Are the FAIR Data Principles fair?

Dunning, Alastair; de Smaele, Madeleine; Boehmer, Jasmin

DOI
10.2218/ijdc.v12i2.567

Publication date
2020

Document Version
Final published version

Published in
International Journal of Digital Curation

IJDC | *General Article*

Are the FAIR Data Principles Fair?

Alastair Dunning
4TU. ResearchData, TU Delft

Madeleine de Smaele
4TU. ResearchData, TU Delft

Jasmin Böhmer
4TU. ResearchData, TU Delft

Abstract

This practice paper describes an ongoing research project to test the effectiveness and relevance of the FAIR Data Principles. Simultaneously, it will analyse how easy it is for data archives to adhere to the principles. The research took place from November 2016 to January 2017, and will be underpinned with feedback from the repositories.

The FAIR Data Principles feature 15 facets corresponding to the four letters of FAIR - Findable, Accessible, Interoperable, Reusable. These principles have already gained traction within the research world. The European Commission has recently expanded its demand for research to produce open data. The relevant guidelines[1] are explicitly written in the context of the FAIR Data Principles. Given an increasing number of researchers will have exposure to the guidelines, understanding their viability and suggesting where there may be room for modification and adjustment is of vital importance.

This practice paper is connected to a dataset (Dunning et al., 2017) containing the original overview of the sample group statistics and graphs, in an Excel spreadsheet. Over the course of two months, the web-interfaces, help-pages and metadata-records of over 40 data repositories have been examined, to score the individual data repository against the FAIR principles and facets. The traffic-light rating system enables colour-coding according to compliance and vagueness. The statistical analysis provides overall, categorised, on the principles focussing, and on the facet focussing results.

The analysis includes the statistical and descriptive evaluation, followed by elaborations on Elements of the FAIR Data Principles, the subject specific or repository specific differences, and subsequently what repositories can do to improve their information architecture.

1 H2020 Guidelines on FAIR Data Management: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Received 20 October 2016 ~ *Revision received* 27 January 2017 ~ *Accepted* 27 January 2017

Correspondence should be addressed to Alastair Dunning, Prometheusplein 1, 2628 ZC Delft, Netherlands. Email: A.C.Dunning@tudelft.nl

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

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/

Introduction

The 4TU.Centre for Research Data[2] (4TU.Research Data) is hosted by the Research Data Services team (RDS) of the Library of the Technical University of Delft. It supports research data management and long-term archiving of research output for three of the four technological universities in the Netherlands. Part of the RDS is providing advice and help on Data Management Plans (DMPs), for general project-planning as well as funding requests.

The Horizon 2020 program (H2020) united all previous European innovation and research funding in 2011, creating a unique strategic framework to elevate research and innovation to a new level in Europe (European Commission, 2017a). Between 2014 and 2020 societal challenges are tackled, industrial leadership established and excellent science sustained (European Commission, 2017b). After the successful conclusion of an Open Research Data Pilot[3], the EU decided that the management of research data, and its publication, would need to be undertaken by all H2020 projects. To encourage researchers to make their data "as open as possible, as closed as necessary" (European Commission, 2016), the European Commission implemented the FAIR principles as guidance[4] to improve the findability and accessibility, interoperability and reusability of their research data.

The FAIR principles intend to give "a minimal set of community-agreed guiding principles and practices" and was created in 2014 (FORCE 11, 2014b). Both machines and humans should be enabled to find (F), access (A), interoperate (I) and re-use (R) research data and metadata in an effortless but confined fashion. Each letter of FAIR is endowed with a subset of facets that focus on technical, information-architectural and knowledge-domain specific requirements (FORCE 11, 2014a).

Being responsible for maintaining the 4TU.Centre for Research Data and striving for excellent research support, made us curious on how our data archive adheres to the principles. Having collaboration and academic knowledge exchange in mind, we set out to examine other data repositories in Europe and gathered information on how easily they can match the requirements and demands in the provided facets. Our conclusions are divided into two sections, one on issues with the FAIR principles themselves, and a second on what repositories can do to respond to the FAIR principles.

Within these sections this practice paper examines how closely existing archives are to meeting the FAIR principles. Relatedly, it looks at how much effort is needed to adjust existing data repository structures to adhere to the FAIR principles, and what can realistically be achieved in current set-ups. Ultimately Research Data Management (RDM) faces a broad spectrum of data and metadata types, and different documentation styles according to research methodologies. Consequently, the type of technical infrastructure required to make datasets compliant to the FAIR principles is very different from making data FAIR compliant.

2 4TU.Centre for Research Data: http://researchdata.4tu.nl/en/home/

3 Open Aire: https://www.openaire.eu/opendatapilot

4 European Commission: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Methodology

Disclaimer: For the sake of convenience, this paper uses the term *repository* for all the varieties of data archives and research data infrastructures covered in this paper. Where necessary more precise terms will be used. In the context of this practice research the following distinction was made in order to categorize the repositories for statistical interpretation:

- **Data-Archive:** Provides persistent identifier, has metadata-record, provides coherent data-sets; self-proclaimed archive in information provided online.
- **Data-Repository:** Does not necessarily provides persistent identifier; metadata-record can be minimal; sometimes no coherent data-sets.
- **Research-Infrastructure:** Offers special features and more services in addition to regular data repository traits.
- **Subject-Based Repositories:** Consists of institutional and subject based repositories, with varying qualities of metadata records, data-sets and persistent identifier, according to the needs of the particular scientific discipline.
- **Online Databases:** Provides an interactive (and predominantly complex) interface and dynamic information creation; the data-set and according metadata is sometimes difficult to determine.

