GEDE DO Topic Group

An Analysis of Scientific Practice towards FAIR Digital Objects

Koenraad de Smedt, Dimitris Koureas, Peter Wittenburg
March 2019, Version 4

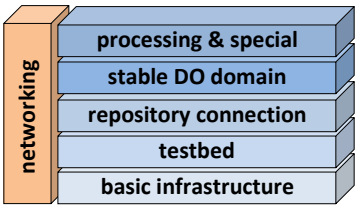
1. Introduction

The GEDE Digital Object Topic Group initiated a request about cases of scientific practice that may indicate the relevance of FAIR[1] Digital Objects (FAIR-DO) beyond which was stated in the two papers on that matter [1, 2]. In total, we received 31 responses from 20 different European research infrastructure initiatives, 2 from US institutions[2] and 1 from a Chinese institution. The main aim of the current paper is to summarise the responses and draw some conclusions.

In addition, we have obtained the commitments from the International DONA Foundation in Geneva about the maintenance of essential components of the FAIR-DO domain. We remain open for additional descriptions of scientific practice from any initiative engaged in GEDE-DO, C2CAMP and international colleagues, since some colleagues have indicated the need for further intensive discussions in order to better understand the potential and limitations of the FAIR-DO concept.

An analysis of all 34 submitted cases leads to the following rough classification:

1. networking interest (many)
2. basic infrastructure, services and operations (5)
3. extendable testbed (1)
4. connection of repositories (5)
5. stable domain of scientific entities (10)
6. automatic processing (7)
7. special wishes (5)



These classes implicitly include dependencies in so far as the higher classes depend on the lower ones. Creating a stable domain of digital scientific entities, for example, will rely on basic services and include repository connections. Testbeds will always be necessary if new features and services will be introduced. This is briefly indicated by the diagram. It also implies that several scientific cases in principle could be mentioned in more than one category.

2. Analysis Results

2.1 Networking Interest

About 150 experts, mainly from European infrastructure initiatives and some international collaborators, have expressed their interest to participate in a network that will organise meetings, education and dissemination activities. The main aims are to exchange methods for scientific practice using Digital Objects, to foster the construction of testbeds supporting such practice, and to disseminate information about relevant solutions in relevant communities. The following major challenges were identified:

[1] https://www.force11.org/group/fairgroup/fairprinciples.
[2] Some contributions include different use case suggestions for DO related activities.

- Bottlenecks in cutting edge data-intensive science use cases
- Missing DO specifications and missing insights about their potential impact on research practices and infrastructures
- The lack of platforms to exchange know-how about DO architecture specifications, their usefulness in various disciplines and relevant shareable code
- The lack of testbed projects that can drive the transformation, accelerate the specification work in RDA and interact with comparable initiatives working with different foci
- The need for mobilisation of relevant communities
- The need for capacity building for specification and implementation work
- The need for training of young data scientists in order to transfer the evolving knowledge and skills from early adopters

This interest in networking at a global scale, but also in particular at the European level, is of great relevance to promote fast bootstrapping actions in various scientific domains. The existing network of mainly scientific organisations must be extended so as to include commercial companies, since a broad infrastructure can only be realised by involving companies in their double roles as technology providers and consumers.

2.2 Basic Infrastructure, Services and Operations

The realization of a FAIR-DO based eco-system of infrastructures can only be enabled by making some essential key components and their services available. These have already been identified by the Digital Object Architecture work, including the DOIP V2.0 specification and the discussions around FAIR-DOs that has been going on in RDA DFT, RDA DF, GOFAIR and EC's FAIR Expert Group. The following developments can be mentioned.

