, machine-actionable licences, and controlled vocabularies are already in use and expressed in some form of linked open data, the repository is probably relatively FAIR. Further increase in FAIR scores requires more planning, perhaps introducing additional metadata sections and elements, and ability to customise the repository platform. If not, rapid progress can be made by introducing some or all these elements.

Regarding certification, if a repository already has a well-documented and appropriate mission, sound practices and a good understanding of its designated community or communities, self-assessment against the CoreTrustSeal Requirements, or applying for certification, should not require much extra effort. If a lot of new content has to be created or old content updated, self-assessment can be a time-consuming effort. However, regardless of the ambition to get certified, most CoreTrustSeal-required documents and practices must be in place and TDRs must meet all CoreTrustSeal requirements to run sustainable data repository operations.

As each repository in the support programme had their own starting point in terms of FAIR maturity or readiness for CoreTrustSeal, the support given was tailored to take into account the starting point. Project based support also sets milestones, interim goals and simply deadlines, which help speed up the process and add more structure to it. However, a support process alone cannot make either FAIRification or self-assessment happen. The repositories need to dedicate time and resources to reach successful results. Our experience is that both content specialists and technical staff are needed to implement changes in metadata or practices and processes in order to make these changes meaningful. Cultural change also takes time and FAIR and repository certification are both still relatively new areas.

The levels of support

During EOSC-Nordic support, the most recurring question was perhaps ‘what is FAIR enough’ or what kind of certification is necessary or ‘enough’. The answer is, perhaps unsatisfactorily, that ‘it depends…’ There is no universal answer - the right level depends on, for example, the repository’s mission, domain, user needs and funder requirements. Making data FAIR should be about increasing the understandability of the content by making data findable, accessible, interoperable and reusable - not about trying to outsmart a FAIR evaluator. For TDRs, the CoreTrustSeal certification is the basic level that comprises the very minimum criteria for trustworthy repositories. Currently there is no formal certification framework for repositories that are not TDRs[47].

The process also supported the assumption made in the EOSC-Nordic planning phase, and observed elsewhere, that there isn’t a direct relationship between FAIR principles and CoreTrustSeal. FAIR principles can be mapped to some CoreTrustSeal requirements, but not on an equal level of detail and not to all. While FAIR benefits from good data management practices and sound digital preservation practices, the focus is on digital objects and FAIR places a lot more emphasis on interoperability and machine actionability than CoreTrustSeal. In turn, the CoreTrustSeal strongly emphasises organisational aspects, documentation and management of processes whereas FAIR often is made visible in implementation. A repository may be

[47] This has been discussed e.g. in CoreTrustSeal Standards and Certification Board. (2021). CoreTrustSeal: Specialists, Generalists, and Repository & Data Service Providers (v02.00). Zenodo. https://doi.org/10.5281/zenodo.4568875

EOSC-Nordic project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857652

CoreTrustSeal certified but still score relatively low on FAIR tests, or an uncertified repository may score high on FAIR tests, even in case it is not in scope for CoreTrustSeal[48].

The EOSC-Nordic support programme had participants from a variety of disciplines and representation of both specialist and generalist repositories. All repositories were able to use the CoreTrustSeal Requirements as a tool for gap analysis regardless of formal certification ambitions. However, it is evident that if the scope of the repository is very wide, the ability to deal with data and metadata formats and migration and managed preservation actions required from a TDR grows very demanding very quickly. This is especially true if the repository is small in size. However, the work showed that there are repositories which can with some support in a fairly short time noticeably increase the FAIRness of their metadata and upgrade or clarify their processes and documentation, and therefore better serve the research communities.

Support and upskilling

EOSC-Nordic WP4 did not operate in a vacuum. At the same time, there were several other initiatives and support programmes underway, most notably in the FAIRsFAIR and SSHOC projects (see ch. *Collaboration and networking*) both of which ended in spring 2022. While the projects had different scopes and the repositories in their support programmes had different starting points, all provided one-to-one support using the CoreTrustSeal Requirements so it is interesting to take a quick look at how the results compare. FAIRsFAIR supported ten repositories that all had formal certification as goal, and by the end of the project, nine out of the 10 repositories had submitted a CoreTrustSeal application[49]. SSHOC supported 14 repositories of which eight had a formal CoreTrustSeal application as their goal, and at the end of the projects six repositories had submitted or were close to submitting their application[50]. In EOSC-Nordic, seven repositories were aiming at CoreTrustSeal applications and at the time of writing, three have submitted and another three are still supported and expected to submit before the end of the project in November 2022. Based on these results, one-to-one support is a very efficient way of guiding repositories in their certification journey. The lessons learnt in these three projects are also very similar. These support approaches provide a good and tested basis to any new programmes.