This practice-based research inductively explores if the FAIR Data Principles are fair by applying a mixed data collection of quantitative analysis, backed up with qualitative commentary. The repositories forming the sample cluster consist of data-repositories affiliated to the Netherlands, as well as international repository popular in the Dutch research community. The registry of research data repositories Re3Data.Org[5] was used as source for Netherland specific repositories. The sample collection of data repositories has been set deliberately broad to include as many data-storing and publishing information-services as possible. Hence, Data Archives (i.e. DANS EASY, 4TU.Centre for Research Data), Research Infrastructures (such as EUDAT B2Share), Institutional Repositories (i.e. SHARE-ERIC), and Subject Based Repositories (such as EDGAR) are part of the population and sample group.

FORCE 11, is a movement of like-minded stakeholders of the research community that wants to advance digital scholarly publishing[6]. The FAIR principles are published as a short overview (FORCE 11, 2014a) and an extended guideline (FORCE 11, 2014b). The contributors and authors of the FAIR principles wrote down their rationale behind the principles and the experiences of implementing them in an article in Nature (Wilkinson et al., 2016).

The applied scoring matrix of this external evaluation ranges from green (compliant), to orange (just about / maybe not), to red (not compliant), to blue (unclear). The scoring is applied accordingly to the information available on the website of the repository, what is written on help pages, and what is visible in the published data-record. The list of attributes used as scoring matrix is displayed in the appendix of this

5 Registry of Research Data Repositories: http://www.re3data.org/
6 About Force 11: https://www.force11.org/about

practice paper and is based on the short version of the FAIR principles (FORCE 11, 2014a).

Analysis

Subsequent to the colour coding, a simple quantitative analysis was executed resulting in the colour code frequencies and compliance proportions for the final sample size of 37 repositories. The level of compliance declines according to the traffic-light rating system, the blue colour indicates that there was no information provided to answer the facet correctly. On the basis of the scoring result, every repository had been contacted and invited to participate and contribute further information. Based on the feedback, it is possible to shape recommendation on how the participating repositories can improve their infrastructure to adhere to the demands in the FAIR principles, and also to see which of the FAIR principles are the easiest to adhere to, and which are more problematic.

The analysis is divided in two sub-chapters, that concern themselves with the statistical appraisal and interpretation, followed by the subject specific or repository specific differences observed, and closing with recommendations for repositories to improve their information architecture. The related spreadsheet including all tables, categories, statistics and graphs, accompanied by the documentation can be accessed via the 4TU.Centre for Research Data (Dunning et al., 2017).

Elements of the FAIR Data Principles

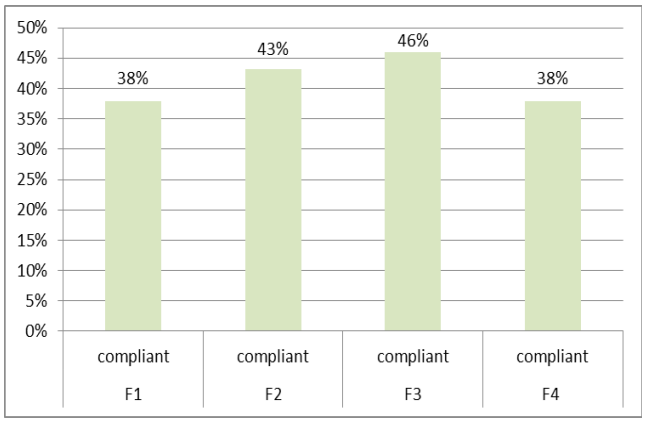

Figure 1. Compliance proportion for every facet assigned a FINDABILITY principle.

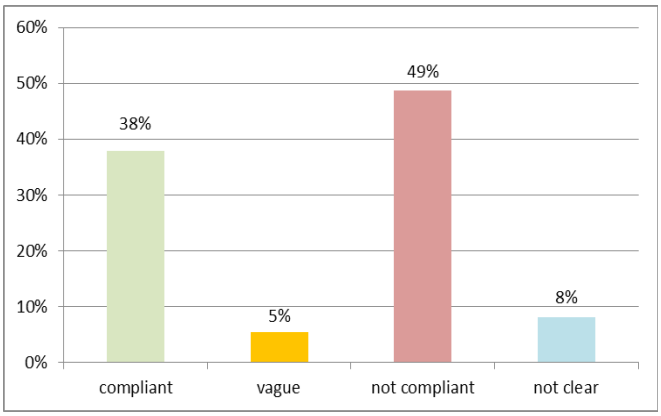

Figure 2. Compliance rate for F1 (meta)data are globally unique and eternally persistent identifier.

Our analysis began with the first FAIR principle – FINDABLE. Facets F1 ((meta)data are assigned a globally unique and eternally persistent identifier) and F4 (metadata specify the data identifier) look reasonably simple to execute. However, for both, compliance rates were only around 40%; and nearly half of repositories were providing persistent identifiers.

We took a persistent identifier to be a DOI or similar system for assigning identifiers. Those repositories using their own minted URLs failed, unless they had specific policies declaring the long-term existence of those URLs. The compliance levels for F1 are in Figure 2. Nearly half of the sample group does not assign a HANDLE[7], DOI[8], or URN[9], or made them visible. A small proportion uses their own system specific identifiers. Recognised data archives, making use of DOIs, tended to have good compliance rates. The number is much lower for subject-based repositories where URL structures based on project names were used to identify the location of data (e.g. EDGAR, the Emissions Database for Global Atmospheric Research or ICTWSS: Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts).