FAIR-DO and DO Concepts: The basis for all work in the FAIR-DO domain consists of the definitions of FAIR-DOs as specified by the RDA DFT group and the DO definition as described in the DOIP protocol. The rationale behind these and their close relationship are explained in URL to come.
DOIP: At the core of DO infrastructures is the **DO Interface Protocol** (DOIP) that defines an interoperability layer between repositories and registries of Digital Objects. The independent, non-profit Swiss DONA Foundation is committed to maintaining this protocol and to keep it free from patents and licences that would hamper its free distribution. The DOIP includes standard operations such as create, delete, move, etc. DOs in a cardinal form.
Handle System: The other major pillar of a global connected FAIR-DO domain is a robust, persistent infrastructure that allows everyone to register and resolve persistent and unique identifiers. The DONA Foundation will also take care of their distribution from a non-profit and self-sustained root system for Handle registration and a resolution domain that will follow open principles.
Handle Service Providers: There are a number of Handle Service Providers including the DOI community which offers DOIs, the ePIC community, and many others. However, there is a need to establish a robust system in Europe that can be used by everyone in science and industry.
CORDRA: This is a prototypical repository that is compliant with the DOIP. It can be seen as a blueprint for the organisation of repositories and for methods to interface with those.
Kernel Types: The RDA Kernel Type group has started to standardise the attributes for Handle records in order to achieve a maximum of interoperability. Maintenance of these standards needs to be put into professional hands with a view towards integration in scientific and industrial applications.
Data Type Registry: The RDA DTR group has specified the mechanisms for Data Type Registries linking Types of FAIR-DOs with operations.

Repositories and **registries** are key pillars in the whole FAIR-DO landscape. Many of these are in operation; huge investments have so far been made to set them up and more recently to make them stepwise FAIR compliant [3]. With respect to registries, not all essential types are clearly specified yet

and not all of the existing ones are suitable to provide services in a FAIR-DO domain. One of the responses to our inquiry explicitly refers to the need for professionally maintained kernel type registries which are prerequisites for interoperability.

Among current contributors are DONA (CH), CNRI (US), CNIC (CN) and GWDG (DE). These organisations offer stable services for others interested in exploring and using the FAIR-DO capabilities.

2.3 Extendable Testbed

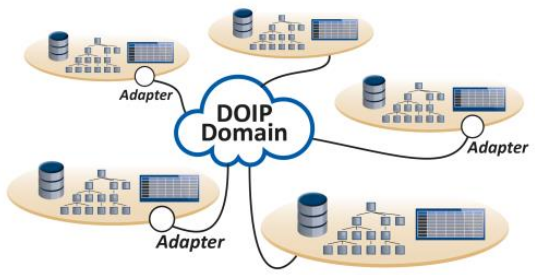
All the components mentioned in Section 2.2 must be integrated and tested in larger configurations based on the needs of scientific communities and their practices. Innovative needs such as FAIR-DO-compliant automated processing guided by workflows, or a stable domain of semantic knowlets (see below), will almost certainly lead to additional requirements for the basic infrastructure. Only extensible testbeds that evolve into real infrastructures will help in overcoming hurdles. One of the biggest gaps that needs to be filled through experience from testing is the adaptation of existing repositories. Furthermore, we need to make this adaptation scalable.

All responses provide input for the testbed and add to our understanding of the needs and bottlenecks in our way forward. One of the responses explicitly mentions the testbed as a goal. The envisaged testbed must integrate all components necessary to test a full Digital Object Architecture and in particular to connect existing repositories with the DOIP.

The contribution explicitly mentioning a testbed is IU (US).

2.4 Connection of Repositories

A few respondents mention explicitly the connection of their repositories, in the form of file systems or clouds, to the DOIP domain. In the end, this is a task for all communities, as indicated in the diagram. The connection effort for various cases may not be underestimated. If, for instance, a repository uses a relational database which contains data and metadata, both the data and the metadata objects will have to be identified clearly. This can only be accomplished by executing queries, and additionally using time stamping, to get the expected results. This implies that queries must be associated with PIDs. In some cases, e.g. the CMIP case in climate modelling, an entire data architecture must be made DOIP compliant.



