Open Call Collection OC-2018-1

Proposal Reference OC-2018-1-22872

Title: **European Digital Objects network**

Acronym: **EUDOn**

Summary

Inadequate data sharing and trading culture, inefficient data management and processing as well as lack of automation are significant impediments in the road towards an effective European Open Science Cloud. The European Digital Objects Network (**EUDOn)** brings together key stakeholders across data science and information systems engineering domains to work towards overcoming the biggest roadblocks on the path towards a seamlessly functioning data domain. It does so by focusing on the concept of Digital Objects (DOs) as the main vehicle for self contained, identifiable and actionable units of information, that can be implemented in cloud systems. The Action wants to organise networks of experts from European research infrastructures, e-Infrastructures, the RDA Europe community and other international research communities.

Key Expertise needed for evaluation

Electrical engineering, electronic engineering, Information engineering
Databases, data mining, data curation, computational modelling
Other engineering and technologies
Databases, data mining, data curation, computational modelling for other engineering and technologies

Keywords

Digital Objects

FAIR Data

Research Infrastructures

European Open Science Cloud

Funded by the Horizon 2020 Framework Programme
of the European Union

TECHNICAL ANNEX — EUDOn

1. S&T EXCELLENCE

1.1. CHALLENGE

1.1.1. DESCRIPTION OF THE CHALLENGE (MAIN AIM)

Scientific Case

Data-intensive science, the systematic analysis of large scale data, is rapidly permeating all areas of research in academia and industry, as it offers promise to revolutionize our understanding of the world. One of the keys for the success of data intensive science is FAIR [1], a globally agreed acronym for Findability, Accessibility, Interoperability and Re-Usability. In turn, one of the keys to FAIR is an interconnected ecosystem of infrastructures. These infrastructures currently serve their scientific communities of practice, curating and annotating data in a domain-specific manner. Arguably, interoperation of research e-infrastructures is a prerequisite for streamlining cross-disciplinary organisational, syntactic and semantic interoperability and largely depends on the adoption of common data systems. Establishing such a complex system will require global cooperation on the introduction of new concepts. We explore in particular one concept called "Digital Object (DO)" and its different flavours to help implementing FAIR compliant infrastructures.

A few examples may indicate the kind of research questions that data-driven scientists want to solve, but they still find many obstacles to carry out such work.

For discovering the causes of for example brain diseases it is widely agreed that machine learning algorithms could be used to find hidden patterns in a variety of data such as genetic data, brain-imaging data and others when being correlated with typical phenomenological patterns. However, large amounts of training data are required which come from many different labs all using different methods and tools. The hurdles for integrating useful data, which in these cases are sensitive data, from different sources are currently so high that such research is hardly doable even for big research centres. Knowing that increasingly more humans suffer from dementia and Alzheimer disease, for example, better solutions for overcoming these hurdles are urgently needed. DOs offer the ground for clear identification and thus findability of data, for making proper usage agreements, for effectively tracing proper re-usage, for associating verified tools with data, and for facilitating interoperability.

Digital language data, as studied in the humanities, cognitive science and related fields, are extremely diverse in their nature, formatting and annotation and also the domain of tools is rather diverse. Integrating data and tools alone on a platform with selection menus would not help since the "normal" researcher working with language data would be completely lost in this heterogeneity and only a few specialists would be capable to manoeuvre in subspaces of the field of acceptable combinations. A large European infrastructure is currently developing a switchboard solution linking specific data types to tools that have proven their usefulness for these types. DOs are exactly the type of solution allowing realising such a switchboard elegantly since they are typed and have other relevant metadata that can automatically be retrieved and these types can be associated with specific tools using the same basic mechanisms. Using these methods would enable a "computer-naive" researcher to create his/her workflows only applying domain knowledge that then can automatically process data of specific types in the intended way.

In material science a trend to make better use of the experimental and simulation data generated in thousands of labs worldwide is clearly visible combined with the expectation that access to massive amounts of results will enable the researchers to come to new categorisations of compound materials applying smart machine learning algorithms. Such multidimensional categorisations would allow researchers and industry to much more quickly find suitable material combinations given a specific new application. Large initiatives in the US and Europe are working in this direction indicating the huge relevance of this data driven approach. Also here the identification of the results, of the included materials and their attributes and the used creation processes, specific relationships and more are of crucial importance to make progress. Due to their binding capacity DOs have the potential to facilitate such research and make the high expectations reality.

In biodiversity, our extensive natural science collections (natural history specimens) have been, for hundreds of years, the focal point of research for new species discovery. Genomic information, morphological and ecological traits, and occurrence records are among the data classes individually extracted by those physical objects. Despite originating from the same physical object, data is currently fragmented and isolated, across small and larger repositories, with no or minimum capacity to bring this information together. The introduction of DOs in the field of biodiversity studies, could provide a technologically and socio-culturally acceptable way through which currently dispersed information is brought back together as a meaningful and machine actionable digital object, which effectively acts as the digital surrogate of the physical object.

Many other cases from different research disciplines could be mentioned and will be studied in the proposed Action. It is important to note that DOs are not just a technical concept, but a way to optimally structure domain data in a way that facilitates Data-Intensive Science.

Efficiency Case

Recent surveys indicate how important increased efficiency in the future handling of data will be in order to tackle unsolved scientific challenges, to include a much broader group of researchers in data-intensive science, to be able to monitor usage, increase trust and foster the path towards automatic processing which will be a must in order to keep a competitive edge.

Different studies show that most of the time of data scientists is wasted with “data wrangling” which includes the steps before the real analysis can be done. The study from RDA Europe mentions 75% of wasted time in 2014 [2], a study from MIT cited by M. Brodie mentions 80% in 2015 [3], a study from CrowdFlower in industry mentions 79% in 2017 [4]. From the research domain we know that many projects cannot be started and that many contributors cannot participate in data science. Huge fragmentation is hampering fast progress. The survey from CrowdFlower does not even include semantic interoperability, which relates to the analysis of knowledge after being extracted from a series of measurements. The main inefficiencies are caused by bad data organisation and quality. For accessible data it is often hard or impossible to find or interpret metadata that enables valid processing of the data streams.

The degree of automation in data-driven science globally, but in particular in Europe, is not adequate given the increasing quantities of data and their inherent complexity. Manual and ad-hoc operations do not scale and in general lead to undocumented and non-reproducible results which have been identified in several publications as a huge problem for scholarly communication. Automation of complex and data-demanding methods such as machine learning requires systematic and systemic approaches to the organisation of data. The development of easy to use workflow systems which are flexible enough to cope with various conditions requires harmonisation of the basic organisation of data. In summary, we can state that one of the major factors preventing the broad take-up of data-intensive science is the lack of broadly agreed basic operating mechanisms such as potentially offered by DOs.

Digital Object Case

Broad interactions at different global platforms such as RDA, FORCE11, C2CAMP [5], GO FAIR, at PIDapalooza and at workshops in Europe, the US and China have resulted in a broad agreement about the crucial role of persistent and globally resolvable identifiers (PIDs) for all data entities as a basic requirement for a fundamental change. The GEDE [6] collaboration which brought together delegates of 47 large European research infrastructures (most ESFRI projects) agreed after a year of intensive discussions on a paper about the need to use PIDs and their patterns of usage [7]. One of

the key messages in this paper is that the granularity with which PIDs should be associated to data is dependent on what is meaningful in a given scientific context.

However, it has been shown recently [8] that simply assigning PIDs is not in itself sufficient to achieve convergence on essential data management principles and thereby overcome the huge fragmentation that obstructs progress. Within global initiatives there is the growing conviction that we must properly exploit the increasing global use of PIDs and the availability of global PID resolving systems. The concept of DO architecture, as suggested already 2006 by Kahn and Wilensky [9] indicated a path towards fundamental changes of data practices. Furthermore, the RDA Data Foundation and Terminology working group [10] describes a DO as a structured bit sequence stored in some repositories, associated with a persistent, unique and resolvable Identifier (PID) and described by metadata. Some kernel metadata are being associated with the PIDs to achieve the high degree of binding different types of information necessary for efficient and especially automatic processing. RDA working groups are working on standardising these kernel attributes [11]. DOs can be simple or complex, i.e. the latter exist of aggregated collections of DOs, and their content can include data of different types, metadata, software code, machine configurations, etc. DOs have a type enabling their association with functions by use of Data Type Registries as has been defined by another RDA working group [12].

The term "object" is widely used in modern software technology since it implies an encapsulation of its internal structure by offering a set of tested functions that can be executed. The term became also very popular through the introduction of cloud stores which are also been called "object stores". Representing each "object" by a locally valid hash value is a step of virtualisation since the user does not need to know anymore how and where exactly the object is stored. Associated with this hash value is also the metadata information needed for finding and processing.

Much work is being done to define the term "research objects" [13] with the intention to capture the complex context of digital object to improve scholarly publishing. Closely related to this concept is the concept of packaging which discusses ways to pack such rich contexts into containers that can be exchanged easily to different environments to be used for further processing [14]. These concepts are complemented by approaches such as Linked Open Data [15] that offer ways to expose and exploit complex semantic relationships.

EUDOn Challenges

The initiatives for Open Science and Open Data request a common understanding which is essential to achieve improvements. The proposed Action will organise networks of experts from European research infrastructures, e-Infrastructures, the RDA Europe community and other international research communities to address the eminent roadblocks by applying the DO concept. We can identify the following specific challenges to be addressed by the Action:

C1. Bottlenecks in cutting edge data-intensive science use cases
We need a common understanding of the bottlenecks in cutting-edge use cases for data intensive science (DIS) across various disciplines. We need a deep analysis of the great inefficiencies that hamper DIS and prohibits the execution of new projects that could lead to new scientific insights. We need a deep analysis of the best way in which the introduction of DOs could help to overcome current limitations. Only a broad network approach respecting the various positions and contributions from the different disciplines will bring us ahead.

C2. Missing DO specifications and missing insights about their potential impact on research practices and on research infrastructures.
The concept of DOs has been introduced by computer science in different flavours to solve specific problems in smart ways such as by introducing cloud stores which are offering the huge stores to data-intensive science that are required to tackle complex problems and a step of virtualisation allowing laymen to use such stores. However, clouds are silos with their own administrative layers. Global data-intensive science needs an open infrastructure that turns this idea of virtualisation into a global one without barriers, i.e. globally registered and resolved DOs will build the bridge between the many silos and repositories all with different data organisations. We need to draw detailed specifications of DOs in a context of trust-worthy practices and technologies, and we need to promote steps for their verification.

