

Report of the



Content: Report of the D4Science World User Meeting

Reporters: Johannes Keizer, Marc Taconet, Anton Ellenbroek,
Alhaji Jallow, Francesco Calderini, Donatella Castelli,
Nicolas Bailly

Prepared by Teresa Iniesta and Erna Klupacs

TABLE OF CONTENTS

| | |
|---|----|
| 1. Acronyms | 3 |
| 2. Meeting Objectives | 4 |
| 3. Sessions Reports | 5 |
| Annex 1: Meeting Agenda | 15 |
| Annex 2: List of Participants | 19 |
| Annex 3: Organization of the user communities' presentations | 22 |
| Annex 4: Summary by needs expressed during the presentations | 23 |
| Annex 5: Needs as reported by participants on D4Science offering | |

1. ACRONYMS

CenSeam - Global Census of Marine Life on Seamounts

CIARD - Coherence in Information for Agricultural Research for Development

CNR - Italian National Research Council

CoML - Census of Marine Life

CWP - Coordinating Working Party on Fishery Statistics

EC - European Commission

EGEE - Enabling Grids for E-science

EGI - European Grid Initiative

FAO - Food and Agriculture Organization

GDI - Geospatial Data Infrastructure¹

GIS - Geographic Information System

IEE - Independent External Evaluation

IFM-GEOMAR - Leibniz Institute of Marine Sciences at the Christian-Albrechts Universität zu Kiel

IOC - Intergovernmental Oceanographic Commission

IODE - International Oceanographic Data and Information Exchange

ISTI - Institute of Information Science and Technologies

JRI - Java/R²

MPA - Marine Protected Area

OAI - Open Archives Initiative

OBIS - Ocean Biogeographic Information System

SDI - Spatial Data Infrastructure³.

UN - United Nations

VME - Vulnerable Marine Ecosystem

VMS - Vessel Monitoring System

VRE - Virtual Research Environment

WFC - World Fish Center

¹ GDI is also used to refer to an SDI on the Grid

² JRI allows calling R from Java

³ SDI is a framework of spatial data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way (http://en.wikipedia.org/wiki/Spatial_data_infrastructure)

2. OBJECTIVES

The Objectives of the D4Science World User Meeting were the following:

- * Follow up on opportunities and innovative projects implementing advanced e-Infrastructures for multi-disciplinary scientific communities;
- * Share experiences, best practices and discuss the most recent advances in e-Infrastructures with emphasis on how to exploit possible synergies;
- * Showcase the latest experiences in building dynamic monitoring progress reports and the most recent results in harmonisation and combination of distributed structured data sources for generation of enhanced indicators;
- * Identify the requirements of the user communities and discuss the road map for D4Science-II Virtual Research Environments.

3. SESSION Reports

Presentations may be downloaded from:

<http://www.d4science.eu/worldusermeeting/presentations>

25 November 2009

Day 1 - Fair for broad audience outreach

Opening remarks

In the introductory session Dr. Anton Mangstl, the Director of FAO's Knowledge Exchange and Capacity Building Division emphasized the collaboration of his division with the Fishery Department through D4 Science. He mentioned the Research4Life partnership, between publishers and the UN systems, CIARD and e-Agriculture as important examples of FAO's activities in this area.

Welcome and meeting objectives

Johannes Keizer also of the Knowledge Exchange and Capacity Building Division and D4Science Project Outreach manager then explained that although FAO is working on very basic needs in developing countries, it also has to be involved in these leading edge initiatives in order to fulfill at global level its role of knowledge broker.

Keynote speech on "Digital Library Infrastructure Support for Virtual Collaboration in Science"

The first Keynote speaker was Peter Young, former director of the Asia Department of the Library of Congress, but more widely active in eScience. In his presentation he emphasized that science is always more data intensive and data driven: Wet science develops to digital science. Looking to old data and linking old, new and intermediate data, keeping data sets alive and available, that is now on the agenda. Articles are only the tip of the iceberg, datasets are the real iceberg.

Vision of the D4Science project, including D4Science-II

Over the next 5 years more scientific data will be produced than has been previously during the entire human history before. This is the challenge of eScience. After the first break, Donatella Castelli, D4Science Scientific Coordinator (National Research Council, ISTI-CNR), spoke about the importance of creating tools for supporting the e-science, for use by a large number of scientists at an affordable price.

She also talked about how initially, digital libraries were thought of not only as a tool for delivering content, but also as a collaborative tool for creating content, hence Virtual Research Environment was the term used.

