

Realizing FAIR Data in an Enterprise Environment

Melanie Hackl¹ and Gökhan Coskun^{1,2}

¹ Bayer AG, Berlin, Germany

² Beuth Hochschule für Technik, Berlin, Germany

FAIR Data for Enterprises

Timely reaction to changing environments like markets and regulations is a severe challenge for globally acting enterprises. Therefore, fast informed decisions are essential for the success and the sustainability. Relevant data needs to be available in the right place at the right time. Exactly this is addressed by the FAIR Data principles by demanding data to be findable, accessible, interoperable and reusable data. However, since its first publication in 2016 [FOR18], it remains an open question, how FAIR Data can be realized.

We are convinced that the set of Linked Data [Tim06] principles is the most promising solution to realize FAIR. Firstly, globally unique identifiers as described in the **Findability** aspect are realized in Linked Data by the consequent adoption of the Uniform Resource Identifier (URI) standard. Secondly, universal **Accessibility** through standard communication protocols is realized by making use of http. Thirdly, **Interoperability** is supported by focusing on leveraging RDF and SPARQL, which are developed for a distributed data space. And finally, as it is emphasized by its name, linking to existing datasets is the most crucial part of Linked Data, which addresses the **Reusability** aspect of FAIR Data.

Required Service in the Enterprise IT Landscape

In the controlled environment of an enterprise, IT systems need to provide an accurate level of service quality and reliability. The four Linked Data principles need to be accomplished by the following five additional aspects to realized FAIR in an enterprise completely.

Persistent Identifier: In order to assign URI-based identifiers to enterprise assets, an appropriate registration service is required. Public domain name systems are not suitable, since enterprise IT solutions need to be independent from outside systems, as much as possible. Furthermore a policy has to exist, which explains what happens if the identifier gets deprecated.

Metadata Service: Due to accountability and responsibility aspects, metadata about existing datasets are crucial. Therefore, a metadata service has to be provided, which stores the metadata separately from the data and is available, even if the dataset itself does not exist anymore.

Metadata Ontology: Metadata is core to FAIR data. To achieve this with the Linked Data principles, a metadata ontology is required. This must include properties like: Security classification, identifier of the resource, information about the creator and about the provenance.

Search Engine: To finally make the (meta)data findable, a search engine is required. In case all of this services are provided, a FAIRness degree of 100% is achieved.

Internal Licensing and Certification Process: The certification should be issued by an recognized body and expresses whether a resource is compliant. A company would have to set up a process for the licences and certifications.

Evaluation and Conclusion

In order to evaluate to what extend Linked Data and the additional services as described above are supporting the realization of the FAIR Data principles we utilized the FAIR Data metrics [Mar18]. The following table illustrates the comparisson of a dataset stored in a sheet, an RDF dataset and a dataset stored in a triplestore supported with the above mentioned services fulfills the FAIR Data metrics.

No. principle	Metric Name	Excel Tables	RDF	With additional Services
F1A	Identifier Uniqueness	Red	Red	Green
F1B	Identifier persistence	Red	Red	Green
F2	Machine-readability of metadata	Red	Red	Green
F3	metadata clearly and explicitly include the identifier of the data it describes	Red	Red	Green
F4	(meta)data are registered or indexed in a searchable resource	Red	Red	Green
A1.1	the protocol is open, free, and universally implementable	Red	Green	Green
A1.2	the protocol allows for an authentication and authorization procedure, where necessary	Green	Green	Green
A2	metadata are accessible, even when the data are no longer available	Red	Red	Green
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	Red	Green	Green
I2	(meta)data use vocabularies that follow FAIR principles	Red	Yellow	Green
I3	(meta)data include qualified references to other (meta)data	Red	Green	Green
R1.1	(meta)data are released with a clear and accessible data usage license	Red	Red	Green
R1.2	(meta)data are associated with detailed provenance	Red	Red	Green
R1.3	(meta)data meet domain-relevant community standards	Red	Red	Green

Fig. 1. Evaluation of the FAIRness degree with the FAIR Data Metrics

This evaluation demonstrates clearly that every aspect of the FAIR Data metrics are fulfilled. However, the metrics themselves are still work in progress. Therefore, this evaluation depends on the acceptance of these metrics. It might worth repeating the evaluation, as soon as the FAIR Data metrics are finalized.