Since DOIP includes prototypical code for the basic DO operations like create, delete, copy etc., the connection of repositories also implies the need for microcode for different repository setups. The creation of a DO, for instance, presupposes a storage location for the bit-sequence, the generation of a checksum, the creation of a digital object to contain descriptive metadata, the registration of a PID including the addition of attributes such as checksum, location of the bit sequence etc., the addition of the PID to this metadata, the creation of a record containing the access rights and many other details. Depending on the repository's data organisation and technological platform, different microcode will be required.

The connection of repositories will be a challenge for all discipline. The effort depends largely on the data organisation and technologies applied in the respective repositories. Connection challenges are explicitly mentioned by ENES, CLARIN, NOMAD, VAMDC and ICOS.

2.5 Stable Domain of Scientific Entities

Seven responses mention the establishment of a stable and persistent domain of Digital Objects as their key goal, although this goal may be implicitly assumed in other contributions as well. Here we want to focus on those that were explicitly mentioned.

The **DISSCO** (Distributed System of Scientific Collections) initiative on natural sciences addresses the challenges towards establishing a stable system of clearly identified Digital Objects representing huge numbers of physical specimens (plants, animals, insects, fossils, rocks, minerals, etc.) in natural science museums and archives. These specimens can be the object of annotations based on multiple information sources, taxonomies, virtual collections, theories and in part also user operations (i.e, scientific analysis) leading to derived data, assertions and other activities. The natural sciences domain thus consists of many different and interconnected, parallel layers of different types of information associated with individual physical specimens and collections. The inherent capabilities of abstraction, binding and encapsulation of FAIR-DOs based on stable identifiers will establish trust relations with researchers working in this digital domain for the next decades. A thorough design of the domain of FAIR-DOs for natural sciences will hundreds offer new capabilities to researchers and is the basis for widening access to collections with lifetimes of hundreds of years.

One of the big challenges of the **ECRIN** initiative is managing highly sensitive clinical trial data coming from different hospitals and making these data available for medical research. A redesign of the data domain using FAIR-DOs with clear identities, bound metadata descriptions and inherent tracing possibilities (via blockchain entries, for example) offers a way to overcome many fundamental hurdles for an easier exchange of trial data which is the basis for their efficient reuse for research purposes across hospitals. Similarly, operations to be carried out on FAIR-DO collections will also consist of FAIR-DOs with identifiers etc. allowing the ECRIN community to effectively trace all reuse of trial data and will enable efficient standard operations such as the identification of useful data or tools for a given purpose.

The **EISCAT** initiative produces terabytes of data from radar systems. Their complex pre-processing and analysis of raw antenna data makes it extremely important to adopt new structural decisions towards FAIR-DOs that can fulfill a key role based on their clear identification and binding capability. Raw data will be reduced due to varying parameter sets and will be channeled into suitable collections ready to be processed by user defined operations. These kinds of data will also be combined with data from other observation sources. The complex relationships between data sets will require a stable identification of the data and their relations. EISCAT sees a huge potential in applying the FAIR-DO concept towards meeting their goals.

The **ELIXIR** initiative in the life sciences is looking for a FAIR compliant interoperability platform across borders (between disciplines, labs, countries, object types, services, etc.) that has the potential to offer a stable domain of digital objects over decades. Such a domain would open the path towards an efficient use of FAIR-DOs in computational systems and advanced workflows. FAIR-DOs with their clear identification and binding capability are seen as basis for such an interoperability domain.

The agricultural research community at **Wageningen University** is increasingly interested in supervising and optimising entire food chains from the agricultural source up to the final consumable product. Different types of actions represented by FAIR-DOs (data, tools, protocols, models, etc.) are engaged in this process. Abstract DOs with clear identities are seen as optimal to come to the needed integrated view and to be able to optimally exploit the rich information. Such a well-structured domain would also be a suitable platform to include other data sources such as about health effects to optimally plan agriculture. New opportunities emerge from the application of FAIR-DOs.

