

Increasing the Reuse of Data through FAIR-enabling the Certification of Trustworthy Digital Repositories

Benjamin Mathers
UK Data Service

Hervé L'Hours
UK Data Service

Abstract

The long-term preservation of digital objects, and the means by which they can be reused, are addressed by both the FAIR Data Principles (Findable, Accessible, Interoperable, Reusable) and a number of standards bodies providing Trustworthy Digital Repository (TDR) certification, such as the CoreTrustSeal. Though many of the requirements listed in the *Core Trustworthy Data Repositories Requirements 2020–2022 Extended Guidance* address the FAIR Data Principles indirectly, there is currently no formal ‘FAIR Certification’ offered by the CoreTrustSeal or other TDR standards bodies. To address this gap the FAIRsFAIR project developed a number of tools and resources that facilitate the assessment of FAIR-enabling practices at the repository level as well as the FAIRness of datasets within them. These include the *CoreTrustSeal+FAIRenabling Capability Maturity model* (CTS+FAIR CapMat), a FAIR-Enabling *Trustworthy Digital Repositories-Capability Maturity Self-Assessment* template, and F-UJI, a web-based tool designed to assess the FAIRness of research data objects. The success of such tools and resources ultimately depends upon community uptake. This requires a community-wide commitment to develop best practices to increase the reuse of data and to reach consensus on what these practices are. One possible way of achieving community consensus would be through the creation of a network of FAIR-enabling TDRs, as proposed by FAIRsFAIR.

Submitted 31 July 2022 ~ Accepted 3 August 2022

Correspondence should be addressed to Ben Mathers, UKDS. Email: bm21346@essex.ac.uk

This article is based on a lightning talk at International Digital Curation Conference IDCC22, online, 13-16 June, 2022

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

The long-term preservation of digital objects, and the means by which they can be reused, are addressed by both the FAIR Data Principles (Wilkinson et al., 2016) and a number of standards bodies providing Trustworthy Digital Repository certification, such as the CoreTrustSeal¹ (Recker et al., 2019). The FAIR Data Principles were first formally conceptualized in 2014 following the original Lorentz Workshop entitled ‘Jointly Designing a Data FAIRport’ (Lorentz center, 2014). The CoreTrustSeal was formed in 2017 following recommendations made by the Research Data Alliance Repository Audit and Certification DSA–WDS Partnership Working Group² (Rickards et al., 2016) through the merging of the World Data Systems’ and the Data Seal of Approval’s certification processes (CoreTrustSeal, 2022).

Though many of the requirements listed in the CoreTrustSeal Extended Guidance (CoreTrustSeal Standards and Certification Board, 2019) address the FAIR Data Principles indirectly, there is currently no formal ‘FAIR Certification’ offered by the CoreTrustSeal or other TDR standards bodies (Consultative Committee for Space Data Systems, 2011; nestor Seal Certification Working Group, 2013; The National Archives, 2018).

Upon application to the CoreTrustSeal, repositories are assessed against 16 requirements which cover organisational infrastructure, digital object management and technical provision, with requirement 14 specifically addressing ‘Data Reuse’ (CoreTrustSeal Standards and Certification Board, 2019). To adequately address (and assess) these areas, both the repository and the CoreTrustSeal require an understanding of the repository’s designated community and the technical environment that it operates in. Compliance with the CTS requirements is assessed on a sliding scale of 1 to 4, with Level 4 being the highest. To achieve a compliance Level 4 for data reuse, applicants must provide evidence of appropriate metadata provision, fully implemented processes to mitigate format obsolescence and an in-depth knowledge of data reuse scenarios to ensure the continued understandability of data (CoreTrustSeal Standards and Certification Board, 2019).

The value of CoreTrustSeal certification in providing assurance that data and metadata become, and remain, FAIR over time has been recognised by initiatives such as the European Open Science Cloud (EOSC),³ which aims to provide the research community with access to FAIR data and interoperable data services (European Commission, Directorate-General for Research and Innovation, 2016). To ensure the trustworthy storage of FAIR data, the EOSC has sought to align its practice, and the practice of participating institutions, with the CoreTrustSeal Requirements (European Commission, Directorate-General for Research and Innovation, 2021).