C3. Missing platform to exchange knowhow about DO architecture specifications, their usefulness in various disciplines and useful software code that could be shared.

Changes in existing infrastructures are not only painful due to the investments needed for adaptations, but in particular since they require learning new workflows and tools from the researchers. These costs result in an utterly conservative behaviour of the researchers which is contrary to the needed changes in data practices. Platforms are needed where researchers can openly exchange ideas about DO architectures, implementations, where compliance criteria are established to ensure the investments made and where useful code can be exchanged. Since data-driven science is crossing disciplinary and geographical boundaries, such a platform needs to be organised globally and cross-disciplinary. Open, technology neutral interactions will establish the trust in taking steps towards adoption of new methods and thus overcome the existing hesitations.

C4. Lack of testbed projects that can drive the transformation, accelerate the specification work in RDA and interact with comparable initiatives working on different foci
A stimulation of testbeds for selected scientific cases implementing the DO concept would be helpful to accelerate the specification work. Progress should be monitored and gaps and bottlenecks should be identified in concrete actions. Embedding such a DO-oriented platform in related initiatives such as GO FAIR will help to position, shape and test the DO concept in the evolving landscape.

C5. Need to mobilise and relevant groups
A big challenge in modern research is getting the attention of stakeholders in a time where people are overloaded with information of all sorts, with solutions full of promises and claims, with a plethora of software programs created by smart people etc. making decision taking so problematic. Only clear and trusted forums where factual and unbiased information is being exchanged will help to overcome hesitation thresholds. Therefore, suitable dissemination strategies of results and progress need to be formulated.

C6. Demand for capacity for the specification and implementation work
Another big challenge is to find ways to engage a young generation of data scientists to take up the concepts and participate actively in the specification, implementation and management work. Ways need to be found to quickly build up knowledge and reduce psychological barriers.

C7. Need for training of young data scientists in order to transfer the evolving knowledge and skills from the early adopters.
When leading data scientists with deep insights from long experience will be retiring, their knowledge and skills need to be transfered to a new generation of experts. There is a need to quickly build up a community of young experts by organising hackathons/datathons which include hands-on session and participants' active contribution towards solution.

1.1.2. RELEVANCE AND TIMELINESS

In the USA there is a drive to fund programs that focus on DOs. In China the establishment of a national PID infrastructure available for science and industry will open the door for innovative developments related with DOs. Therefore, in Europe there is an urgent need for a network which will bring together stakeholders in science and technology in order to drive innovation to gain a competitive edge in data driven science. The proposed Action will bring together these stakeholders across disciplines, while linking with global players. At European level we can refer to activities to improve data practices in the following arenas:

- EOSC which tries to create a momentum towards a harmonised European data landscape fostering a single data market;
- the GO FAIR initiative with its three FAIR oriented pillars to change culture, to stimulate training and to bring together developers;
- RDA and the GEDE collaboration working on an agreement about specifications of common components of a future data infrastructure;
- the Industrial Data Space, which tries to come up with an abstract reference architecture guiding the implementation work towards an efficient data infrastructure in industry.

On the brink of a major step towards convergence [8], it is time for a common understanding of how this can be achieved. The proposed Action is timely for bringing about an alignment movement and critical mass with a broad European dimension, to push the conceptualisation and implementation work in close collaboration with comparable initiatives at the global level. Crucial for the Action is that it will give networking a clear direction by putting DO in focus. Adding the concept is not risky, since: a) it

is general enough to act as a platform for substantial growth (as TCP/IP did for the Internet); b) it is a strong concept that has been subject of discussions for almost 20 years; and c) it has already shown its potential strength in the cloud concept having been implemented by several companies. In contrast to many other networks, the Action is defining the conceptual nucleus which per definition is FAIR compliant and is adding an implementation direction for open science.

1.2. OBJECTIVES

1.2.1. RESEARCH COORDINATION OBJECTIVES

The proposed action will bring together, for the first time at this scale and breadth, a network of experts from different research fields to jointly look into the innovation and consolidation potential of the DO concept for data science and its direct applications to the data challenges of different scientific disciplines. This network includes European research and data infrastructure initiatives and their experts to interact about the concept, its required components, implications and impact. In particular early career data scientists need to be engaged since they will determine data-intensive science in the coming decades. The network bridges research infrastructures and e-Infrastructures, since the notion of DOs offers a unifying level of abstraction for data management, access, interoperability (at the level of data organisation) and reuse, and thus a way to implement FAIR. The Action will support work already undertaken by a few global initiatives in testing the DO concept, its flavours and its current implementations against jointly developed criteria and requirements. Of great importance will be to disseminate the acquired knowledge in suitable ways, organise a variety of capacity building events to attract early career data scientists and transfer the evolving knowledge and technology from the early adopters, and to closely interact with and contribute to European initiatives such as EOSC, PRACE, OpenAIRE, EUDAT, GO FAIR, FREYA, the ESFRI projects etc. The overall objective is to achieve a momentum towards implementing a more coherent infrastructure landscape and to stimulate and accelerate the specification work in RDA by providing clear directions. Participation in international forums which will bring together relevant initiatives from for example the US, China, South Africa, Australia and Russia need to be guaranteed, to exchange the results and commonly define future directions.

The following concrete objectives will be pursued:

O1. Collectively carry out a SWOT/GAP analysis for the need of DOs and DO based architectures, drawing on a collection of relevant scientific use-cases from various disciplines.

Collect relevant scientific use cases of data-intensive projects from various disciplines and at different scales. Analyse them with respect to their potential of increasing efficiency and effectiveness for the researchers in form of a SWOT/GAP analysis. Extract salient components from these use cases and identify gaps. Identify their specifications and compare them with components already specified to formulate generic requirements where applicable. The results of this work will be regular updates of online documents/wiki with snapshots every six months to indicate progress and remaining bottlenecks. The Action can build on results from different initiatives such as RDA and will include industry.

O2. Propose specific approaches to DO specifications and chart their potential impact on research practices and on research infrastructures, including e-Infrastructures.

Propose specific approaches to DO implementations and chart their potential impact on research practices and on research infrastructures, including e-Infrastructures. Discuss possible solutions to the fragmentation at the level of data organisation and in facilitating semantic interoperability. Based on the needs of data-driven projects it must be understood what the problems are which DOs can solve, which efficiency effect can be achieved and what the limitations are. This objective will also lead to an understanding of how the introduction of DOs would change data culture, for example by combining DOs with smart contracts implemented in blockchain technology to increase trust. It must also be understood what the transformational implications of introducing generic components are and which amount of work would be required to adapt existing systems. For interested communities, implementation plans will have to be worked out that can give directions and set priorities. Also this objective will result in regular updates of online documents/wiki with snapshots every six months and where applicable in a stimulation of the RDA specification work.

O3. Propose specific approaches to DO architecture specifications and chart their potential impact on research practice and on research infrastructure, including e-Infrastructures.
Propose, agree upon, and disseminate specifications for DO architectures. Integrate ideas, define compliance criteria, work out interface, linkage and harmonisation specifications, and exchange code of components from existing implementation initiatives. EUDOn will not implement software, but will establish an interaction platform to keep a momentum and to achieve increasing convergence with respect to DO related implementations. DO-related concepts are currently being discussed in Europe in initiatives such as GO FAIR, various research infrastructures, in implementation initiatives such as EOSC, EUDAT and PRACE, but also in the US, China, South Africa and other countries. EUDOn will maintain a close interaction with these networks to converge ideas. In these interactions EUDOn will also identify roadblocks that need to be tackled with a certain priority and stimulate specification and development work. The results of this work are regular updates of online documents/wiki with snapshots every 6 months and where applicable in a stimulation of the RDA specification work.

O4. Stimulate and contribute capacity for the specification work in RDA or the implementation work in networks such as GO FAIR.
Develop implementation concepts for selected cases, focusing on versatile architecture models as an output to demonstrate the potential of DOs to make data-driven science more efficient and to put the FAIR principles into practice. Continuously monitor progress in filling gaps and bottlenecks and stimulate concrete action, be it at the specification level by RDA groups or at implementation level within initiatives that participate in the shared programming effort such as for example GO FAIR. The effort needed for oversight and steering new specification and implementation initiatives should not be underestimated, since the endeavour is organised across disciplines and across regions. The results of this work are regular updates of online documents/wiki with snapshots every 6 months to indicate agreements, shared tasks and successful delivery and the initiation of specification and implementation groups.

1.2.2. CAPACITY-BUILDING OBJECTIVES

O5. Dissemination of results and progress.
Disseminate the results of EUDOn to stakeholders. Exchange successful implementations that have shown the usefulness in concrete data-intensive science projects. Promote the capacity building actions for early career people (see below) to the many research infrastructures that have already indicated support to EUDOn and beyond. Relevant audiences at conferences and other events will be targeted by presentations, written materials, lightning talks, exhibits etc. in order to build an active and engaged community, to openly evaluate solutions and to achieve broad consensus.

O6. Stimulate and contribute capacity for the specification and implementation work
Encourage learning and engagement of Early Career Investigators (ECIs) through mobility in Short-Term Scientific Missions (STSMs). Use STSMs to build bridges between research infrastructures and e-Infrastructures and between academic and industrial research. These STSMs should offer opportunities to develop different types of skills such as participating in the definition of data-intensive science projects, in the specification of DO related concepts, in corresponding implementation work or in managerial type of actions. All these skills will be needed for an efficient and competitive data science ecosystem.

O7. Organise training courses and hackathons in particular to attract young data scientists and transfer the evolving knowledge and technology from the early adopters
Plan and implement a training program for ECIs and data managers from a wide range of scientific communities together with IEEE BDGMM [16] and other powerful organisations, in order to secure successful changes of research practice and promote convergence. EUDOn will use various channels such as webinars, tutorials, training courses with hands-on sessions and hackathons. Where possible EUDOn will organise these events together with specific scientific communities or domains, since this has proven to be most effective. Participants will more easily understand the terminology and more easily see the effect on the science they are familiar with.

1.3. PROGRESS BEYOND THE STATE-OF-THE-ART AND INNOVATION POTENTIAL

1.3.1. DESCRIPTION OF THE STATE-OF-THE-ART

Current disadvantages and costs in the data domain are caused by the huge fragmentation creation of legacy data which will be difficult to integrate into an easily sharable and reusable data domain. The phase we are in can be described as creolisation with mixed suggestions and solutions at all levels.