The **E-RIHS** initiative in the cultural heritage domain is confronted with a wide variety of different data types coming from many different sources and often representing physical artefacts. Therefore, it is an urgent necessity to host all data on a stable platform where FAIR-DOs with their inherent clear identification, types and binding capacities will be the key entities. Relationships to ontologies, different views and annotations could all be captured in a stable domain for research. FAIR-DOs would pave the way to make digital humanities a data-driven discipline where provenance is tightly linked with the objects.

The **MIRRI** initiative is connecting a large variety of microbial databases, each having a different structure, using different systems and having different export capabilities. MIRRI's primary objective is to make these databases FAIR compliant. The implementation of a FAIR-DO domain is seen as a way to implement FAIRness. Making the MIRRI databases FAIR is the path to improved exploitation of the joint knowledge captured in them to the benefit of controlling the stability of our circumstances of life.

The **GESIS** organisation in social science wants to combine social science survey data with geological data to study the effects of geo phenomena on health conditions, well-being, attitudes and other social indicators. But this is only the first example of bringing together data from different sources. The creation of a linked domain of data is considered now and from the beginning it should be based on FAIR-DOs being able to implement FAIRness. Also in this application it is the inherent capability of assigning clear identities with all data items and the capability to store all relationships in a stable way allowing re-purposing the emerging data domain in a flexible way that make FAIR-DOs a natural solution.

The **ForumX i**nitiative wants to build an interoperable domain of FAIR DOs within the German Research Infrastructure program. It will start with a DO-based integration of data from various experimental sciences (economics, psychology, neurology, etc) and then extend the DO-domain to other interested disciplines active in this German infrastructure program. The work will include the design of a stable domain of digital entities, an adaptation of existing repositories and include advanced analytical processing which need to be worked out in more detail.

The **INSTRUCT** initiative wants to establish a research infrastructure in structural biology that will be based on the DO approach which has the needed capacity of integrating the many datasets that currently exist in that domain. It will also support the distribution of data more easily and opn the path to an increased exploitation with help of workflows. It is intended to start applying this approach in the context of the EOSC-LIFE project.

There is quite some overlap between the intentions and suggestions of these 10 responses.

2.6 Automatic Processing

Seven responses express an interest in applying the FAIR-DO concept in the context of their increasing needs for automating data processing with the help of workflows.

The **CLARIN** infrastructure for language data and technologies has long been working on workflow orchestration, allowing researchers to perform complex linguistic analyses of texts in different languages. They are currently working on a Language Resource Switchboard (LRS) that can easily identify which tools are compatible with which text types. Using only MIME type indicators is not sufficient for their effective profile matching. The use of FAIR DOs that binds the texts with its rich metadata descriptions of complex types is a richer basis for finding useful connections between datasets and tools. The LRS opens the path towards easier integration of data and tools coming from various contributors.

The **ENES** community produces large amounts of data from the various climate simulations. Their contribution suggests three workflow types of activities where FAIR-DOs will play a role. The sheer mass of data (100PB) resulting from different simulations requires a transition to automated data management which profit from DO capabilities such as abstraction, binding and encapsulation – the latter in particular showing the path towards automation. This is followed up by the required automated support of all processing stages (from production to quality control and publication) where full provenance tracking becomes increasingly important. A third, more challenging project is the step towards actionable digital object collections. These extend PID-based "shopping baskets for data" with the identification of possible operations (which will also be FAIR-DOs) on the data.

The **NOMAD** initiative active in the area of computational material science aggregates large amounts of simulation calculations about compound materials from many labs worldwide. They are produced by about 40 different software packages and are curated, cleansed and normalised. The goal is to find new patterns in the entire collection, as well as new types of classifications, by means of advanced analytic tools currently provided through a tool kit. Intended steps for NOMAD are to increase the FAIRness of their methods by implementing a FAIR-DO domain and to extend their analytic toolkit concept with a workflow framework allowing users to build collections and then execute a sequence of analytic operations on the selected data.