Keynote speech on "The Data Conservancy: A Digital Research and Curation Virtual Organization"

The second Keynote speaker was Sayeed Choudhury, of the National Science Foundation DataNet project at Johns Hopkins University. He explained the problems of Data conservancy: a digital research and curation organization; curation is about the data trans-disciplinary use).

He pointed out that preservation is a precondition: even if we have had many cases of data lost, we still don't know what means preserving on a long term. Print material has to be interpreted by our eyes; and digital data has to be interpreted by machines, and this raises many questions.

The road map is not having a rigid plan: there is no one way to design a cyber infrastructure. He used a metaphor with the railroad in the US: initially, there were local networks of railroads with their specific characteristics, and when it came to linking East and West US, it was not possible. Only by then was a standard established after which all local peculiarities disappeared.

Sayed Choudhury's work has many similarities with D4Science, but also some complimentary objectives. The intellectual merit of his team comes from infrastructure research and development to information science and computer science research. He pointed out that broader impacts need educational requirements, and that sustainability implies some business requirements in this field.

He spoke about Technical architecture with emphasis on modularity, and about the need for common conceptualization (semantic technologies). As an example, he used the concept of observations which are the foundation of research studies.

OAI protocol is used to access underlying data, and prevalence means access down to raw data used.

Multi site user research methods are a blend of case study and domain comparisons, depth and breadth, local and global.

The sustainability aspects were described as relying on a combination of tactics: diversified portfolio of funding and perspectives; alignment with existing institutional priorities; the leveraging on partners' sustainability mechanisms; the need of focus on economies of scale; the need of considering business requirements as equal to other requirements; and finally he pointed out the importance of integrating the findings of Blue Ribbon Task Force on sustainable digital preservation and access.

Sayed Choudhury showed his interest on linking various communities with international initiatives, but doesn't know how this will work in terms of protocols and financial arrangements.

Presentation on "Scientific Data e-Infrastructures in the European Capacities Programme"

Krystina Marek from the European Commission also spoke about the migration from wet labs to digital labs. She underlined the necessity of data scientists in a world in which science becomes always more data-centric. Communities have established their infrastructure, but these are heterogeneous because of their nature and of history. So the goal now is to promote virtual scientific cross-disciplinary infrastructures, and she stressed the significant investments EU is making in support of e-infrastructures.

In the discussion which followed the question was raised if the EU could not come to more substantial Infrastructure projects throughout different disciplines. The difficulty underlined is that infrastructure projects need more time for completion. Mr. Carlos Morais-Pires responded that the Commission is taking steps in this direction with new opportunities offered to 8th Framework Programme (FP8): funding programmes can now go directly to the European parliament to advocate for more funding, and to look for possibility of joint programmes with EU member states. Calls for greater funding will also target more aggregated communities.

Keynote speech on "Fishery science and policy-making: Connecting information and decision"

Serge Garcia, former director of the Fisheries Management Division (Fisheries and Aquaculture Department of FAO) focused his presentation on fishery science and policy. He pointed that the fishery system is one of the most complex systems on earth. It can be studied in multiple time units, it has a complex governance and its policy is influenced by some cross-sectoral connections like the socio economic policy, the environmental policy, the maritime policy and the foreign policy. It also has some cross-scale connections, national, regional and global, with science, market, government and civil society.

After providing an overview of the actual situation on access, processing and diffusion, Serge Garcia finished his presentation by pointing to the need to use Information and Communication Technology to

foster the development of a community of practice around fishery science and management, with perhaps many interconnected, smaller and more specialized communities around sub-sectors (e.g. artisanal fisheries) or themes (e.g. roles games). Within such effort, to develop open source platforms that will accelerate the collaborative development and diffusion of inter-disciplinary, bio-economic, behavioral, and ecosystem models as well as participatory role games; and to develop pool computing for large model runs.

The discussion which followed stressed two points. First concern was expressed as to how developing countries can benefit from such tools: many of these run risks to be one-way tools to pass data from developing to developed countries; so what can be done to reverse this, and make sure there is enough capacity building. Second, instead of first focussing on sectoral issues, and then cross sector approaches, considerations should be given to directly benefit from an the horizontal approach which the infrastructure will enable.

Community fair – Introduction of the D4Science communities, their challenges and the synergies

The Community fair – “Introduction of the D4Science communities, their challenges and their synergies” was hosted by Marc Taconet, Senior Fishery Information Officer at FAO with the participation of Nicolas Bailly, FishBase Project Manager (The WorldFish Center), Luigi Fusco, Earth Observation Programme Directorate (European Space Agency), and Richard Grainger, Chief of the Fisheries and Aquaculture Information and Statistics Service (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations).

