

# FAIR Data, FAIR Services and the FAIR Data Action Plan

© Simon Hodson

ICSU Committee on Data for Science and Technology (CODATA),

Paris, France

simon@codata.org

Due to the work of initiatives such as CODATA, Research Data Alliance (RDA), those who resulted in the FAIR principles and some others we can observe a trend towards increased agreements on how we need to change our data practices to improve data sharing and reuse and make data-intensive work much more efficient. In 2014, the FAIR principles were published in the realm of FORCE11 [1]. They summarise a longer discussion in a perfect way and are being accepted globally as guidelines in science. Important to note is that they not only discuss human usage of data & metadata, but in particular address the need to make data & metadata ready for machine usage. According to the F-A-I-R Principles data needs to be

- **Findable** (data & metadata need to have a globally unique persistent identifier, need to be described with rich metadata, should be searchable via portals, and the identifier needs to be findable in the metadata),
- **Accessible** (data & metadata needs to be retrievable by their identifier using standard communication protocols, protocols need to be free, open and support authentication and authorisation, and metadata needs to be accessible even when the data is not available anymore),
- **Interoperable** (data & metadata needs to be encoded using agreed representation standards, make use of FAIR compliant vocabularies and include relevant references),
- **Re-Usable** (data & metadata are associated with relevant attributes, are released with clearly defined usage licenses, are associated with their provenance and meet community standards)

There is now a broad agreement about these principles and a number of implicit implications such as data should be as open as possible and be preserved where necessary for future generations. Based on the broad support in science the European Commission and the member states have given the FAIR principles a central role for their plans to establish the European Open Science Cloud - a stepwise evolving eco-system of research and data infrastructures. The EC established an expert group to develop plans for "Turning FAIR data into reality". This expert group published recently its interim report [2] for open discussion by the community and is currently including the many comments which

were received.

The report is structured in seven sections reporting on the urgency of making data FAIR, the need to change culture with respect to data, the implications of the principles for establishing a FAIR compliant ecosystem, the urgency for skills development and capacity building, the need for instruments to measure the change towards FAIRness in community practices and the implications for funding programs. For all these sections recommendations and actions have been extracted. From the total 34 recommendations 14 have been indicated as primary ones and major stakeholder groups were associated with them.

1. A broader definition of the FAIR principles is given and their relevance is being stressed. Data should be FAIR even if it is not open.
2. It is stressed data should be as open as possible and as closed as necessary. This holds in particular for data that is being created as part of publicly funded research.
3. FAIR Objects as being defined in RDA [3] are introduced as a way to organise data compliant with the FAIR principles where the use of persistent identifiers (PIDs) plays a crucial role.
4. A number of components are identified to implement a FAIR compliant ecosystem such as repositories, registries, identifier resolution systems, standards for structures and semantics, policies and data management plans.
5. It is stressed that there is a need of sufficient and sustainable funding to maintain all these components.
6. Funding of services should be tied to FAIR metrics and depend on impact and community adoption.
7. Further support should be given to research communities to continue the development and maintenance of their disciplinary interoperability frameworks including principles and policies for data management and sharing, data formats/structures, semantics, tools etc.
8. The need to open ways for cross-disciplinary FAIRness by developing and adopting common standards where possible is stressed to enable interdisciplinary research. It is important that these are being developed in a international context.
9. The development and implementation of robust FAIR metrics is important to assess progress in the FAIRification of data in the research

- communities.
10. An utterly important role is assigned to trustworthy repositories since they need to support access to and reuse of the data. The tasks range from managing the stored bit sequences up to the stewardship of structures and embedded semantics. Repositories are motivated to participate in regular quality assessments according to standards such as CoreTrustSeal [4] which is already applied worldwide by many institutions.
  11. A special concern is the assessment of the FAIRness of services in addition to data. New certification standards have to be developed based on existing models.
  12. Data Management Plans should be made and regularly updated by all funded projects that include data. They help to make arrangements with the required services providers such as repositories, to plan sufficient resources for data management and stewardship and developing plans how data can be made FAIR compliant.
  13. Measurements are required to develop two cohorts of professionals to support FAIR data. Data scientists who have insights in the intended scientific work and data stewards who have deep knowledge to ensure proper management and curation of data.
  14. Finally it is requested that there is more recognition of the professions of data stewards and that the efforts in creating FAIR data are rewarded.

Funders in Europe seem to be willing to intensively discuss these recommendations and to anchor them in their funding programs. The publication of the FAIR principles and the broad support they receive can thus be seen as a milestone to improve data sharing and re-use across disciplinary and regional borders. They will help to reduce the huge inefficiencies and thus costs that can be found in data-driven projects and that are estimated at around 80% of waste of time due to what is called the data wrangling.

## References

- [1] The FAIR Data Principles.  
<https://www.force11.org/group/fairgroup/fairprinciples>
- [2] Hodson, S., Jones, S., Collins, S., Genova, F., Harrower, N., Laaksonen, L., Mietchen, D., Petrauskaitė, R., Wittenburg, P. Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data. 2018.  
<https://zenodo.org/record/1285272#.W5y7uPmYTb0>
- [3] Berg-Cross, G., Ritz, R., Wittenburg, P. RDA DFT Core Terms and Model. 2018.  
<http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318>

- [4] CoreTrustSeal Data Repository.  
<https://www.coretrustseal.org/>