Since the WP4 support included both FAIRification and certification, it became clear, especially in the 1-on-1 discussions and the Q & A sessions of the online events, that the repositories are also looking for peer support and examples of best practices in basic data management, the use of licences, defining access levels, solving privacy issues and making data curation decisions, such as migrations. Neither the FAIR Principles nor the CoreTrustSeal requirements directly provide this guidance. For example, both have clear requirements that the (meta)data must have licences, but neither have instructions on which licence(s) should be chosen since this decision needs to be done by the repository based on e.g. the needs of its designated community.

Research data repositories provide the organisational context for FAIR research data and are key actors in the FAIR ecosystem and in supporting open science. It is not sufficient to ensure that these repositories enable FAIR data. The context in which these repositories operate must also be recognised. There are similarities in policies, culture and incentives in repository activities and sharing research data in the Nordic and Baltic region, which we believe has contributed positively to commitment, communications, peer support and general community building. Sustained efforts to coordinate national, regional and international initiatives and funding would be essential.

[48] Although, as reported in EOSC Nordic D4.3, in the EOSC-Nordic sample, CoreTrustSeal certified repositories got somewhat higher FAIR scores.

[49] https://www.fairsfair.eu/articles-publications/repository-reflections-fairsfair-repository-support-programme-part-1

[50] Kleemola, Mari, Ala-Lahti, Henri, Alaterà, Tuomas, L'Hours, Hervé, Mathers, Benjamin Jacob, Broeder, Daan, van Horik, René, Jeriehag, Birger, Degl'Innocenti, Emiliano, Sanesi, Maurizio & Koski, Niko. (2022). D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories. https://doi.org/10.5281/zenodo.6530203

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Finally, the task team members themselves gained a lot of knowledge both on FAIR data and certification during the project. The team had senior and junior data specialists working together and this allowed learning by doing in this quickly evolving area. A great benefit in this respect was also the collaboration with other pan-European projects. At the end of the EOSC-Nordic, there are more capable data specialists who can continue to implement FAIR in the Nordic and Baltic region.

Conclusion

It is fair to say, based on our efforts, that any repository benefits from periodic screening of their practices against a set of guidelines. This applies to most seasoned repositories, too. While there might not be universally applicable formal frameworks or recommendations to each challenge, solutions do exist. Overall, the FAIR and certification support, along with the efforts of other WP4 tasks, suggest that there are clear benefits to be gained from investing in more FAIR metadata and improving repository practices through self-assessment and/or formal certification. The support model described in this deliverable was generally considered a useful way to implement a targeted change quickly, identify gaps, or produce up-to-date documentation of the repository's processes, policies or practices under expert guidance. It can be used as a basis for future programmes although as FAIR data assessment and repository frameworks develop, support programmes will need to be adjusted and expanded accordingly.

The webinars provided room for discussion on best practices in very specific data management issues, and it is clear that the repositories benefit from deeper discussions with peers with similar challenges. Regional peer groups or networks could be an optimal solution for sharing knowledge. In addition, the repositories showed interest in a higher-level trust network so EOSC-Nordic has together with FAIRsFAIR and SSHOC initiated discussion towards the development of a European network of FAIR enabling trustworthy digital repositories.

The Nordic and Baltic research data repository landscape is diverse. Several data repositories are small in terms of staff or budget, and rely on project-based funding. They are aware of open science, but may not have had the resources to implement FAIR metadata or invest in documented practices, or they may need to depend on outsourced technical expertise that might not allow much flexibility. From a sustainability viewpoint, this is a source of concern.

The repositories have different organisational or functional outsourcing or insourcing partners. Many belong to larger research infrastructures, and there are relatively rich relationships not only between organisations, but also metadata that are harvested, repurposed or linked from different locations. Trust and an effective FAIR-enabling exchange of information between these different service providers is an essential building block. There is a continuous need to measure and support data repositories in certification and FAIRification. These actions should also be widened to different service providers, from suppliers of technical platforms to software vendors and from administrators of research information systems to universities – many at different levels of maturity. The more modular the system becomes, the more important it is that there are ways to ensure the quality and interchangeability of these modules in providing reliable and FAIR research data curation services.