References

- [AH11] Dean Allemang and Jim Hendler. *Semantic Web for the working Ontologist*. Morgan Kaufmann, USA, 2. edition, 2011.
- [AIe09] Sicilia MA, Astrova I. and Lytras M.D. (eds). *Metadata and Semantics*. Springer, Boston, MA, Boston, 1. edition, 2009.
- [BEH15] Bernhard Humm BÄ¶rteÄ§in Ege and Anatol Reibold (Hr). *Corporate Semantic Web*. Springer Vieweg, Heidelberg, 1. edition, 2015.
- [BHB09] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1â22, 2009.
- [BLHL01] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.
- [CHLR⁺16] Goekhan Coskun, Ralf Heese, Markus Luczak-RÄ¶sch, Radoslaw Oldakowski, Ralph SchÄfermeier, and Olga Streibel. Towards corporate semantic web: Requirements and use cases. *Technical report, Freie Universitat Berlin*, 2016.
- [FOR18] FORCE11. <https://www.force11.org/group/fairgroup/fairprinciples>, 2018. Accessed: 2018-06-01.
- [FÄ16] Christian FÄrber. *Data Quality Management with Semantic Technologies*. Springer Gabler, Wiesbaden, 1. edition, 2016.
- [GGHC⁺16] I. Grangel-GonzÄlez, L. Halilaj, G. Coskun, S. Auer, D. Collarana, and M. Hoffmeister. Towards a semantic administrative shell for industry 4.0 components. *10.1109/ICSC.2016.58*, 2016.
- [GOF18] GOFAIR. <https://www.go-fair.org/fair-principles/>, 2018. Accessed: 2019-09-12.
- [Kel13] Thomas Kelly. How semantic technology drives agile business. *Cognizant*, 2013.
- [LSH⁺09] Atif Latif, Anwar Us Saeed, Partick Hoefler, Alexander Stocker, and Claudia Wagner. The linked data value chain: A lightweight model for business engineers. In Adrian Paschke, Hans Weigand, Wernher Behrendt, Klaus Tochtermann, and Tassilo Pellegrini, editors, *Proceedings of I-Semantics 2009. 5th International Conference on Semantic Systems*, pages 568–577. Journal of Universal Computer Science, 2009.
- [Mar18] Mark D. Wilkinson, Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos and Michel Dumontier. <https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen1/ALL.pdf>, 2018. Accessed: 2019-09-12.
- [MNV⁺17] Barend Mons, Cameron Neylon, Jan Velterop, Luiz Olavo Bonino da Silva Santos Michel Dumontier, and Mark D. Wilkinson. Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud. *Information Services Use, vol. 37, no. 1, pp. 49-56, 2017*, 2017.
- [SBF98] Rudi Studer, V. Richard Benjamins, and Dieter Fensel. Knowledge engineering: Principles and methods, 1998.
- [SS09] Steffen Staab and Rudi Studer. *Handbook on Ontologies*. SpringerVerlag Berlin Heidelberg, Berlin, 2. edition, 2009.
- [Tim06] Tim Berners-Lee. <https://www.w3.org/DesignIssues/LinkedData.html>, 2006. Accessed: 2019-09-12.
- [W3C18] W3C. <https://www.w3.org/2001/sw/RDFCore/Schema/200203/>, 2018. Accessed: 2019-09-12.

- [Wil16] M. D. et al. Wilkinson. The fair guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18, 2016.
- [Yu11] Liyang Yu. *A Developer's Guide to the Semantic Web*. Springer, Heidelberg, 1. edition, 2011.