RDA IG From Observational Data to Information (OD2I IG)

Charter

Name of Proposed Interest Group

From Observational Data to Information (OD2I)

Introduction

At the GEO-XIV Plenary, GEO Secretariat Director Barbara Ryan stated that the "link from data provider research infrastructures to users will become more important" [1]. This statement opens several interesting questions. Who are the users? What data are provided by research infrastructures (RIs)? What kind of links are there between data provider RIs and users? How do RIs provide the data needed by users? Do users produce data? Are such derived data acquired by RIs, further curated, published, processed and used?

Users are of different kinds, ranging from individual researchers and research communities to industry, decision makers and the general public. Data are of different kinds as well, ranging from primary observational, experimental, or computational data to data derived in numerous activities performed by users along contextualized value chains. Data provided by RIs can be primary or derived. While users consume data provided by RIs, many groups of users surely also produce data. Such derived data should be acquired by infrastructures.

In this complex landscape, there is an important constant across infrastructures and domains that lies at the core of the OD2I IG. Along value chains, primary data are interpreted for their meaning in determinate contexts of scientific, industrial, or broadly societal relevance. Within the context of a particular value chain, primary data are uninterpreted. In contrast, meaningful data resulting from data interpretation are information, interpreted data. Primary data thus evolve to become contextually meaningful information further used for both scientific and nonscientific purposes.

With primary focus on (i.e., not exclusively for) observational data and environmental research infrastructures, the OD2I IG studies this constant. Building on collected use cases and existing conceptual frameworks, the OD2I IG advances understanding for how observational data evolve to information, ultimately integrated into bodies of knowledge about natural and human worlds.

In other words, the OD2I IG studies and advances understanding of the relatively unexplored interface between users and infrastructures in the data use phase of the research data lifecycle. It studies and models the roles of, and interactions among, human and computer agents; the data and information consumed and produced by agents in this phase; the performed activities and the systems supporting their execution.

The notion that primary data evolve to information (and knowledge) is increasingly common. Research infrastructures emphasize there is knowledge to gain through observation [2]. Earth observation satellites "provide critical information for global food security" [3]. The European Open Science Cloud (EOSC) is envisioned as an environment that enables turning ever increasing amounts of data "into knowledge as renewable, sustainable fuel for innovation in turn to meet global challenges" [4]. At the 2016 Fall Meeting of the American Geophysical Union (AGU), Rebecca Moore (Google) envisioned the possibility of monitoring a changing planet and "generating precise, actionable information and knowledge".

Following a proposal by members of the OD2I IG that suggests to adopt the Floridi framework [5, 6] with the notion of data interpretation borrowed from Aamodt and Nygård [7], the OD2I IG considers technical aspects of (semantic) information representation in systems and the management of explicit and formal semantics. Connecting data to users relies on systems capable of acquiring, curating and processing the meaning of data generated by users along value chains. A critical aspect is the mechanism for representing information in ways suitable for both machine-to-machine interaction and for presentation to and use by users. Since the ability to exploit any given information demands a specific knowledge on the part of the user, presentations need to consider both user type and intended purpose. Data provide a basis for building information that will lead to decision making. The transition from data to information involves processes of interpretation, in which meaning is attached to data. It is information, and its use against a background of prior knowledge that provides sufficient understanding to allow consequences of decisions to be foreseen.

The OD2I IG aligns to the mission and vision of RDA through the specific concern of socio-technical support for the extraction of information from primary observational data, activities that are primarily carried out by research communities as they make use of data in their everyday work. The OD2I IG will add value by working to realize information and knowledge-based systems layered above the current data systems, resulting in improved usability of data as information by both humans and machines. Of specific emphasis is the outcome that machines are enabled in automated processing of information. The OD2I IG is committed to make a difference in this regard.

User scenario(s) or use case(s) the IG wishes to address

The main motivation for the OD2I IG is the frequent and surely widespread notion that information can be gained from data, specifically from observational data acquired, curated, and published by environmental research infrastructures. The proposition is that observational and other kinds of data ultimately serve to inform and advance our collective understanding of human and natural worlds and, thus, support decision making.