The **DEWCom** initiative, operating in the realm of the IEEE, is extending the classical centralised cloud computing paradigm to a distributed scenario where on-premises computational facilities make use of cloud computing frameworks. While on-premise computers may provide functionality independently, they can also tap the huge capacities of cloud computing to support modern AI methods such as deep learning. Thus DEWCom computing is meant to efficiently support data intensive science. It appears that the concept of DOs is the optimal way to support the many data driven operations to be executed in DEWCom frameworks. The DEWCom software stacks should be amended to include the capacities of FAIR-DOs in the sense of abstraction, binding and encapsulation. One result of such an approach would be systematic correct provenance recording.

The **CNRI EAGER** project studies and tests how data types identified by PIDs can unleash the rich domain of possible operations. Relationships between DO types and operations open the path towards simpler workflow orchestration. Furthermore, the use of additional context and provenance information in metadata opens the path towards automatic processing without human involvement. Studies are urgently required to understand the full potential of this paradigm and to identify the mechanisms required.

The contributions which address automatic processing address scientific intentions as well as motivations for adding infrastructural elements.

2.7 Advanced Proposals

Five suggestions contain desiderata that cannot easily be classified under the above headings.

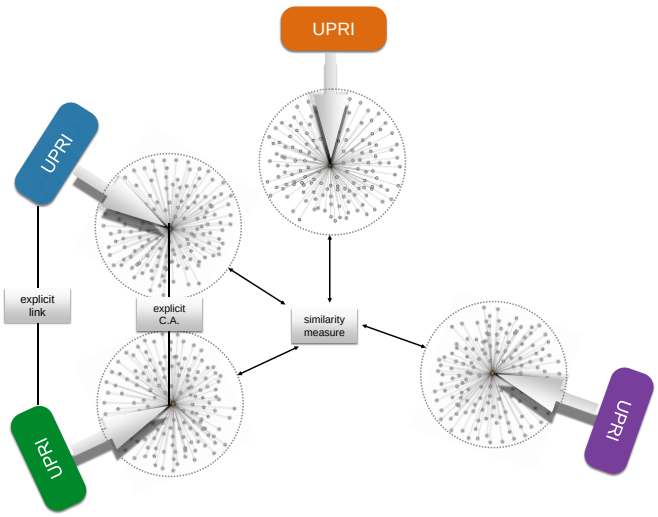
The **CLARIN** initiative is providing a collection builder tool that allows researchers to collect data from various sources into virtual baskets that can be used for management tasks, scientific analyses or simply as a collection to be referenced in papers. This innovative tool is currently relying on PIDs, but should evidently embrace the FAIR-DO concept fully. Especially the binding capacity of FAIR-DOs would be beneficial in this respect. This tool would not only be applicable to CLARIN, but to other data domains as well.

The **ENES** initiative is designing a Virtual Research Environment (VRE) for climate modelling studies with tailored services that can also be offered to non-professional users. Such a VRE needs to support certain levels of abstractions which the FAIR-DO domain offers. This implies that the researcher

should only need to address FAIR-DOs through their PIDs and metadata, without needing to know their type, internally structure or place of storage.

The **NOMAD** initiative on experimental studies about compound materials will extend their scope with respect to the large number of experimental labs, all working with different experimental settings and equipment. Such experiments are often described in electronic lab books which are currently disconnected from other processes. The intention is to turn lab book entries into FAIR-DOs with all their inherent capabilities and create blockchain entries linked to the FAIR-DOs' identities which document the exact time, place, results etc. of each experiment. It is the binding capacity of the FAIR-DOs that opens such new linking of secure information.

