1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.309

Published: 07 Sept. 2023

Ten Simple Rules for Designing and Building a FAIR Research Infrastructure

Sharif Islam[1][https://orcid.org/0000-0001-8050-0299]

[1] Naturalis Biodiversity Center, Leiden, The Netherlands

Abstract. One of the key priorities of The European Strategy Forum on Research Infrastructures (ESFRI) is to build sustainable and FAIR (Findable, Accessible, Interoperable, Reusable) infrastructures. However, designing and building such infrastructures requires careful consideration of various factors, such as data interoperability, operational sustainability, and governance. This poster proposes ten simple rules, inspired by Ten Simple Rules for scientific research, for designing and building a research infrastructure drawing from existing initiatives particularly from experiences in preparation of DiSSCo (Distributed System of Scientific Collections) – a new research infrastructure that was in the ESFRI 2018 roadmap. While these rules are not comprehensive, they highlight a few essential traits that can be applied across different disciplines. For each rule, we highlight how within DiSSCo we accomplished the specific aspect.

Keywords: FAIR Digital Objects, Natural Science Collections

1. Introduction

One of the key priorities of The European Strategy Forum on Research Infrastructures (ESFRI) is to build sustainable and FAIR (Findable, Accessible, Interoperable, Reusable) infrastructures. However, designing and building such infrastructures requires careful consideration of various factors, such as data interoperability, operational sustainability, and governance. This poster proposes ten simple rules, inspired by Ten Simple Rules for scientific research, for designing and building a research infrastructure drawing from existing initiatives particularly from experiences in preparation of DiSSCo (Distributed System of Scientific Collections) – a new research infrastructure that was in the ESFRI 2018 roadmap. While these rules are not comprehensive, they highlight a few essential traits that can be applied across different disciplines. For each rule, we highlight how within DiSSCo we accomplished the specific aspect.

1.1 Rule 1: Ensure a Clear Purpose and Vision

Before designing and building a research infrastructure, it is important to have a clear purpose and vision. This should include a definition of the research infrastructure's scope, its target user community, and the expected outcomes. DiSSCo's focus on European natural science collections at the centre of data-intensive scientific excellence and innovation provided a clear purpose and vision to work towards [1].

1.2 Rule 2: Adopt FAIR Principles from the beginning

FAIR principles should be adopted from the beginning of the design process. This will enable the research community to find and use the data effectively, leading to discoveries and insights.

During the ICEDIG project (which delivered the DiSSCo design blueprint) and DiSSCo Prepare, the FAIR implementation plan was at the forefront. DiSSCo's involvement in the ENVRI FAIR poject helped in definning a FAIR implementation plan. DiSSCo's decision to use Digital Object Architecture also aligns with the ongoing work within the FAIR Digital Object specification [2, 3]. Looking into specific workflows from multiple organisations and understanding the data lifecycle (see Figure 1) helped us to scope different aspects of FAIR.

