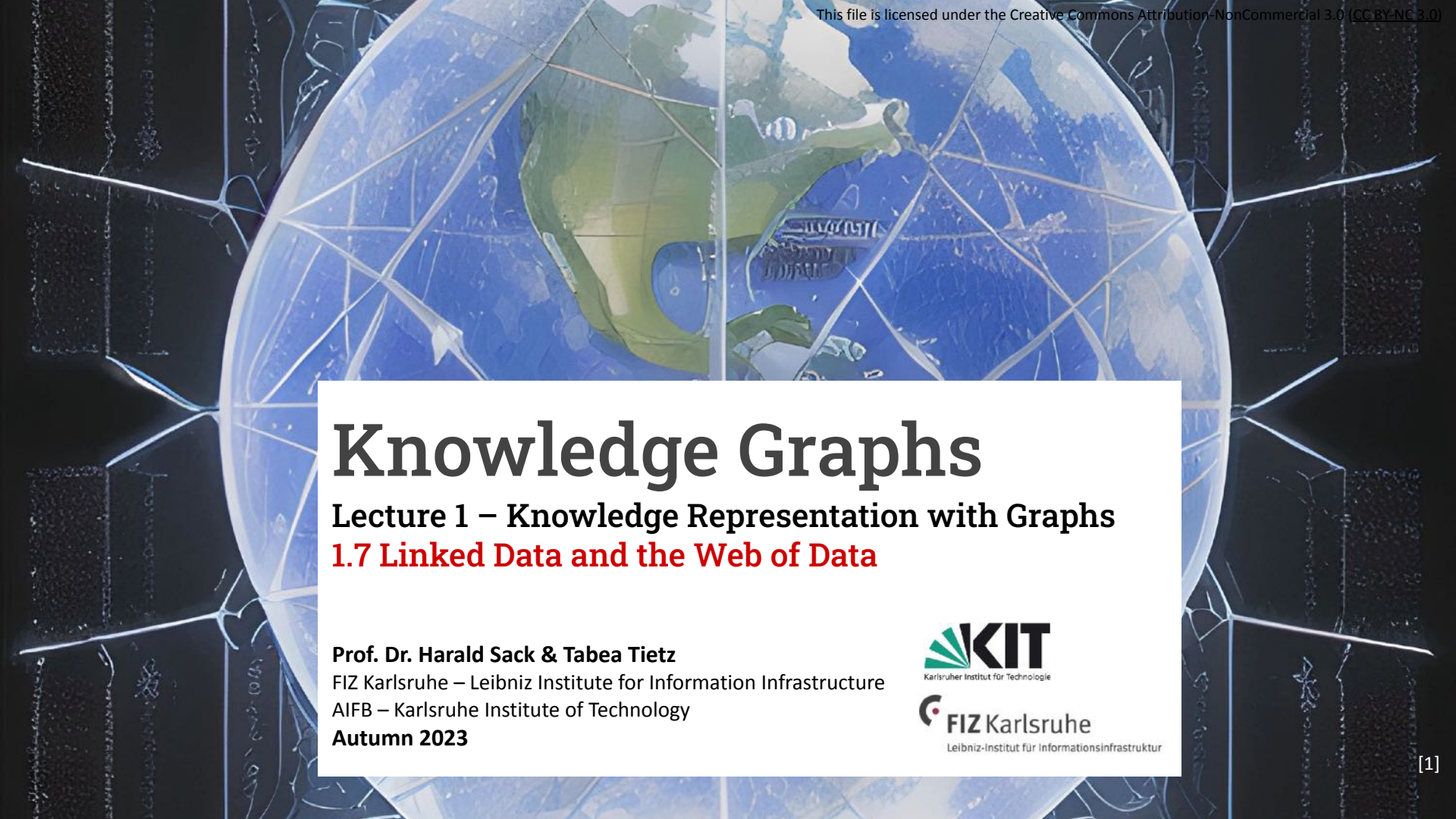


Knowledge Graphs

Lecture 1 – Knowledge Representation with Graphs

1.7 Linked Data and the Web of Data

Prof. Dr. Harald Sack & Tabea Tietz
FIZ Karlsruhe – Leibniz Institute for Information Infrastructure
AIFB – Karlsruhe Institute of Technology
Autumn 2023

