Report Paris Meeting

Organisers: Edit Herzcog, Barend Mons, Larry Lannom, Robert Quick, George Strawn, Peter Wittenburg, Carlo Zwölf, Assisted by Zsuzsanna Szeredi

This is a first draft report about the FAIR Digital Object (FDO) Meeting that took place at 28/29.10 in Paris. Please, correct and add so that we can put a final report on the web-site asap.

This meeting brought together 28 experts from different countries to discuss

- requirements for the FDOs and the FDO Framework from various scientific communities, from special themes such as FAIR maturity indicators and citation, and from technical points of view
- an open statements about FDOs as agreed result of the meeting as a starting point to organise a FDO community and
- the next steps for moving ahead.

The agenda and all slides can be found here: https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects/Paris-FDO-workshop

Results

The results of the meeting met the expected goals:

1. Based on a draft consensus document presented to the participants, it was agreed to work on an improved version which can be used then to contact other committed experts to get their agreement and thus to form stepwise a FDO community. A group was formed to improve the document.
2. It was agreed that we should also create a document which is addressing other stakeholders such as funders and industry to make them aware and draw their interest. A group was formed that will produce this document.
3. A coordination group was formed that will work on the next steps, synchronise activities and suggest possible light governance models to organise the FDO community.
4. A group of technical implementation experts should start working that includes interested experts and that initiate RDA groups where necessary to find broad consensus.

The GEDE platform will be used for the FDO interactions until a new structure has been defined.

Session 1: Introduction to FAIR Digital Objects

Chair: Carlo Zwölf, Speakers: Peter Wittenburg, Luiz Bonino

The idea of an information management architecture based on Digital Objects began with R. Kahn and colleagues at CNRI in the late 1980s[1], with the handle system as one early result. In 2013, the Research Data Alliance started with several groups working on topics related to the Digital Objects inspired by the CNRI work. The Data Foundation & Technology Group worked out a Core Model for Digital Objects based on various use cases. The PID Information Type Group worked on standardising the set of attributes being used in PID records and was followed up by the Kernel Information Types Group which developed a first core set. The Data Type Registry Group presented a mechanism to

[1] Kahn, Robert E. "The Architectural Evolution of the Internet". Corporation for National Research Initiatives, November 17, 2010. http://hdl.handle.net/4263537/5044

register types and to relate types with operations. All these results allowed the active people to build a large user community on Digital Objects and to initiate first implementations.

Recently, a discussion was started about the degree of FAIRness being supported by the DO concept. Basic FAIRness was built into the DO concept such as the resolution to attributes that point to the locations of the DO's bit sequences, to a PID of the DO's metadata, etc. However, it was also identified that current practices are far away from the capacity of the Core DO Model. To fully comply with the FAIR principles the DO model had to be extended by making some specifications mandatory. All attributes being used in the PID record need to be semantically typed using a type registry or ontology allowing machines to act on them. In addition, it is not sufficient to simply point to the metadata, but it needs to be requested that metadata is designed in such a way that they are machine actionable. Also collections which are DOs need to be constructed in a way allowing machines to understand the relations between the components of the collection.

This extended concept which is making use of Linked Data mechanisms to include explicit semantics is now being called a FAIR Digital Object (FDO). This term was already coined by the "Turning FAIR into Reality" report and subject of a few papers, but it has now received momentum by bringing together the DO concept with mechanisms from the Linked Data domain. Additional work needs to be carried out to work on further specifications.

Session 2: Community Experts

Chair: Francoise Genova; Speakers: Dimitris Koureas, Tobias Weigel, Koenraad de Smedt, Barend Mons, Roberto di Cosmo

This session was meant to listen to requirements and expectations of scientific communities which are already discussing the (F)DO concept with clear intentions towards implementations. The DiSSCo community wants to create one large virtual domain with digital surrogate objects that can be used instead of the physical objects stored in 119 facilities in 21 countries. The intention is to convince users to use this domain of FAIR Digital Objects for annotations, studies to gain new insights and reproducibility actions. This will only happen if users will have trust in the sustainability of the FDO landscape where scalable, agile and persistent IDs are crucial pillars. In this digital specimen domain about 3 billion specimens organised in about 2 million classes will be linked to all kinds of information and classified using different criteria. These Digital Specimen FAIR Digital Objects are the envelopes to implement a stable binding to all these different digital entities using Handles to create the persistent bindings. Success will not only depend on the availability of technical components (Handles, type registries, etc.) but on stakeholder buy-in and massive investments in capacity building.