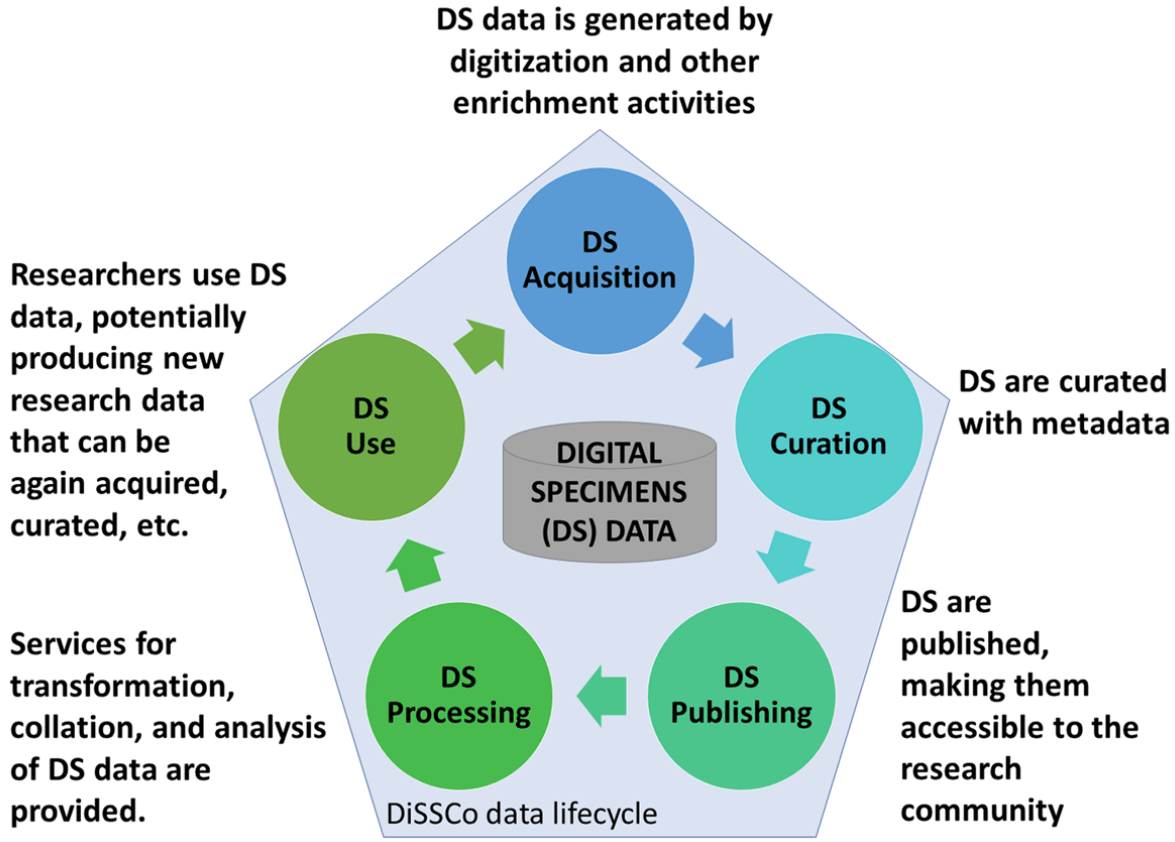

Figure 1: Digital Specimen (DS) lifecycle as the core data element for DiSSCo

1.3 Rule 3: Build on Existing Collaborations and Projects

By leveraging existing resources and expertise, it is possible to avoid duplicating efforts and to build on what has already been achieved. In DiSSCo's case, different digitisation projects over the past decades provided the foundation for building a data-driven ecosystem. EU funded Projects deliverables from ICEDIG, SYNTHEYES+ were valuable for the DiSSCo Prepare project. DiSSCo's involvement in other EU projects also provided opportunities to work and collaborate with other research infrastructures (GBIF, LifeWatch, eLTER, ELIXIR).

1.4 Rule 4: Design for Interoperability

Interoperability is key to the success of any system. The infrastructure should be designed to enable the seamless exchange of data and metadata between different systems and platforms. This will ensure that the data is reusable and can be combined with other data to create new insights. DiSSCo's decision to use the FAIR Digital Object framework provides the building blocks for creating interoperable systems (see Figure 2; also see [4]) The current development effort also focuses on open source solutions that are better suited for delivering interoperable research infrastructure.