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Knowledge Graphs

Lecture 1: Knowledge Representation with Graphs

1.1 From Data to Knowledge

1.2 Knowledge and how to represent it

1.3 The Art of Understanding

1.4 Graphs and Triples

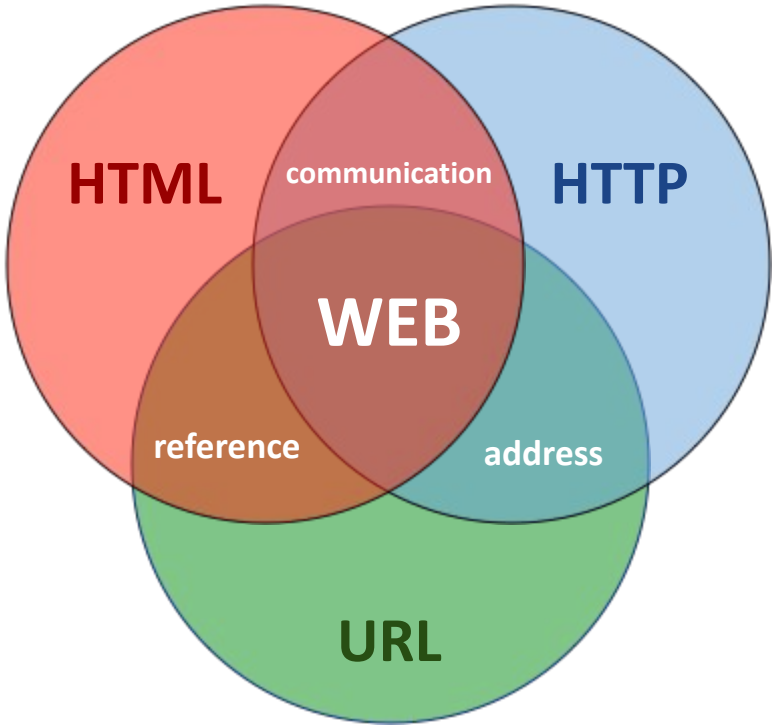
1.5 Knowledge Graphs

1.6 The Semantic Web

1.7 Linked Data and the Web of Data

[1]