Climate modelling experts working together globally are confronted with an extreme increase in data volumes and heterogeneity making manual data management increasingly impossible. One result of the current management practice is, for example, that the state of objects is often unclear and the current hierarchical structures are increasingly inadequate to clarify states. What is needed instead is a generic workflow design that guides all digital objects from their creation, mostly as results from HPC calculations, up to their publication. In such workflows data and metadata need to be seen as closely related couple as the FAIR DO model implements it. Machine actionability at all stages, scalability of all mechanisms and the availability of a library of reusable procedures that can be easily integrated in handy workflow frameworks such as Jupyter would help to achieve the required transformation towards automated workflows, would reduce the workload of data scientists and managers and thus guarantee user buy-in. A big challenge will be how to transform the legacy.

The language community built up an infrastructure (CLARIN) of data and tools covering 50 certified centres mainly in Europe, but some also from abroad (US, SA) all connected via single sign on and all

using Handles already. Their Virtual Language Observatory is harvesting metadata from many more centres worldwide and gives access to rich metadata which can be combined flexibly to virtual collections by using a Virtual Collection Builder. Within the CLARIN network of centres the flexible Component Metadata (CMDI) framework is accepted as a standard having registries for concepts, components and profiles as core elements. It is enabling every researcher to design his/her metadata scheme without losing semantic interoperability. A key application being used by many researchers is a workflow tool which is amended by a switchboard facility. Users select a set of language resources and the switchboard tool matches metadata profiles of data and tools to do the workflow orchestration. This automatic matching enables non-specialists to carry out complex operations such as for example named entity recognition. A step over to FAIR DOs is intended since it would give CLARIN the advantage of a uniform interface to all repositories, to the switchboard, virtual collection registries etc. and thus help simplifying the workflow mechanisms. One open question to be addressed is the effort needed to adapt the many existing repositories.

In biomedical research one of the big challenges is to extract conclusions from the huge amount of papers produced (doubling every 6 months) that, for example, study the effect of medications on patients. Nano-publications, which are augmented RDF assertions, are a promising way to master this unmanageable lake of evidences (10^{14} assertions). Cardinal assertions (CA) can be derived from the mass of nano-publications when their core assertions are identical already reducing the space by factors. Knowlets can be derived from all CA having the same subject resulting in graphs representing key concepts and their relationship patterns in this reduced space of evidences (10^{6} knowlets). Using graph matching techniques knowlets emerging from different sources or the change of the patterns over time can be studied. These kinds of operations will help in getting deeper insights and improve medication. In formal terms, Knowlets, being a speecific type of collection, are FAIR Digital Objects when we assign PIDs to them and thus create a domain characterised by referential integrity over many decades. FDOs are in the centre of the biomedical knowledge domain and generic and domain specific metadata associated with the FDOs open the way to draw relevant conclusions. FDOs should be in the centre of the EOSC partnership plan.

Archiving software source code has been identified as part of human cultural heritage, since it is involved in all aspects of societies and embodies collective knowledge. More than 6 billion source code files have been collected systematically and assigned with billions of internal IDs being hashes on the content. Much curation effort to create useful metadata is necessary, yet there are no broadly agreed standards for metadata for software and citations. Software objects are complex since they evolve over decades with continuously new versions and their understanding is dependent on many external software objects. Sophisticated developer communities have developed smart mechanisms to cope with this complexity. As other digital objects software objects stored in archives need to be FAIR, i.e. Findable, Accessible and Re-usable with Interoperability being implemented in special ways. Therefore, the FDO model is attractive for a software archive.

Discussion & Summary

Questions addressed were whether software objects are FAIR Digital Objects (FDO) and whether the internal hashes can be transformed to globally resolvable and technology independent identifiers. It was obvious that indeed software projects can be seen as highly complex collections and thus FDOs. The internal hash codes could be added as suffix to Handles, for example, to make them globally resolvable, yet there is no need for the software archive to take that step. Also other communities use hash values of the DO's content already as suffixes enabling direct checks whether the content is the one expected.

Other questions addressed the issue of what is meant by FDO "types". The term "type" is heavily used in IT for a variety of different aspects. What is meant in the case of FDOs is that the content of it is encoded according to a set of rules which are normally described in a variety of different metadata assertions allowing machines to interpret a given bit sequence. These multiple structure indicators

can be summarised as a "type" which is a shortcut which then can be associated with operations. Types can be associated with several operations for example for processing, for visualisation, for modification, etc., i.e. in addition to specifying a "type" with the help of a PID a client needs to also provide a selection information.