The **GOFAIR** initiative suggests that the FAIR-DO concept be applied in order to fundamentally improve operations in the semantic domain. So-called knowlets (see diagram) are attractive ways to organise the semantic domain of assertions (augmented RDF, nano-publications) and identify clusters of concepts, the central concepts of such clusters, their internal relationships and their relationships to other clusters. This analytical approach to a domain of concepts used in assertions about observations may help unleashing unseen patterns. However, domains of assertions are not stable as new assertions may be introduced, new insights may lead to changes, etc. The systematic use of the FAIR-DO concept with its binding to metadata, relationships and in particular provenance builds a stable fundament for analyses that, for instance, examine older states and their evolutions. In this domain, each assertion, each concept and each central concept will be identified by a PID, and the PID records will have pointers to their relationships. Knowlets are thus virtual structures on top of the huge number of assertions. There will be different sorts of these layered virtual structures.



Finally, the **VAMDC** initiative is working on an interoperable e-infrastructure that provides the international research community with access to atomic and molecular (A&M) data. The broad range of these data in various databases prompted recognition of the urgent need for the assessment of data provenance, data authenticity and scientific quality. Three possible improvements of current practices through the adoption of the FAIR-DO concept are indicated: (1) a generic platform for the evaluation of scientific FAIR-DO and for error tracking, based on AI methods; (2) a system for tracing the provenance (i.e. production workflow) of each FAIR-DO; 3) a mechanism joining 1 and 2 for the recursive propagation of warnings to all entities which used FAIR DOs containing errors.

3. Conclusions

First, we can distinguish the following areas of activities according to the type of main actors whose involvement is crucial:

	domain scientists	IT specialists
networking interest		
basic infrastructure, services and operations		
extendable testbed		
connection of repositories		
stable domain of scientific entities		

automated processing		
advanced proposals		

- Networking is of interest to all. Close interactions keep a clear focus on the needs of the scientific communities and their practices.
- The construction and testing of a basic infrastructure by bringing components together in a testbed and connecting the repositories is mainly an task for IT experts.
- The design of a domain of scientific entities is mainly a task for the scientists, with the help of IT experts.
- Automated data processing is still in its infancy and will require a close collaboration between domain scientists and IT experts.
- Advanced proposals will require combinations of advanced expertise.

The results as a whole suggest that a 3-way funding strategy is necessary:

1. funds for networking;
2. funds for establishing and maintaining the infrastructure;
3. funds for focusing on the direct scientific needs.

4. Contributors

The following initiatives and organisations responded to our inquiry (see also Appendix):

- Europe: Agriculture WU, CLARIN, DEWCom, DISSCO, DONA, ECRIN, E-RIHS, EISCAT, ELIXIR, ENES, EUDOn collaboration, ForumX, GOFAIR, GWDG, ICOS, INSTRUCT, MIRRI, NOMAD, VAMDC
- US: CNRI, CNRI EAGER, IU EAGER
- China: CNIC-CAS

5. References

[1] P. Wittenburg, G. Strawn: Commenting on “Digital Object” Aspects; http://doi.org/10.23728/b2share.2317b12321764f669c92ebbcf7518164
[2] 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
[3] M.D. Wilkinson, M. Dumontier, et. al: The FAIR Guiding Principles for scientific data management and stewardship; https://www.nature.com/articles/sdata201618

Appendix: GEDE/C2CAMP DO Contributions

Version 23.3.2019

This table lists use cases, suggestions and potential contributions from various disciplines towards the creation of a DO-based data infrastructure. At this stage only a few communities seem to anticipate the needs in a 10 year perspective. We distinguish several states: active (green) means that actual work is going on; waiting (yellow) means that a decision on funding is expected soon; in preparation (orange) means that concrete plans are being developed with a clear plan for implementation; suggested (blue) means that discussions are in an early stage. More detailed descriptions can be found at two places in the GEDE DO Share:

A) written descriptions: https://datashare.mpcdf.mpg.de/index.php/apps/files?dir=/GEDE/digital%20objects/DO%20Participation%20Requests&fileid=68600770,