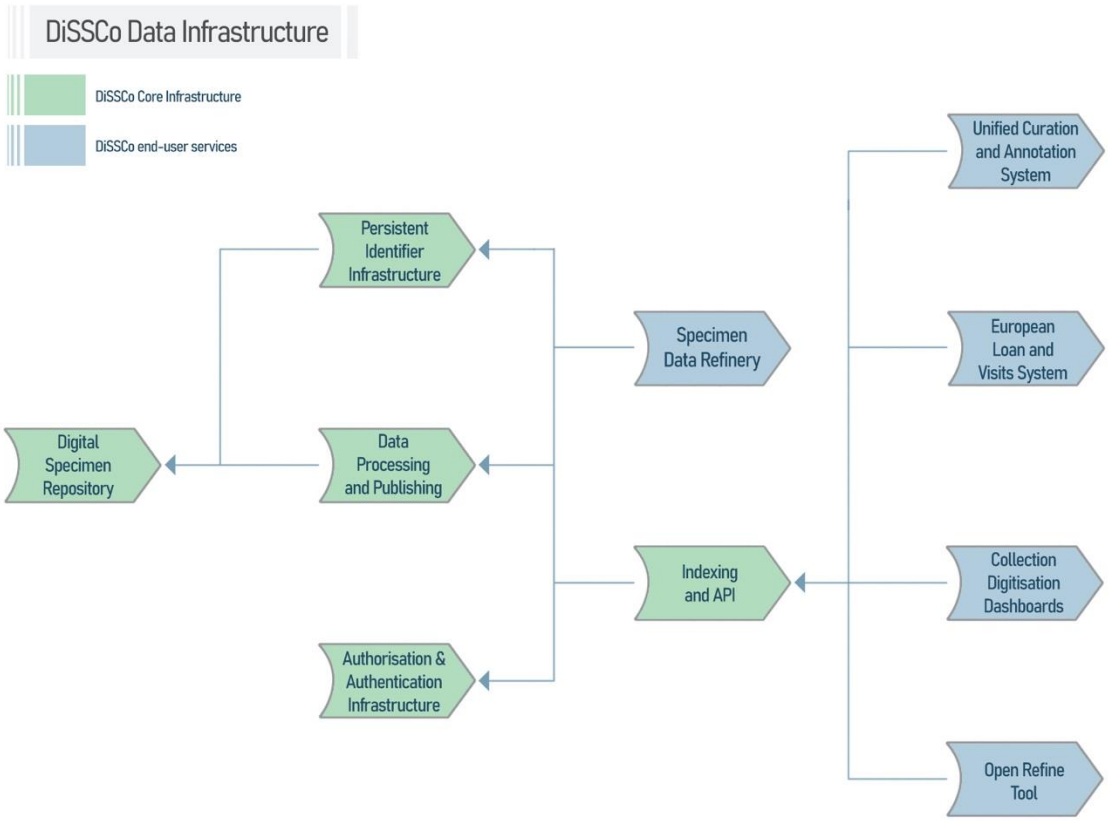

Figure 2: DiSSCo data infrastructure building blocks

1.5 Rule 5: Consider Sustainability

The infrastructure must be planned to guarantee long-term technical and financial viability. The technical team at DiSSCo adopted a framework for creating modular products that could be changed and expanded because of the use of FAIR principles, open source software, and the adoption of Agile and DevOps practices. Interoperability and sustainability also enable the system to be designed to be flexible and future-proof. For the financial aspect, extensive discussion and planning went on in the preparation phase of DiSSCo towards building a funding structure supported by the nation states and a plan towards forming an ERIC (European Research Infrastructure Consortium) [5].

1.6 Rule 6: Develop Data Governance and Policy Frameworks

The infrastructure should be designed to include policies and guidelines that govern data access, sharing, and reuse. This will ensure that the data is used ethically and in compliance with legal and ethical standards. Even though DiSSCo is not operational yet, our pilot is already considering the policy implications of different data workflows (from sample collections to publishing data). Some of these are also aligned with global policy discussions. At the same time, the aforementioned ERIC roadmap is contributing towards a robust discussion around governance mechanisms. These discussions have significant implications for how resources can be allocated to implement the wider technical vision.

1.7 Rule 7: Promote Openness and Collaboration

The infrastructure should be designed to promote openness and collaboration between different stakeholders, including researchers, data providers, and infrastructure providers. This will ensure that the infrastructure is used to its full potential and that discoveries and insights are created. During DiSSCo's design and preparation phase, we used various modes for collaboration and communication (GitHub, newsletters, webinars for instance).

1.8 Rule 8: Plan for User Support and Training

The infrastructure should be designed to include user support services that help researchers to use the infrastructure effectively. This includes training, documentation, and technical support. Both SYNTHESYS+ and DiSSCo Prepare projects had a specific focus on user support and training.