A related topic that was addressed in the discussion was whether indeed FDOs have the inherent capacity of increasing trust. Trust includes many dimensions. What FDO can do is to improve the trust level by persistent context binding through its inherent mechanisms, by supporting systematic modularisation and thus reducing complexity and by using the encapsulation mechanism as a path towards increased automatization. FDO is an integrative technology reducing the N*N effort to a 1*N effort when integrating the large number of repositories. However, the effort to do the adaptation of these repositories to FDO should not be underestimated.

The presentations from scientific communities indicated that FAIR Digital Objects play already an important role in the ways they structure the increasingly large and complex domains of data, metadata, assertions, software and other types. It is the increasing volumes and complexity in the relationships that require new conceptual approaches. The conceptually simple FDOs with their inherent abstraction, persistent binding and possibility to do encapsulation are the missing building blocks that motivate some advanced scientific communities to invest in building complex relational structures representing knowledge and in building automatic workflows not only improving FAIRness but also solving the reproducibility crisis.

Session 3: Requirements Specifications

Chair: Bob Hanisch; Speakers: Edit Herczog, Carlo Zwölf, Larry Lannom

An RDA working group is working hard to define FAIR maturity indicators and a framework for applying them. The FDO model needs to demonstrate that it has the capacity to comply with all criteria. The WG is bringing together a variety of stakeholders to discuss core assessment criteria, a FAIR Maturity Model and specifications for a tool set and a checklist for data to finally write an RDA recommendation, despite the differing views of what FAIRness exactly means. 15 indicators have been widely agreed in the group, 2 are still under discussions. The group also determined 3 prioritisation classes: mandatory, recommended and optional and these are distributed across the 4 FAIR dimensions. There are still deep discussions about questions such as: (1) Should a PID refer to data or a landing page? (2) Should the information support machine as well as human usage? (3) Should two different speeds of FAIRness be defined? Should summary scores being defined and how should they be presented? The group sees it as necessity to test the indicators in practice and use the experiences to refine the outcomes.

In the discussion it became evident that the concept of FDO must meet all criteria and cannot make a difference between mandatory, recommended and optional. Also, some of the open questions are clearly addressed by the FDO concept. The result of the PID resolution step must be machine actionable and may not be a weakly defined landing page. The issue of supporting human readability is a matter of visualisation tools. The assessment of FDOs must go beyond the assessment of individual data sets, since relationships inherent to FDOs need to be tested as well.

FDOs have a great relevance for proper referencing and citation. The first question addressed is whether referencing and citation are different. Crossref defines the act of citation as something special since it needs to be associated with quality standard for the content. But this raises more questions in the digital domain, since who defines the quality standards for the big data volumes we are creating where traditional peer-to-peer review mechanisms don't work anymore. Citations, in general, serve two functions requiring different granularity levels and long-term availability of the data: (1) Create trust and reproducibility and (2) give credits to the contributors. This raises the

second question whether automatically created scores derived from citations of collections still make sense. Over-acknowledgments can be expected and errors will be propagated without corrections. The third question addressed is whether the context of FDOs need to be downloaded to enable work in the absence of an Internet connection.

The discussion revealed that there are indeed different opinions about the question whether referencing and citation are the same. In general, citations are for human consumption (authors, dates, etc.). However, an FDO includes all information that is necessary to create a useful citation, since the available metadata could be transformed into DublinCore semantics, for example. It is a matter of tooling to immediately download the citation context to be independent of an Internet connection.

Finally, we discussed the requirements for FDO based on what was presented so far and what has been worked out beforehand (see appendix). Why are we talking about Digital Objects and FAIR Digital Objects right now and why does it suddenly get so much attention? There are different reasons for this. Until recently making all of the data available simply wasn't feasible but now we understand that we can get new scientific insights from the rich data which is available. However, this will only work if we have an architecture to support the integration efficiently and seamlessly, i.e. reducing the 80% data wrangling substantially and allowing many researchers to participate in data science. Such an architecture would be the core building block of a data science supporting infrastructure as intended for example by EOSC. FAIR has been widely agreed as a common guideline for working with data. The next logical step was to invent a core model for digital objects that supports the FAIR principles by design. Earlier concepts such as formulated by Kahn & Wilensky and later by RDA can be integrated with concepts from the linked data community bringing us very close to specify such an architecture.

The proposed FDO Framework (FDOF) as sketched in the appendix is an excellent start towards more complete specifications and opens the way towards implementations. Many aspects which have been mentioned by the speakers such as referential integrity over decades, persistent binding of informational entities crucial for FAIRness, the possibility to do encapsulation as a path towards improved automation are all built-in features of FDOs leaving it to the scientific communities to define granularity, versioning strategies, etc. But we need to also admit that much more needs to be done to further specify FDOs in a way that user communities can apply the concept. The "type" topic was already indicated as one of the areas where work is needed. How many types do we expect? How to prevent a useless proliferation of types? How to classify types to make it easy for users to navigate? There are many more such aspects that need to be worked out by networks of experts - most probably using RDA.