Among the actors making reference to this notion we find research infrastructures, research projects, organizations and other institutions. Software systems are said to support the acquisition (extraction) of information from data and the “transfer of information into knowledge” (e.g. [8]). Inspired by the Helmholtz Motto 2017 [9], the RDA 11th Plenary [10] meeting is held under the theme "From Data to Knowledge". The Integrated Carbon Observation System[1] (ICOS) research infrastructure uses the tagline "Knowledge through observations" [2]. The European Multidisciplinary Seafloor and water column Observatory[2] (EMSO) suggests that the research infrastructure plays "a major role in supporting the

[1] https://www.icos-ri.eu/
[2] http://www.emso-eu.org/

European marine sciences and technology [...] to enter a new paradigm of knowledge in the XXI Century" [11]. The European Cloud Initiative aims at building "a competitive data and knowledge economy in Europe" [12]. Supporting "sense making" of data is argued as the key aim for the European Open Science Cloud (EOSC) [13].

However, we lack a solid shared understanding for what this concretely means. The terms data, information and knowledge are often used casually. Data and information are frequently used interchangeably. A specific concern is socio-technical support for the extraction of information from primary data, activities primarily carried out by research communities; the systematic acquisition and curation of formal meaning of data; the construction and maintenance of information and knowledge-based systems; as well as further processing and use of information. Earth and environmental sciences, in particular, lack the required infrastructural elements for such activity. This is the case even though the concepts of data, information, and data interpretation have been studied in philosophy, information science, and other disciplines for considerable time and, over several decades, the computer science and informatics disciplines have developed related technology. Yet, with principles such as FAIR [14], critical aspects such as formal languages for knowledge representation or interoperable and reusable vocabularies are only just beginning to gain broad recognition among research infrastructures. Advancing research infrastructure *information systems* that systematically integrate the human infrastructure of researchers that make sense of primary data and systematically acquire and curate the derived information will require substantial resources and expertise, and contribute substantially toward implementing the FAIR principles.

The OD2I IG aims at being a platform for discussion and advancement on this subject matter. The IG deals with both technical and social aspects, relevant technical and human infrastructure, and the processes within and between infrastructures needed to acquire, curate, publish, process and reuse information extracted from primary data via data interpretation.

User scenarios and use cases in which infrastructural actors process data to information are of interest to the OD2I IG. Several case study areas / scenarios can already be identified, including: unmanned aerial systems for climate and land-use monitoring; essential variables for understanding the state of biodiversity; aerosol science and the study of atmospheric new particle formation [15]; precision agriculture for the monitoring of disease outbreaks [16]; and intelligent transportation systems detecting vehicles and describing their properties [17]. There exist numerous other user scenarios where primary (observational) data are turned into information about the human and natural worlds, e.g. early warning systems for earthquakes and tsunami, weather information systems, etc.

In the remaining of this section we highlight two use cases the OD2I IG has addressed in more detail, so far. The first use case is in aerosol science. Using observational data for the size distribution of (polydisperse) aerosol - gas-particle mixtures sampled and measured at regular time intervals several times per hour by an instrument deployed at some fixed location - the aerosol science research community studies particle formation processes. In this context, measured particle size distribution are observational data, uninterpreted data, while descriptions about particle formation processes are information, interpreted data or meaningful well-formed data. Descriptions about particle formation processes are the result of interpreting observational data. Descriptions include the spatial location, temporal duration, growth rate and other attributes of formation processes. Individual descriptions are further processed, e.g. to compute summary statistics - ultimately published in literature. An interesting challenge in this research community is that, while observational data are used along the same value chain, the syntax and semantics of gained information are not harmonized across research groups in the community. The

community lacks community standards for the description of particle formation processes. As a result, information is difficult to integrate and communities are hindered in their ability to perform further integrated use of information. Aspects primarily relevant to the OD2I IG are i) the distinction between observational data and information, including different data syntaxes and the representation of meaning; ii) the process of observational data interpretation, in particular the activities and agents involved; iii) and the systematic acquisition, curation, publishing and further processing and use of information in infrastructures.