Assuming all expected CoreTrustSeal certification submissions take place during the remaining EOSC-Nordic project period and result in certification, EOSC-Nordic has - together with the commitment and efforts of the supported repositories - produced a sharp increase in the number of certified repositories in the region. The increase in FAIR maturity is not so easy to measure, but all supported repositories were able to increase their FAIR scores. In addition, there is a significant demand for EOSC-Nordic FAIR expertise in other projects, research infrastructures and as speakers at events.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

References

CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022 (v02.00-2020-2022). https://doi.org/10.5281/zenodo.3632533

CoreTrustSeal Standards and Certification Board. (2021). CoreTrustSeal: Specialists, Generalists, and Repository & Data Service Providers (v02.00). https://doi.org/10.5281/zenodo.4568875

Davidson, Joy, Dillo, Ingrid, Grootveld, Marjan, Hodson, Simon, & Pittonet Gaiarin, Sara. (2021). D5.6 Report 3 of the Synchronisation Force (V1.0). https://doi.org/10.5281/zenodo.5595863

European Commission (2018). Turning FAIR into reality - Final report and action plan from the European Commission expert group on FAIR data. https://doi.org/10.2777/1524

European Commission, Directorate-General for Research and Innovation, Jones, S., Aronsen, J., Beyan, O., et al., Recommendations on certifying services required to enable FAIR within EOSC, Genova, F.(editor), Publications Office, 2021, https://data.europa.eu/doi/10.2777/127253

FAIR Data Maturity Model Working Group. (2020, June 25). FAIR Data Maturity Model. Specification and Guidelines (Version 1.0). http://doi.org/10.15497/rda00050

Jaunsen, Andreas Ortmann, Kleemola, Mari, Alaterä, Tuomas J., Lehväslaiho, Heikki, Hasan, Adil, Nordling, Josefine, & Assinen, Pauli. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). https://doi.org/10.5281/zenodo.4045402

Kleemola, Mari, Ala-Lahti, Henri, Alaterà, Tuomas, L'Hours, Hervé, Mathers, Benjamin Jacob, Broeder, Daan, van Horik, René, Jeriehag, Birger, Degl'Innocenti, Emiliano, Sanesi, Maurizio & Koski, Niko. (2022). D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories. https://doi.org/10.5281/zenodo.6530203

L'Hours, Hervé, Kleemola, Mari, von Stein, Ilona, van Horik, René, Herterich, Patricia, Davidson, Joy, Rouchon, Olivier, Mokrane, Mustapha, & Huber, Robert. (2022). FAIR + Time: Preservation for a Designated Community (02.00). https://doi.org/10.5281/zenodo.5797776

L'Hours, Hervé, Verburg, Maaike, de Vries, Jerry, Cepinskas, Linas, von Stein, Ilona, Huber, Robert, Davidson, Joy, Herterich, Patricia & Mathers, Benjamin. (2022). Report on a maturity model towards FAIR data in FAIR repositories (D4.6) (V2.0). Zenodo. https://doi.org/10.5281/zenodo.6699520

L'Hours, Hervé, van Horik, René, Kleemola, Mari, Recker, Jonas, Štebe, Janez, & Jerlehag, Birger. (2020). CESSDA Trust Group: Overview of Support Approaches (v02.00). https://doi.org/10.5281/zenodo.4020399

L'Hours, Hervé, von Stein, Ilona, Huigen, Frans, Devaraju, Anusuriya, Mokrane, Mustapha, Davidson, Joy, de Vries, Jerry, Herterich, Patricia, Cepinskas, Linas & Huber, Robert. (2020). D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures (1.0). https://doi.org/10.5281/zenodo.5360937

L'Hours, H., Kleemola, M., & de Leeuw, L. (2019). CoreTrustSeal: From academic collaboration to sustainable services. IASSIST Quarterly, 43(1), 1–17. https://doi.org/10.29173/iq936

Nordling, Josefine, Mihai, Hannah, Meerman, Bert, Alaterä, Tuomas, Kleemola, Mari, & Livenson, Ilja. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. https://doi.org/10.5281/zenodo.6880904

von Stein, Ilona, Dillo, Ingrid, Cuciniello, Christian, Cepinskas, Linas, Verburg, Maaike, L'Hours, Hervé, Alaterä, Tuomas & Ala-Lahti, Henri. (2022, January 14). Towards a network of European FAIR-enabling Trustworthy Digital Repositories. https://doi.org/10.5281/zenodo.5849658

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Verburg, Maaike, von Stein, Ilona, Cepinskas, Linas, L'Hours, Hervé, Herterich, Patricia, Davidson, Joy, Ashley, Kevin, Rouchon, Olivier, Greco, Andrea, Muradore Gallas, Serenella, & Pittonet Gaiarin, Sara. (2021). D4.3 Report on the certification support and guidance for repositories and reviewers (V1.0). Zenodo. https://doi.org/10.5281/zenodo.6656437