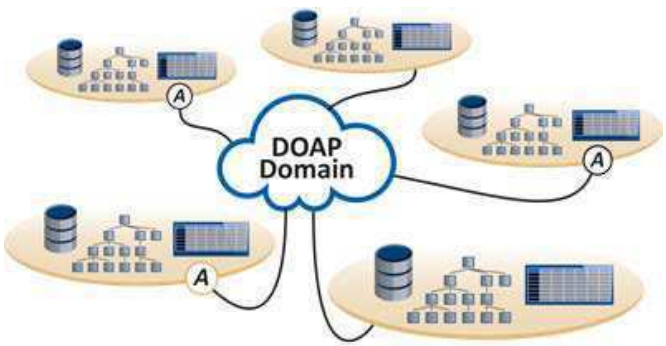
Various initiatives listed below have been started from different stakeholders - some more top-down, others more bottom-up – which help to better understand the challenges and possible approaches and the FAIR principles, while broad discussions at OECD, G8+5, and RDA are milestones towards convergence. Yet we lack the core concept to turn such principles into practice, and none of the following initiatives include a broad network of practitioners to exchange views on DO as intended by the proposed Action.

- Stimulated by the availability of huge amounts of information in the web, the Semantic Web initiative has created many excellent solutions and standards to exploit this rich world. However, raw data first needs to be managed and transformed in order to enable semantic processing.
- The European Open Science Cloud is another huge initiative that wants to create a framework that facilitates the transformation towards more efficient and FAIR compliant data intensive science, but it does not yet discuss any implementation concept.
- The RDA has brought together communities that focus on scholarly communication aspects related to data and those that focus on the data fabric aspects, so as to define best practices and standards to overcome the inefficiencies. Like FORCE11, which also focuses on aspects of scholarly communication, RDA does specification work and not implementation.
- GO FAIR is a European initiative sponsored by some member states to foster implementation work towards FAIR compliance including initiatives with different foci. However, GO FAIR offers facilities to promote the work of the implementers, but not funds for a broad interaction platform.
- C2CAMP is a global initiative of some strong centers and research communities that want to start implementing and testing the DO concept and since it is an implementation network it is acting as part of the GO FAIR initiative, however, not having the financial capacity to establish a broad network

1.3.2. PROGRESS BEYOND THE STATE-OF-THE-ART

The Action is meant to fill this gap and offer a platform to facilitate and fund broad interactions. Due to its complementary role, members need to interact in appropriate ways with the above mentioned initiatives.

The main benefit for data science of the suggested DO-based approach can be seen in its universal way of binding essential information about data entities, either atomic or complex in form of collections. In this way the DO concept addresses aspects that are causing major inefficiencies described by the term "data wrangling". The DO model could be seen as defining a new commodity layer allowing growth again as J. Hendler, a well-known Semantic Web expert, put it recently at a US workshop [17]. G. Strawn, one of the Internet pioneers, expressed his convictions at a follow-up workshop [18] by indicating that a standardised and simple digital object access/interface protocol combined with persistent identifiers may have a potential and impact that can be compared to the one of TCP/IP some decades ago. This DO access protocol can be implemented by repositories either directly or indirectly by providing adaptors as is indicated by (A) in the diagram depending on the state of their data organisation.



The Action will not reinvent the notion of DO but discuss its implications for infrastructure building with all interested data scientists which are often collaborating with research infrastructures and e-Infrastructures and beyond. The Action wants to create the momentum that is needed to make big steps towards convergent specifications. Since this can only be achieved globally The Action is committed to collaborate closely with experts in other regions such as the US, China, South Africa, etc. It is now time to act, discuss, implement and test approaches in global synchronisation, in

synchronisation with experts working on other strands such as semantic interoperability for example and by including knowledge about the already existing solutions in the different scientific disciplines.

1.3.3. INNOVATION IN TACKLING THE CHALLENGE

In computer science the concept of "objects", as in object-oriented programming, has led to an enormous improvement in software engineering since its encapsulation principles facilitate building complex systems. The concept of "objects" has also been realised in modern cloud systems which is why some experts call cloud stores "object stores" where all relevant information is encapsulated. This can be seen as a virtualization step since the user does not have to deal with different types of data organisations (file structures, clouds, sql-db, nosql-db, etc.), with the question where the metadata and data can be found, what kind of operations the special object type is supporting etc. However, cloud solutions are proprietary and pointers can only be resolved within the limitations of the cloud.

EUDOn wants to extend this concept to an open global space to facilitate data driven research and to make data reuse independent on the specific nature of local data organisations in the various repositories worldwide. Scientific users should be able to focus on the questions they are interested in: they want to select a certain digital object (in general a collection) identified by a PID and described by metadata (including data provenance) and carry out some operations on it or even engage agents to select a digital object based on profile descriptions and include it in workflow executions. The move towards DOs would allow us to build an interoperable domain where an agreed *Digital Object Access Protocol* links all kinds of different repositories some being connected via adaptors. This is very similar to the way the early Internet itself was built. Data scientists from various disciplines expressed their interest in DOs as follows:

- Faster selection of useful tools and services based on types of Digital Objects and their rich metadata will facilitate data re-use and their automated processing effectively.
- Scientists working in an increasingly complex data landscape want to work at higher levels of abstraction enabling scalability.
- Scientists and managers need wide scale data tracing and reproducibility to facilitate trust and verification, which are not achievable in a scalable way with today's means.
- DOs abstract from concerns of location, storage method and community-dependent services.
- DOs will achieve a higher degree of interoperability and overcome fragmentation.
- Improved ways to automatically create scientific annotations and assertions to capture and exploit knowledge will enable machines to understand the object context and take appropriate action.
- DOs allow creators to associate mechanisms for controlled transactions and tracing of reuse which is crucial for sharing sensitive or commercial data and which finally is crucial for trust building.

1.4. ADDED VALUE OF NETWORKING

1.4.1. IN RELATION TO THE CHALLENGE

The concepts that will bring us to a change of data practices require large efforts in networking. The GEDE discussions about PIDs involving delegates and experts from 47 different European research infrastructure projects took about a year to result in a broad agreement about their relevance and their usage. This may give an indication of how much effort is required in consensus building. In iterations the commonalities of data science in different fields need to be extracted so that convergence can be achieved.

A European data infrastructure needs to be based on broad support from policy makers and data scientists being active in the research infrastructures and e-Infrastructures. A momentum towards a higher degree of interoperability and thus efficiency can only be achieved if this community will become actively engaged. At the specification level, RDA has been established and can be seen as successful. At the implementation level GO FAIR has been initiated and C2CAMP, putting the digital objects in its centre, is participating in this initiative, but neither GO FAIR nor C2CAMP offer funds for active participation from a broad group of people. The voluntary GEDE work showed that a voluntary basis alone is not sufficient to organise meetings, to support travels, to organise training courses etc.

EUDOn can build on the GEDE experience and the required trust in fair interactions has been established. But it will add funds to support networking activities and to facilitate participation of a

broad group of engaged European experts. In doing so, it will also foster the work of initiatives such as GO FAIR and C2CAMP.

1.4.2. IN RELATION TO EXISTING EFFORTS AT EUROPEAN AND/OR INTERNATIONAL LEVEL

EUDOn fits well into the current European and global landscape in so far as it

- adds an innovative and operative dimension to the FAIR principles because the DO model is not only FAIR compliant but can also directly facilitate data intensive science;
- could help to realise the EOSC which has been focusing on important aspects of governance, business models, etc., but currently needs a common concept that facilitates data intensive science;
- adds an implementation dimension to the global RDA specification work by transforming RDA endorsed outputs into a larger testbed, so that some work in RDA is given more momentum and the speed and density of specification work will increase;
- adds a global network focussing on a specific approach to the GO FAIR implementation networks working on related challenges such as semantic interoperability;
- integrates experts from various data science communities in Europe, linking also to experts from other countries such as the US, China, Australia, South Africa and Russia.

EUDOn will enable European experts to take a leading position and attract young experts to participate in their work. It will allow EU data scientists to not only invite experts from abroad, but also to steer the global discussion, the specification and implementation work. It will allow EU data scientists to participate in corresponding discussions going on in other countries. It will help selected experts to contribute to the ongoing specification work in RDA and the implementation work in GO FAIR.

In summary, European data scientists looking for innovation in data practices will get an enormous impulse and take over a leading position. Without EUDOn, Europe risks falling behind countries such as the US and China which are already planning large investments in work related with DOs.

2. IMPACT

2.1. EXPECTED IMPACT

2.1.1. SHORT-TERM AND LONG-TERM SCIENTIFIC, TECHNOLOGICAL, AND/OR SOCIOECONOMIC IMPACTS

The impact of a DO based global data infrastructure ecosystem on scientific progress will be immense:

- FAIR data will be put into practice and data sharing/trading will become the accepted practice;
- data integration from different silos will be much more efficient;
- data intensive research in all disciplines will be facilitated so that more scientific questions can be addressed;
- many more researchers will be able to participate in data intensive research since efforts and costs will be reduced;
- sensitive data can be integrated into the shared data domain since it will be possible to restrict operations, to make specific "smart contracts" regulating re-usage, to improve tracing of data usage;
- automated data processing will be enabled – the only practicable way to go, given the huge amounts of data being generated in the coming decade from smart sensors, the digital marketplace, etc.

Comparable to the invention of the Internet, we cannot foresee which types of brilliant applications will be invented in a few years. No one could foresee for example the emergence of the Web application which changed the world when TCP/IP was born. However, we can imagine a situation where new data becomes available from a trustworthy source, where crawlers tuned by scientists detect the relevance of this new data for a given task, where smart contracts are available indicating the type of operations allowed on the data and where the data will immediately be included in useful processing

tasks. It would free the researcher to spend time on finding hidden patterns, on checking the trustworthiness and reliability of the results, etc.

In summary, there will be considerable short-term benefits for the researchers by adopting the DO concept since it will reduce the wasted time on data wrangling. In addition, there will be even larger long-term impacts for the data volumes produced by the billions of future smart devices.

2.2. MEASURES TO MAXIMISE IMPACT

2.2.1. PLAN FOR INVOLVING THE MOST RELEVANT STAKEHOLDERS

Data Scientists: The most important stakeholders are data scientists, including those who are experts in applying state of the art methods such as machine learning and those who are experts in formats, metadata, etc. Many are involved in research data infrastructures; they know the scientific challenges and have a deep understanding of the roadblocks. With GEDE, a collaboration between such data scientists has been established which can serve as a basis, but this collaboration needs to be extended.