Most important, however, is also that we need testbeds to try out new ways and mechanisms, to test actively the evolving FAIR Maturity Indicators and the compliance of the FDOs, to indicate gaps in our specifications, to understand where we may have to revise specifications and where we can learn from each other. Providing for example an incrementally growing library of FDO tools would help to simplify the adaptation of repositories to the now suggested unifying DO Interface Protocol. European RI are often using tools such as DSpace or Fedora to set up repositories. Having a well-tested adapter for these tools would easily integrate hundreds of existing repositories. It is obvious from the discussions that we needed all the time spent in RDA, GOFAIR, DONA and other initiatives to come to this point of agreement, but that we now need to accelerate supported by some governance where data scientists need to be in the driving seat.

Session 4: Comments from Technologist View

Chair: Erik Schultes; Speakers: Christophe Blanchi, Jonathan Clark, Ray Plante, Klaas Wierenga

The DONA Foundation is care-taker of the Handle System which is also used to resolve DOIs which is often not known to people. Nevertheless, it has an enormously crucial role in maintaining stability, in taking care of sustainability of the Handle System and in guaranteeing that it is free of commercial interests. It is governed by an international board of experts coming from different regions and different applications areas. ITU is represented as well as the DOI Foundation and various national/regional members from science and industry. The National PID Coalition of China, for example, runs industrial projects in Food Supply Chain projects where billions of Handles are being issued to enable quality control. The members of the board are committed to the goals. DONA also takes responsibility of the Global Handle Resolution System which is a network of currently 10 root nodes called Multiple Primary Agencies. They act as registration authorities issuing prefixes and as GHR service providers. In most cases they also act as local service providers which are out of the scope of DONA. DONA is strictly neutral with respect how Handles are being used and how registration authorities and service providers define their business models. Currently, there are about 4000 Handle Servers worldwide having their own prefix domains. DOIs use the prefix 10. Handles allow many delegation levels within the prefixes, current practice is to have three (<level1>.<level2>.<level3>/<suffix>). Choosing schemes for suffixes is up to the local service provider.

The discussion showed that the DONA Foundation is still learning and adapting. Developing a global PID infrastructure that is free of commercial interests, balanced, self-sustained and open for all kinds of application scenarios is a challenge and can only succeed if it evolves smoothly. Different organisational forms are being tested in various countries/regions and will be evaluated. Currently, strict rules for the functioning of root nodes are in operation, it may be that more rules will be required. Also the board which has been initiated by including committed and respected experts will change in the coming years. If EOSC would make use of the Handle System this would certainly lead to interesting interactions with the DONA Foundation and eventually to new models.

The DOI world that evolved from identifying electronic publications led to a fairly comprehensive social infrastructure enabling 20 years of cooperation and competition. It is an excellent example to indicate that open standards drive trust. Such infrastructures can be successful if they have a tendency towards "lowest energy state solutions", i.e. create value by relieving pains or providing gains for the end-user. Due their abstraction, FDOs follow at first instance a machine-centered design principle. Questions to be addressed in the further work is whether FDO based infrastructures can lead to creating value to the end users as envisioned by the presentation of the domain scientists and whether we have examples to learn from.

Use cases as explained in the second session need to be implemented as soon as possible since they indicate what researchers should be able to accomplish and thus can be used to extract feedback for the requirements and specifications of the FDOF. Only use cases will help in overcoming the natural scepticism of funders and researchers. Collecting such use cases is of great relevance to convince a critical mass. Even in early documents we need to address the researchers' goals and describe how FDOs help to achieving them. FDOF needs to be open for the current heterogeneity and needs to make clear how to integrate for example the domain of non-Handle identifiers. Here FDO can indeed learn from the LDP domain where mechanisms evolved over decades. The specification of types is crucial, and therefore it needs to be simple for scientific communities to participate. Only a federation of type registries will create the required level of acceptance.

Establishing a functioning FDO domain and integrating already existing mechanisms to achieve a broad buy-in will require substantial investments. Therefore, it makes sense to identify classes of users and look for scenarios enabling "easy wins" for these classes. Users and stakeholders with different backgrounds need to be approached with different types of information. The question was raised what kind of governance structure will be required to advance the FDOF and possible implementations. A thin layer will be required to bring people together to advance the work and

maintain a roadmap of activities. The relation to existing initiatives needs to be clarified. The discussion showed that the suggested FDOF needs additional work, for example, to emphasize the metadata aspects, needs to indicate different compliance levels and carefully chose footnotes and terminology. It is obvious that the current FDOF specification document is directed to the experts and that other documents need to be prepared for other stakeholders.

