

Sphere 7 Data: LOUD and FAIR Data for the Research Community

Florian Thiery^a

^arse@fthiery.de - ORCID: 0000-0002-3246-3531
Research Software Engineer, Römisch-Germanisches Zentralmuseum, Mainz, Germany

Abstract

The world of data modelling, publishing and sharing has changed rapidly in the last years. Starting from the invention of the World Wide Web in 1989 by Sir Tim Berners-Lee this Web is on an evolution to the so called Web 4.0. Before we will reach that we have to fill the Semantic Web as part of the Web 3.0 with resources and research content. Therefore the 5 star open data principles, the Linked Open Data principles, the 5 star LOUD principles and the FAIR principles are necessary. I will merge all of that principles in the so called **Sphere 7 Data principles**, described in this paper.

Key words: Linked Data; Linked Open Data; Five Star Data; FAIR Data; Semantic Web.

1 Introduction

In some old maps (Fig. 1) of the ancient world the phrase '**Hic sunt dracones**' (engl. *Here be Dragons*)^[1] is used to describe an area, that is unknown to the map creator. As Ulrike Wuttke¹ mentioned, this is quite appropriate to summarise the ambivalence of humanists towards data and these 'fancy' new concepts^[2] in this digital new thing, called the World Wide Web and the Internet.



Figure 1. Nature and Science, <https://t1p.de/wmtz>

The WWW give researchers a new possibility to share there special research data with other (non subject-specific) communities, link them to other available data on the internet and enable the creation of new unknown

knowledge. Maybe a good technique to fight against the data dragons (Fig. 2) in research!



Figure 2. Fight against the data dragons

Here the techniques of the Semantic Web^[3] as part of the Web 3.0 comes into play. Using the idea of Open Data in mind, combined with RDF, several semantically described links, usable interfaces and applications for data, as well as FAIR data principles, will be possible to create a huge research data cloud (Fig. 3).

The aim should be the creation of a huge data graph (Fig. 4) as part of the *Linked Open Data Cloud*² or *Giant Global Graph* (GGG) using the principles described in

¹ <https://twitter.com/uwuttke>

² <http://lod-cloud.net>



Figure 3. Linked Data Cloud Scheme, Florian Thiery [CC BY 4.0]

the next sections: (1) the five stars of Open Data, (2) the four Linked Data principles, (3) the five stars of Linked Open Usable Data (LOUD) and (4) the 15 FAIR data principles to end up in a model, what I will call the **Sphere 7 Data** principles.

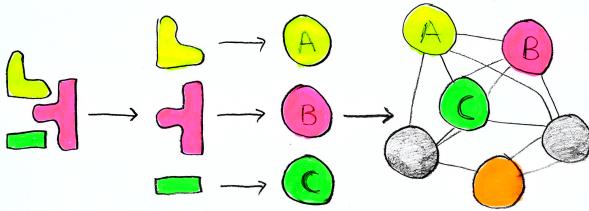


Figure 4. Graph From Data Pieces, Florian Thiery [CC BY 4.0]

2 Linked Data Principles

Referring to Tim Berners-Lee³ »*The Web does not just connect machines, it connects people.*« it is the same with data and research data:

The Web does not just connect machines, it connects data and thus knowledge of people!

But to do this, some technique and 'standards' had to be build. In his first publication on Linked Data, Tim Berners-Lee wrote:

»*The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.*«[4]

»*Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike*

³ https://twitter.com/timberners_lee

the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF, the URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow.«[4]

Tim Berners-Lee ended up in defining »*imperfect but published*«[4] four simple rules for Linked Data (on 2006-07-27), the so called Linked Data principles[4], illustrated in Fig. 5:

- (1) Use URIs as names for things
- (2) Use HTTP URIs so that people can look up those names.
- (3) When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
- (4) Include links to other URIs. so that they can discover more things.