Regarding their objectives in D4Science, Marc Taconet pointed out that FAO is also responsible for coordinating the user community inputs to the project. In this respect, it is clear that the more communities get involved the more synergies are identified, and the ensuing cross-fertilization of data and application among communities feeds positively into sustainability considerations. It was also evident that building larger communities has to be considered an investment, since this process takes time and has a relatively steep learning curve and low immediate return: the current D4Science community must be considered as a precious achievement, both for the project and the community actors themselves.

Luigi Fusco mentioned that building an earth observation portal is a challenging task and that ESA's main objective in the project is building shared user digital libraries on the fly, relying on the GRID-based services provided by the infrastructure.

Nicolas Bailly reported that WFC primary objective is implementing the AquaMaps modelling approach for aquatic species distribution maps, relying on the e-infrastructure services and computational resource capabilities: it currently takes around 3 days for a normal PC to calculate the 200.000 computations for elaborating 9000 species distributions. Secondly, they are definitively looking for synergies with other virtual research environments (e.g. the Environmental Monitoring community providing environmental data as input for their modelling approach).

Richard Grainger, chief of the Fishery Information and Statistics Service represented the Food and Agriculture Organization of the UN (FAO). FAO is involved with D4Science since January 2008 and is contributing with two scenarios, within the so called Fishery and Aquaculture Resource Monitoring community: the Integrated Capture Information System (ICIS) and the Fishery Country Profile Production System (FCPPS).

Richard Grainger stated that FAO's initial internal driver was the Independent External Evaluation (IEE) the organization is undergoing, which pushed FAO towards improving knowledge dissemination management and activating international partnerships. This translated into the materialization of the two concrete scenarios within the fisheries and aquaculture department.

Another point of discussion was to what extent does D4Science facilitate the needs of user communities and what are the weaknesses encountered. Marc Taconet stated that 2 years is too short

a period for fairly answering the question, more iterations would be needed. Good progress has been shown, although the current VREs are not yet at a stage to be delivered to the final users.

Luigi Fusco added that the software is not yet completely mature, although one should separate infrastructure from user operability.

Richard Grainger reported that the members of the Coordinating Working Party (CWP), the standardization body in charge of global fishery statistics, agreed to take part to the project, hoping for phase II to reach more tangible results.

Community fair – Showcasing “Building dynamic research environments using D4Science technology”

The Community fair - showcasing “Building dynamic research environments using D4Science technology” started with Anton Ellenbroek who gave a brief overview of the development process followed: starting from the users’ vision to the developers, through the requirements tracking.

Veronica Guidetti demonstrated how to interactively setup and develop a VRE, as an aggregator of contents and functionalities, utilizing the VRE management wizard application.

Kristin Kashner showed how to operate the AquaMaps VRE for generating a species distribution map, following the AquaMaps modelling approach.

Anton Ellenbroek assembled and generated a report, collecting and integrating heterogeneous resources, such as the previously generated AquaMaps species map.

Panel discussion “How to exploit projects’ synergies”

The last session of the day was a Panel discussion on how to exploit projects’ synergies, chaired by Carlos Morais-Pires, Head of Sector Scientific Data Infrastructure (Information Society and Media Directorate-General, European Commission). He outlined that Infrastructure projects need more time to implement their outcomes since they need to satisfy the needs of large groups of potential customers and deliver production quality instruments. Communities that may benefit from infrastructures should be more aggregated and present their requirements collectively. This approach will increase their chances of being funded for the realization of such infrastructures.

Sujevan Ratnasingham (Barcode of Life Initiative, Canadian Centre for DNA Barcoding) presented the Barcode of Life Initiative. This initiative has developed a system, BOLD, which supports the collection, management, analysis, and use of DNA barcodes. It provides: (i) a repository for barcode records coupled with analytical tools; this repository supports the organization and upload of experimental data as it’s generated and serves as an online workbench for the DNA barcode community; (ii) a species identification tool that accepts DNA sequences from the barcode region and returns a taxonomic assignment to the species level when possible; (iii) a REST interface that allows access to public sequence and specimen data via Web developers in order to build tools and workflows that can be integrated with the BOLD framework.

The main objectives of the Barcode of Life Initiative are: (i) to support collaboration, i.e. share confidentially with others or openly with the community and (i) to support unification, i.e. connect molecular data with specimen data.

Data comes from 100000s researchers belonging to many micro-projects, each having their own IPs.

The Barcode of Life Initiative plans in the future to make extension in each of the BOLD functionality areas: (i) add functionality for annotating data; (ii) support analytical pipelines; (iii) integrate heterogeneous data; (iv) move from a dedicated infrastructure (340 CPUs, 100TB) to a cloud infrastructure (1000 CPUs, 100sTB).