E-Infrastructure-Providers: Computing and data centres are currently revising their services to data scientists. They need to facilitate data workflows and to offer a variety of services on data. Computing and data cannot be separated anymore in most data intensive applications. Members from PRACE and EUDAT will be included in the Action.

Computer Scientists: Large computer science departments work on specific methods and also create cross-disciplinary tools. It is important for success to bind them into the discussions as well to integrate their deep methodical knowledge, to adapt their software where required and help educating a new generation of experts.

Global Service Providers: Another relevant group are those who offer global services such as the DONA Foundation, the DOI Foundation, the ePIC Foundation, companies offering cloud services, etc. since these are providing crucial services such as for the functioning of a global DO domain that is widely based on stable and persistent services for PIDs or by providing advanced data services.

Global Initiatives: A close collaboration with initiatives such as RDA, W3C, IEEE, CODATA and others is important to consider different aspects related with data.

Policy Makers: The Action will have a close collaboration with policy makers organised in scientific organisations and with funders at European and member state level. Here the commitment to contribute to EOSC and to the GO FAIR initiative will help EUDOn to reach out to policy makers.

Industry: An outreach to industry will be important because a change of practices will only be successful if industry will take up the new methods and build appropriate tools. We will reach out to start-ups and SMEs which agree with the DO model and which already started to work out innovative technology models. We will participate in industry forums and invite industry experts to specific meetings.

2.2.2. DISSEMINATION AND/OR EXPLOITATION PLAN

The WG on Dissemination will be responsible for producing a Dissemination and Exploitation Plan (to be ratified by the Action MC) that will guide the EUDOn outreach to various stakeholder groups described above. Data intensive ESFRI projects and other mature e-infrastructures will be included as first level targets, because they will assist in reaching out to data scientists. This has to be complemented by two directions: 1) more disciplines need to be approached and 2) more data scientists within the scientific domain of each participating ESFRI project need to be addressed. They are the drivers of change and need to be actively involved in the interactions which is not trivial given the usual workload. The ESFRI and national representatives have to take an important role in EUDOn as ambassadors and as facilitators to get DO related issues on the agendas of discipline and national meetings. Training courses and webinars driven by the scientific use cases will also help conveying the main messages and showing the implications of applying the concept.

This will be complemented by strategies to involve the young generation. Here in particular hackathons and web-based competitions turned out to be excellent mechanisms to reach out and

attract young people. For this work partners have to be looked for such as IEEE BDGMM, RDA and the ESFRI projects since the effort to prepare such actions is high. We can build on some experiences made by RDA Europe and IEEE BDGMM for example. Industry will be contacted using the started collaborations by RDA Europe, using also the collaborations with the IoT Forum and various national meetings. Industry can appear in different flavours (IT service providers, sector specific service providers). In particular, innovative start-ups and SMEs working on challenging new ideas will be the target. Policy makers will be addressed e.g. by regular flyers with essential messages. Through an annual event, we will try to attract many of the actors mentioned above and bring them together using attractive formats such as interactive panels.

2.3. POTENTIAL FOR INNOVATION VERSUS RISK LEVEL

2.3.1. POTENTIAL FOR SCIENTIFIC, TECHNOLOGICAL AND/OR SOCIOECONOMIC INNOVATION BREAKTHROUGHS

Every approach that reduces the huge fragmentation and thus the inefficiencies in the data domain globally will lead to “revolutionary changes” that will not only influence science, but also industry and society as a whole. Some argue that the urgent changes needed will probably involve disruptive steps and they compare the expected transformations with those after the invention for example of the Internet.

A recent paper [8] points at global agreements on important but comparatively simple specifications that ultimately generated a boost of new services and businesses impacting all aspects of society. The question is which simple specifications will have the potential to be accepted by many as a new platform in the data domain to invest and adapt current strategies. We believe that DOs are such a fundamental enabling technology.

In order to maximize the chance of success of this promising but potentially disruptive concept, we need to agree on a roadmap and build testbeds with more investments by global actors as C2CAMP is starting it in collaboration with initiatives such as GO FAIR. Such initiatives will only succeed if we are creating networks of stakeholders around it that drive discussions based on scientific use cases and if we succeed in attracting the young generation.

3. IMPLEMENTATION

3.1. DESCRIPTION OF THE WORK PLAN

We can identify the following matrix of objectives and methods:

	O6: dissemination	O7: mobility program	O8: training
O1: SWOT/GAP Analysis	X	X	X
O2: Approaches to DO specifications and their impact	X	X	X
O3: Approaches to DO Architecture specifications and their impact	X	X	X
O4: stimulate specification and implementation work	X	X	X

3.1.1. DESCRIPTION OF WORKING GROUPS

WG1. Use case based SWOT/GAP analysis
Objectives. Carry out a SWOT/GAP analysis to assess the DO model based on scientific use cases
Tasks **T1.1.** Establish an Activity Plan

T1.2. From the participating data scientists practical use cases of data driven science will be collected and analysed in terms a SWOT/GAP analysis and of salient components that can be identified. **T1.3.** Also use cases collected within RDA by various groups will be analysed **T1.4.** The synthesis work will extract the essential characteristics of common components and their interactions with other components. **T1.5.** In addition, the compatibility with the Digital Object Model will be assessed. **T1.6**. Organise the necessary dissemination and participate in important forums (O6), the mobility program (O7) and training (O8) together with WG5
Deliverables **D1.1.** Activity Plan (M3) **D1.2**. Use Cases Analysis and DO model midterm assessment (M16) **D1.3.** Use Cases Analysis and DO model final assessment (M32)

WG2. DO specifications and their impact
Objectives. To specify the DO model and to chart their potential impact
Tasks **T2.1.** Establish an Activity Plan **T2.2.** Work out and suggest DO model based alternatives to current practices described in the use cases **T2.3.** Work out the potentials and the limitations of the DO Model given the use cases **T2.4.** Work out the impact and implications for the current data practices to adopt the DO model where applicable **T2.5**. Organise the necessary networking and participate in forums (O6), the mobility program (O7) and the training (O8) together with WP 5
Deliverables **D2.1.** Activity Plan (M3) **D2.2.** Impact & Implication of DO model midterm report (M18) **D2.3.** Impact & Implication of DO model final report (M34)

WG3. DO architecture specifications and their impact
Objectives. To specify DO architectures and to chart their potential impact
Tasks **T3.1.** Establish an Activity Plan **T3.2.** Document the state of the art with respect to DO architecture specifications and discuss the requirement and specification work within the GO FAIR networks and the international collaborations devoted to the DO model **T3.3.**Based on the analysis of the use cases and the synthesis work derive requirements for the further work on DO architectures and push the specification work for which also RDA will be used as a platform **T3.4**. Organise the necessary networking and participate in forums (O6), mobility program (O7), and training (O8) activities together with WP 5
Deliverables **D3.1.** Activity Plan (M3) **D3.2.** Document the State of DO architecture specifications (M6) **D3.3.** Developed specifications midterm report (M21) **D3.4.** Developed specifications final report (M34)

WG4. Stimulating specification and implementation work
Objectives. To stimulate and analyse specification and testbed activities
Tasks

T4.1. Establish an Activity Plan **T4.2.** Document the state of the art with respect to DO model testbeds and discuss the testbed development work within the GO FAIR networks and the international collaborations devoted to the DO model **T4.3.** Stimulate where feasible testbed activities in the realm of the global DO implementations and give help where needed and stimulate the inclusion of testbed projects in the global DO testbed being implemented by many different groups in various countries **T4.4**. Stimulate where required additional specification work in RDA **T4.5.** Organise the necessary networking and participate in forums (O6), mobility program (O7), and training (O8) activities together with WP 5
Deliverables **D4.1.** Activity Plan (M3) **D4.2.** Document the State of DO Model testbed development (M6) **D4.2.** Testbed development midterm report (M21) **D4.3.** Testbed development final report (M34)

WG5. Dissemination
Objectives. To coordinate networking and outreach work of all WGs
Tasks **T5.1.** Establish an Dissemination and Exploitation Plan **T5.2.** Setup the required Wiki-type spaces to enable flexible communication between the participants enabling global participation **T5.3.** Coordinate and supervise the networking and forum participation (O6), mobility program (O7), and training (O8) activities of the various WGs **T5.4.** Develop additional outreach activities such as flyers, summarizing statements for various stakeholders etc. **T5.5.** Organise the annual EUDOn conferences and help in other meeting activities
Deliverables **D5.1.** Dissemination and Exploitation Plan (M3) **D5.2.** Wiki Setup Report (M6) **D5.2.** Dissemination midterm Report (M19) **D5.3.** Dissemination final Report (M35)

List of Milestones

M1.1 First Set of 15 Use cases analysed (M6)
M1.2 First Assessment of DO Model against analysed Use Case set (M10)
M1.3 Extended Use case set (+10) analysed and DO model compliance check (M14)
M2.1 First Analysis of Impact and Implications of required changes on 5 Use cases (M10)
M2.2 Extended set of Impact and Implications on +5 use cases (M14)
M3.1 First extended set of component specifications based on use cases analysis (M10)
M4.1 First round of stimulations (5) of testbed developments (M10)
M5.1 Summary report on outreach and communication activities

