Research Ideas and Outcomes 8: e95268
doi: 10.3897/rio.8.e95268

Conference Abstract

The Vision of the FAIR Digital Object Machine and Ubiquitous FDO Services

Peter Wittenburg[‡], Christophe Blanchi[§], Claus Weiland[|], Ivonne Anders[¶], Karsten Peters[¶], Ulrich Schwardmann[#], George Strawn[¤]

‡ Unaffiliated, Berlin, Germany
§ DONA, Geneva, Switzerland
| Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt am Main, Germany
¶ DKRZ, Hamburg, Germany
# GWDG, Göttingen, Germany
¤ BRDI, Washington, United States of America

Corresponding author: Peter Wittenburg (peter.wittenburg@mpcdf.mpg.de)

Reviewable v 1

Received: 23 Sep 2022 | Published: 12 Oct 2022

Citation: Wittenburg P, Blanchi C, Weiland C, Anders I, Peters K, Schwardmann U, Strawn G (2022) The Vision of the FAIR Digital Object Machine and Ubiquitous FDO Services. Research Ideas and Outcomes 8: e95268. https://doi.org/10.3897/rio.8.e95268

Abstract

In addition to the previous intensive discussion on the “Data Deluge” with respect to enormous increase of available research data, the 2022 Internet-of-Things conference confirmed that in the near future there will be billions if not trillions of smart IoT devices in a very wide range of applications and locations, many of them with computational capacities. This large number of distributed IoT devices will create continuous streams of data that will require a global framework to facilitate their integration into the Internet to enable controlled access to their data and services, to name but a few aspects. This framework would enable tracking of these IoT devices to measure their resource usage for instance to globally address the UN Sustainable Development Goals. Additionally, policy makers are committed to define regulations to break data monopolies and increase sharing. The result will be an increasingly huge domain of accessible digital data which on the one hand allows addressing new challenges especially cross-sector ones. A key prerequisite for this is to find the right data across domain boundaries supporting a specific task.

Digitisation is already being called the fourth industrial revolution and the emerging data and information is the 21st century's new resource. Currently this vision is mostly

unrealised due to the inability of existing data and digital resources to be findable, accessible, interoperable, and reusable despite the progress in providing thematic catalogs. As a result, the capacity of this new resource is latent and mostly underutilized. There is no Internet level infrastructure that currently exists to facilitate the process by which all data and digital resources are made consistently and globally accessible. There are patchworks of localized and limited access to trusted data on the Internet created by specific communities that have been funded or directed to collaborate.

To turn digital information into a commodity, description, access to, validation, and processing of data needs to become part of the Internet infrastructure we call the Global Integrated Data Space (GIDS). The main pillars of this approach require that data and services be globally identified and consistently accessed, with predictive descriptions and access control to make them globally findable.

Currently researchers are relying partly on informal knowledge such as knowing the labs and persons to maximize the chance to access trustworthy data, but this method is limiting the use of suitable data. In the future data scenario, other mechanisms will become possible. In the public information space Google-like searches using specific key terms have become an accepted solution to find documents for human consumption. This approach however, does not work in the GIDS with large numbers of data contributors from a wide range of institutions, from millions of IoT devices worldwide, and where a wide range of data types and automatic data processing procedures dominate. Indeed, successful labs that apply complex models describing digital surrogates can automatically leverage data and data processing procedures from other labs. This makes the currently often operationally applied manual stitching of data and operations too costly both in time and resources to be a competitive option. A researcher looking for specific brain imaging data for a specific study has a few options:

- Rely on a network of colleagues.
- Execute Google-like searches in known registries looking for appropriate departments and researchers.
- Execute Google-like searches on suitable data.
- He/she engages an agent to execute profile matching in suitable sub-spaces.

We assume that data creators will have the capability and be interested to create detailed metadata of different types and that the researchers, who are looking for specific data, will be able to specify precise profiles for data they are looking for. Two of the key characteristics of the future data space will be operations that can carry out profile matching at ultra-high speeds and that will lead to various subspaces according to some facets using self-organizing mechanisms. Of course, this poses high requirements on the metadata quality being used and that creators and potential consumers share knowledge about the semantic space in which they operate, and available semantic mappings used by brokers or self-provided. Metadata must be highly detailed and suitable schemas have been developed already in many communities. In addition to the usual metadata, potential

users will need to specify their credentials in the form of trusted roles and their usage purposes to indicate access opportunities.

Changing current metadata practices to yield richer metadata as presribed by the FAIR principles will not be simple, especially since we seem to be far away from formalizing roles and usage purposes in a broadly accepted way, but the pressure to create rich and standardized metadata will increase. It should be noted of course that for data streams created by IoT sensors, defining proper metadata is an action that is only requested once or a few times.

Why are FDOs special in this automatic profile matching scenario? FDOs are bundling all information required for automatic profile matching in a secure way, i.e., all metadata information are available via the gloablly unique resolvable and persisten identifiers (PID) of the FDO and the PID security mechanisms are at the basis to establish trust. FDOs will be provided with a secure method that is capable of computing efficiently coded profiles representing all properties of an FDO relevant for profile matching. This would speedup profile matching enormously.

We will address two major questions characterizing the “FDO Machine” we are envisioning:

1. Which kinds of representations could make profile matching much more efficient?
2. How could FDO-based mechanisms be used to efficiently create sub-spaces that would help the emerging layer of information brokers to offer specialized services addressing specialized needs as for example requested by UN’s Sustainable Development Goals?

Brokers might want to use specialized agents to create subspaces along many different important facets such as domains, trustworthiness, roles, etc. Such subspaces are ephemeral virtual structures on top of the huge global integrated data space.

Keywords

FAIR, FAIR Digital Objects, Future Data Space, Mining

Presenting author

Peter Wittenburg

Presented at

First International FAIR Digital Objects Conference, presentation