Facets F2 (data are described with rich metadata) and F3 ((meta)data are registered or indexed in a searchable resource): F2 is vague in its description (how does one decide what makes metadata rich?), but we interpreted a data repository as being compliant if its datasets tended to be accompanied by a variety of attributes - not just title, creator and date, but additional information on contributors, keywords and temporal and spatial coverage that can help its findability (as opposed to R1 which looks at metadata in the context of re-use). In order to determine F3 we used Google as a proxy to determine whether metadata had been indexed. In case of doubt the alternative search engine DuckDuckGo[10] was used to double check the findability of (meta)data. To confirm F3 a data-set title in quotation marks was entered in the search field of Google or DuckDuckGo to enable the specific search of only these words. Datasets of nearly half of the repositories can be searched and found via one of the two online search engines.

As with F1 and F4, there was surprisingly low compliance figure given how essential metadata is for finding research data on the web. Around 40-45% of repositories were compliant for both F2 and F3. For some repositories, the quantity of

7 Handle System Overview: https://www.ietf.org/rfc/rfc3650.txt
8 Digital Object Identifier System: https://www.doi.org/
9 Uniform Resource Names Name Space Definition Mechanisms: https://www.ietf.org/rfc/rfc3406.txt
10 DuckDuckGo Search Engine: https://duckduckgo.com/

metadata could vary with different datasets, meaning that a quarter of repositories partially complied with F2. That is primarily due to the different community demands on metadata in the dataset and purposes of the repository: the motivation to provide a reusable dataset is different than just to publish and share data in an appropriate repository.

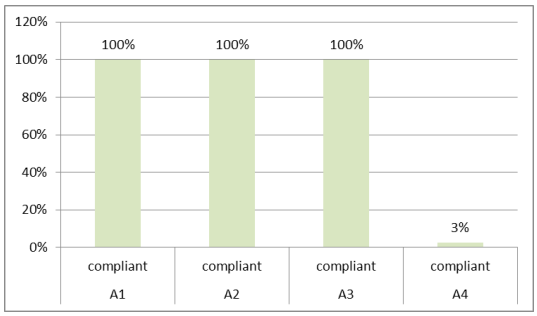

Figure 3. Compliance proportion for every facet of the ACCESSIBILITY principle.

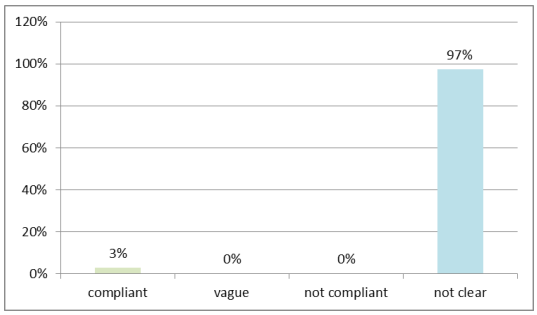

Figure 4. Compliance rate for A4 metadata are even when the data are no longer available.

The second FAIR principle contains four facets related to ACCESSIBILITY (Figure 3). The 100% scores for A1, A2 and A3 can be explained by the fact that all visited repositories are using (at least) the HTTP communication protocol for delivering data on the World Wide Web. The scoring was perhaps generous - some repositories made partial metadata available via HTTP but the data itself was available via email. HTTP provides a standardized way for computers to communicate with each other, and by implication, share information about research data in the ways suggested by the FAIR principles.

However, facet A4 (metadata are accessible, even when the data are no longer available) is hardly met at all. To be compliant requires a clear policy statement (or various examples of data this has actually happened to) indicating that metadata is still available even if the data is removed. As the repositories do not seem to provide information or examples that shows if and how they comply with this particular facet. Figure 4 shows that only 3% of repositories are compliant (i.e. only one repository from the sample!). Even well-established data archives do not publish this information in their data or preservation policies. However, this number could easily be changed; the publication of even a short policy on how data is treated if deleted / removed would be enough to meet this FAIR guidelines.

Despite this, the facet itself seems rather odd in comparison to the other facets - facets A1, A2 and A3 require a technical implementation and are easy to judge. A4, however, is dependent on specific policies being created (and published).

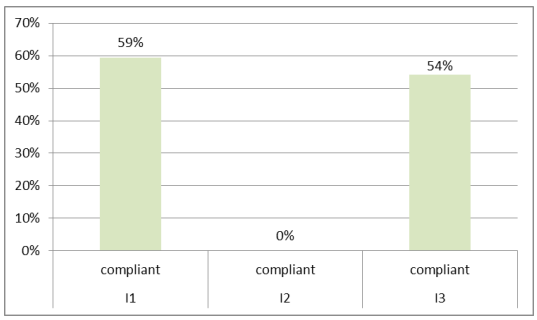

Figure 5. Compliance proportion for every facet of the INTEROPERABILITY principle.

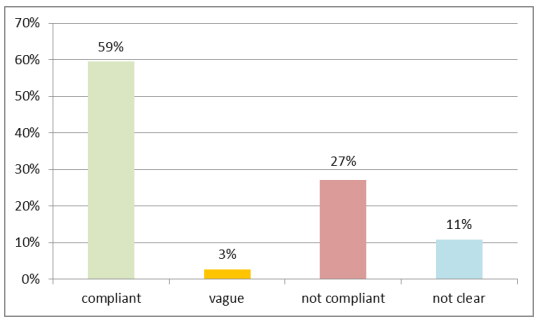

Figure 6. Compliance rate for I1 knowledge for (meta)data.

Analysis of the third principle - INTEROPERABLE - shows that two of the three facets are well implemented (Figure 5). For I1 ((meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation), More than half of the repositories represent their metadata in a structured, sometimes standardized way, that is easily comprehensible to the end user. Many of the compliant archives seem to be based on the core attributes of Dublin Core (e.g. including title, creator, date etc.). The results are demonstrated in Figure 6.