The Basic Architecture of the Web

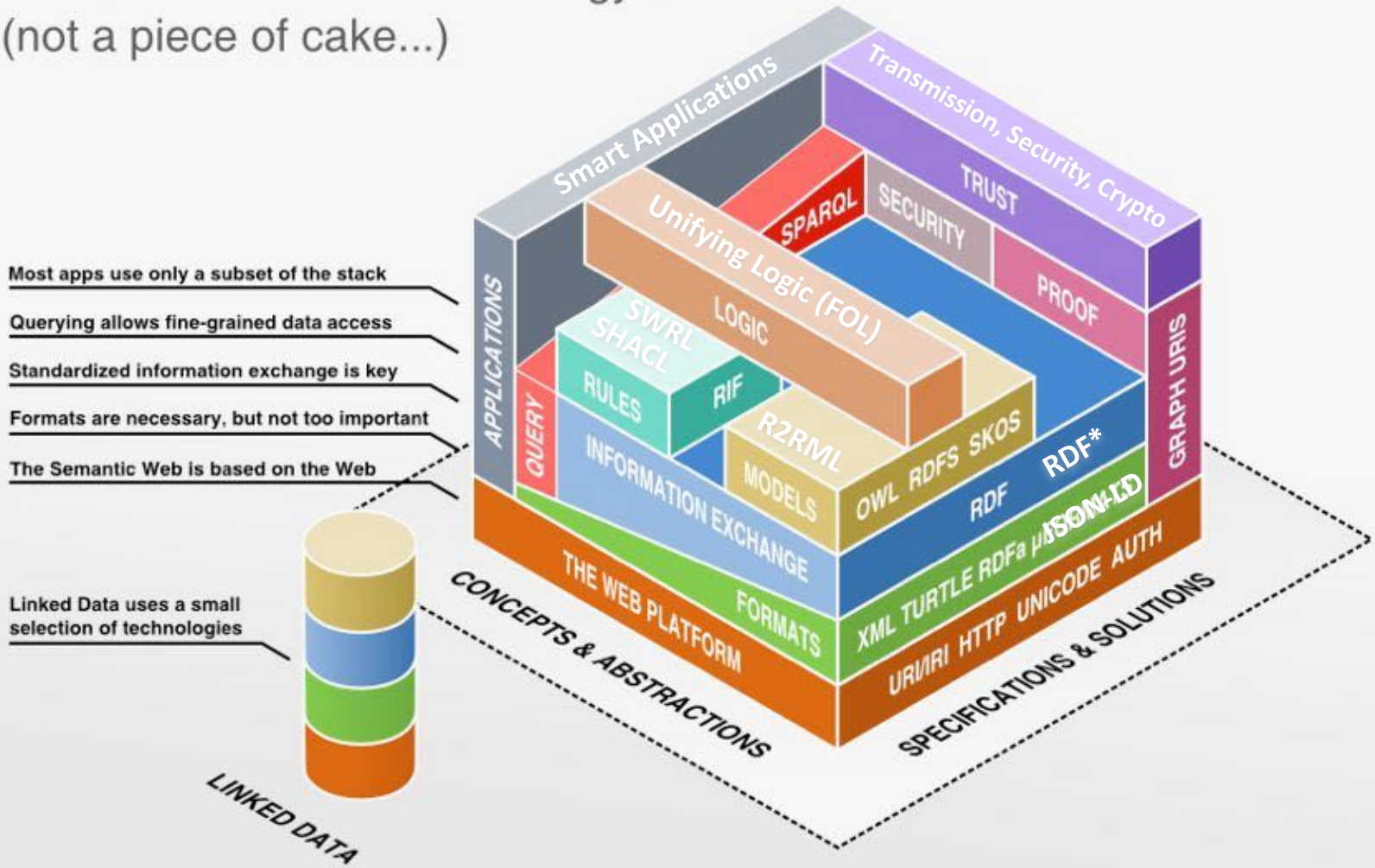


1. Identification (URI) & address (**URL**)
   e.g. `http://fiz-karlsruhe.de`

2. Communication / protocol (**HTTP**)
   ```
   GET /index HTTP/2
   Host: fiz-karlsruhe.de
   ```

3. Representation language (**HTML**)
   ```
   Tabea works at
   <a href="http://fiz-karlsruhe.de">FIZ</a>.
   ```

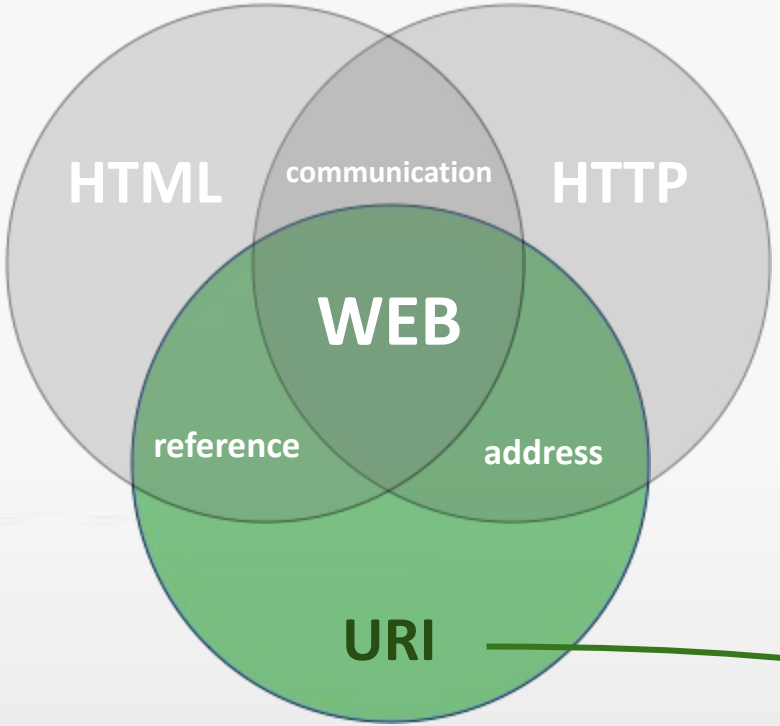
The Semantic Web Technology Stack
(not a piece of cake...)

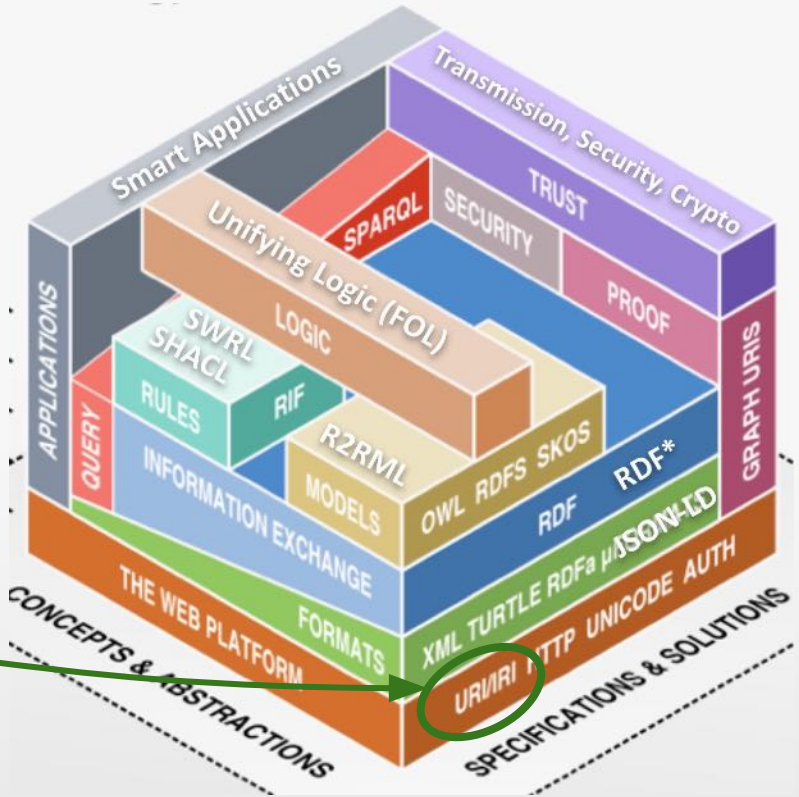


[2]