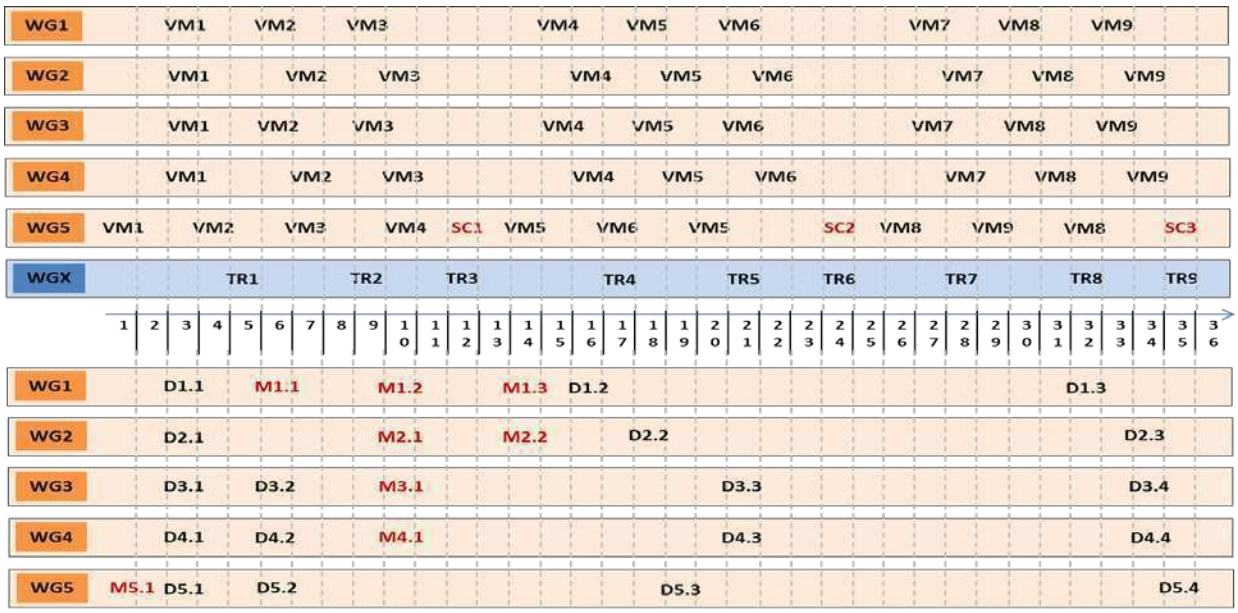
3.1.2. GANTT DIAGRAM

This project will depend on continuous interaction between the various activities with short and fast feedback loops. It thus mimics the modern planning style and due to the global embedding we only ask for 3 years project time, since the pressure is high to deliver concrete outcomes. For the same reason, there is substantial overlap in time between the activities of the Working Groups. All intermediary insights and results need to be communicated with the global and European partner initiatives in a bi-directional way to keep European data scientists organised in EUDOn in the driving and influential roles, since some work is already ongoing outside of Europe. The Gantt chart below includes an indication of the deliverables (green stars) and milestones (red stars).

The timeline below indicates how the various activities need to be synchronised expecting that proper time management for such a highly active process will be crucial as the GEDE collaboration has shown.

Working Groups will need to meet regularly, normally through virtual meetings (VM). There will be annual conferences (SC) linked to MC meetings where also the different WGs will have f2f meetings and where training courses will be held as well. Each WG needs to define its own rhythm and select meetings dates dependent on availability of the experts, i.e. the schedule below can only be seen as a first indication. With respect to the training courses (TR) (be it traditional training courses and/or hackathons) each WG needs to define a schedule. The intention is to have one hackathon per year per WG. In addition, 2 training courses should be organised per WG per year. WGs can decide to organise these jointly to reduce the effort.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
WG1	★	★		★	★	★					★	
T1.1												
T1.2												
T1.3												
T1.4												
T1.5												
T1.6												
WG2	★			★	★	★						★
T2.1												
T2.2												
T2.3												
T2.4												
T2.5												
WG3	★	★		★			★					★
T3.1												
T3.2												
T3.3												
T3.4												
WG4	★	★		★			★					★
T4.1												
T4.2												
T4.3												
T4.4												
T4.5												
WG5	★ ★	★				★						★
T5.1												
T5.2												
T5.3												
T5.4												
T5.5												



3.1.4. RISK AND CONTINGENCY PLANS

Risk	Probability	Impact	Prevention/contingency plan
the DO model turns out to be not matching the requirements	Low	Very High	the risk is low since e.g. RDA DFT and DF work was based on more than 30 use cases from different disciplines, also the engagement of a 6 research infrastructures in C2CAMP indicates the solidity.
participants do not provide sufficient use cases	Very low	High	this risk is low since data scientists from very many fields of science are involved; if necessary the Action will look for even more active data scientists.
adaptation of different practices towards the DO model turns out as too costly	Medium	Medium	already now 6 research infrastructures have drafted plans how to integrate the DO model, thus we can conclude that from all participants at least 20 will carry out concrete adoption work which would be a sufficient basis for the global testbed given that also in the US, China, Australia, South Africa and Russia teams will receive funds to contribute; in case of problems additional funding via other projects would have to be solicited.
too many meetings prevent active participation	High	Medium	The Action will apply a balanced strategy to involve different types of experts and to distribute the load on many different shoulders; experts are in general overloaded requiring a smart strategy which is in the hands of the MC and WG5; sensitive interactions and active project leadership (incl. WG leaders) will be applied.

3.2. MANAGEMENT STRUCTURES AND PROCEDURES

The proposed Action organisation and management structure will follow COST rules and procedures. The overall management of the project will be the mandate of the **Management Committee** (MC), presided by the Action Chair and Vice-Chair. The MC will ensure the democratic process for decision-making within the project. MC Members are nominated by the COST Member Countries and can join the Action during the first three years. There will be one Action MC meeting every year.

A **Core Group** (CG) will assist the MC and act as the main driving force to ensure that the Action achieves its technical objectives. The CG will include at least the Action Chair, Vice-Chair, and the WGs Chairs. The CG will appoint an **STSM Coordinator** and include this person in the CG.

The Action is composed of five **Working Groups** (WGs) each dedicated to achieve a specific set of Action objectives and follow the specific, strategically important, technical direction of the work plan. Every WG will have a WG Chair and a quality assessment committee. The particular management structure and procedures of each WG will be decided by itself at their own meetings. The participation of WG members and ad hoc participants is decided by the MC. The WGs and other aspects of the management structure can be adapted by the MC of the Action during its lifetime. WGs will present their work during their intermediate and final meetings and other dissemination channels (conferences, website, etc.).

3.3. NETWORK AS A WHOLE

The main objective of this Action is to create a pan-European open network to promote synergies between data science researchers, ESFRI research infrastructures, e-Infrastructures, data associations and initiatives, the data industry and other stakeholders in industry and society, to develop a common understanding, standards, and best practices in supporting data science in Europe. After the first round of invitation to join the COST Action proposal we engaged 87 experts from about 17 different disciplines coming from 26 different COST countries. The initial proposers cover many distinct fields, are involved in many ESFRI infrastructures and include experienced and well-known research managers. Amongst them are 40 early career experts.

In order to further increase the scope of the network in terms of covered languages and cultural exchange, the Action will promote the inclusion of researchers coming from Near Neighbour Countries (NNC) and International Partner Countries (IPC). To that end some members of non-COST countries have been already included in the initial network and more will be invited in the future.

References

EUDOn References:

[1] https://www.force11.org/group/fairgroup/fairprinciples

[2] http://hdl.handle.net/11304/6e1424cc-8927-11e4-ac7e-860aa0063d1f

[3] M. Brodie: Understanding Data Science: An Emerging Discipline for Data-Intensive Discovery, keynote, Data Analytics and Management in Data Intensive Domains (DAMDID'2015), Obninsk, Russia, October 13-16, 2015

[4] https://visit.crowdflower.com/WC-2017-Data-Science-Report_LP.html

[5] https://github.com/c2camp

[6] https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda

[7] https://zenodo.org/record/1116189

[8] http://doi.org/10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0

[9] https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf

[10] http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318

[11] https://www.rd-alliance.org/groups/pid-kernel-information-wg

[12] https://rd-alliance.org/group/data-type-registries-wg/outcomes/data-type-registries

[13] http://www.researchobjects.org/

[14] https://www.slideshare.net/carolegoble/what-is-reproducibility-gobleclean

[15] https://en.wikipedia.org/wiki/Linked_data

[16] https://standards.ieee.org/develop/indconn/BDGMM_index.html

[17] https://www.nap.edu/catalog/25015/international-coordination-for-science-data-infrastructure-proceedings-of-a-workshop

[18] https://www.rd-alliance.org/workshop-registration-now-open-%E2%80%9Cmanaging-digital-research-objects-expanding-science-ecosystem-1

COST Mission and Policies

EUDOn addresses the huge inefficiencies in data-intensive science in the various research disciplines and also in industry which prevents tackling many open research questions and broader participation. It is the huge fragmentation that leads to 80% waste of time of data scientists in research and industry that needs to be overcome urgently. Several initiatives worldwide started to address this issue. In Europe the EOSC framework has been started, but does not include infrastructural suggestions, the FAIR principles are widely accepted now, but do not include an implementation concept. Research infrastructures and e-Infrastructures are working on harmonising their data landscape, but widely isolated from each other. Cross-disciplinary and global initiatives such as the Research Data Alliance are successful in working out specifications and engaging European data scientists from various disciplines as the GEDE collaboration has shown but on a voluntary basis which support acceleration towards convergent solutions as it is required. EUDOn will fill this gap by putting the promising concept of Digital Objects into the focus, bringing together data scientists from many different disciplines and many countries in Europe and interacting with and embedding its work in other relevant initiatives not only in Europe but globally.

According to COST Mission and Policies, EUDOn will:

- build capacity by accelerating the work on reducing the many roadblocks towards an efficient and democratised data intensive science and train especially a new generation of early career data scientists
- encourage and enable especially researchers from less research-intensive countries across Europe by including their data scientists
- support countries just becoming aware of the need of data infrastructures to participate in advanced concepts for data-intensive science by inviting early career experts to visit advance labs
- act as the highly visible network of experts in Europe and beyond to attract experts who are interested in discussing and pushing the concept of Digital Objects to overcome inefficiencies supported by appropriate dissemination and training activities
- act as a platform for experienced data scientists, infrastructure builders, computer scientists and industry experts with a focus on involving Early Career Investigators (ECIs) to advance the RDA specification work and to stimulate implementation work in the realm of existing networks such as GO FAIR
- addressing the sustainability of these EUDOn efforts by embedding the specification and implementation work in initiatives with long-term perspectives
- foster cooperation across research infrastructure and e-Infrastructure experts from COST Member Countries and Cooperation Countries
- raise policy makers awareness on the need to urgently overcome the inefficiencies, of the great potential of the DO concept and on the need to remain competitive by taking a leading role
- use the broad engagement to accelerate the DO-related specification and implementation work and include industrial experts, in particular SMEs
- strengthen the European expertise in a decisive moment to create a momentum with global impact
- help in accelerating the developments towards a flourishing data economy based on simple and powerful mechanisms of facilitating interoperability and thus with an enormous impact on society and social growth
- contribute through concrete work to the ongoing discussions on open science and open data