The discussion also made clear that we need an accepted authorisation structure that can define what FDOF should contain and what not and that we need to accelerate the specification process now. The RDA GEDE umbrella could be used to further indicate use cases, specifications and other activities around FDOs until a new structure has been defined. RDA groups are an excellent way to find broad agreements. The FDO community is at a tipping point and smart strategies are required without ignoring that the scientific communities are faced with a range of challenges.

Session 5: Comments on Joint Statement and Governance Structure

Chair: Jean Francois Abramatic; Speakers: Barend Mons, Robert Quick, Edit Herczog, Mark Leggott, Michel Schouppe

The session was opened by describing the development and intentions of the Web and by assessing the needs of the data community. The Web with its well-known fundamental standards HTTP, HTML and URI was primarily meant to quickly exchange text, graphics, images, sound and video. The appearance of killer applications such as Mosaic and later Google made it a globally used infrastructure. With respect to data roughly 3 phases of the Web can be identified: (1) exchange of XML structured documents and access to SQL queries; (2) the appearance of the Semantic Web with its many frameworks (RDF, OWL, SPARQL); and (3) the presentation of the Linked Data Platform (LDP). But it was concluded that data scientists were not convinced resulting in a low usage of the LDP framework.

Still eminent challenges such as machine-actionability, heterogeneity, scalability, versioning, disambiguation, reproducibility, etc. need to be tackled. The FAIR principles can be seen as milestones on the way to overcome some of the hurdles. The FAIR Digital Object Framework (FDOF) allowing different implementations and turning FAIR into reality can become another milestone. Therefore, publishing the FDOF is a contribution to the Open Science agenda and opens a gate to an infrastructure ecosystem for research data management across disciplines and borders. Since it seems to offer genericity, scalability, a path towards automation, prospects for reproducibility and proper referencing, it is difficult to see what an alternative could be. The vision included in FDOF needs to be communicated actively, champions have to be identified and expertise needs to be aggregated to support rapid uptake. These processes need to happen at a global level and the further use of RDA as an interaction platform to further specify missing details is needed.

Three themes were dominant in this session:

- How to come from rough consensus to running code?
- How to get an efficient and lean organisation going?
- What are the most urgent gaps to be addressed?

Obviously we need to create an **implementation group** with technologists on board or revitalise/merging existing groups as from C2CAMP, GEDE-DO and the persons indicated at an earlier meeting. It was obvious that such a group should improve and detail the FDOF specifications as guidelines for participation, involve industry at an early stage without risking vendor lock-in, organise hackathons, work on early demonstrators appealing to young technologists, start RDA groups for broad consensus finding, etc. For building first demonstrators it makes sense to start with basic

functions such as applying and adapting to DOIP. Scalability, long-term sustainability and feasibility of the solutions need to be demonstrated.

A **governance structure** is required which is interacting with organisations to join, takes care that the FDO/FDOF domain will speak with one voice, to protect the FDO/FDOF specifications from any commercial dominance, takes care of global consensus finding, interacts with industry, setting priorities, organises the outreach, engages intermediates to talk with different stakeholders and initiates standardisation activities allowing funders to join in. The most prominent task of governance is to create trust that there is more than a bunch of enthusiasts working on a new idea, but that there will be road mapping and continuity. A coordination group should be installed to work out a governance structure amongst others.

With respect to the third aspect many answers need to be found as soon as possible to questions such as (1) when will there be more specifications and who is driving the process? (2) What needs to be done to allow massive take-up? (3) How to make adaptation simple and get a critical mass of adopters? (4) What kind of testbeds can be started when? How does this all link to EOSC? Only a clear governance with committed experts will be able to convincingly address these questions.

Session 6: Conclusions

Chair: George Strawn

Three essential phases of computing were described at the beginning:

(1) Many computers, many data sets: nothing being connected;
(2) One computer, many data sets: from simple terminal connection to computers to the Internet connecting all computers and to Grids/Clouds enabling distributed processing; and
(3) One computer, one data set: a development we are currently designing by integrating data sets.