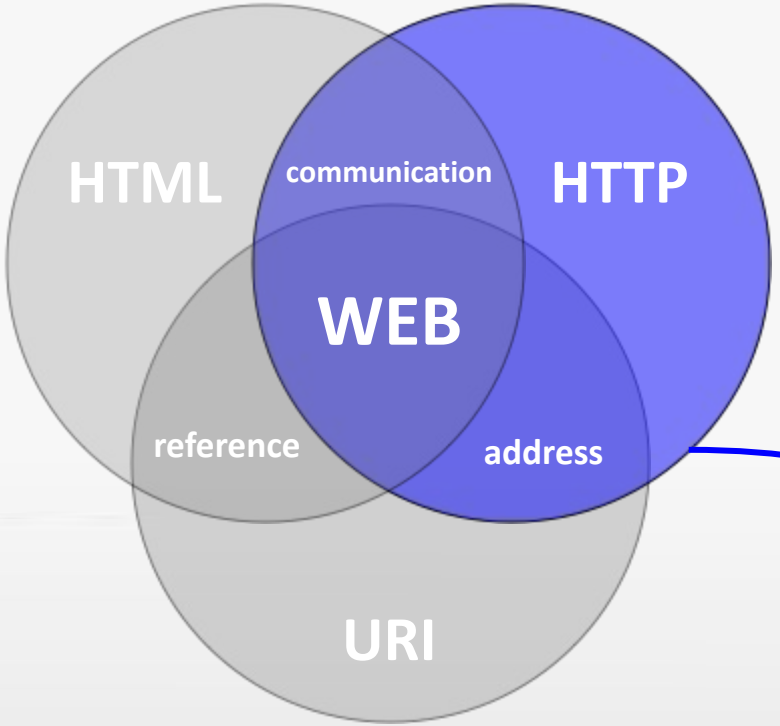
Basic Architecture of the Web of Data

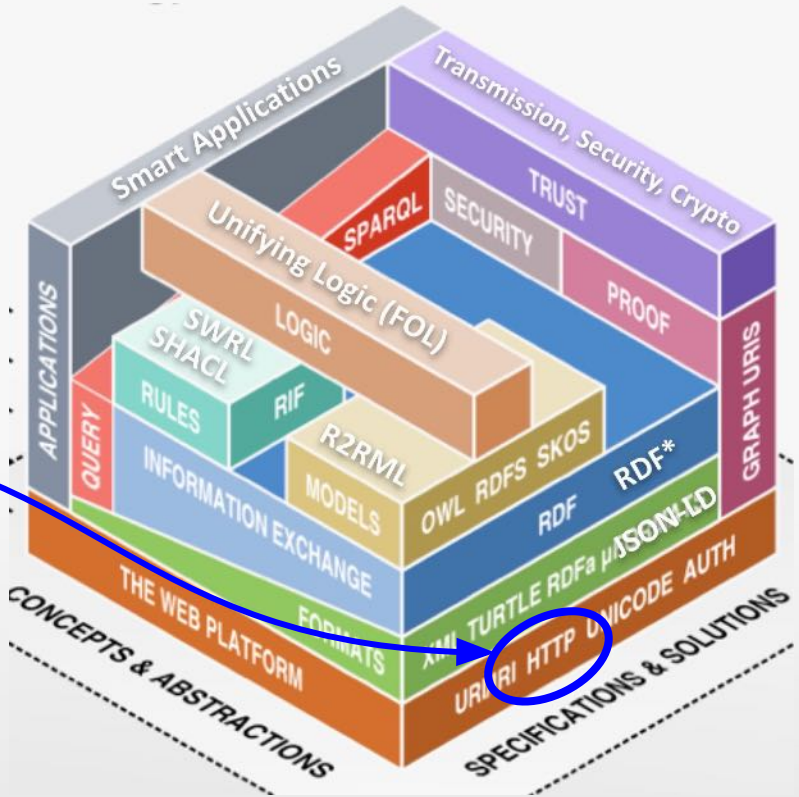
Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure





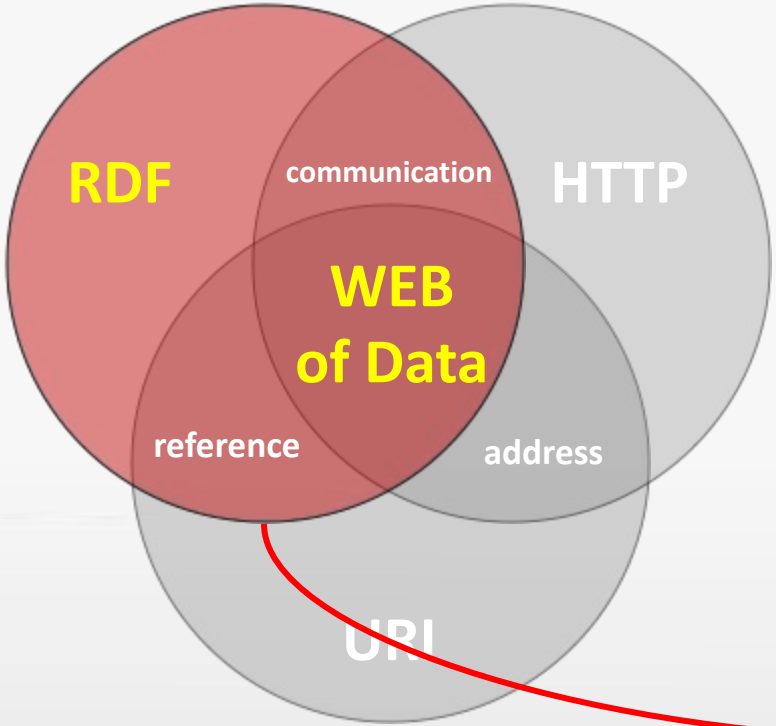
Basic Architecture of the Web of Data

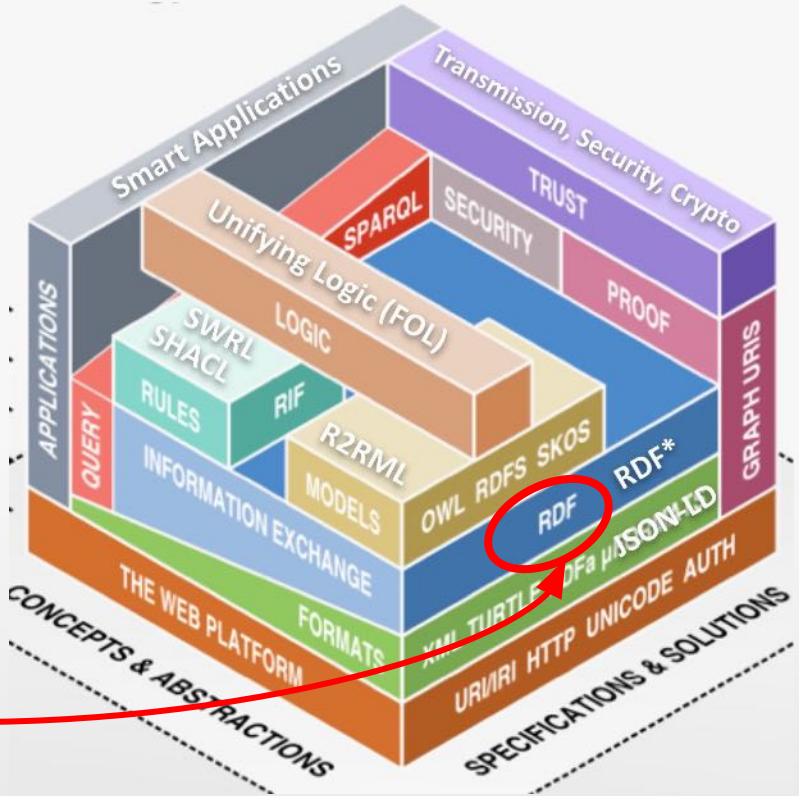




Basic Architecture of the Web of Data

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure





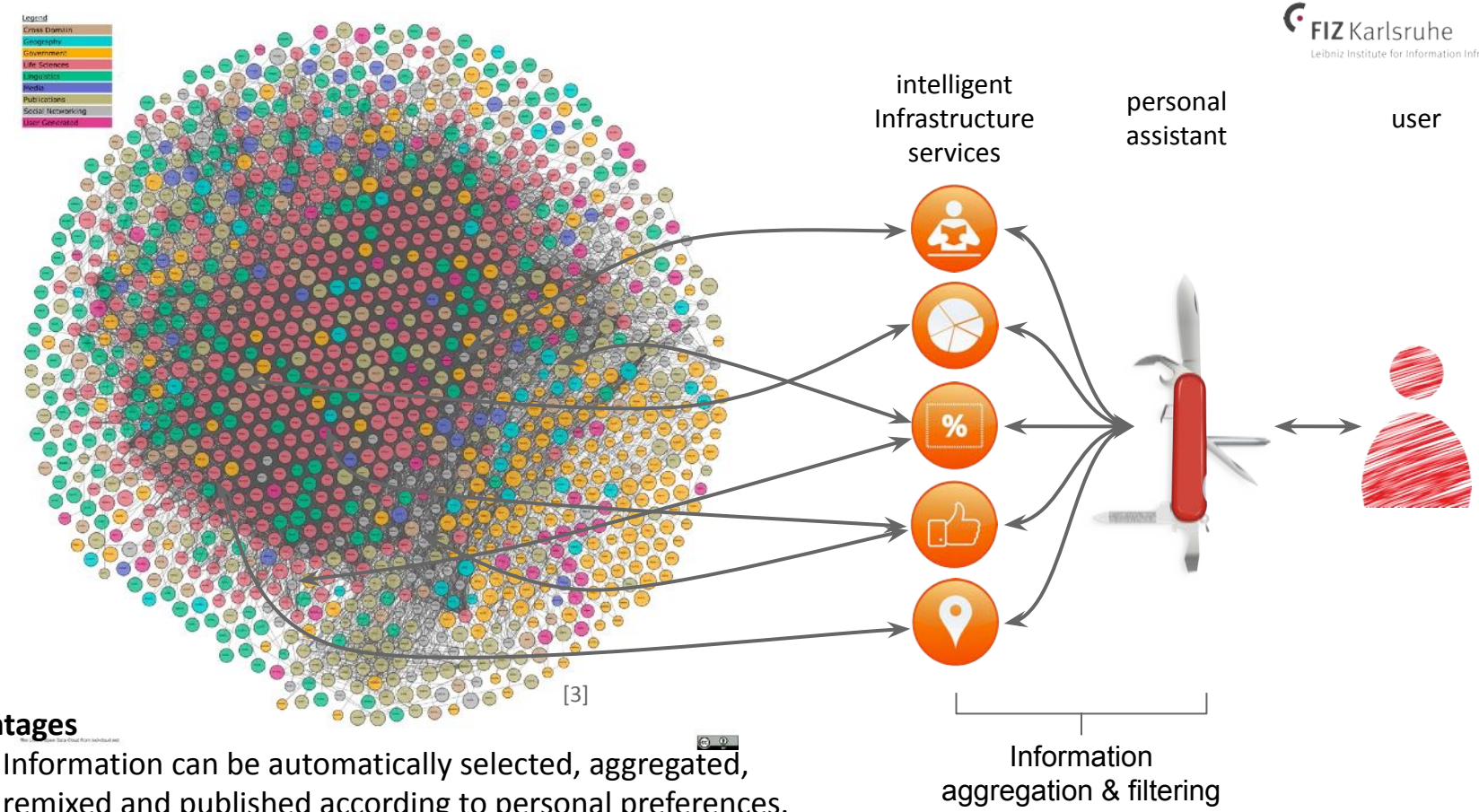
The Semantic Web – A Web of Data

- The **Web of Data** is an upgrade of the traditional Web of Documents.
- It is the Web as a huge decentralised database (knowledge base) of **machine-understandable data.**

„The web of **human-readable document** is being merged with a web of **machine understandable data**. The potential of the mixture of humans and machines working together and communication through the web could be immense.“

***Tim Berners-Lee**, The World Wide Web: A very short personal history, May 1998*

How to Access the Web of Data



Advantages

Information can be automatically selected, aggregated, remixed and published according to personal preferences.