Given the objective of the project and its plans for the future there is much potential for integration with D4Science.

Hendrik Segers (BioFresh) presented the BioFresh project. The objective of this EU co-funded project, which will start in February 2010, is to improve capacity to protect and manage freshwater biodiversity in the face of ongoing changes to global climate and socioeconomics. This objective will be achieved by: (i) Building a dedicated freshwater biodiversity Information Platform; (ii) Integrate tools and models to predict the responses of multiple stressors over scales; (iii) Increase awareness amongst scientists, policy makers and the public, and thereby improve conservation strategies and support the work of the EU and of international environmental agreements.

The expected challenges and prospects are: (i) deal with open access, open source, open standards in the area; (ii) integrating, operationalizing diverse distributed resources (data and tools) - for this would have to rely on initiatives such as D4Science; (iii) retain focus in the niche Freshwater biodiversity information, penetrate into the data sphere; (iv) mainstream data publishing; (v) address sustainability in order to support archiving and the maintenance of the portal and the network.

Marc Goovaerts (Hasselt University, IOC/IODE Expert) introduced a video presentation on Ocean Data Portal (ODP) prepared by Greg Reed and Peter Pissierssen from IOC-IODE. The IODE programme, which provides the institutional framework, has developed a network of 80 national nodes (NODCs). IODE is responsible for the quality control, archive and dissemination of ocean data, and it encourages free and open access to data and focuses on all ocean related data including chemical, physical, and biological. IODE's instrument is the Ocean Data Portal which facilitates and promotes the exchange and dissemination of marine data and services. Through ODP, the network of NODCs provide seamless access to their collections and inventories of marine data, and the full range of processes including data discovery (based on ISO19115), viewing (with data querying, access to remote data, tabular and map visualization), analysis (with composition of multiple GIS layers and joint analysis of data to obtain conclusions about the spatial variability of marine processes), and download (in NetCDF, ASCII, XML).

The key principle behind the IODE ODP is its interoperability with existing systems, including the WMO information system (WIS), SeaDataNet and IMOS. Standards are the key to interoperability and include common Metadata standard vocabulary, WMS and WFS protocols for maps visualization, and common file formats such as in NetCDF, ASCII, XML. A process has been set-up in order to develop, and adopt common standards related to Ocean data management and exchange.

Frank Bisby presented 4D4Life (Distributed Dynamic Diversity Databases for Life), the Catalogue of life Ecosystem of e-Services. The Catalogue of Life (CoL) aims at gathering all current accepted names and main synonyms (possibly all) for all species on Earth from viruses to whales, which represents about 3-4 million species described so far with 1.9 million considered as valid (Chapman, 2009). CoL is the result of a partnership between Species 2000 and the Integrated Taxonomic Information System (ITIS). The principle of the constitution of this global index is the aggregation of Global Species Databases (GSD) that are managed by various custodians around the world, and dedicated to entire taxonomic groups of organisms. The 2010 Annual Checklist contains contributions from 77 databases with information on 1,257,735 species and 98,075 infra-specific taxa, and also includes 886,882 synonyms and 343,586 common names (65% of coverage so far).

4D4Life e-infrastructure therefore intends to provide taxonomic backbone services to users worldwide, by facilitating structured information exchange, synthesises from a globally significant resource for science, and disseminating this with modern web services.

Sayed Choudhury (Johns Hopkins University, The Data Conservatory project, NSF DataNet project) outlined that the Infrastructure should add value added to science and work seamlessly by hiding complexity, and reliably. Regarding funding, he gave the view that donor agencies should not support such infrastructures, which instead should be supported by the user communities, and that funding should be sought across continental initiatives.

Discussion

The discussion was centered on how the presented initiatives could be placed in the e-Infrastructure landscape, and whether a global network of scientific information sources ("ecosystem") in which these initiatives are nodes, or node of nodes, could be built.

Johannes Keizer raised the importance of working towards the achievement of more satisfactory solutions to interoperability issues, and in particular, towards the establishment of data exchange standards.

Sujevan Ratnasingham highlighted that interoperability is a complex problem, new issues are emerging that pose new requirements on standards.

Sayed Choudhury stressed that we have to not only identify organizational solutions, but also exchange technological solutions. The organization of joint projects' workshops addressing technological open issues might contribute to achieve this objective.