Some policy decisions have been made on this way to come to an integrated data area. Europe has launched the EOSC program and the FAIR principles have been accepted as guidelines. With the Digital Object Architecture and the Linked Data Platform two architecture proposals have been identified that urgently need to be further specified and tested. The Internet development was possible since DARPA had both time and money and it started without having application goals. We need to accept that specifying such complex infrastructures as EOSC implies learning by implementing. Rough consensus and running code are the best principles in such scenarios to achieve fast advancements.

In this sense it was unanimously agreed that the proposed draft FDO consensus paper and the first FDO Framework is an excellent start, but that it needs editing and improvement. It was also agreed that the consensus and the FDO Framework specifications should not be separated, although the FDOF specifications will be extended continuously in the future while the consensus paper will describe the message form the workshop. Since the FDOF need to be kept at an abstract level it makes sense to include short illustrations of possible implementations.

An **editing team** was built to work out an improved version of the joint consensus document until end November: Mark Leggott, Bonnie Carrol, Alex Hardisty, Carlo Zwölf and Peter Wittenburg.

It was agreed that in addition to this consensus paper we need a **flyer** until end of November that can be presented to funders and industry. The editing team includes Barend Mons, Dimitris Koureas and Jonathan Clark.

The current landscape of initiatives that are involved in the discussions of FDOs is heterogeneous and lacks steering which is capable to advance the FDO domain in the required speed and precision. The following was therefore agreed:

- A **Coordination Group** is required to start with steering activities and to work out a suitable governance structure.
- A **Technical Implementation Group** will be established that will intensify the specification work, work out roadmaps and initiate RDA Groups where necessary.
- **RDA activities** should be initiated to work on FDO topics and include a broad community in finding specifications.

The **coordination group** consists of Francoise Genova, Edit Herczog, Dimitris Koureas, Larry Lannom, Barend Mons, Rob Quick, Koenraad de Smedt, George Strawn, Peter Wittenburg, Carlo Zwölf [2]

An initial group of technical experts formed at a meeting in Washington should be extended by other committed experts to form the **Technical Implementation Group**.

It was agreed that all this should be organised in the realm of the RDA GEDE group as long as there is no other structure.

[2] The current co-chairs of the GEDE-DO activity, partly not around anymore at the end of the meeting, should join to demonstrate continuity to the broad group interested in FDOs.

Appendix

FAIR Digital Object Framework

Version 1.01, October 2019
Luiz Bonino, Peter Wittenburg

This document will be a dynamic one, i.e. it will evolve and be changed dependent on the comments and insights. This version is just a snapshot being presented and discussed at October and November meetings in 2019. It contains some framework principles which should be as technology neutral as possible. Due to this neutrality we added two illustrations of possible implementations and a glossary.

"We need a set of principles that are sufficiently specific to be useful but sufficiently abstract to exclude specific software stacks, i.e., a document that will still make sense and still be useful ten years from now."

General Guidelines

We can mention a few overall guidelines that need to be met by the FAIR DO Framework (FDOF). FDOF needs to

- **G1**: show a path for infrastructure investments for **many decades**
- **G2:** demonstrate **trustworthiness** to researchers and developers to become engaged
- **G3**: offer compliance with the **FAIR principles** being turned into indicators of FAIRness
- **G4**: support **machine actionability** which includes referential integrity, which states that all references need to be valid without temporal limitation, and explicitness of semantic relationships
- **G5**: support the **abstraction principle**, i.e. abstract away from details that are not needed at a specific layer such as at management layer there is no difference to be made between data, metadata, software, semantic assertions, etc.
- **G6**: support **stable binding** between all informational entities that are required for machines to act
- **G7**: support **encapsulation** which means that operations can be specified by types
- **G8**: support **technology independence** allowing implementations using different technologies[3]

Requirements for FDOF

Here we mention the requirements that emerged during the recent discussions. They need to be verified and checked on completeness.

- **FDOF1**: A PID, standing for a globally unique, persistent and resolvable identifier, is assumed to be the basis of the Internet of FAIR Data and Services.
- **FDOF2**: A PID is resolved to a structured record with attributes which are semantically defined and semantically defined within an ontology[4].
- **FDOF3**: The structured record includes at least a reference to the locations where the bit-sequences encoding the content of a FAIR-DO[5] (FDO) can be accessed, a PID pointing to the metadata FDO(s) describing properties of it, including the DO's type.

[3] Currently, we can refer to three technologies that have ambitions to implement the FDOF requirements: DO Architecture, Linked Data Platform, Database Technology.
[4] Different options such as a type registry can be thought of.
[5] In the LDP domain this is called a resource.