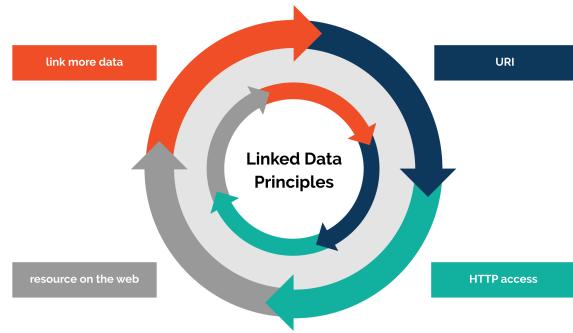


Figure 5. Linked Data Principles, Florian Thiery [CC BY 4.0]

3 Five stars of Open Data

In an ideal world, research data and all other kind of data are published and available for everyone, from every place on this planet. To reach this, data must be published OPEN.

The idea and most important facts of Open Data⁴ can be summarised⁵ as follows: (1) availability and access, (2) re-use and redistribution and (3) universal participation. So it is all about interoperability!

This idea in mind, Sir Tim Berners-Lee started to ask in 2010: 'Is your Linked Open Data 5 star?':

»*This year, in order to encourage people – especially government data owners – along the road to good linked data, I have developed[sic!] [a] star rating system. [...]*

⁴ <https://opendefinition.org/od/2.1/en/>

⁵ <http://opendatahandbook.org/guide/en/what-is-open-data/>

Linked Open Data (LOD) is Linked Data which is released under an open licence, which does not impede its reuse for free. [...]. Linked Data does not of course in general have to be open – there is a lot of important use of linked[sic!] data internally, and for personal and group-wide data. You can have 5-star Linked Data without it being open. However, if it claims to be Linked Open Data then it does have to be open, to get any star at all.«[4]

The five star rating system for 5 Star Open Data is shown in Fig. 6. The five star data website^[5] introduces in the principles as follows:

- ★ make your stuff available on the Web (whatever format) under an open license
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)
- ★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)
- ★★★★ use URIs to denote things, so that people can point at your stuff
- ★★★★★ link your data to other data to provide context

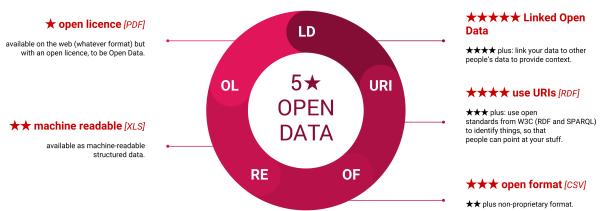


Figure 6. 5 Star Open Data, Florian Thiery [CC BY 4.0]

4 Five stars of LOUD

On top of the before mentioned 5 Star Linked Open Data modelling summit, the (semantic) Information Architect at The Getty Trust, Rob Sanderson,⁶ shouted his further A-B-C-D-E ideas out LOUD at his *EuropeanaTech* keynote^[6] in 2018⁷.

»Linked Open Data, with its five stars of excellence, changed the way that we publish data on the web. It gave us a short and very practical checklist that we could use to go from thinking about semantics to publishing actual data. It promotes openness as a necessity for reuse. It promotes standards as a necessity for reuse. It promotes linking between systems as a necessity for reuse. But it comes exclusively from a publishing perspective, and does not make any recommendations about how to publish data in a way that is usable by potential consumers. In the intervening time, the web community has recognized that we need 5 more

⁶ <https://twitter.com/azaroth42>

⁷ <https://youtu.be/r4afi8mGVAY>

*stars for data consumption. If our data isn't used, then no value is gained from the resources that were invested in its creation, publication, maintenance and improvement. If we want our data to be used, then it needs to be **usable**: Linked Open Usable Data.«[7]*

»Meaning that Usability is dependent upon the audience. And the Audience for Data is Developers. The user interface for developers is the API, and the API for LOD is based on the ontology. Ontologies are design primarily for semantic or theoretical correctness, which is the least practical concern. Instead, we need to manage the complexities between completeness of expression, and usability of the resulting data constructs. JSON-LD allows for some mapping of ontological constructs into JSON, the lingua-franca of modern developers and is a cornerstone technology of LOUD. Five further stars, or design principles, govern usability.«[7]