There tended to be very specific reasons for repositories not complying with Facet I1. Those applying data access restrictions, or those that used their own interface to generate dynamic data-sets, tended not to have (visible) structured metadata.

Facet I2 ((meta)data use vocabularies that follow FAIR principles) is very challenging, both to assess and to meet. As many repositories do not fully use FAIR principles themselves, it seems unlikely that many will make use of external vocabularies that use FAIR. It might be feasible for some metadata attributes, but certainly not all. Additionally, how is vocabulary defined? Could an external service such as ORCID be a vocabulary? Or does the principle simply mean vocabularies that define terminologies? Because of this, we considered that 0% of the repositories comply according to the external view on the information provided in the web interfaces.

Regarding Facet I3 ((meta)data include qualified references to other (meta)data), we interpreted this as the metadata providing additional links to related publications. Over half of the repositories achieved this. Later conversation with the FAIR team indicated

that ‘qualified references’ meant links to information that would help disambiguate terms used (e.g. a link to a source that could identify the difference between Paris, France and Paris, Texas).

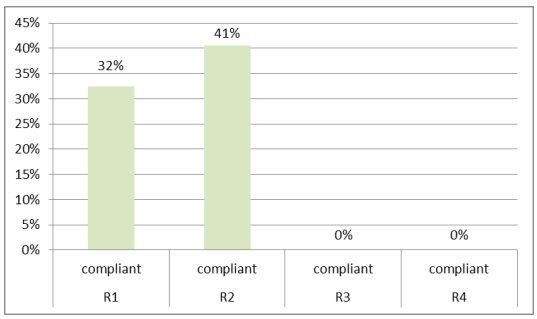

Figure 7. Compliance proportion for every facet of the accurate and relevant attributes.

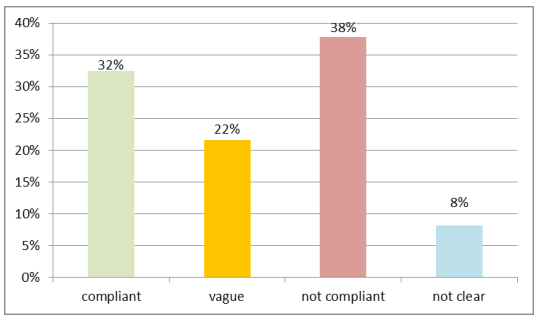

Figure 8. Compliance rate for R1 meta(data) have a plurality of REUSABILITY principle.

The last principle - RE-USABLE - seems the most difficult to meet, as Figure 7 indicates. A third of repositories comply with R1 (meta(data) have a plurality of accurate and relevant attributes). But, as Figure 8 demonstrates, nearly one third of the repositories (30%) provide vague or unclear metadata attributes. However, the largest number of repositories (38%) were still sorely lacking.

Analysis of this principle is further confused by the seeming similarity between R1 and F2 (data are described with rich metadata). We took F2 to define descriptive metadata that could help with findability, and R1 to be about the aspects of metadata that help one to evaluate how reusable a dataset is (i.e. once a dataset has been found) Interestingly both facets have the exact same amount of vague (22%) and unclear (8%) scoring, but that isn't necessarily based on the same repositories.

Facet R2 ((meta)data are released with a clear and accessible data usage license) is largely easy to assess. The results indicate that there is still a fair way to go before all repositories clearly indicate how data may or may not be used. Only 41% of the sample had a clear licence, and a similar percentage having no clear licence. The category of repositories defined as being Research Infrastructures (including Europeana, SeaDataNet, Zenodo and EUDAT-B2Share) is the only category where all repositories state clear data usage licenses.

Facet R3 ([meta]data are associated with their provenance) was difficult to determine. The term ‘provenance’ means the origin or creator information; the explanation in the FORCE 11 guidance talks about “provenance of the Data Elements to

their original Data Object and subsequently to the underlying resources" (FORCE 11, 2014b). That explanation led us to the decision to not just be satisfied if the creator name or information about involved institution, but added documentation on how the data was created. In the light of this, we are currently reviewing the rating given to each repository.

Finally, facet R4: whereas other facets are clearly visible and brief investigation, R4 requires detailed subject knowledge to know whether (meta)data meet domain-relevant community standards. As authors we could have taken a guess, but it would have remained that - a guess. As is highlighted below, many repositories seem to meet domain-specific practices for the general sharing of data, but not necessarily for domain-specific metadata. Without added input from community experts - the authors decided to place award every repository in the sample an 'unclear' rating.

Subject Specific and Repository Specific Differences

The FAIR guidelines are a fairly recent invention, published in 2014. Many of the data repositories have histories longer than that, and draw on discipline-based practices that have well established protocols for how data should be shared.

It was notable that datasets in the social sciences suffered particular. There was only compliance with around a quarter of the FAIR principles were found to be compliant for the seven repositories in our sample that are part of the social sciences, see Figure 9.

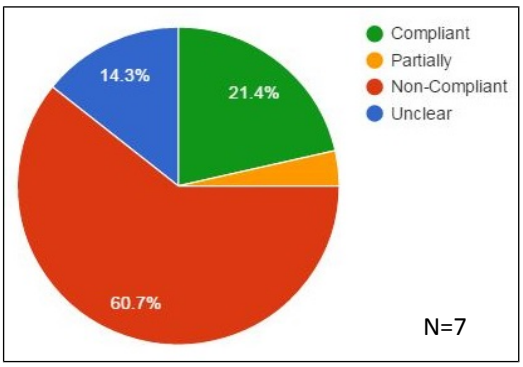

Figure 9. Compliance of Social Science data repositories to FAIR Principles.