The second use case is in biodiversity and conservation science. Essential Biodiversity Variables (EBVs) are part of an information supply chain, conceptually positioned between raw data (i.e., primary data observations) and indicators (i.e., synthetic indices for reporting change). EBV-useable data, EBV-ready data, and EBV-derived data respectively represent raw observational data, harmonized/aligned data sets, and products derived from processing with statistical models. When measured as, for example species distributions and abundances such products can be used, to quantify changes over space and time. Primary biodiversity observations record the presence (or absence) of a species at specific geographical locations and points in time. Metadata often also include the name of the person recording the observation, as well as many other attributes, such as the institution owning the observation, the collection catalogue number, the reference to the journal article that is the taxonomic basis of the observation, etc. When such raw data are structured, well-formed and based on comparable measurement units they are usable for producing EBV data products, hence 'EBV-useable'. EBV-ready data consist of subsets of EBV-usable data, harmonized and bounded by species, space and time; often with supplemental attributes providing context for further interpretation e.g., native / alien status. EBV-ready data are first-level information products that permit analysis of, for example invasiveness, i.e., changes in the distribution and abundance of a species over time. Derived EBVs are also information products. They are secondary products, with a higher information content derived from processing (interpreting) EBV-usable or EBV-ready data with statistical models. As with the preceding use case in aerosol science, aspects relevant to the OD2I IG centre around i) the distinction between the primary data and the supplemental attributes that add meaning; ii) the interpretation process itself; and iii) systematization of the process such that information products are complementary and comparable.

Objectives

The OD2I IG aims at being a development platform for the subject of interest to the group, as detailed above. It acts as an umbrella, aiming to engage stakeholders; collect use cases, solutions and challenges as reported by stakeholders; develop solutions to unresolved challenges; and transfer solutions across stakeholders.

Stakeholders include individuals interested in the subject, e.g., individual researchers and engineers, as well as individuals representing larger groups or organizations, in particular research infrastructures and e-Infrastructures that publish observational data and support their analysis in one or more research communities, operational infrastructures such as those operating early warning and emergency systems, or individuals representing small and medium enterprises that specialize on aspects of the subject of interest to the OD2I IG.

The IG aims at identifying a framework that can act as a reference conceptualization for the subject of interest. The framework by Floridi [5, 6] with the notion of data interpretation by Aamodt and Nygård [7] was already discussed and will be evaluated against others. The agreed reference conceptualization will

ground our understanding of the distinction of observational data and information as well as the relevant activities and agents in between. Furthermore, it will build a foundation on which the OD2I IG will operate and develop solutions for its stakeholders.

The OD2I IG recognizes the importance of engaging stakeholders in order to learn from a wide range of communities and practices as well as to devise solutions that are viable and practical across stakeholders. Engaging stakeholders will be a running activity that relies on representing the IG at workshops, conferences and other relevant events as well as on peer solicitation.

As a bottom-up activity, we have started collecting comparable use case descriptions. The OD2I IG intends to continue this activity using a use case description template that captures the needed information for downstream use case analysis aimed at identifying common patterns for which the IG could decide to develop solutions.

Development of solutions for specific identified and unresolved challenges will be tasked to specific working groups, that may possibly be RDA WGs with a corresponding Case Statement designed to address a particular concern over the WG's lifetime. However, working groups relevant to the OD2I IG are not limited to formally endorsed RDA WGs and may include other efforts. The role of the OD2I IG is to consider the proposed solutions and integrate them into the overall IG framework, and advocate adoption across stakeholders.

The theme and work of the OD2I IG relates to activities within and outside of RDA. Within RDA, of particular interest are the Data in Context IG[3], the Data Fabric IG[4], the WDS/RDA Assessment of Data Fitness for Use WG[5], the Data Foundation and Terminology WG[6], the Research Data Provenance IG[7], the Virtual Research Environment IG[8], the Reproducibility IG[9], the Long tail of research data IG[10], the Global Water Information IG[11], the Geospatial IG[12]. A detailed analysis for how the OD2I IG relates to and differs from these RDA groups, as well as relationship management, will be tasks of the OD2I IG. Here we limit the description to a brief summary.