The Web of Data

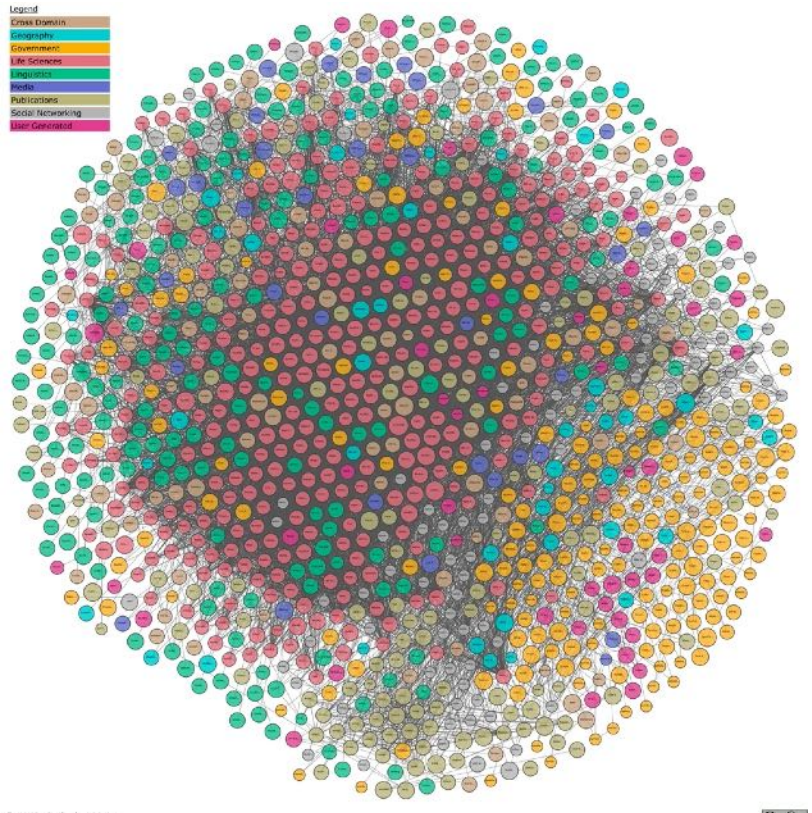
Linked Data

Linked Open Data (LOD) denote publicly available (RDF) Data in the Web, identified via URI and accessible via HTTP. Linked data connect to other data via URI.

The Web of Data

- Currently (01/2023) lod-cloud.net visualizes 1588 LOD datasets
- 2021 Common Crawl reported (JSON-LD usage)
  - 8,342,031 Web Sites
  - 793,347,572 URLs
  - 7,952,535,579 Entities
  - 37,872,880,504 Triples

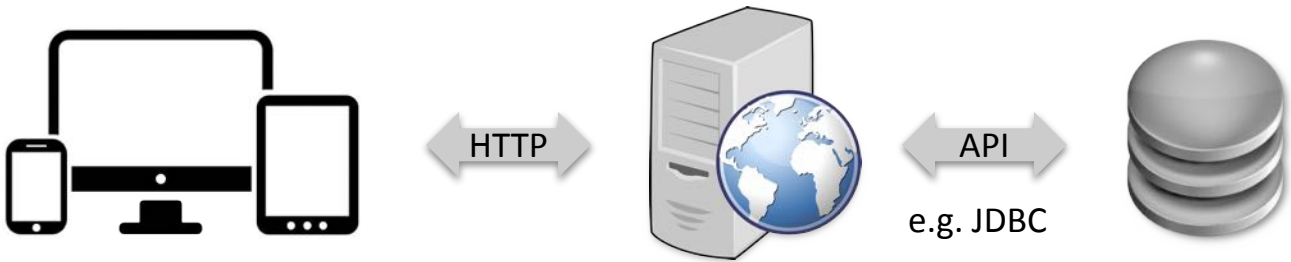
http://webdatacommons.org/structureddata/2021-12/stats/stats.html



http://lod-cloud.net/

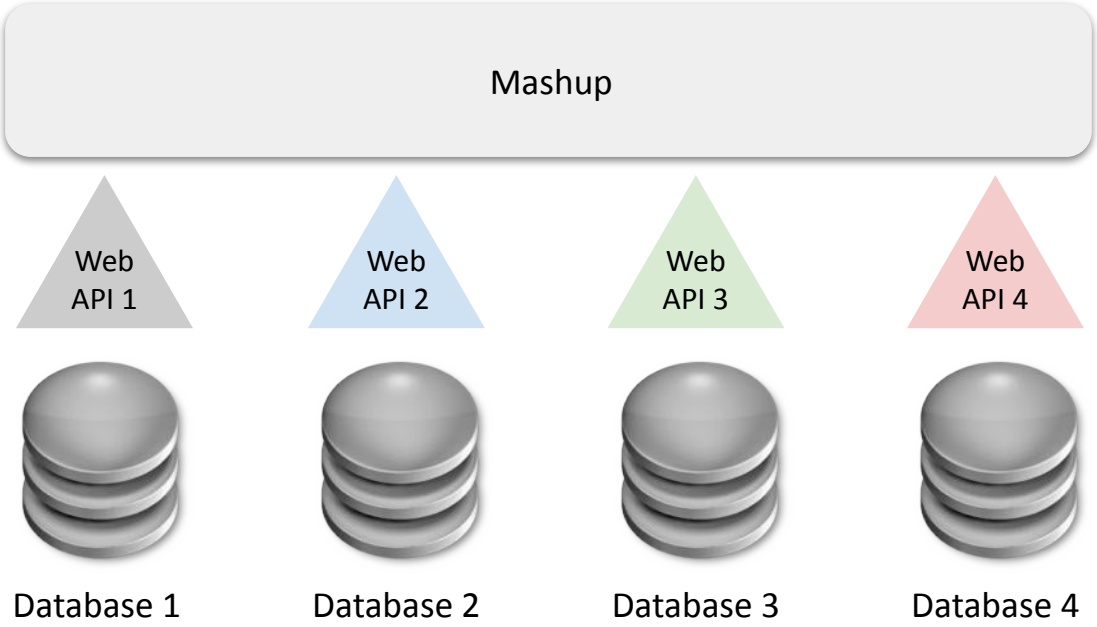
[2]

Data Access in the Traditional Web



Data Access in the Traditional Web

There is a number of different (proprietary) **Web APIs**, data exchange formats, and **Mashups** on top of that.