For example, the Longitudinal Aging Study Amsterdam contains a rich seam of data relating to "physical, emotional, cognitive, and social functioning in late life." Running since 1992, it has a healthy publication list associated with it and data continues to be gathered.

Yet because of the importance of data protection, nearly all the data is available only on request, and not made available on the LASA website. Therefore, the site fails many of the FAIR principles. Even the more straightforward Findable principles were not passed, e.g. no rich metadata description, no global identifier.

But while no structured metadata is made available online about the LASA datasets there is free-text documentation about the datasets. Users can find out using what methodologies the data is collected and there tables indicate in what year data was collected. For the user who has a methodological background in the social sciences, it should not be too difficult to reach and access the data.

Therefore, there is an argument to be made that domain-relevant community standards are being met. It should emphasise that the importance of non-compliance should not be over emphasised. Data from this research project is not being hidden without adequate justification. If one sees the FAIR guidelines as a means to extend and refine how data is shared, the LASA website is doing well.

However, this paper argues that this is not enough. The methods used in the social sciences repositories means that the metadata are not machine readable. Sharing the data, or even finding the data via a search engine is precluded by the limited technical infrastructure in use.

There is no crucial reason why more of the FAIR guidelines cannot be met. Findable metadata is feasible; making the datasets available via http (behind authentication of course), would also be possible. Likewise, creating metadata that is interoperable for each dataset can be achieved, and can be achieved quite easily.

The story with the TRAILS (Tracking Adolescents' Individual Lives Survey) database appears to be much the same. A pdf document gives the extent of the data collected each year and the different variables. Thus the user has a basic indication of what data is available, but each dataset comes without structured metadata. And to actually access the data, a bespoke licence form needs to be filled out and then this is shared with the organisers of the site. Again, this manual approach to data sharing means that the data did not pass the FAIR principles.

However, there is an interesting coda to the TRAILS data. Further investigation reveals that the same datasets had been deposited in the DANS data archive. Here, they appear with persistent identifiers, structured metadata etc. So although the TRAILS repository of data did not comply with many of the FAIR guidelines the individual datasets could have a much higher FAIR compliance score, thanks to their simultaneous publication in a data archive with more mature policies for data curation.

Despite being not made very explicit on the TRAILS website, this kind of arrangement could provide a model for very focussed research projects to work in the future. The researcher focussed website provides the interface, but the underlying data and metadata is created and imported from (or linked to) a FAIR-compliant archive.

As with personal data in the social sciences, climate data is another field of study with an established tradition of sharing data that predates the FAIR principles. Here though, data protection is not the crucial issue, rather it is the sheer quantity of data-sets that are made available for dissemination. And again disciplinary norms have developed specific ways of making data accessible that do not always tally with the FAIR principles.

The Southeast Asian Climate Assessment and Dataset (SACA) contains series of daily observations at meteorological stations throughout Southeast Asia, with details on temperature, precipitation, air pressure. Datasets are sometimes dynamically created from queries from the user, or they are simply presented as a grid of links. This plurality of datasets means that there is not structured metadata associated with each dataset. As with the social science examples, there is free-text documentation that explains the collection of the data, and helps with understanding the codes, acronyms it uses. Documentation is also embedded in the dynamically created datasets. Licensing conditions are mentioned, but not until documentation is downloaded.

This lack of structured data again had a significant impact on the FAIR principles. While the SACA climate data was largely ACCESSIBLE (in making use of http to share), it was not FINDABLE and missed many of the qualities that would make it INTEROPERABLE or RE-USABLE.

Much the same is true of the WorldClim website, that makes available 'free climate data for ecological modelling and GIS'. Again, there are a grids of data that can be downloaded (although this time with a clear Creative Commons licence.) There is no metadata, but a documentation page provides details on the data's creation and interpretation. As with the social sciences datasets, it seems likely that this dataset is meeting discipline specific norms - after all, the data is easily accessible once users have found it, it comes with rich internal documentation, and a clear licence. But it lacks many of the attributes that would help with the optimal sharing of the data via the Internet.

Conclusion

The Fair Data Principles Themselves

The 15 facets of the FAIR principles are all short sentences. Their brevity gives the impression that they are all items that can be checked off. However, our analysis shows that the FAIR principles are much trickier than this. Some facets appear to overlap (e.g. the plurality of attributes in R1 and rich metadata in F2). Some are vague (e.g. the qualified references of I3), others are open ended (the recursive request of I2 that '(meta)data use vocabularies that follow FAIR principles'), while others require interpretation from external parties (e.g. the domain relevant community standards of R4). Some appear to be technical in scope (A1, A2 and A3, for example) whereas others are more policy driven (the policy on the retention of metadata in A4).

When it comes to working with the guidelines, researchers, data librarians, funders and other stakeholders must acknowledge this variation. Compliance should not be seen as a stick, but rather a desirable goal, with the recognition that some of the guidelines are open to interpretation and debate. Indeed, the term compliance, at least for some of the facets, is misguided. Rather the facets provide targets that will help with getting recognition and reward for the publication of data. Given that the EU is now including FAIR guidance as part of the its H2020 programme, it is important that funders and peer reviewers take heed of this, so the principles do not get misused as sanctions.

Implementing the Fair Data Principles

Our analysis reflects the difficulties in interpreting the FAIR guidelines, and also putting them into practice. For many facets, less than half the sampled repositories were compliant. The Interoperable and Re-usable facets were, in particular, the most difficult to adhere to.

But for many of the repositories sampled, implementing basic policies can help achieve compliance. If a repository implements policy and practice in the four following areas...