The five further stars to produce LOUD instead of LOD will be [7]:

1. The right Abstraction for the audience

As we all know »developers do not need the same level of access to data as a ontologist. Here, some use cases and requirements should drive the interoperability layer between systems, not ontologies. «

2. Few Barriers to entry

For developers it should »be easy to get started with the LOD and build something «immediately. »If it takes a long time to understand the model, ontology, sparql query syntax and so forth, then developers will look for easier targets. Conversely, if it is easy to start and incrementally improve, more [and more] people will use the data. «

3. Comprehensible by introspection

Published and shared »data should be understandable by looking at it, rather than requiring the developer to read the ontology and vocabularies. «By »using [e.g.] JSON-LD[, it] lets us to talk to the developer in their language, that they already understand. «

4. Documentation with working examples

Data have to be comprehensible and well documented. »You can never «get all the 'hidden assumptions' and »all of the rules for the data. Documentation clarifies the patterns that the developer can expect to encounter, such that they can implement robustly. Example use cases allow contextualization for when the pattern will be encountered, and working examples let you drop the data into the system to see if it implements that pattern correctly. «

5. Few Exceptions, many consistent patterns

»Every exception that you have in an API [or ontology] is another rule that the developer needs to learn in order to use the system. Every exception is jarring, and requires additional code to manage. While not everything is homogenous, a set of patterns that manage exceptions well is better than many custom fields. «

Check Fig. 7 so see quickly how your LOD can be LOUD.

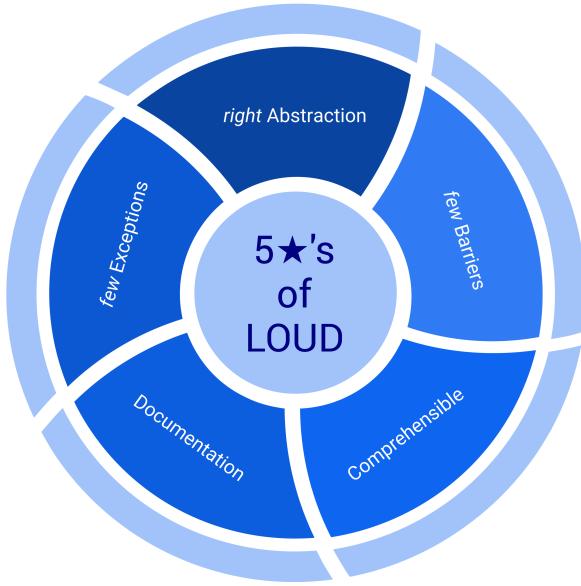


Figure 7. 5 Star LOD, Florian Thiery [CC BY 4.0]

5 FAIR data

LOUD data are not enough. In 2014 at the Lorentz workshop⁸ the term *FAIR* (Fig. 8) [Findable, Accessible, Interoperable, Reusable] was launched and the resulting 15 FAIR principles were published[8] in 2016 [9].

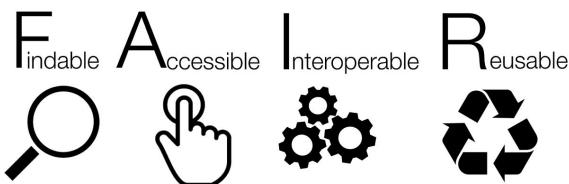


Figure 8. FAIR data principles, SangyaPundir [CC BY-SA 4.0]

The FAIR Guiding Principles are[8]:

To be Findable

F1. (meta)data are assigned a globally unique and persistent identifier

⁸ <https://www.force11.org/group/fairgroup/fairprinciples>

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be 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

To be Reusable

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

6 Sphere 7 Data

Merging all the mentioned principals above, reaching not only LOD, not LOUD, not FAIR but LOUD and FAIR research data, we need a new stairway to the Olympus of semantic data modelling. Therefore we can have a look into history and Aristotles idea of seven spheres of heaven, see Fig. 9.