Tools and Resources for Increasing the Reuse of Data

CoreTrustSeal+FAIRenabling Capability Maturity model

To this end, the FAIRsFAIR project⁴ has developed the *CoreTrustSeal+FAIRenabling Capability Maturity model* (CTS+FAIR CapMat) (L’Hours et al., 2022a) as a first step in aligning the

¹ CoreTrustSeal: <https://www.coretrustseal.org/>

² Research Data Alliance Repository Audit and Certification DSA–WDS Partnership Working Group: <https://rd-alliance.org/group/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg/outcomes/dsa-wds-partnership>

³ European Open Science Cloud: <https://eosc-portal.eu/about/eosc>

⁴ FAIRsFAIR - Fostering Fair Data Practices in Europe: <https://fairsfair.eu/>

CoreTrustSeal Requirements and the FAIR Data Principles. The *CTS+FAIR CapMat* provides a detailed mapping of the CoreTrustSeal requirements to the applicable FAIR Data Principles which gives repositories a way to simultaneously assess their current capability and maturity as well as the FAIRness of their practices. The *CTS+FAIR CapMat* achieves this using a simplified, three tier scale for assessing FAIR-enabling practice: (1) Initial; (2) Managed; and (3) Defined (L'Hours et al., 2022a). Alongside the other core requirements, the *CTS+FAIR CapMat* addresses a number of areas relating to the reuse of data including the provision of License Information (R02), Data Integrity and Authenticity (R07) and Data Reuse (R14). In order for repository practice to be classed as FAIR-enabling in relation to 'Data Reuse', the repository must have identified the needs of the designated community and ensured that the characteristics of the data and metadata contained in the digital objects within its collection(s) address these needs (L'Hours et al., 2022a).

A *FAIR-Enabling Trustworthy Digital Repositories-Capability Maturity Self-Assessment template* (L'Hours et al., 2022a) has also been developed to allow repositories to ascertain their current levels of capability and maturity in relation to both the FAIR Data Principles and TDR best practices using a single resource. By providing specific mappings for data reuse, alongside a readily available resource for repositories to self-assess their current practice, the *CTS+FAIR CapMat* can help increase potential data reuse through repository capability and maturity planning and an increased awareness of the FAIR Data Principles.

Automated FAIR Data Assessment

Yet, repository assessment is not the only method for increasing data reuse; assessing digital objects themselves according to how FAIR they are is another valuable tool in this regard. To this end, the FAIRsFAIR project created F-UJI,⁵ a web-based tool designed to programmatically assess the FAIRness of research data objects at the dataset level. The tests utilized by F-UJI assess both the technical and content-related information embedded within the data and metadata of a given dataset, with the tests using metrics developed by the FAIRsFAIR project (Devaraju et al., 2020). Similar to the FAIR Data Principles, the F-UJI tool was developed in a domain-agnostic way and can be used to assess publicly available datasets, with plans to further develop it during the FAIR-IMPACT project.⁶

Using the tests developed within F-UJI, the FAIR Data Assessment Toolset and Badging Scheme (Huber et al., 2021) was also created, which provides the associated FAIRsFAIR metrics and practical tests giving repositories a further resource to test the FAIRness of the data they preserve, in turn providing repositories with a publicly available means to increase the reusability of their data, amongst other benefits. The *CTS+FAIR CapMat*, the F-UJI tool and FAIR Data Assessment Toolset are all underpinned by the mappings conducted in the FAIR Principles: Baseline Comments (L'Hours et al., 2022b) working paper, which maps the FAIR Data Principles, RDA Indicators and the Metrics and Practical tests developed by FAIRsFAIR, providing a further resource for repositories who wish to develop their FAIR-enabling practises.

Network of FAIR-enabling Trusted Digital Repositories

However, increasing the reusability of data across the digital preservation landscape is not only dependent on the availability of adequate methods, but also their uptake. This requires a community-wide commitment to develop best practices to increase the reuse of data and community consensus on what these practices are. Once reached, best practices can be incorporated within the requirements of the different TDR certification bodies such as the CoreTrustSeal. A possible way of achieving such a consensus would be through the creation of a network of FAIR-enabling TDRs, as proposed by FAIRsFAIR (von Stein et al., 2021). Such a network could help ensure that specific and minimum measures are in place to provide evidence

⁵ F-UJI: Automated FAIR Data Assessment Tool: <https://www.f-uji.net/>

⁶ FAIR-IMPACT: <https://fair-impact.eu/>

that the relevant designated community (as a subset of the wider community of users) are consulted on the necessary requirements to ensure the availability of data for reuse. To this end, two possible mechanisms that could be utilised would be a Technology Watch function and Community Watch function, with the latter focusing on contextual information about the digital objects through metadata, thus providing a means for increasing the reusability of data. Such a network could also facilitate the maintenance of the current *CTS+FAIR* mappings and provide a means for the community to engage with the CoreTrustSeal on further integrating the FAIR Data Principles within the CoreTrustSeal Requirements.

Conclusion

In summary, the practical complexities of implementing the FAIR Data Principles and increasing data reuse necessitates a multi-pronged approach. The *CTS+FAIR CapMat*, alongside the other outputs discussed above, provides a clear and accessible means for repositories to increase the reusability and FAIRness of their data. But rather than being a singular answer to these objectives, they should be viewed as part of a wider movement towards FAIR Data, with data reuse being one key component. The successful implementation of the FAIR Data Principles depends upon collective community action, uptake, input and consensus, alongside the continued availability of the necessary tools, resources, standards, and practices to make FAIR Data a reality.