- creating a lasting policy for deploying PIDs
- insisting on a minimum set of metadata, ideally coupled with the preferred used of semantic terms
- having a clear licence

- using HTTPS

…then are well on the road to achieve working in accordance with the FAIR principles. The principles also demand that repositories are transparent about the implication of such policies.

However, this is more than simply policy or technical implementation. There is also a social element to this. In our analysis data archives fared better than the subject-based repositories. Established data archives were much more likely to have put policies related to the above four areas into place. Subject-based repositories, which had more of a focus on the sharing of data but with less concern for the long-term archiving, were less likely to have adopted such policies. In many cases, as our conversations with repository managers indicated, this was simply because they did not have the time and resources to do such a thing; they did appreciate the importance. Therefore following FAIR principles is access to the time, money and skills do implement the necessary policies

Our analysis leads us to the three following conclusions:

1. The FAIR principles are not just about compliance. Some of their facets need to be seen as being open-ended guidelines that can be interpreted in different ways; and varying interpretations can all be within the spirit of the original guidelines.

2. Implementing some basic policies (and publishing details of these policies) on identifiers, metadata, licensing and protocol will help all repositories align with the FAIR principles.

3. And finally closer alliances between data archives and researchers building subject-based repositories should be sought. Archives can be bring the policy and long-term expertise, whereas researchers understand tools and their domains. Satisfying the FAIR principles requires both sets of skills to be brought together.

Acknowledgements

We would like to thank all participants for their feedback and insights on how to perceive the different facets according to the specific field of research.

References

DataCite. (2017). *DataCite - Mission*. Datacite.org. Retrieved from https://www.datacite.org/mission.html

Dunning, A.C., de Smaele, M.M.E., Böhmer, J.K. (2017). Evaluation of data repositories based on the FAIR Principles for IDCC 2017 practice paper. TU Delft. Dataset. doi:10.4121/uuid:5146dd06-98e4-426c-9ae5-dc8fa65c549f

European Commission. (2017a). *What is Horizon 2020?* European Commission - Horizon 2020. Retrieved from https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020#Article

European Commission. (2017b). *History of Horizon 2020.* European Commission - Horizon 2020. Retrieved from https://ec.europa.eu/programmes/horizon2020/en/history-horizon-2020

European Commission. (2016). *Guidelines on FAIR data management in Horizon 2020 (3rd ed.)*. European Commission - Directorate-General for Research and Innovation. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

FORCE 11. (2014a). *The FAIR data principles*. FORCE11. Retrieved from https://www.force11.org/group/fairgroup/fairprinciples

FORCE 11. (2014b). *Guiding Principles for findable, accessible, interoperable and re-usable data publishing version b1.0. FORCE11*. Retrieved from https://www.force11.org/fairprinciples

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Nature.* Retrieved from http://www.nature.com/articles/sdata201618#author-information

Appendix

List of FAIR Principles and Corresponding Facets According to FORCE 11

FINDABLE

- **F1** (meta)data are assigned a globally unique and eternally persistent identifier
- **F2** data are described with rich metadata
- **F3** (meta)data are registered or indexed in a searchable resource
- **F4** metadata specify the data identifier

ACCESSIBLE

- **A1** (meta)data are retrievable by their identifier using a standardized communications protocol
- **A2** the protocol is open, free, and universally implementable
- **A3** the protocol allows for an authentication and authorization procedure, where necessary
- **A4** metadata are accessible, even when the data are no longer available

INTEROPERABLE

- **I1** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*
- **I2** (meta)data use vocabularies that follow FAIR principles
- **I3** (meta)data include qualified references to other (meta)data

RE-USABLE

- **R1** meta(data) have a plurality of accurate and relevant attributes
- **R2** (meta)data are released with a clear and accessible data usage license
- **R3** (meta)data are associated with their provenance
- **R4** (meta)data meet domain-relevant community standards

List of Repositories in the Sample Size of 37

Table 1. Reviewed Data Repositories

Name of the Data Repository	Data Repository URL in January 2017
DANS-EASY	https://easy.dans.knaw.nl/ui/home
EUDAT-B2Share	https://b2share.eudat.eu/
Zenodo	https://zenodo.org
PseudoBase	http://www.ekevanbatenburg.nl/PKBASE/PKB.HTML
OpenML	http://www.openml.org/
Profiles-Registry	http://www.profilesregistry.nl/
Mendeley-Data	https://data.mendeley.com/
4TU.Centre for Research Data	http://data.4tu.nl/
CancerData.org	https://www.cancerdata.org
DHS Data Access	http://www.dhsdata.nl
WorldClim	http://worldclim.org/
World Data Centre for Soil	http://www.isric.org/
Infrared Space Observatory	http://www.cosmos.esa.int/web/iso/access-the-archive
Longitudinal Aging Study Amsterdam	http://www.lasa-vu.nl/index.htm
Southeast Asian Climate Assessment & Dataset	http://saca-bmkg.knmi.nl/
TRAILS	https://www.trails.nl/
ICOS Carbon Portal	https://www.icos-cp.eu/node/1
CESSDA	http://cessda.net/

Name of the Data Repository	Data Repository URL in January 2017
DANS-EASY	https://easy.dans.knaw.nl/ui/home
SeaDataNet	http://www.seadatanet.org/
LISS	https://www.lissdata.nl/lissdata/
ORGIDS / RodRep	http://www.orgids.com/ / http://www.rodrep.com/
eartH20bserve	http://www.earth2observe.eu/
EDGAR	http://edgar.jrc.ec.europa.eu/
KNMI	https://data.knmi.nl/datasets
STITCH	http://stitch.embl.de/
ECA&D	http://www.ecad.eu/
Europeana	http://www.europeana.eu/portal/en
MycoBank	http://www.mycobank.org/
AlgaeBase	http://www.algaebase.org/
Amsterdam Cohort Studies	https://www.amsterdamcohortstudies.org/acsc/index.asp
ICTWSS	http://uva-aias.net/en/ictwss
Share ERIC	http://www.share-project.org/
LOVD3	http://databases.lovd.nl/whole_genome/genes
CARIBIC	http://www.caribic-atmospheric.com/
EIDA	http://www.orfeus-eu.org/data/eida/
Sound and Vision	http://www.beeldengeluid.nl/en
Figshare	https://figshare.com/