1.9 Rule 9: Ensure Data Security and Privacy

The infrastructure should be designed to include security and privacy measures that protect sensitive data and comply with legal and ethical standards. Even though DiSSCo does not deal with human research data, we will have the researcher's profile and other personal information that needs to adhere to GDRP. Along with that, global procedures and steps that are in place need to be incorporated within DiSSCo for handling endangered species data.

1.10 Rule 10: Monitor and Evaluate Progress

Monitoring and evaluating progress can be achieved by regularly reviewing metrics such as usage and impact. The EU project framework for milestones and deliverables provides a structure for such monitoring and evaluation. However, DiSSCo internally also developed processes to ensure we are working towards our vision and allow opportunities for modification and improvement.

2. Conclusion

It is essential to recognise that the proposed set of ten "simple rules" for designing a FAIR research infrastructure, while inspired by the FAIR principles, extends beyond them. Rather than providing easy-to-follow rules for addressing the 16 FAIR principles, the poster focuses on a broader approach towards designing and building a sustainable and interoperable research infrastructure like DiSSCo. The findings emphasise that FAIR is not solely about making data discoverable, accessible, interoperable, and reusable. Instead, it encompasses a more holistic perspective, considering the entire data lifecycle, governance structure, and other digital objects (such as workflow management, algorithms, models, and research software [6]). This approach advocates for creating a modular, loosely coupled, yet integrated design that can serve the user community's needs.

The poster advocates for going beyond mere checkbox compliance with the FAIR principles and instead adopting a comprehensive, integrated approach. This forward-thinking, FAIR-by-design perspective will advance research, collaboration, and knowledge dissemination in various disciplines.

Author contributions

Sharif Islam [CRediT roles: Conceptualization, investigation, writing—original draft, review & editing].

Competing interests

The author declare that there is no competing interests.

Funding

The DiSSCo Prepare project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871043.

Acknowledgement

The authors thank and acknowledge valuable feedback offered by Wouter Addink, Eva Alonso, and the reviewrs.

References

1. A. Hardisty, H. Saarenmaa, A. Casino, M. Dillen, K. Gödderz, Q. Groom, H. Hardy, D. Koureas, A. Nieva de la Hidalga, D.L. Paul, V. Runnel, X. Vermeersch, M. van Walsum, and L. Willemse, "Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1," *Research Ideas and Outcomes*, vol. 6, p. e54280, 2020. [Online]. Available: https://doi.org/10.3897/rio.6.e54280
2. A.R. Hardisty, E.R. Ellwood, G. Nelson, B. Zimkus, J. Buschbom, W. Addink, R.K. Rabeler, J. Bates, A. Bentley, J.A. Fortes, and S. Hansen, "Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet," *BioScience*, vol. 72, no. 10, pp. 978-987, 2022. [Online]. DOI: https://doi.org/10.1093/biosci/biac060
3. L. Lannom, D. Koureas, and A.R. Hardisty, "FAIR data and services in biodiversity science and geoscience," *Data Intelligence*, vol. 2, no. 1-2, pp. 122-130, 2020. [Online]. DOI: https://doi.org/10.1162/dint_a_00034
4. T. Loo, M. Benvenuti, L. Cecchi, G. Innocenti, M. Biaggini, L. Bellucci, V. Moggi Cecchi, and F. Di Vincenzo, "DiSSCo Prepare Deliverable D9.6 - Compilation of Construction Masterplan," 2023. [Online]. DOI: https://doi.org/10.34960/cy1m-b238
5. S. Scory, C. Paleco, A. Casino, E. Alonso, D. Koureas, T. Loo, P. Mergen, A. Nivart, F. Dusoulier, and V. Demanoff, "MS-7.2 Analysis of the legal entity models and their suitability for achieving DiSSCo objectives," 2022. [Online]. URI: https://know.dissco.eu/handle/item/473
6. M. Barker, N.P. Chue Hong, D.S. Katz, et al., "Introducing the FAIR Principles for research software," Sci Data, vol. 9, p. 622, 2022. [Online]. DOI: https://doi.org/10.1038/s41597-022-01710-x