Initiated early in the RDA history, the intention of the Data in Context IG was to encourage rich metadata so that assets could be understood in context. Focusing on rich description of assets, a primary difference between Data in Context IG and OD2I IG is in the data interpretation aspect of evolving observational data to information with formal meaning. Originally a joint initiative of the first RDA WGs, the Data Fabric IG proposed to develop the notion of "data fabric" as a concept overarching the work of the individual WGs. According to the Infrastructure Component View, a data fabric consists of the infrastructure used to manage data, information, and knowledge. The Data Fabric IG underscored that primary data are in a "continuous process of being enriched and analyzed creating new derived data." The OD2I IG focuses on the information content of derived data and on the formal representation of that

[3] https://www.rd-alliance.org/groups/data-context-ig.html
[4] https://www.rd-alliance.org/group/data-fabric-ig.html
[5] https://www.rd-alliance.org/groups/assessment-data-fitness-use
[6] https://www.rd-alliance.org/groups/data-foundation-and-terminology-wg.html
[7] https://www.rd-alliance.org/groups/research-data-provenance.html
[8] https://www.rd-alliance.org/groups/vre-ig.html
[9] https://www.rd-alliance.org/groups/reproducibility-ig.html
[10] https://www.rd-alliance.org/groups/long-tail-research-data-ig.html
[11] https://www.rd-alliance.org/groups/global-water-information-interest-group.html-0
[12] https://www.rd-alliance.org/groups/geospatial-ig.html

content. As some use cases show, rather than tabular data, the result are semantic descriptions. Since the quality of such semantic descriptions is superior compared to tabular data (with generally implicit semantics), the OD2I IG relates to the WDS/RDA Assessment of Data Fitness for Use WG which aims at better understanding of and the possibility to assess data quality. The Data Foundation and Terminology WG (completed) has resulted in Core Term Definitions, including for Digital Object. The bitstream representing a Digital Object may represent information. However, the Core Term Definitions do not expand on this aspect. The OD2I IG focuses on information encoded as bitstreams (data) whereby information is primarily about phenomena of interest to research. Since information is derived from primary data, the OD2I IG has a clear relation with the Research Data Provenance IG (and related WGs). Tracking the provenance of information extracted from data in data interpretation is important and is expected to be an aspect the OD2I IG will touch upon, likely in collaboration with the Research Data Provenance IG. The OD2I IG also relates to Virtual Research Environment (VRE) IG because data interpretation implemented as software workflows is typically embedded in VREs. This aspect will also be elaborated in collaboration with the Virtual Research Environment IG. By focusing on the information resulting from data interpretation and the acquisition and curation of such information in systems, the OD2I IG contributes to the reproducibility of scientific claims from primary data and, hence, relates to the Reproducibility IG. As data and their information content derived in interpretation performed by research communities are seldom systematically acquired by research infrastructures, derived data are generally part of the "long tail", typically resting on the hard drives of researchers and sometimes acquired by the corresponding universities and research institutions. In this aspect, the OD2I IG strongly relates to the Long tail of research data IG. The OD2I IG also relates to domain specific groups that focus on domain information, such as the Global Water Information IG. Finally, the geospatial character of observational data and information about human and natural worlds spans a relation to the Geospatial IG.

Participation

The following communities are relevant to the OD2I IG, a subset of which will be involved: Research infrastructures, research communities, ICT specialists, data infrastructures, data centers.
Research infrastructures, in particular those building on observatories and operating a "portal" through which research communities can access data that are uninterpreted in the context of the value chain of interest to the community, are important communities for the OD2I IG to engage with. In general, research infrastructures are providers of the data that serve as input to data interpretation and from which information is extracted. Example infrastructures include the pan-European research infrastructures on the ESFRI roadmap such as ICOS, EMSO, LifeWatch[13] as well as national research infrastructures, such as FRontiers in Arctic marine Monitoring[14] (FRAM) [18] or the Station for Measuring Ecosystem-Atmosphere Relations[15] (SMEAR) [19]. Such research infrastructures specialize on the acquisition, curation, publishing of observational data, and may provide services for the processing and use of data [20]. Research infrastructures will be engaged in the OD2I IG via representative individuals that are either involved in or collaborate with the infrastructures. These representatives will thus liaise between the IG and research infrastructures.