Network of Proposers - Features

COST Inclusiveness target countries
50.00 %

Number of Proposers
87

Geographic Distribution of Proposers

Country	ITC/ non ITC/ other	Number of institutions from that country	Number of researchers from that country	Percentage of the proposing network
Australia	other	1	1	1.15 %
Austria	non ITC	1	1	1.15 %
Bulgaria	ITC	1	1	1.15 %
Cyprus	ITC	1	1	1.15 %
Czech Republic	ITC	4	4	4.6 %
Finland	non ITC	2	3	3.45 %
France	non ITC	4	4	4.6 %
Germany	non ITC	11	11	12.64 %
Greece	non ITC	1	1	1.15 %
Hungary	ITC	2	2	2.3 %
Ireland	non ITC	1	1	1.15 %
Italy	non ITC	11	11	12.64 %
Lithuania	ITC	3	3	3.45 %
Netherlands	non ITC	9	9	10.34 %
Norway	non ITC	4	4	4.6 %
Poland	ITC	1	1	1.15 %
Portugal	ITC	3	3	3.45 %
Romania	ITC	2	2	2.3 %
Russian Federation	other	1	1	1.15 %
Serbia	ITC	1	1	1.15 %
Slovakia	ITC	2	2	2.3 %
Slovenia	ITC	4	4	4.6 %
Spain	non ITC	1	1	1.15 %
Sweden	non ITC	3	5	5.75 %
Switzerland	non ITC	2	2	2.3 %
Turkey	ITC	1	1	1.15 %
United Kingdom	non ITC	6	6	6.9 %
fYR Macedonia	ITC	1	1	1.15 %

Gender Distribution of Proposers

73.6% Males
26.4% Females

Average Number of years elapsed since PhD graduation of Proposers with a doctoral degree

15.4

Number of Early Career Investigators

40

Core Expertise of Proposers: Distribution by Sub-Field of Science

18.4% Computer and Information Sciences
12.6% Physical Sciences
10.3% Biological sciences
9.2% History and Archeology
8.0% Earth and related Environmental sciences
37.6% Other
3.4% Unspecified

Institutional distribution of Network of Proposers

73.6% Higher Education & Associated Organisations
17.2% Government/Intergovernmental Organisations except Higher Education
8.0% Private Non-Profit without market revenues, NGO
1.1% Business enterprise

Private Non-Profit without market revenues, NGO:7

- Number by Type
  Charity:1
  Other:5
  Advocacy/Membership Organization:1
- Number by Level
  National:5
  International or European:2

Higher Education & Associated Organisations:64

- Number by Field of Science of Department/Faculty of Affiliation
  Computer and Information Sciences:16
  Languages and literature:2
  Physical Sciences:9
  History and Archeology:6
  Earth and related Environmental sciences:5
  Other agricultural sciences:2
  Biological sciences:2
  Other medical sciences:3
  Other humanities:3
  Chemical sciences:3
  Mathematics:2
  Interdisciplinary:3
  Clinical medicine:1
  Electrical engineering, electronic engineering, Information engineering:1
  Other engineering and technologies:1
  Political Science:1
  Media and communications:1
  Sociology:1
  Other social sciences:1

- Number by Type
  Research Oriented:49
  Education Oriented:15
- Number by Ownership
  Fully or mostly public:56
  50-50 Public and Private:2
  Fully or mostly private:6

Government/Intergovernmental Organisations except Higher Education:15

- Number by Level
  Central and Federal Government:9
  International:2
  Local government:4
- Number by Type
  Government department or government-run general public services:7
  Other Public Non-Profit Institution:2
  R&D Funding and/or R&D Performing bodies:6

Business enterprise:1

- Number by Market sector of unit of affiliation
  Information And Communication:1
- Number by Type
  Private enterprises:1
- Number by Ownership and International Status
  Independent Enterprise:1
- Number by Size
  SME (EU Definition provided underneath after selection):1

COST Country Institutions(26) : Austria , Bulgaria , Cyprus , Czech Republic , Finland , France , Germany , Greece , Hungary , Ireland , Italy , Lithuania , Netherlands , Norway , Poland , Portugal , Romania , Serbia , Slovakia , Slovenia , Spain , Sweden , Switzerland , Turkey , United Kingdom , fYR Macedonia
Near-Neighbour Country Institutions(1) : Russian Federation
COST International Partners(1) : Australia
European Commission and EU Agencies(0)
European RTD Organisations(0)
International Organisations(2)

Network of Proposers - Details

Main Proposer's Details

Title:	Dr	**Gender:**	M
First Name:	Dimitrios	**Year of birth:**	30/01/1977
Last Name:	Koureas	**Years from PhD:**	6
Email:	dimitris.koureas@naturalis.nl	**Telephone Number:**	+31717519251
Institution:	Naturalis Biodiversity Center	**Type of Institution:**	Private Non-Profit without market revenues, NGO
Address of the Institution:	Postbus 9517, 2300 RA Leiden, Netherlands		
Sub-field of Science of Department:		**Core Area of Expertise:**	Electrical engineering, electronic engineering, Information engineering (Biodiversity Informatics)

Secondary Proposers' Details

Australia

Dr Andrew Treloar (Australian National Data Service)

Participating as Secondary Proposer
E-mail: andrew.treloar@gmail.com
Telephone: +61399020572
Core Expertise: Computer and Information Sciences: Research data management, e-research infrastructures
Gender: M
Years from PhD: 19

Austria

Dr Edeltraud Aspöck (Austrian Academy of Sciences - Institute for Oriental and European Archaeology)

Participating as Secondary Proposer
E-mail: edeltraud.aspoeck@oeaw.ac.at
Telephone: +431515816129
Core Expertise: History and Archeology: Prehistory and protohistory
Gender: F
Years from PhD: 9

Bulgaria

Ms Nadezhda Kecheva (National Archaeological Institute with Museum at the Bulgarian Academy of Sciences [Interdisciplinary Research and Archaeological Map of Bulgaria])

Participating as Secondary Proposer
E-mail: n.kecheva@gmail.com
Telephone: +359884721799
Core Expertise: History and Archeology: Archaeology, archaeometry, landscape archaeology
Gender: F
Years from PhD: 0

Cyprus

Dr SORIN HERMON (The Cyprus Institute [STARC])

Participating as Secondary Proposer
E-mail: sorin.hermon@gmail.com
Telephone: +35722208600
Core Expertise: History and Archeology: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 16

Czech Republic

Mr Matej Bozik (Czech University of Life Sciences Prague (CULS))

Participating as Secondary Proposer
E-mail: bozik@af.czu.cz
Telephone: 00420728235114
Core Expertise: Other agricultural sciences: Sustainable production
Gender: M
Years from PhD: 0

Dr David Novák (Institute of Archaeology of the Czech Academy of Sciences, Prague, v.v.i.)

Participating as Secondary Proposer
E-mail: novak@arup.cas.cz
Telephone: +420257014379

Core Expertise: History and Archeology: Archaeology, archaeometry, landscape archaeology
Gender: M
Years from PhD: 1

Dr Monika Sabolová (Czech University of Life Sciences Prague [Department of Microbiology, Nutrition and Dietetics])
Participating as Secondary Proposer
E-mail: sabolova@af.czu.cz
Telephone: 00420608766516
Core Expertise: Health Sciences: Nutrition and dietetics
Gender: F
Years from PhD: 1

Dr Jana Vrbková (Faculty of Medicine and Dentistry, Palacký University Olomouc [Institute of Molecular and Translational Medicine])
Participating as Secondary Proposer
E-mail: jana.vrbkova@upol.cz
Telephone: +420 585 632 057
Core Expertise: Mathematics: Statistics
Gender: F
Years from PhD: 0

Finland

Dr Ari Asmi (University of Helsinki)
Participating as Secondary Proposer
E-mail: ari.asmi@helsinki.fi
Telephone: +358407708729
Core Expertise: Earth and related Environmental sciences: Climatology and climate change
Gender: M
Years from PhD: 7

Dr Ewan O'Connor (Finnish Meteorological Institute)
Participating as Secondary Proposer
E-mail: ewan.oconnor@fmi.fi
Telephone: +358505907973
Core Expertise: Earth and related Environmental sciences: Meteorology, atmospheric physics and dynamics
Gender: M
Years from PhD: 15

Dr Simo Tukiainen (Finnish Meteorological Institute)
Participating as Secondary Proposer
E-mail: simo.tukiainen@fmi.fi
Telephone: +358443405385
Core Expertise: Physical Sciences: Atmospheric retrievals
Gender: M
Years from PhD: 2

France

Mr Damien Boulanger (CNRS)
Participating as Secondary Proposer
E-mail: damien.boulanger@obs-mip.fr
Telephone: 05-61-33-27-71
Core Expertise: Computer and Information Sciences: Theoretical aspects of data curation, data mining and database handling
Gender: M
Years from PhD: 0

Prof Laurent Romary (Inria)
Participating as Secondary Proposer
E-mail: laurent.romary@inria.fr
Telephone: +33630653374
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 29

Mr Kai Salas Rossenbach (French National Institute for Preventive Archaeological Research)
Participating as Secondary Proposer
E-mail: kai.salas-rossenbach@inrap.fr
Telephone: +0664330206
Core Expertise: History and Archeology: Archaeology, archaeometry, landscape archaeology
Gender: M
Years from PhD: 0

Dr Nicolas Sauger (Sciences Po - Fondation Nationale des Sciences Politiques)
Participating as Secondary Proposer
E-mail: nicolassauger@gmail.com
Telephone: +33671255242
Core Expertise: Political Science: Political systems and institutions, governance
Gender: M
Years from PhD: 15

Germany

Dr Peter Wittenburg (Max Planck Gesellschaft - Max Planck Computing and Data Facility)
Participating as Secondary Proposer
E-mail: peter.wittenburg@mpcdf.mpg.de
Telephone: +49 15141858784
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 0

Dr Nazaret Bello Gonzalez (Kiepenheuer-Institut fuer Sonnenphysik)
Participating as Secondary Proposer
E-mail: nbello@leibniz-kis.de
Telephone: +497613198210
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: F
Years from PhD: 0

Dr Kay Graf (Friedrich-Alexander Universität Erlangen-Nürnberg [ECAP - Department of Physics])
Participating as Secondary Proposer
E-mail: kay.graf@fau.de
Telephone: +4991318527265
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: M
Years from PhD: 10

Dr Andreas Haungs (Karlsruhe Institute of Technology [Institut für Kernphysik])
Participating as Secondary Proposer
E-mail: andreas.haungs@kit.edu
Telephone: 004972160823310
Core Expertise: Physical Sciences: High energy and particles astronomy, X-rays, cosmic rays, gamma rays, neutrinos

