

FAIR Data Principles - improving the re-usability of NBN data holdings

The FAIR data principles are a set of guidelines, developed primarily in the research and academic sector, to encourage and enable better sharing and reuse of data. The principles were first published in 2016 (*Wilkinson et al. [The FAIR Guiding Principles for scientific data management and stewardship](#). *Sci Data* 3, 160018 (2016) doi:10.1038/sdata.2016.18*) and are now a standard framework for the storage and sharing of scientific information. FAIR data are Findable, Accessible, Interoperable and Reusable. The Principles emphasise the importance of machine-readable and actionable information by the use of identifiers (e.g. DOIs), data and metadata standards and controlled vocabularies.

Two sessions were held to provide an overview of the principles followed by questions. Delegates were asked whether the NBN should formally adopt and align itself to the FAIR data principles and, if so, should FAIR be considered in the NBN's strategy refresh in 2020. Also, delegates were asked to consider linked data as the ultimate expression of FAIR. Most delegates had not heard of linked data in any detail and apart from describing what it was, it was not a focus of the session.

Delegates were very supportive for the adoption of FAIR within the NBN and the majority agreed that FAIR should be considered in the NBN's strategy refresh. General feedback was that, with the exception of persistent and resolvable identifiers (e.g. DOIs) for datasets uploaded to the NBN Atlas, the NBN is already quite advanced in following the FAIR principles:

1. We have a single data repository (NBN Atlas), which is available over the internet;
2. Our licences are clear and machine readable;
3. We have the species inventory (UKSI), which includes a unique and persistent ID for each taxon concept; and
4. The NBN Atlas already follows the Ecological Metadata Language (EML) standard for metadata and Darwin Core for the data.

A common thread in the sessions was: are there any reasons for not following FAIR? The answer depends on the extent to which the NBN decides to implement FAIR. Organisations can decide individually how FAIR they wish to make their data. One very simple option is to just make the metadata available, so that the dataset is findable, and make the data available at later date. Rating datasets on the NBN Atlas by their level of FAIRness by using a traffic light scheme was discussed.

There were two particularly important aspects of FAIR for the delegates:

- 1) The distinction between Open data and FAIR data; and
 - 2) That recognition and assessment of data-use is much easier when the data is FAIR.
- FAIR data provides the exact conditions under which the data are accessible, for example by the licence or because an authorisation protocol is required. Data can be FAIR and shared under restrictions (data should be *as open as possible and as closed as necessary*). The use of personal or institutional ORCID IDs (<https://orcid.org/>) was discussed as a way of claiming ownership of work (data, publications, presentations etc.). DOIs and ORCID facilitate automatic tracking of citations and data-use. Sandy Knapp, in her keynote NBN conference address, introduced Bloodhound (<https://bloodhound-tracker.net/>), which via the use the DOIs and ORCID enables anyone to claim and track the use of natural history specimens that they have collected or identified. This is hugely valuable information for grant applications and funding bodies.

The following next steps were agreed:

1. Delegates would discuss FAIR within their organisations
2. NBN Trust will provide DOIs for datasets on the NBN Atlas as soon as possible
3. The provision of metadata for datasets on the NBN Atlas will be improved and data providers will be given support, where necessary, in completing metadata
4. The NBN Trust will start developing controlled vocabularies for terms in Darwin Core. This will be done together with TDWG (Biodiversity Information Standards: <https://www.tdwg.org/>) where possible.

The session organisers are grateful for the support of Rachel Drysdale from ELIXIR (<https://elixir-europe.org/>) and Peter McQuilton from FAIRsharing (<https://fairsharing.org/>) for providing information for the presentation.

Useful Background References:

The FAIR Guiding Principles for scientific data management and stewardship
<https://www.nature.com/articles/sdata201618>

FAIRsharing as a community approach to standards, repositories and policies
<https://www.nature.com/articles/s41587-019-0080-8>

Force11 website: <https://www.force11.org/fairprinciples>