[13] https://www.lifewatch.eu/
[14] https://www.awi.de/en/science/biosciences/deep-sea-ecology-and-technology/observatories/fram-ocean-observing-system.html
[15] https://www.atm.helsinki.fi/SMEAR/

Research communities form the social infrastructure that interprets the primary data published by research infrastructures, thereby generating information. The OD2I IG will engage representatives of research communities relevant to the studied use cases in order to understand the value chains along which research communities interpret observational data to extract information about natural and human worlds. Research communities are key to learn about the relevant data interpretation workflows as well as the required data and resulting information objects. For example, as it interprets observational data published by the SMEAR research infrastructure, the aerosol research community is relevant to a use case studied by the OD2I IG. Similarly to research infrastructures, the OD2I IG will engage research communities through representatives, either directly involved in the IG or indirectly through personal contacts.

A third community relevant to the OD2I IG is the one formed by ICT specialists. ICT specialists provide the IG with the expertise required to develop information systems that acquire, curate, and publish information collected by research communities in analysis of primary data published by research infrastructures. In OD2I IG, information is generally assumed to be machine actionable, i.e. the meaning of data must be formal. In the context of the IG and in collaboration with other communities, ICT specialists gather and analyse use cases to identify commonalities and model the relevant activities, entities, and agents involved in primary data interpretation, extraction of information, information representation. Since it is (currently) primarily driven by ICT specialists, the OD2I IG will directly involve numerous representatives from this community.

Data e-Infrastructures such as EUDAT[16] or EGI[17] - as well as overarching infrastructure such as the EOSC - are a fourth relevant community. Of primary interest to the OD2I IG are their capabilities for executing data interpretation workflows, possibly embedded in Virtual Research Environments (VREs). Such e-Infrastructures expose research communities with software workflows designed to analyse and interpret data for specific purposes along value chains from primary data to information. Tailored for individual value chains, workflows can be designed so that research communities can focus on the data use phase of the research data lifecycle [20] leaving the specifics of data conversions, access and acquisition as well as information representation to the software systems of research infrastructure. Workflows, the VREs in which they are embedded, and the underlying infrastructures interact to provide the right data in the right syntax to research communities for analysis, and to acquire the resulting information as meaningful well-formed data in information systems.

e-Infrastructures such as EUDAT and data centers such as PANGAEA[18] - and, more abstractly, systems such as the EOSC, or similar - are relevant for the OD2I IG in their role for acquiring and curating information resulting from data interpretation. Such a configuration and distribution of roles among the various actors may be a topic for discussion and development within the OD2I IG.

Outcomes

Supported by the objectives listed above, the OD2I IG aims toward the following main outcomes. A first important outcome is the systematic acquisition by infrastructures of information generated by research communities. As we discussed, information is meaningful well-formed data and (to varying degree) should be intelligible to both human and computer agents. The acquisition of visual data products, such as

[16] https://eudat.eu/
[17] https://www.egi.eu/
[18] https://www.pangaea.de/

maps, resulting from analysis of primary data is thus not sufficient. The acquisition and curation of the information content in visual data products is what matters primarily to this IG.

As concrete examples: In the aerosol use case this outcome could mean that a research infrastructure, e.g. ACTRIS[19] or SMEAR, or an e-Infrastructure, e.g. EUDAT, provides a service for the acquisition of information describing particle formation processes. As a result, the data encoding such descriptions will not merely remain on the hard drives of those who perform data interpretation. Furthermore, such data will conform to a community agreed metadata schema and be represented using formal knowledge representation language. In the biodiversity use case, n-dimensional layers representing information are the norm and services for manipulating, visualizing and relating them one to another increasingly will be in demand.

Linked to this, a second outcome is for infrastructure to support research communities with workflows that abstract from the specifics of how observational data are obtained and how information are encoded and acquired as data by infrastructures. Infrastructure should enable research communities to focus on the core activity of data interpretation.

These two outcomes are measured by the number of research communities and the number of observational data interpretation workflows that are implemented in this or a comparable manner.

As a result of these outcomes, we should see systems with improved capabilities for the acquisition, curation, publishing and further processing and use of information. In other words, a third outcome is the realization of information and knowledge-based systems layered above the current data systems. Recall that, among other information such as about relevant instrumentation or organizations and people, of primary interest to the OD2I IG are systems that curate information about the natural and human worlds of interest. Success is measured by the number of developed information systems.