FDOF4: Each FDO identified by a PID can be accessed or operated on using an interface protocol by specifying the PID of a registered operation and the PID of the access point.
FDOF5: This protocol offers the typical CRUD operations on FDOs and a possibility to use extended operations.
FDOF6: The relations between FDO Types and operations are maintained by a type ontology.
FDOF7: Metadata descriptions being FDOs and describing the properties of the FDO are made available as semantic assertions[6].
FDOF8: A collection[7] of FDOs is an FDO and semantic assertions are to be used to describe their construction.
FDOF9: The "Deletion" of a FDO leads to standardised and thus machine interpretable tombstone notes in the metadata and PID records, i.e. only the bit-sequences of the FDO will be deleted.

Appendix: Digital Object Domain

Here we briefly summarise what is available, what is missing and what can be read.

Req	available	missing
G1	basic construction & intention is meant to survive for many decades	broadest uptake to ensure survival
G2	trustworthiness is a social concept and depends on reliability and uptake	solidity of all components is given, broad mobilisation is to be achieved
G3	FAIR compliance to a certain extent (see below)	some higher level specifications are missing
G4	all aspects of DOs specified yet are machine actionable	some higher level specifications are missing
G5	abstraction is at the core of the DO concept	no
G6	stable binding is realised by using the PID record	no
G7	encapsulation is intended and implemented through DOIP	no
G8	DOs are one implementation of FDOF	
FDO1	based on clearly defined PID systems such as Handle	no
FDO2	resolution is a structured record and attributes should be defined and registered[8]	miss an authority to maintain registry
FDO3	should be specified by FDO service providers (repositories)	miss an authority to define best practices
FDO4	model allows access through PIDs, DOIP allows association between types and operations	no
FDO5	DOIP has these features	no
FDO6	RDA specified a type registry, is being used	more complex ontologies might be necessary
FDO7	metadata are indeed FDOs; metadata availability as assertions is possible	miss specifications associated with the DO model
FDO8	collections are DOs; construction of collections not defined	miss specifications associated with the DO model
FDO9	tombstone notes are possible, yet not defined	miss an authority to define best practices

[6] Currently, RDF assertions are generally used.
[7] In the LDP domain this is called a container.
[8] It should noted that some Handle Services do not support machine actionable types in the Handle record.

The digital object approach provides a framework for the various components needed for a FAIR DO Framework but any implementation will require further specification and one unspecified piece is how to get to the explicit semantics needed for machine understanding of structured metadata and collections.

Available readings

- R. Kahn, R. Wilensky (1995): A Framework for Distributed Digital Object Services; https://www.cnri.reston.va.us/k-w.html
- R. Kahn, R. Wilensky (2006): A Framework for Distributed Digital Object Services; https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf
- RDA DFT Group: DFT Core Terms and Model; http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318
- RDA DTR Group: Data Type Registry, https://www.rd-alliance.org/group/data-type-registries-wg/outcomes/data-type-registries
- RDA Kernel Group: Recommendation on PPID Kernel Information; https://www.rd-alliance.org/group/pid-kernel-information-wg/outcomes/recommendation-pid-kernel-information
- RDA Research Collection Group: Recommendations; https://www.rd-alliance.org/group/research-data-collections-wg/outcomes/rda-research-data-collections-wg-recommendations
- DONA: DOIP V2.0, https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf
- P. Wittenburg, G. Strawn: Common Patterns in Revolutionary Infrastructures and Data; http://doi.org/10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0
- P. Wittenburg, G. Strawn, B. Mons, L. Bonino, E. Schultes: Digital Objects as Drivers towards Convergence in Data Infrastructures; http://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11
- S. Hodson et. al.: Turning FAIR into Reality; https://doi.org/10.2777/1524
- E. Schultes, P. Wittenburg: FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure; http://doi.org/10.23728/b2share.166a074bff614a31b05e9df5bfd9809d
- G. Strawn: Open Science, Business Analytics, and FAIR Digital Objects; http://doi.org/10.23728/b2share.6ceeed13eb6340fcb132bcb5b5e3d69a
- K. de Smedt, D. Koureas, P. Wittenburg: Analysis of Scientific Practice towards FAIR Digital Objects; http://doi.org/10.23728/b2share.e14269d07ce84027a7f79ee06b994ef9

Appendix: Linked Data Platform

Here we briefly summarise what is available, what is missing and what can be read.