Compliance Proportion for each FAIR Facet for 37 Repositories

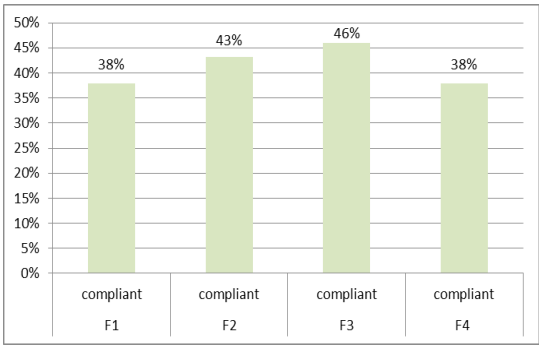

Figure 10. Findable.

 doi:10.2218/ijdc.v12i2.567

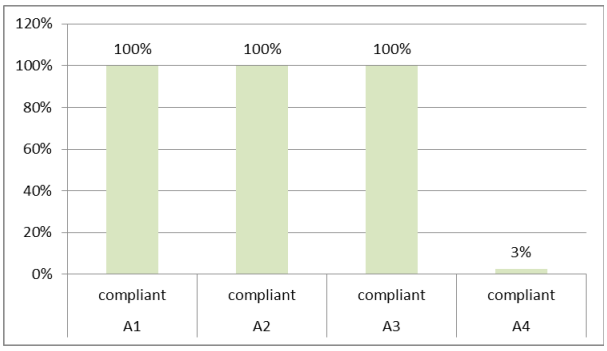

Figure 11. Accessible.

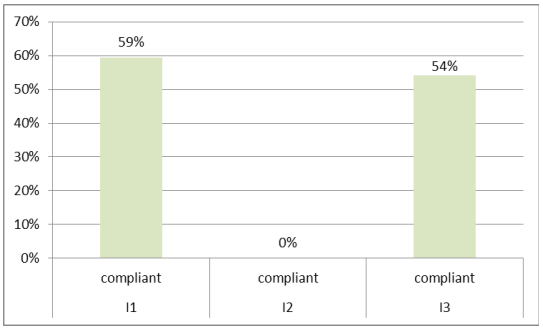

Figure 12. Interoperable.

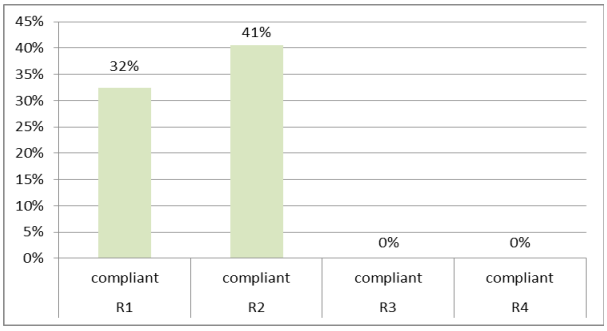

Figure 13. Reusable.

Compliance Proportion to the FAIR Principles for 37 Repositories

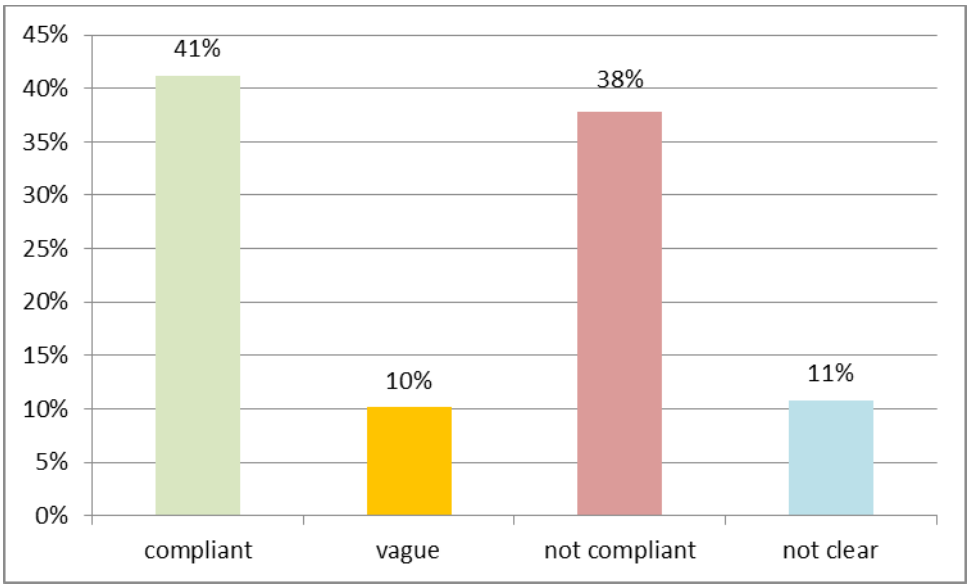

Figure 14. Findable.