The objectives of the OD2I IG are expected to result in improved usability of data as information by both humans and machines. As scientific information is largely encoded for human experts to consume, of particular emphasis is the outcome that machines are enabled in automated processing of information. The OD2I IG is committed to make a difference in this regard. Explicit access to formal meaning is expected to improve the interoperability and usability of data curated by infrastructures. The OD2I IG will thus contribute to implementing FAIR, in particular some of the aspects on interoperability and reusability of *data*, more than metadata.

Ultimately, the OD2I IG aims to contribute to improved collective understanding of processes and best practices of data analysis and, in particular, human-in-the-loop data interpretation activities, including the relevant input and output data and information objects and the involved agents and activities. The outcome here is that infrastructures and research communities will better appreciate the plurality of research data, that primary observational data are to be distinguished from derived data that encode information about phenomena that are not directly observable, or data encoding statistical information about phenomena. The differences are not just in encoding and syntax; they are principally in semantics, the distinction between data resulting in observation - the activity of estimating the value of a property of a feature - and information about complex phenomena that we aim to understand.

[19] http://actris2.nilu.no/

Mechanism

The OD2I IG will propose sessions for physical meetings at RDA Plenaries. Between Plenaries, the IG will meet in regular monthly conference calls fixed on an agreed day and time. To facilitate participation from different regions, we will consider introducing European/Australian/American-friendly calls. We will also make use of regular conferences (e.g. EGU) or workshops to organize informal meetings for members of the OD2I IG who attend the conference. This will be organized accordingly through event planning and polling for attendance. Collaboration on shared documents occurs via Google Drive. The folder was already set-up and shared with IG members. Unless better approaches are proposed, we will continue this practice. Furthermore, we will use the RDA IG webpages to archive documents, e.g. meeting minutes.

Timeline

The following milestones and goals are envisioned for the first 12 months:

1. OD2I IG kick-off session at Plenary 11 in Berlin
2. Liaise with related RDA IGs and WGs as well as groups outside RDA (e.g. GEO/GEOSS)
3. Develop the OD2I IG's reference conceptualization
4. White paper on developed reference conceptualization
5. Finalize the template for use case descriptions
6. Update existing use cases to reflect possible changes to the template
7. Collect new use cases and align them with the reference conceptualization
8. Analyse the use cases for commonalities and differences
9. Identify and report common challenges
10. Collect feedback from teams implementing use cases

Potential Group Members

FIRST NAME	LAST NAME	EMAIL	AFFILIATION
Markus	Stocker (co-chair)	markus.stocker@tib.eu mstocker@marum.de	German National Library of Science and Technology (TIB) and MARUM Center for Marine Environmental Sciences (PANGAEA)
Stefano	Nativi (co-chair)	stefano.nativi@cnr.it	Coordinator of Earth and Space Science Informatics laboratory of CNR-IIA
Jay	Pearlman (co-chair)	jay.pearlman@ieee.org	Director, FourBridges
Ari	Asmi (co-chair)	ari.asmi@helsinki.fi	University of Helsinki
Steven	Browdy	steveb@omstech.com	President, OMS Tech

Alex	Hardisty	hardistyar@cardiff.ac.uk	Director of Informatics Projects, School of Computer Science & Informatics, Cardiff University, UK
Jacco	Konijn	j.l.konijn@uva.nl	Project Manager University of Amsterdam
Frank	Schluenzen	frank.schluenzen@desy.de	HPC/Scientific Computing Staff

References

[1] ".@BarbaraRyanGEO at #ENVRIcomm side event: "The link from data provider RIs to users will become more important" #GEOWeek17" (23 Oct 2017). https://twitter.com/ICOS_RI/status/922483317471109121

[2] ""ICOS is a basic brick in the overall global measurements system" Interview w/ ICOS ERIC GA Chair Jean-Marie Flaud: http://www.icos-ri.eu/news/2016/icos-eric-1st-year-anniversary-interview-jean-marie-flaud" (28 Nov 2016). https://twitter.com/ICOS_RI/status/803156982349729793

[3] "Earth observation satellites will provide critical information for global food security @esa @ESA_EO @CarstenLinz @SAPB1HANA @SAPLeonardo" (17 May 2017). https://twitter.com/AschbacherJosef/status/864707075334692864

[4] Commission High Level Expert Group on the European Open Science Cloud (2016). Realising the European Open Science Cloud. https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

[5] Floridi L. (2010). Information—A Very Short Introduction. Oxford University Press.