In the Traditional Web…

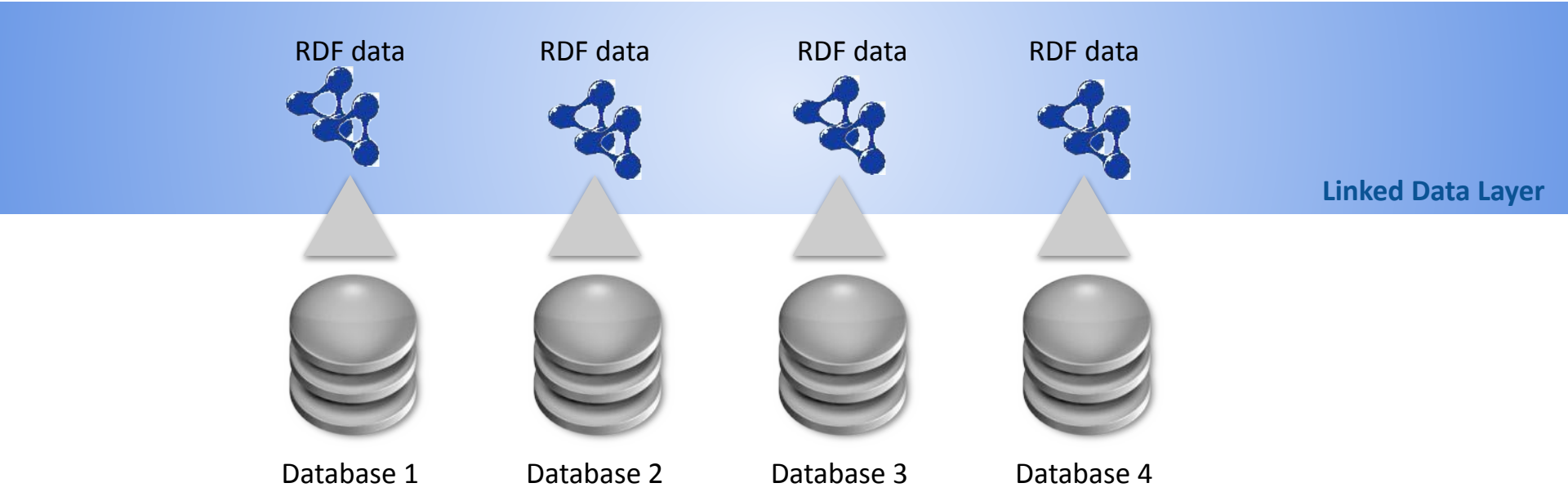
- Data is locked up in **small data islands.**
- Other applications usually cannot access this data…

[3]

How to Avoid Isolated Data Islands

Apply Linked Data technology

- to publish (structured) data on the Web
- to draw connections from one data source to data from other data sources



Linked Data Principles

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.

[4]

(Tim Berners-Lee, Linked Data, 2006, http://www.w3.org/DesignIssues/LinkedData.html)

5-Star Linked Open Data

- Public Linked Data resources on the Web, licensed as Creative Common CC-BY
- Tim Berners-Lee's 5-Star Criteria for Linked Open Data

★ Available on the Web (whatever format) but with an **open licence**, to be Open Data

★★ Available as **machine-readable structured data**
(e.g. excel instead of image scan of a table)

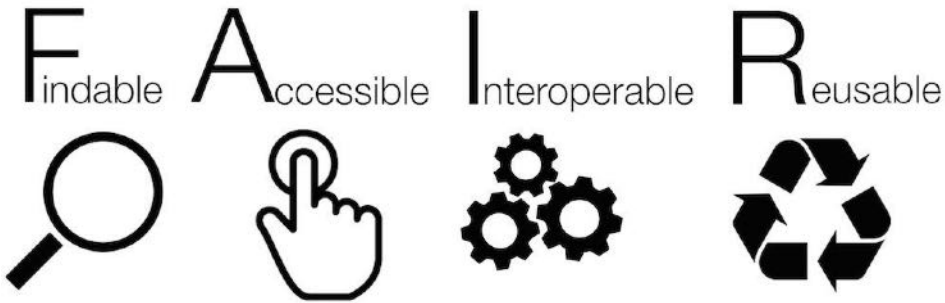
★★★ as (2) plus **non-proprietary format** (e.g. CSV instead of excel)

★★★★ All the above plus: use **open standards from W3C**
(RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: **link your data to other people's data** to provide context

(Tim Berners-Lee, Linked Data, 2006, http://www.w3.org/DesignIssues/LinkedData.html)