Gender: M
Years from PhD: 22

Dr Tibor Kálmán (GWDG)
Participating as Secondary Proposer
E-mail: tibor.kalman@gwdg.de
Telephone: +495512011541
Core Expertise:
Gender: M
Years from PhD: 0

Dr Mike Mertens (University of Göttingen [State and University Library, University of Göttingen])
Participating as Secondary Proposer
E-mail: mmerten1@gwdg.de
Telephone: +49 551 39-20475
Core Expertise: Other humanities: Comparative History
Gender: M
Years from PhD: 18

Dr Christoph Quix (Fraunhofer Institute for Applied Information Technology FIT [High Content Analysis and information intensive Instruments])
Participating as Secondary Proposer
E-mail: christoph.quix@fit.fraunhofer.de
Telephone: +49 2241 14 1525
Core Expertise: Computer and Information Sciences: Theoretical aspects of data curation, data mining and database handling
Gender: M
Years from PhD: 15

Dr Tobias Weigel (Deutsches Klimarechenzentrum GmbH [Data management])
Participating as Secondary Proposer
E-mail: weigel@dkrz.de
Telephone: +4940460094431
Core Expertise: Computer and Information Sciences: Geoinformatics, distrubuted systems and cyberinfrastructure
Gender: M
Years from PhD: 2

Dr Daniel Bangert (Georg-August-Universität Göttingen [Niedersächsische Staats- und Universitätsbibliothek Göttingen])
Participating as Secondary Proposer
E-mail: bangert@sub.uni-goettingen.de
Telephone: +4915771489739
Core Expertise: Arts: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 6

Dr Oya Beyan (RWTH Aachen University [Informatik 5])
Participating as Secondary Proposer
E-mail: beyan@dbis.rwth-aachen.de
Telephone: 49 241 80 21512
Core Expertise: Computer and Information Sciences: Data Management, Medical Informatics
Gender: F
Years from PhD: 8

Ms Brigitte Hausstein (GESIS Leibniz Institute for the Social Sciences [Data Archive for the Social Sciences])

Participating as Secondary Proposer
E-mail: brigitte.hausstein@gesis.org
Telephone: +49 221 47694596
Core Expertise: Sociology: Data management; data archiving
Gender: F
Years from PhD: 0

Greece

Dr Fotis Karayannis (ATHENA RESEARCH AND INNOVATION CENTRE - Information Management Systems Institute)
Participating as Secondary Proposer
E-mail: fotis.karayannis@gmail.com
Telephone: +302106208794
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Communications engineering and systems (select for additional explanation)
Gender: M
Years from PhD: 19

Hungary

Dr Attila Kreiter (Hungarian National Museum)
Participating as Secondary Proposer
E-mail: attila.kreiter@gmail.com
Telephone: +205765660
Core Expertise: History and Archeology: Archaeology, archaeometry, landscape archaeology
Gender: M
Years from PhD: 12

Ms Judit Gárdos (Hungarian Academy of Sciences [Centre for Social Sciences])
Participating as Secondary Proposer
E-mail: gardos.judit@tk.mta.hu
Telephone: 0036204188767
Core Expertise: Sociology: Sociology of science
Gender: F
Years from PhD: 0

Ireland

Mr Eoghan Ó Carragáin (University College Cork)
Participating as Secondary Proposer
E-mail: eoghan.ocarragain@ucc.ie
Telephone: 00353214902926
Core Expertise: Computer and Information Sciences: Theoretical aspects of data curation, data mining and database handling
Gender: M
Years from PhD: 0

Italy

Dr Cristiano Bozza (University of Salerno [Department of Physics])
Participating as Secondary Proposer
E-mail: cbozza@unisa.it
Telephone: +39089969132
Core Expertise: Physical Sciences: High energy and particles astronomy, X-rays, cosmic rays, gamma rays, neutrinos
Gender: M
Years from PhD: 18

Dr Giuseppe D'Amico (Consiglio Nazionale delle Ricerche - CNR - Istituto di Metodologie per l'Analisi Ambientale (CNR -IMAA))

Participating as Secondary Proposer
E-mail: giuseppe.damico@imaa.cnr.it
Telephone: +390971427297
Core Expertise:
Gender: M
Years from PhD: 0

Dr Emiliano Degl'Innocenti (Consiglio Nazionale delle Ricerche (CNR) [Istituto Opera del Vocabolario Italiano - OVI])
Participating as Secondary Proposer
E-mail: emiliano.deglinnocenti@cnr.it
Telephone: +393334945358
Core Expertise:
Gender: M
Years from PhD: 9

Dr Paolo Mazzetti (National Research Council of Italy (CNR) - Institute of Atmospheric Pollution Research (IIA-CNR) [Division of Florence])
Participating as Secondary Proposer
E-mail: paolo.mazzetti@cnr.it
Telephone: +39 055 5226591
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Development of scientific computing, data processing, simulation and modelling tools
Gender: M
Years from PhD: 0

Dr LUCIA MONA (CNR - CNR-IMAA)
Participating as Secondary Proposer
E-mail: lucia.mona@imaa.cnr.it
Telephone: +390971427257
Core Expertise: Earth and related Environmental sciences: Meteorology, atmospheric physics and dynamics
Gender: F
Years from PhD: 15

Dr Luca Pezzati (CNR [INO])
Participating as Secondary Proposer
E-mail: luca.pezzati@cnr.it
Telephone: +3905523081
Core Expertise: Physical Sciences: Optics, non-linear optics (theory)
Gender: M
Years from PhD: 23

Dr Paolo Romano (IRCCS AOU San Martino IST)
Participating as Secondary Proposer
E-mail: paolo.romano@hsanmartino.it
Telephone: +390105558288
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 31

Dr Antonio Rosato (University of Florence)
Participating as Secondary Proposer
E-mail: rosato@cerm.unifi.it
Telephone: +390554574267
Core Expertise: Chemical sciences: Databases, data mining, data curation, computational modelling

Gender: M
Years from PhD: 20

Dr Milena Stefanova (ENEA - Italian National Agency for New Technologies, Energy and Sustainable Territorial Development [Department for sustainability of production and territorial systems/Biotechnology and Agro-industry division/Laboratory for agro-food products sustainability, quality and security])
Participating as Secondary Proposer
E-mail: milena.stefanova@enea.it
Telephone: 0039 051 6098 409
Core Expertise: Other agricultural sciences: Sustainable production
Gender: F
Years from PhD: 19

Prof Silvio Tosatto (Universita' degli studi di Padova [Department of Women's and Children's Health])
Participating as Secondary Proposer
E-mail: silvio.tosatto@unipd.it
Telephone: +390498276269
Core Expertise: Biological sciences: Bioinformatics
Gender: M
Years from PhD: 16

Dr Giovanna Zappa (ENEA - Italian National Agency for New Technologies, Energy and Sustainable Economic Development [Biotechnology and Agroindustry Division - Department for Sustainability -])
Participating as Secondary Proposer
E-mail: giovanna.zappa@enea.it
Telephone: +390630483436
Core Expertise: Chemical sciences: Analytical chemistry
Gender: F
Years from PhD: 0

Lithuania

Prof Ruta Petrauskaite (Vytautas Magnus University)
Participating as Secondary Proposer
E-mail: r.petrauskaite@hmf.vdu.lt
Telephone: +37060102117
Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology
Gender: F
Years from PhD: 16

Mr Michael Crusoe (Všį "Darbo eigos")
Participating as Secondary Proposer
E-mail: mrc@commonwl.org
Telephone: +14806279108
Core Expertise: Biological sciences: Bioinformatics
Gender: M
Years from PhD: 0

Ms Ingrida Kelpšienė (Vilnius University)
Participating as Secondary Proposer
E-mail: ingrida.vosyliute@gmail.com
Telephone: +37062045193
Core Expertise: Media and communications: Media and communications, social aspects of information science and surveillance, socio-cultural communication
Gender: F

Years from PhD: 0

Netherlands

Mr Wouter Addink (Naturalis Biodiversity Center)
Participating as Secondary Proposer
E-mail: wouter.addink@naturalis.nl
Telephone: +31717517364
Core Expertise: Biological sciences: Bioinformatics
Gender: M
Years from PhD: 0

Dr Jan-Willem Boiten (Lygature)
Participating as Secondary Proposer
E-mail: janwillem.boiten@lygature.org
Telephone: +31 618639236
Core Expertise: Other medical sciences: Databases, data mining, data curation, computational modelling for other medical sciences
Gender: M
Years from PhD: 23

Dr Luiz Olavo Bonino da Silva Santos (GO FAIR International Support and Coordination Office [GO BUILD])
Participating as Secondary Proposer
E-mail: luiz.bonino@go-fair.org
Telephone: 0624619131
Core Expertise: Computer and Information Sciences: Artificial intelligence, intelligent systems, multi agent systems
Gender: M
Years from PhD: 7

Dr Rob Hooft (Dutch Techcentre for Life Sciences)
Participating as Secondary Proposer
E-mail: rob.hooft@dtls.nl
Telephone: +31627034319
Core Expertise: Chemical sciences: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 25

Mr Dieter Van Uytvanck (CLARIN ERIC)
Participating as Secondary Proposer
E-mail: dieter@clarin.eu
Telephone: +31-(0)850091363
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 0

Dr Rene van Horik (KNAW - DANS - Data Archiving and Networked Services)
Participating as Secondary Proposer
E-mail: rene.van.horik@dans.knaw.nl
Telephone: +31623297389
Core Expertise: Computer and Information Sciences: Theoretical aspects of data curation, data mining and database handling
Gender: M
Years from PhD: 13

Dr Cees Hof (Royal Netherlands Academy of Arts and Sciences (KNAW) - Data Archiving and

Networked Services (DANS))
Participating as Secondary Proposer
E-mail: cees.hof@dans.knaw.nl
Telephone: +31 70 349 44 50
Core Expertise: Biological sciences: Biodiversity, comparative biology
Gender: M
Years from PhD: 20

Dr Enrico Camporeale (Centrum Wiskunde & Informatica [Multiscale Dynamics])
Participating as Secondary Proposer
E-mail: e.camporeale@cwi.nl
Telephone: +31 20 592 4240
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: M
Years from PhD: 10