[6] Floridi, L. (2011). The Philosophy of Information. Oxford University Press.

[7] Aamodt, A. and Nygård, M. (1995). Different roles and mutual dependencies of data, information, and knowledge – An AI perspective on their integration, Data & Knowledge Engineering, vol. 16, no. 3, pp. 191-222. https://doi.org/10.1016/0169-023X(95)00017-M

[8] Ocean Network Canada. Oceans 2.0. http://www.oceannetworks.ca/innovation-centre/smart-ocean-systems/ocean-observing-systems/oceans-20

[9] Helmholtz. Das Motto 2017 lautet "Von Daten zu Wissen". https://www.helmholtz.de/ueber_uns/die_gemeinschaft/mission/motto_2017/

[10] Research Data Alliance. RDA Eleventh Plenary Meeting, Berlin, Germany. https://www.rd-alliance.org/plenaries/rda-eleventh-plenary-meeting-berlin-germany

[11] European Multidisciplinary Seafloor and water-column Observatory Development. Work packages. http://www.emsodev.eu/work_packages.html

[12] European Commission (2016). Communication: European Cloud Initiative - Building a competitive data and knowledge economy in Europe. https://ec.europa.eu/digital-single-market/en/news/communication-european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe

[13] Nature Editorial (2017). Don't let Europe's open-science dream drift. Nature 546, 451 (22 June 2017) https://doi.org/10.1038/546451a

[14] Wilkinson, M. D., Dumontier, M., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3. https://doi.org/10.1038/sdata.2016.18
[15] Stocker, M., Baranizadeh, E., Portin, H., Komppula, M., Rönkkö, M., Hamed, A., Virtanen, A., Lehtinen, K., Laaksonen, A., Kolehmainen, M. (2014). Representing situational knowledge acquired from sensor data for atmospheric phenomena. Environmental Modelling & Software, 58:27-47. https://doi.org/10.1016/j.envsoft.2014.04.006
[16] Stocker, M., Nikander, J., Huitu, H., Jalli, M., Koistinen, M., Rönkkö, M., Kolehmainen, M. (2016). Representing Situational Knowledge for Disease Outbreaks in Agriculture. Journal of Agricultural Informatics, 7(2):29-39. https://doi.org/10.17700/jai.2016.7.2.290
[17] Stocker, M., Rönkkö, M., Kolehmainen, M. (2014). Situational Knowledge Representation for Traffic Observed by a Pavement Vibration Sensor Network. IEEE Transactions on Intelligent Transportation Systems, 15(4):1441-1450. https://doi.org/10.1109/TITS.2013.2296697
[18] Soltwedel, T., Schauer, U., Boebel, O., Nöthig, E.-M., Bracher, A., Metfies, K., Schewe, I., Klages, M., Boetius, A. (2013): FRAM - FRontiers in Arctic marine Monitoring: Permanent Observations in a Gateway to the Arctic Ocean. OCEANS - Bergen, 2013 MTS/IEEE. https://doi.org/10.1109/OCEANS-Bergen.2013.6608008
[19] Hari, P., Nikinmaa, E., Pohja, T., Siivola, E., Bäck, J., Vesala, T., Kulmala, M. (2013). Station for Measuring Ecosystem-Atmosphere Relations: SMEAR. In Hari, P., Heliövaara, K., Kulmala, L. (eds.): Physical and Physiological Forest Ecology. Springer Netherlands. https://doi.org/10.1007/978-94-007-5603-8_9
[20] Hardisty, A., Nieva de la Hidalga, A., Lear, D., Magagna, B., Atkinson, M., Jeffery, K.G., Martin, P., Zhao, Z. (2016). Deliverable 5.2: A definition of the ENVRIplus Reference Model. http://www.envriplus.eu/wp-content/uploads/2015/08/D5.2-A-definition-of-the-ENVRIplus-Reference-Model.pdf