Wilkinson, M.; Dumontier, M.; Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 https://doi.org/10.1038/sdata.2016.18

Appendices

Appendix 1: Recommendations for FAIR Data Standards

NR.	FAIR STANDARD	FAIR Principle	RECOMMENDED ACTION	FURTHER RECOMMENDATION
1	UNIQUE and PERSISTENT IDENTIFIER	F1	Assure the data and metadata resources have a unique and persistent identifier (at least a globally unique ID (GUID) like URL or DOI).	Provide the (meta)data with a Unique Identifier, ideally a Global Unique Persistent Resolvable Identifier (GUPRI)
2	STRUCTURED, GROUNDED METADATA	F2	Assure existence of structured metadata (for instance in RDF or in embedded JSON).	Provide 'rich' metadata in RDF or embedded JSON – ideally by following existing metadata templates, schemas and controlled vocabularies - multilingual if appropriate.
3	DATA IDENTIFIER EXPLICITLY IN METADATA	F3	Assure the metadata contains the unique identifier to the data.	Separate metadata from the data and assure that the metadata explicitly includes the identifier of the related data.
4	METADATA IDENTIFIER EXPLICITLY IN METADATA	F3	Assure the metadata has a Global Unique Persistent Resolvable Identifier (GUPRI).	Separate metadata from the data and assure that the metadata explicitly includes the identifier of the related metadata.
5	SEARCHABLE IN MAJOR SEARCH ENGINE	F4	Assure a machine is able to discover the resource by indexing, using a machine actionable search engine.	Assure your resource is findable in a registered /indexed searchable resource.
6	USES OPEN FREE PROTOCOL FOR DATA RETRIEVAL	A1.1	Assure data may be retrieved by an open and free protocol by testing data GUID for its resolution protocol	Assure your data resources can be properly resolved (given that they can be made openly available).
7	USES OPEN FREE PROTOCOL FOR METADATA RETRIEVAL	A1.1	Assure metadata may be retrieved by an open and free protocol by testing metadata GUID for its resolution protocol	Assure your metadata resources can be properly resolved
8	DATA AUTHENTICATION AND AUTHORIZATION	A1.2	Assure a discovered data GUID has the availability to implement authentication and authorization in its resolution protocol.	Assure your data is compliant with an explicit data access policy.
9	METADATA AUTHENTICATION AND AUTHORIZATION	A1.2	Assure the metadata GUID has the availability to implement authentication and authorization in its resolution protocol.	Assure your metadata is compliant with an explicit data access process.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

10	METADATA PERSISTENCE	A2	Assure that the metadata contains an explicitly identified and machine readable persistence policy.	Assure your metadata is compliant with a machine-readable Persistence Policy Key.
11	METADATA KNOWLEDGE REPRESENTATION LANGUAGE	I1	Assure that the metadata uses a formal language broadly applicable for knowledge representation.	Use a (community defined) ontology and machine-actionable language to structure your metadata or use an existing metadata template
12	DATA KNOWLEDGE REPRESENTATION LANGUAGE	I1	Assure the data uses a formal language broadly applicable for knowledge representation.	Use a (community defined) ontology to structure your data or use an existing data template
13	METADATA USES FAIR VOCABULARY	I2	Assure the linked data metadata uses terms that resolve to linked FAIR data.	Assure the metadata identifier resolves and uses FAIR vocabulary (check F1, A1.1. and I1) .
14	METADATA CONTAINS QUALIFIED OUTWARD REFERENCES	I3	Assure that metadata (as Linked Data) link outward to third-party resources.	Assure your dataset can be represented as Linked Data and has a reference to other metadata.
15	METADATA INCLUDES LICENCE	R1.1	Assure that metadata contains an explicit pointer to a licence.	Assure an explicit pointer to the licence or use existing schemas that include licence terms.
16	(META)DATA INCLUDES PROVENANCE	R.1.2	Assure structured metadata that describes provenance for instance in RDF or in embedded JSON).	Assure there is an explicit pointer to a schema that has machine readable provenance.
17	(META)DATA INCLUDES DOMAIN COMMUNITY STANDARDS	R.1.3	Assure that (meta) data is associated with community / domain standards and is based upon agreed vocabularies and semantic models.	Assure an explicit pointer to the templates, schemas, ontologies, vocabularies, variables that the community has defined as the standard. (Ideally published by the community as their FIP, their FAIR Implementation Profile)

VBME2601

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652