B) workshop slides: https://datashare.mpcdf.mpg.de/index.php/apps/files?dir=/GEDE/digital%20objects/GEDE-DO%20meetings/workshop-september-18&fileid=63958403

Community/ Project	Proposer/Contact Person	Institution	Category	Status	Short Description
CNRI	Giridhar Manepelli, Larry Lannom	CNRI US	basic	active	CNRI continues to do the technical development behind DOA, in support of DONA, and is one of the DONA MPAs.
DONA	Christophe Bianchi	DONA HS	basic	active	Swiss Foundation responsible for overall governance and direction of the Digital Object Architecture, including the identifier/resolution system and the Digital Object Interface Protocol.
ENES	Tobias Weigel	DKRZ DE	basic	suggested	the need for registries of different types (kernel information, etc.) needs to be supported either at generic or community level
GWDG	Ulrich Schwardmann	GWDG DE	basic	active	offering Handle prefixes, PID services and DTR services to register types; organising the ePIC consortium for redundant PID services
various	Liu JIA	CNIC-CAS	basic	suggested	CNIC as the major information service enter of the Chinese academy wants to extend their current PID offer to offer DO/DOIP services to their customers. CNIC sees the great relevance of the DO approach and will look now actively to start testbed projects with interested communities

Agriculture	Willem Jan Knibbe	U Wageningen NL	domain	suggested	agriculture is looking for more optimal methods to combine data of various types and organizations with tools in order to optimise food production chains, so a systematic use implementation of the DO concept offers new chances to solve challenges which are currently impossible, and thus a fundamentally new design step would be useful
DISSCO	Alex Hardisty	Cardiff U UK	domain	suggested	design and implementation of a stable domain of digital entities for natural sciences
ECRIN	Wolfgang Kuchinke	U Düsseldorf	domain	suggested	the complex domain of sensitive clinical trial data requires a fundamental reorganisation to make it possible to make more efficient use of the data for research purposes, the DOs are the best way to structure the clinical data domain
EISCAT	Ingemar Häggström	Kiruna SE	domain	suggested	the complexity of EISCAT's landscape of digital objects, in particular data, requires a new approach and design and the systematic use of DOs seem to be a good candidate to manage the necessary change, most important will be to maintain all relationships between the different raw and derived entities stable over years.
ELIXIR	Nick Juty	U Man UK	domain	suggested	looking for a FAIR compliant interoperability platform across borders (disciplines, labs, countries, object types, services, etc.) that has the potential to offer a stable domain of digital objects over decades. Such a domain would open the path towards an efficient use of DOs in computational systems and advanced workflows.
E-RIHS	Sorin Hermon	Cyprus Inst	domain	suggested	design and implementation of a stable domain of digital entities for archaeology and cultural heritage to implement FAIR data, establishing a stable domain of identities and relations between the related data sets making use of the binding function of DOs
ForumX	Dirk Betz, Claudia Biniossek	U Magdeburg	domain	suggested	the ForumX project wants to build an interoperable domain of FAIR DOs within the German Research Infrastructure program; it will start with integrating data from various experimental sciences (economics, psychology, neurology, etc) and then extend to other disciplines; the work will include the design of a stable domain of digital entities and also an adaptation of existing repositories
GESiS	Peter Mutschke	GESiS	domain	suggested	bringing together data from social and geo sciences into a new linked data domain based on the stable and binding DO concept
Instruct	Antonio Rosato	CERM/CIRMMP	domain	suggested	as a research infrastrucutre in structural biology, INSTRUCT is now convinced that a DO approach can facilitate the integration of the many datasets that currently exist and their dissemination and exploitation with help of workflows; they want to start applying this approach in the context of the EOSC-LIFE project.