Figure 9. Ptolemaic System, Fastfission [Public domain]

This model in mind, I propose the **Sphere 7 Data Model**, or **Model of Seven Data Spheres**, as displayed in Fig. 10. This sphere model combines all the mentioned principles to reach the seventh sphere of heaven via LOUD FAIR research data.

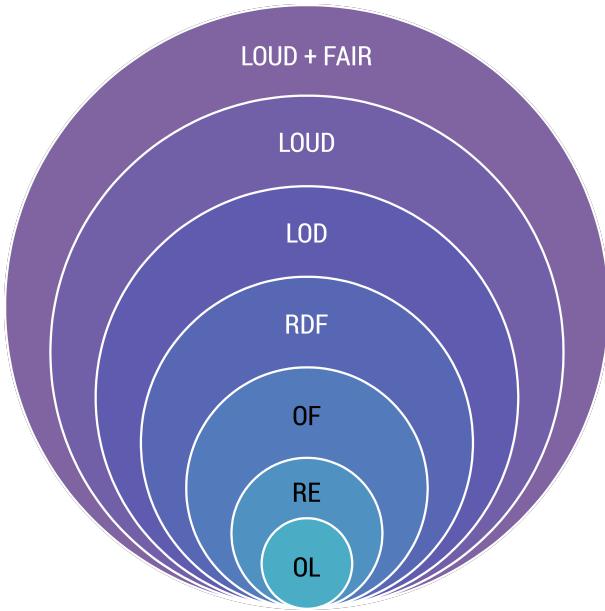


Figure 10. Seven Sphere Data, Florian Thiery [CC BY 4.0]

The **single spheres** are defined as follows:

1. Sphere: OL

Data is available in the Web under a open license.

2. Sphere: RE

Data is available as structured machine readable data.

3. Sphere: OF

Data is available in an open, non-proprietary, structured machine readable data format.

4. Sphere: RDF

Data is available as URIs and semantically modelled as RDF.

5. Sphere: LOD

Data is available as resource, semantically modelled as RDF and semantically linked to other resources.

6. Sphere: LOUD

LOUD is available as usable data, according to the LOUD principles.

7. Sphere: LOUD + FAIR

LOUD is available as findable, accessible, interoperable and reusable data, according to the FAIR principles.

Acknowledgements

I would like to thank the Mainz Centre for Digitality in the Humanities and Cultural Studies (mainzed) and Römisch-Germanisches Zentralmuseum for supporting me.

References

- [1] U. Wuttke, “Here be dragons”: Open Access to Research Data in the Humanities,” Apr. 2019. [Online]. Available: <https://zenodo.org/record/2630154>
- [2] ——, “Here be dragons: Open Access to Research Data in the Humanities,” Apr. 2019. [Online]. Available: <https://ulrikewuttke.wordpress.com/2019/04/09/open-data-humanities/>
- [3] T. Berners-Lee, J. A. Hendler, and O. Lassila, “The semantic web” in scientific american,” 2001.
- [4] T. Berners-Lee, “Linked Data - Design Issues,” Jul. 2006. [Online]. Available: <https://www.w3.org/DesignIssues/LinkedData.html>
- [5] M. Hausenblas and J. G. Boram Kim, “5-Sterne Offene Daten,” Aug. 2015. [Online]. Available: <https://5stardata.info/de/>
- [6] R. Sanderson, “Shout it Out: LOUD by Rob Sanderson - EuropeanaTech Conference 2018,” May 2018. [Online]. Available: <https://t1p.de/j171>
- [7] ——, “LOUD: Linked Open Usable Data,” 2018. [Online]. Available: <https://linked.art/loud/>

- [8] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, p. 160018, Mar. 2016. [Online]. Available: <http://www.nature.com/articles/sdata201618>
- [9] force11, “The FAIR Data Principles | FORCE11,” 2016. [Online]. Available: <https://www.force11.org/group/fairgroup/fairprinciples>