FAIR Principles



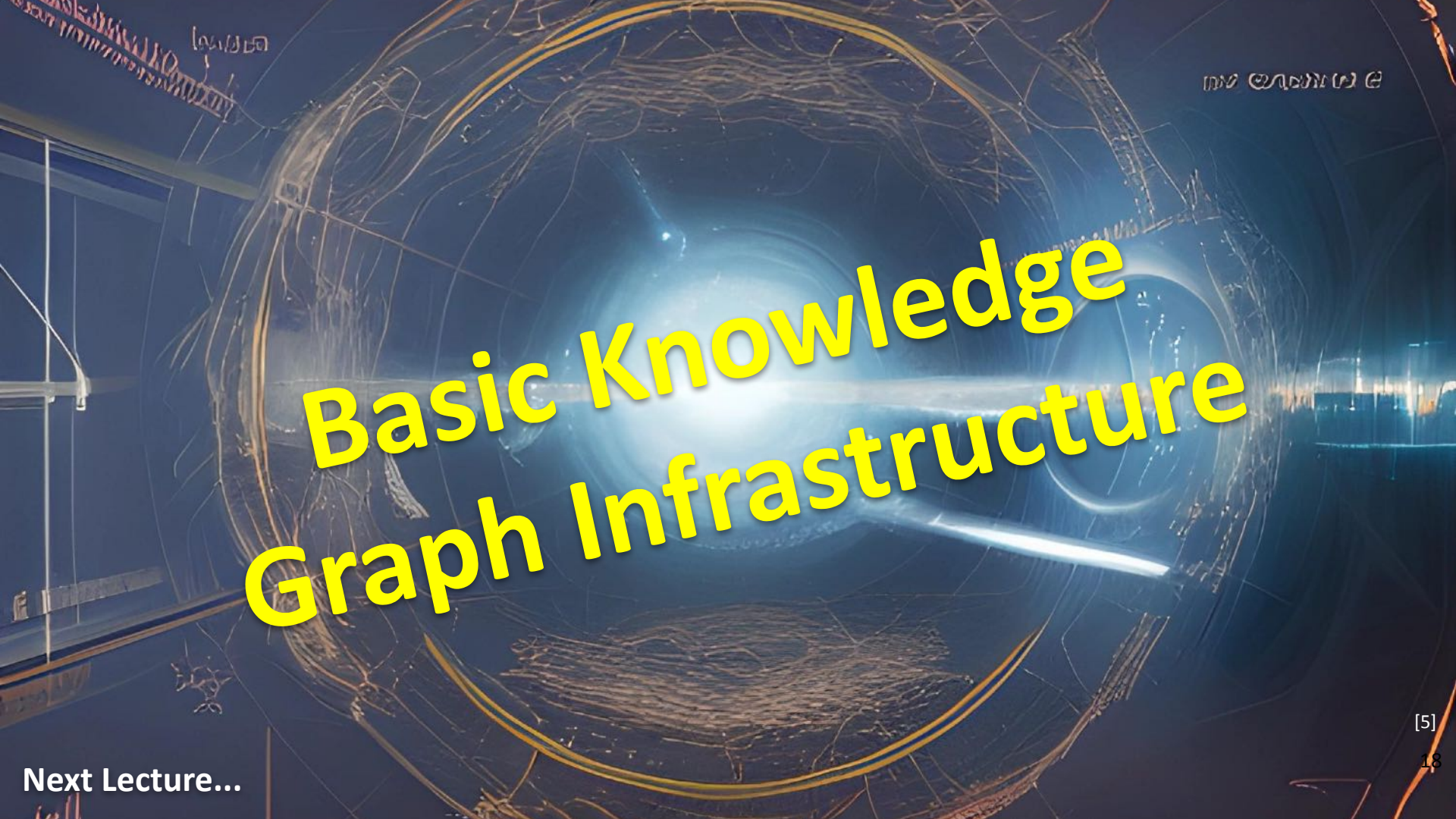
https://www.go-fair.org/fair-principles/

- **Findable**
  The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.

- **Accessible**
  Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

- **Interoperable**
  The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- **Reusable**
  The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.



[5]

Next Lecture...

Bibliographic References:

- Tim Berners-Lee (May 1998), *The World Wide Web: A very short personal history*.
- Christian Bizer, Robert Meusel, Anna Primpeli, Alexander Brinkmann, Web Data Commons - Microdata, RDFa, JSON-LD, and Microformat Data Sets 2021, http://webdatacommons.org/structureddata/2021-12/stats/stats.html
- Jeremy Debattista, Judie Attard, Rob Brennan, and Declan O'Sullivan (2019). *Is the LOD cloud at risk of becoming a museum for datasets? Looking ahead towards a fully collaborative and sustainable LOD cloud*. In Companion Proceedings of The 2019 World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 850–858.
- Tim Berners-Lee (2006), Linked Data, 2006, http://www.w3.org/DesignIssues/LinkedData.html
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016), *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018 (2016).

Picture References:

[1] "An image of the Semantic Web which is an extension of the World Wide Web…" , created via ArtBot, Dreamlike Diffusion, 2023, [CC-BY-4.0], https://tinybots.net/artbot

[2] Benjamin Nowack, The Semantic Web - Not a Piece of cake…, at bnode.org, 2009-07-08 , [CC BY 3.0], https://web.archive.org/web/20220628120341/http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake

[3] The Linked Open Data Cloud, https://lod-cloud.net/

[4] "An image of the Semantic Web which is an extension of the World Wide Web…" , created via ArtBot, Dreamlike Diffusion, 2023, [CC-BY-4.0], https://tinybots.net/artbot