MIRRI / EOSC-Life	Alexander Vasilenko	VKM IBPM RAS RU	domain	active	MIRRI wants to use DO as a basic harmonised interoperability layer between the biological databases and microbial culture collections system hub, which includes the latest databases integration technology, FAIR, variant taxonomies, creation of adapters for their db
EUDOn/GEDE	Koenraad de Smedt, Dimitris Koureas, Peter Wittenburg	U Bergen, Naturalis, MPCDF	network	suggested	networking on DO matters, workshops, documenting, etc.,
CLARIN	Dieter van Uytvanck	CLARIN-EU NL	repository	suggested	adopt the DOIP, investigating a suitable approach for CLARIN repositories and develop the necessary microcode and setup the needed structures
ENES	Tobias Weigel	DKRZ DE	repository	active	entering basic DO aspects into the cmip6 model, making repositories etc. DO compliant
ICOS	Maggie Hellström	U Lund SE	repository	suggested	the integration of the ICOS repository into a DO domain is a first step of interest
NOMAD	Petr Wittenburg, Carsten Baldauf	FHI DE	repository	suggested	making the NOMAD repository, archive, and encyclopaedia DO compliant with support for mirrors and distributed/federated operation
VAMDC	Carlo M Zwölf	Paris FR	repository	suggested	Extending the current VAMDC Query Store facilities (elements stored in those service are atomic and molecular data which may be assimilated to DO) for a full DOIP compliance.
CLARIN	Dieter van Uytvanck	CLARIN-EU NL	special	in prep	extending the existing Collection Builder to support DOs
ENES	Tobias Weigel	DKRZ DE	special	suggested	a VRE for climate modelling studies offered also to non-professional users with taylored services is planned requiring levels of abstractions which DOs offer
GOFAIR	Erik Schultess	GOFAIR NL	special	suggested	knowlets stabilisation
NOMAD	Petr Wittenburg, Carsten Baldauf	FHI DE	special	suggested	experiments in material science are described in electronic labbook entries; we suggest tools that transfer such labbook entries including the measurement results to DOs and making use of the possibility to integrate a blockchain entry that documents the experiment
VAMDC	Carlo M Zwölf	Paris FR	special	suggested	Using advances DO interoperability facilities for building crucial mechanisms for data-driven science and FAIR principles: 1) a generic platform for evaluation of scientific DO and error tracking (based on AI); 2) A system which trace the provenance (i.e. production workflow) of each DO. 3) By joining 1 and 2 a mechanisms for propagating recursively warning to all the entities which used DOs containing errors.

IU EAGER	Robert Quick, Larry Lannom	IU US	testbed	active	Extending the existing DOA testbed (RPID) to include DOIP and new domains. Will develop outreach/educational material.
CLARIN	Dieter van Uytvanck	CLARIN-EU NL	workflow	in prep	turning to DO-based Switchboard component for CLARIN's workflow orchestration
CNRI EAGER	Giridhar Manepelli	CNRI US	workflow	active	prototyping the use of data types to disclose the APIs that software clients can execute on contextual slices or combinations of data, potentially with limited or (potentially) no human intervention
DEWCom	Karolj Skala	RBI HR	workflow	suggested	advanced development, implementation, support in Art and Humanities. Usage DO in Dew-Fog-Cloud Computing and Service hierarchy. Implementing DO in new platforms, services, tools and applications.
ENES	Tobias Weigel	DKRZ DE	workflow	suggested	implementing an automated digital object management is urgently necessary given the increasing amount of data
ENES	Tobias Weigel	DKRZ DE	workflow	suggested	implementing "actionable digital object collections" is of great priority, i.e. collections that are associated with operations that can be executed automatically
EOSC Pilot - ENES	Tobias Weigel	DKRZ DE	workflow	in prep	implementing automated data workflows with automatic provenance generation is planned to better inform different user groups
NOMAD	Petr Wittenburg, Carsten Baldauf	FHI DE	workflow	suggested	extending the NOMAD analytics toolkit to support advanced workflows that treat analytics results as DOs to manage hierarchies of interdependent material science knowledge