Norway

Prof Koenraad De Smedt (University of Bergen)
Participating as Secondary Proposer
E-mail: desmedt@uib.no
Telephone: +47 55582052
Core Expertise: Languages and literature: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 28

Mr Ron Dekker (CESSDA ERIC)
Participating as Secondary Proposer
E-mail: ron.dekker@cessda.eu
Telephone: +4755583643
Core Expertise: Economics and business: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 0

Dr Markus Fiebig (NILU - Norsk Instituu for Luftforskning - NILU - Norsk Institutt for Luftforskning [Dept. Atmospheric and Climate Research (ATMOS)])
Participating as Secondary Proposer
E-mail: Markus.Fiebig@nilu.no
Telephone: +47 6389-8235
Core Expertise: Earth and related Environmental sciences: Atmospheric chemistry and composition
Gender: M
Years from PhD: 0

Dr Cathrine Myhre (NILU -Norwegian Institute for Air Research [ATMOS - Dept. Atmospheric and Climate Research])
Participating as Secondary Proposer
E-mail: clm@nilu.no
Telephone: +47063898042
Core Expertise: Earth and related Environmental sciences: Atmospheric chemistry and composition
Gender: F
Years from PhD: 14

Poland

Dr Andrzej Sobecki (Gdansk University of Technology)

Participating as Secondary Proposer
E-mail: andrzej.sobecki@pg.edu.pl
Telephone: +48500137857
Core Expertise: Computer and Information Sciences: Machine learning algorithms
Gender: M
Years from PhD: 1

Portugal

Dr Isabel Castanheira (Instituto Nacional de Saúde Doutor Ricardo Jorge [Food and Nutrition Department/ Laboratory of Reference Materials])
Participating as Secondary Proposer
E-mail: Isabel.castanheira@insa.min-saude.pt
Telephone: +351217519288
Core Expertise: Biological sciences: Biodiversity, comparative biology
Gender: F
Years from PhD: 27

Dr Alexandre Francisco (INESC-ID Lisboa [IDSS Lab])
Participating as Secondary Proposer
E-mail: aplf@tecnico.ulisboa.pt
Telephone: +351213100272
Core Expertise: Computer and Information Sciences: Algorithms, distributed, parallel and network algorithms
Gender: M
Years from PhD: 8

Dr António Santos Silva (Laboratório Nacional de Engenharia Civil [Materials Department])
Participating as Secondary Proposer
E-mail: ssilva@lnec.pt
Telephone: +351 21 8443674
Core Expertise: Materials engineering: Durability of building materials
Gender: M
Years from PhD: 13

Romania

Dr Gabriel Mustatea (National Research and Development Institute for Food Bioresources - IBA Bucharest [Food Packaging Laboratory])
Participating as Secondary Proposer
E-mail: gabi.mustatea@bioresurse.ro
Telephone: +40762675684
Core Expertise: Chemical sciences: Analytical chemistry
Gender: M
Years from PhD: 4

Dr Roxana Radvan (National Institute of R&D for Optoelectronics)
Participating as Secondary Proposer
E-mail: radvan@inoe.ro
Telephone: +40314056398
Core Expertise: Other engineering and technologies: Heritage science, Heritage investigation
Gender: F
Years from PhD: 22

Russian Federation

Mr Alexander Vasilenko (Russian Academy of Science - G.K.Skryabin Institute of Biochemistry and Physiology of Microorganisms Russian Academy of Science (IBPM RAS) [VKM])
Participating as Secondary Proposer
E-mail: vanvkm@gmail.com

Telephone: +7 (916) 714-54-51
Core Expertise: Biological sciences: Bioinformatics
Gender: M
Years from PhD: 0

Serbia

Dr Nevena Veljkovic (Institute of Nuclear Sciences Vinca University of Belgrade - Institute of Nuclear Sciences Vinca [Centre for multidisciplinary research])
Participating as Secondary Proposer
E-mail: nevenav@vinca.rs
Telephone: +381113408471
Core Expertise: Biological sciences: Molecular biology and interactions
Gender: F
Years from PhD: 17

Slovakia

Dr Peter Gömöry (Astronomical Institute of the Slovak Academy of Sciences)
Participating as Secondary Proposer
E-mail: gomory@ta3.sk
Telephone: +421527879182
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: M
Years from PhD: 12

Dr Sergio Javier Gonzalez Manrique (Slovak Academy of Sciences - Astronomical Institute Slovak Academy of Sciences [Solar Physics Department])
Participating as Secondary Proposer
E-mail: smanrique@ta3.sk
Telephone: +421-52-7879137
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: M
Years from PhD: 1

Slovenia

Dr Barbara Koroušić Seljak (Jožef Stefan Institute)
Participating as Secondary Proposer
E-mail: barbara.korousic@ijs.si
Telephone: +38614773900
Core Expertise: Computer and Information Sciences: Algorithms, distributed, parallel and network algorithms
Gender: F
Years from PhD: 21

Dr Brane Leskosek (University of Ljubljana, Faculty of Medicine)
Participating as Secondary Proposer
E-mail: brane.leskosek@mf.uni-lj.si
Telephone: +38615437775
Core Expertise: Other medical sciences: Databases, data mining, data curation, computational modelling for other medical sciences
Gender: M
Years from PhD: 16

Dr Polonca Ropret (Institute for the Protection of Cultural Heritage of Slovenia)
Participating as Secondary Proposer
E-mail: polona.ropret@zvkds.si
Telephone: +38612343118
Core Expertise: Chemical sciences: Spectroscopic and spectrometric techniques

Gender: F
Years from PhD: 11

Dr Benjamin ŠTULAR (Research Centre of Slovenian Academy of Sciences and Arts)
Participating as Secondary Proposer
E-mail: bstular@zrc-sazu.si
Telephone: +38614706387
Core Expertise: History and Archeology: Archaeology, archaeometry, landscape archaeology
Gender: M
Years from PhD: 11

Spain

Prof Alex Sanchez-Pla (Institut Català de la Salut - VHIR [Statistics and Bioinformatics Unit])
Participating as Secondary Proposer
E-mail: alex.sanchez@vhir.org
Telephone: +934405613
Core Expertise: Mathematics: Statistics
Gender: M
Years from PhD: 22

Sweden

Dr Ingemar Häggström (EISCAT Scientific Association)
Participating as Secondary Proposer
E-mail: ingemar.haggstrom@eiscat.se
Telephone: +4698079155
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: M
Years from PhD: 28

Dr Margareta Hellström (Lund University [Physical Geography and Ecosystem Science])
Participating as Secondary Proposer
E-mail: margareta.hellstrom@nateko.lu.se
Telephone: +46462229683
Core Expertise: Earth and related Environmental sciences: Databases, data mining, data curation, computational modelling
Gender: F
Years from PhD: 26

Mr Birger Jerlehag (University of Gothenburg - Swedish National Data Service)
Participating as Secondary Proposer
E-mail: Birger.Jerlehag@snd.gu.se
Telephone: +46 31 7761206
Core Expertise: Political Science: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 0

Dr Anders Tjulin (EISCAT Scientific Association)
Participating as Secondary Proposer
E-mail: anders.tjulin@eiscat.se
Telephone: +46 980 79157
Core Expertise: Physical Sciences: Astrophysics, astronomy, space sciences
Gender: M
Years from PhD: 15

Dr Carl-Fredrik Enell (EISCAT Scientific Association)
Participating as Secondary Proposer

E-mail: carl-fredrik.enell@eiscat.se
Telephone: +46706608657
Core Expertise: Earth and related Environmental sciences: Ozone, upper atmosphere, ionosphere
Gender: M
Years from PhD: 0

Switzerland

Dr Karl Presser (Premotec GmbH)
Participating as Secondary Proposer
E-mail: karl.presser@premotec.ch
Telephone: 0041 79 456 57 78
Core Expertise: Computer and Information Sciences: Theoretical aspects of data curation, data mining and database handling
Gender: M
Years from PhD: 6

Prof Torsten Schwede (SIB Swiss Institute of Bioinformatics [Computational Structural Biology Group])
Participating as Secondary Proposer
E-mail: torsten.schwede@unibas.ch
Telephone: +41 61 207 15 81
Core Expertise: Biological sciences: Bioinformatics
Gender: M
Years from PhD: 20

Turkey

Prof Osman Ugur Sezerman (Acibadem University [Medical Informatics])
Participating as Secondary Proposer
E-mail: ugur.sezerman@acibadem.edu.tr
Telephone: +902164839513
Core Expertise: Biological sciences: Bioinformatics
Gender: M
Years from PhD: 25

United Kingdom

Prof Carole Goble (The University of Manchester)
Participating as Secondary Proposer
E-mail: carole.goble@manchester.ac.uk
Telephone: +441612756195
Core Expertise: Computer and Information Sciences: Theory of scientific computing and data processing
Gender: F
Years from PhD: 0

Mr Alex Hardisty (Cardiff University)
Participating as Secondary Proposer
E-mail: hardistyar@cardiff.ac.uk
Telephone: +442920874761
Core Expertise: Computer and Information Sciences: Distributed systems
Gender: M
Years from PhD: 0

Mr Joseph Padfield (The National Gallery [Scientific])
Participating as Secondary Proposer
E-mail: joseph.padfield@ng-london.org.uk
Telephone: 02077472553

Core Expertise: Other humanities: Heritage Scientist
Gender: M
Years from PhD: 0

Prof Julian Richards (University of York)
Participating as Secondary Proposer
E-mail: julian.richards@york.ac.uk
Telephone: +4401904323930
Core Expertise: History and Archeology: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 32

Dr Mark Hedges (King's College London)
Participating as Secondary Proposer
E-mail: mark.hedges@kcl.ac.uk
Telephone: +442078481970
Core Expertise: Computer and Information Sciences: digital humanities
Gender: M
Years from PhD: 32

Dr Lucy Bastin (Aston UNiversity - Lucy Bastin)
Participating as Secondary Proposer
E-mail: l.bastin@aston.ac.uk
Telephone: +441212043560
Core Expertise: Computer and Information Sciences: Spatial/temporal/geospatial modelling and software development
Gender: F
Years from PhD: 23

fYR Macedonia

Prof Zorica Arsova Sarafinovska (Institute of Public Health of the Republic of Macedonia [Laboratory Reference Centre])
Participating as Secondary Proposer
E-mail: zarsova2002@yahoo.co.uk
Telephone: +389 2 312 50 44 (ext. 106)
Core Expertise: Health Sciences: Public and environmental health
Gender: F
Years from PhD: 10