References

- Consultative Committee for Space Data Systems. (2011). *Audit and certification of Trustworthy Digital Repositories* (Magenta Book CCSDS 652.0-M-1). Retrieved from <https://public.ccsds.org/pubs/652x0m1.pdf>
- CoreTrustSeal (2022). *History*. Retrieved from <https://www.coretrustseal.org/about/history/>
- CoreTrustSeal Standards and Certification Board. (2019). *CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020-2022 (v02.00-2020-2022)*. Retrieved from <https://doi.org/10.5281/zenodo.3632533>
- Devaraju, A., Mokrane, M., Cepinskas, L., Huber, R., Herterich, P., de Vries, J., Akerman, V., Davidson, J., LHours, H. and Diepenbroek, M. (2020). *M4.9 Report on Fair Data Assessment Mechanisms to Develop Pragmatic Concepts for Fairness Evaluation at the Dataset Level*. Retrieved from <https://doi.org/10.5281/zenodo.4118405>
- European Commission, Directorate-General for Research and Innovation. (2016). *Realising the European Open Science Cloud: Realising the European Open Science Cloud of the Commission High Level Expert Group*. Retrieved from <https://data.europa.eu/doi/10.2777/940154>
- European Commission, Directorate-General for Research and Innovation. (2021). *Recommendations on certifying services required to enable FAIR within EOSC*. Retrieved from <https://data.europa.eu/doi/10.2777/127253>
- Huber, R., Cepinskas, L., Davidson, J., Herterich, P., LHours, H., Mokrane, M., von Stein, I., and Verburg, M. (2021). *D4.5 Report on FAIR Data Assessment Toolset and Badging Scheme*. Retrieved from <https://doi.org/10.5281/zenodo.6656444>

- L'Hours, H., Verburg, M., de Vries, J., Cepinskas, L., von Stein, I., Huber, H., Davidson, J., Herterich, P. and Mathers, B. (2022a). *Report on a maturity model towards FAIR data in FAIR repositories (D4.6)*. Retrieved from <https://doi.org/10.5281/zenodo.6699520>
- L'Hours, H., von Stein, I., Mokrane, M., Parland-von Essen, J., de Vries, J., Huigen, F., Devaraju, A., Davidson, J. and Mathers, B. (2022b) *FAIR Principles: Baseline Comments*. Retrieved from <https://doi.org/10.5281/zenodo.6472149>
- Lorentz center. (2014). *Jointly designing a data FAIRPORT*. Retrieved from <https://www.lorentzcenter.nl/jointly-designing-a-data-fairport.html>
- nestor Certification Working Group. (2013). *Explanatory notes on the nestor Seal for Trustworthy Digital Archives*. Retrieved from <http://d-nb.info/1047613859/34>
- Recker, J., Dillo, I., L'Hours, H., Kleemola, M., Crabtree, J., Downs, R.R., Hugo, W., Jenkins, R., Lin, D., Mokrane, M., Trilsbeek, P., von Stein, I., Callaghan, L. (2019). *An Introduction to the CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022*. Retrieved from <https://doi.org/10.5281/zenodo.3638211>
- Rickards, L., Vardigan, M., Diepenbroek, M., Dillo, I., Genova, F., L'Hours, H., Hugo, W., Khodiyar, V., Li, G., Minster, J.B., Trilsbeek, P., Wyborn, L., Edmunds, R., Mokrane, M. and Panagou, E. (2016) *DSA–WDS Partnership Working Group Catalogue of Common Procedures for Certification*. Retrieved from <https://doi.org/10.15497/RDA00019>
- The National Archives. (2018). *Archival Service Accreditation Standard*. Retrieved from <https://cdn.nationalarchives.gov.uk/documents/archives/archive-service-accreditation-standard-june-2018.pdf>
- von Stein, I., L'Hours, H.; Cepinskas, L.; Mathers, B.; Dillo, I.; Verburg, M.; Mokrane, M.; Herterich, P. and Rouchon, O. (2021). D4.4 Coordination Plan for a sustainable network of FAIR-enabling Trustworthy Digital Repositories. Retrieved from <https://doi.org/10.5281/zenodo.5726690>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W. Bonino da Silva Santos, L., Bourne, P.E. Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, E., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Alasdair Gray, A.J.G, Groth, P., Goble, C., Grethe, J.S. Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016). The Fair Guiding Principles for scientific data management and stewardship, *Scientific Data*. Retrieved from <https://doi.org/10.1038/sdata.2016.18>