Req	available	missing
G1	The corpus of LDP-related recommendations provide a clear investment path for a reasonable amount of time.	Increase the adoption base of the technology.
G2	LDP is an W3C recommendation and W3C is recognised as a trustworthy standardization organization.	
G3	LDP and associated W3C standards facilitate the compliance with the FAIR Principles, but users need to behave in certain ways to do so.	More strict guidelines on how to use LDP in ways to better follow the FAIR Principles
G4	LDP and Linked Data provides the technological ground for users to provide explicit semantics with qualified references.	
G5	Layers of abstractions can be introduced using	More strict guidelines on how to

	LDP/RDF.	use LDP in ways to better support the abstraction principle.
G6	Once the relations/bindings are defined, they are there until the resource is removed or the users make changes.	
G7	LDP supports the HTTP methods/operations	
G8	LDP is a technology based on RDF and, therefore, is technology dependent.	
FDO1	LDP adopts URI as a globally unique, persistent and resolvable identifier.	A consistent resolution behaviour that doesn't depend on user's best practices is still missing.
FDO2	What is resolved from the URI is up to the creator of the resource. It is possible to define the resolution to this structured record and LDP/RDF provides infrastructure for semantically describe this structured record.	Instructions on how to use LDP to return the URI's structure record including the semantic references to the record's elements.
FDO3	Same as above	Same as above
FDO4	LDP/RDF can be used to provide the semantic description of the operations supported by each DO type.	The type ontology, including the description of the operations need to be defined.
FDO5	LDP supports HTTP methods that provide CRUD functionality.	Extended operations need to be defined.
FDO6	LDP/RDF the definition of the semantic descriptions required by the type ontology.	no
FDO7	LDP/RDF supports semantic descriptions of metadata elements through qualified references to existing volcabularies/ontologies.	no
FDO8	LDP defines the concept of container, including three types of containers and the relations between container and its member elements.	no
FDO9	Tombstone notes can be semantically described using LDP/RDF.	Instructions on how to construct the tombstone notes using LDP/RDF and update the identifier's structure record to point to this note.

Available readings

- Linked Data Platform 1.0 - https://www.w3.org/TR/ldp/
- Linked Data Platform 1.0 Primer - https://www.w3.org/TR/ldp-primer/
- Linked Data Platform Best Practices and Guidelines - https://www.w3.org/2012/ldp/hg/ldp-bp/ldp-bp.html
- RDF 1.1 Primer - https://www.w3.org/TR/rdf11-primer/

Appendix: Glossary

A short glossary with explanations about crucial terms such as "repository", "encapsulation" etc. will help in clarifications, since some terms may be interpreted differently by the participants.

Term	Explanation
abstraction	Abstraction is a conceptual process where general rules and concepts are derived from the usage and classification of specific examples. literal signifiers, first principles or other methods (Wikipedia)
binding	With binding we mean the possibility for humans and machines to find other relevant entities of a DO when being exposed to another, i.e. when an actor receives a PID of a DO it must find the PID of the corresponding metadata DO and the access rights information, since otherwise interpretation and access is impossible
Collection	A collection is a complex DO consisting of other DOs, that have a PID and metadata.
CRUD operations	These are the usual primary type of operations such as create, read/retrieve, update and delete
encapsulation	Encapsulation is known from abstract data types and oo programming where internals of data objects are hidden to the user and where the user can only influence the internal state by using defined methods
	Note: in the FDO case DO types can be associated with registered operations that can be used to operate on DO's content
machine actionability	With machine actionability the capacity of computational systems is meant to find, access, interoperate and reuse data and services without human intervention (GOFAIR)
metadata	Metadata descriptions of DOs are sets of assertions describing properties of DOs content which are required for finding, accessing, interpreting and reusing, these assertions can cover a wide range such as descriptive to support finding, deep scientific to support science, systemic to support management, rights to prevent unauthorized access, etc.
	Note: Yet the domain of metadata is not structured very well, i.e. terminology is not well-defined.
	Note: Basic interoperability assumptions are that the schemas are registered and the concepts defined and registered.
repository	**DO View**: from the perspective of Digital Objects repositories are nothing else than a complex DO associated with a PID, metadata of different kinds and functions to offer DOs
	Common View: from the most common point of view repositories are entities that host data, metadata etc., apply trustworthy management procedures, offer a search and access interface, have a team of experts taking care and have a sustainability plan
	Note: repositories can be associated with research organisations, communities or projects, they can be small or big in terms of the collections they hold.
type	"Type" is an attribute of digital objects which tells computational actors how the content of the DO needs to be parsed, i.e. it defines the operations that can be done on the data, the meaning of the data, and the way values of that type can be interpreted
	Note: A MIME type is a standard that indicates the nature and format of a document, file, or assortment of bytes, i.e. it is a restricted concept of type.
	Note: A type of a DO implies a summary of otherwise complex metadata assertions describing the format, encoding etc. of a content.