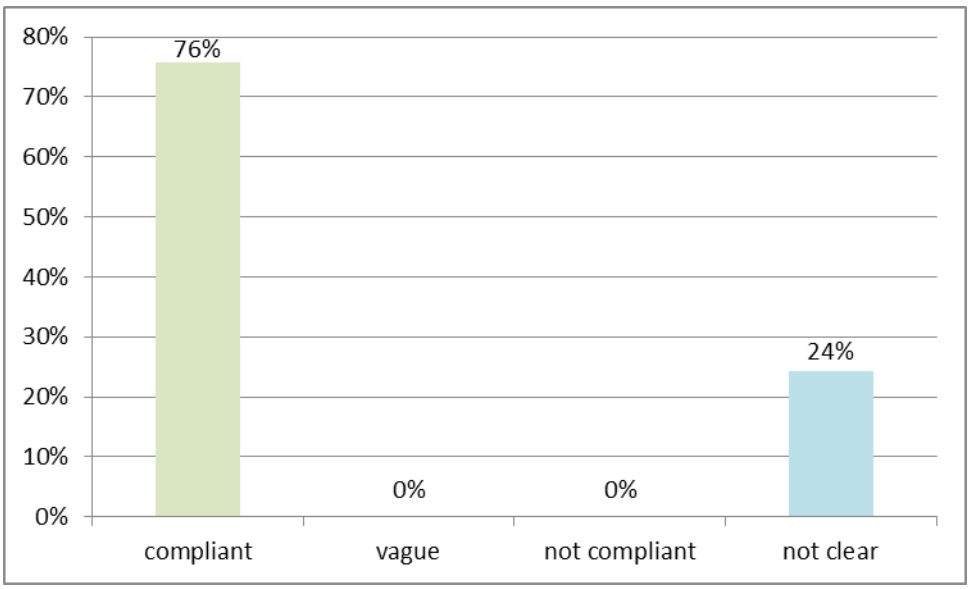

Figure 15. Accessible.

 doi:10.2218/ijdc.v12i2.567

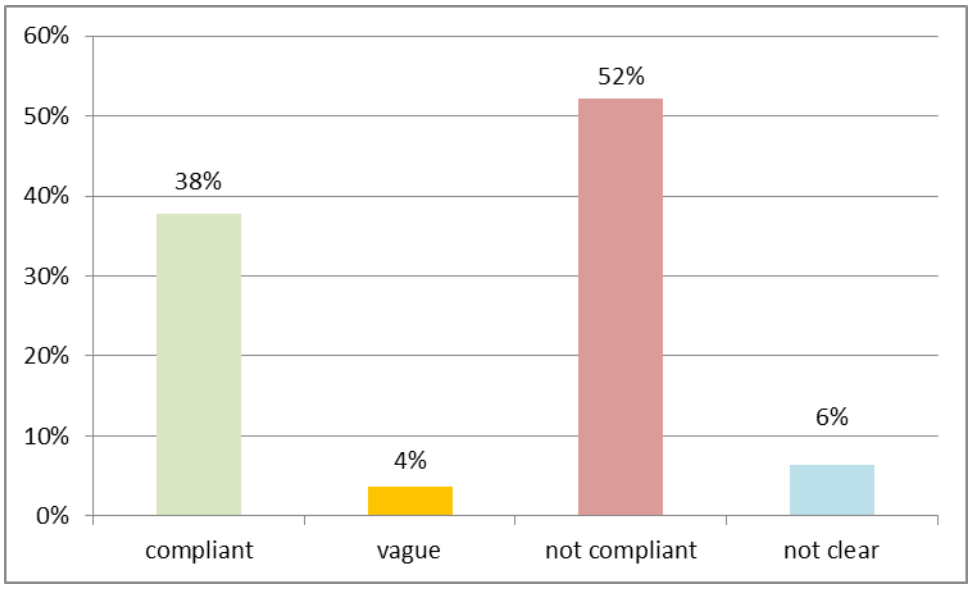

Figure 16. Interoperable.

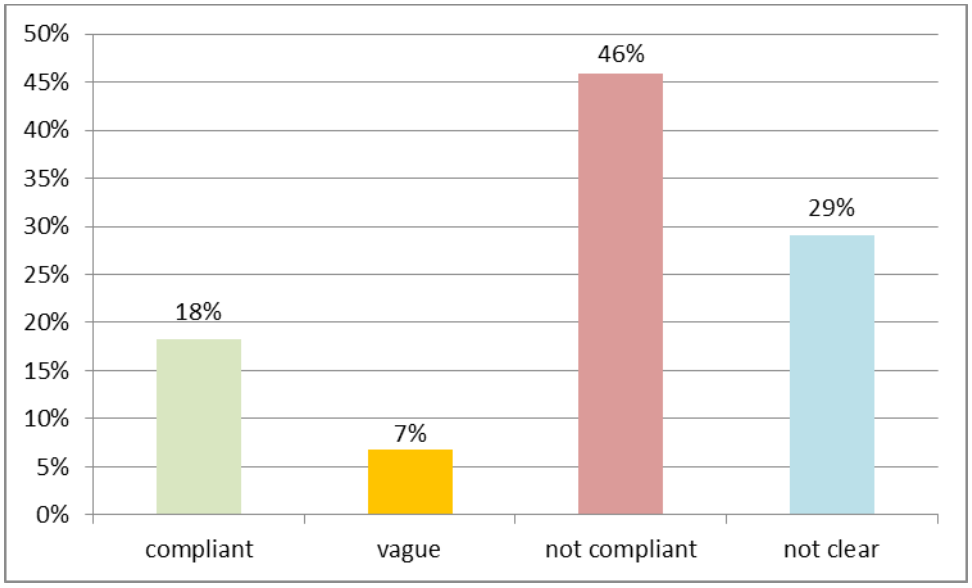

Figure 17. Re-usable.