Alhaji Jallow (fishery officer at the FAO regional office for West Africa, Ghana) asked for suggestions from on what D4Science can do to facilitate access to published work for developing countries. Marc Goovaerts referred to the Ocean Data Portal and OceansDocs initiatives, and Franck Bisby to the Catalogue of life, to explain that their infrastructure assist developing countries in publishing their science through data flows started from national levels and channelled through regional hubs, to global web-publishing facilities. D4Science can build on those models by further playing the role of broker and facilitating free access to literature. Andrea Manieri (from NKUA, Athens) added that D4Science can also offer sharing of software.

Donatella Castelli asked to the panellists if they could envisage synergies between D4Science and the projects they were representing. A number of potential collaborations were identified: Catalogue of Life might be made available to an infrastructure like D4Science; D4Science might mediate access to earth observation data for Barcoding; it might facilitate access to published work for developing countries; it might facilitate free access to literature; it might help in supporting national and regional hubs for Catalogue of life, etc.

26 November 2009

Day 2- Working session for D4Science members, partners and interested parties

The first part of the morning was dedicated to technical issues. Late morning and early afternoon was dedicated to user communities (half of them being regional bodies).

D4Science technical features and opportunities of the grid infrastructure for large scale data management

The early morning was devoted to a comprehensive but high-level presentation of the technical aspects of the infrastructure (Pasquale Pagano). Participants realized that while the infrastructure is firmly rooted in current initiatives (e.g., EGI), tools (e.g., gLite) and standards (especially metadata), its basic technical aspects are complex and require highly qualified staff to develop, install, and maintain it. But the maintenance of a VRE as long as the infrastructure is maintained by a third party is quite simple, even the creation of a new one as demonstrated the day before.

Through the succession of EC funded project from Diligent to D4Science-II, the various concepts (e.g., Infrastructure, Virtual Organization, Virtual Research Environment, Infrastructure ecosystem) and current or potential data flows were explained. It was demonstrated that the increase of the adaptation

level of an infrastructure decreases the cost when resources are put in common like in the GRID or cloud computing: all temporary extra-needs in terms of computing capacity and storage, which would increase the cost on one machine, are absorbed. Few technical questions were asked: one on if and which is the use of some GRID standards which D4Science takes into account; another one about cloud computing infrastructures provided by other companies or initiatives: the fact is that D4Science could use these resources if needed, and not see them as competitors.

Community fair - Discussion on D4Science technical features

Most of the questions from communities were about the sustainability, the possible cost to maintain a VRE in the infrastructure, and if/how these aspects are communicated. Examples of questions were: if a community is to join D4Science what happens at the end? are GDI (Geospatial Data Infrastructure) or SDI (Spatial Data Infrastructure) considered? Java based technologies (OGC, JRI) are supported but what about IBM, Oracle, VMWare? can developers from user communities access to inner parts of the software? can the infrastructure support workflows? are services cost effective, and how is cost-effectiveness measured?

Answers were the ones which Donatella Castelli synthesized for the afternoon session 'Long-term sustainability and maintenance'

After the break, Rainer Froese, Leader of the Fishbase Project (IFM-GEOMAR, Leibniz Institute of Marine Sciences) Chaired the session about fishery and marine biodiversity communities. Nine presentations were delivered under this session for an overview of user needs in the field of Fishery reports, Fishery statistics, Vessels monitoring, Biodiversity and Environmental monitoring, Fishery digital libraries. The presentations were delivered following a logical thematic sequence (see Annex 3) and the following presenters intervened:

- Mette Bertelsen: International Council for the Exploration of the Sea, ICES
- Bob Branton: Ocean Biogeographic Information System, OBIS, and Ocean Tracking Network, OTN
- Hamady Diop: Research and Information Systems Sub-Regional Commission of Fisheries, CSRP
- Ricardo Federizon: Northwest Atlantic Fisheries Organization,NAFO
- Jerome Guitton: Agrocampus
- Moi Khim Tan: WorldFish Center, WFC
- Kjartan Hoydal: North East Atlantic Fisheries Commission, NEAFC
- Peter Pissiersen: International Oceanographic Data and Information Exchange, IOC-IODE
- Dominique Greboval and Marc Taconet, FAO
- Pouchamarn Wangsanga: Southeast Asian Fisheries Development Center, SEAFDEC

Almost all of them require a GIS system, which comforts the idea that GIS functionalities are necessary in all VREs. This was confirmed by the written remarks gathered from the participants at the end of the conference. One activity that seems relevant to D4Science is the development of VMSs (Vessel Monitoring System) as they generate huge datasets, and thus require storage capacity and computing power. Coupled with data on MPAs (especially in high seas), or with environmental data, they could help to explain certain patterns both of fishes and fishers, and thus help to increase the good management in the framework of the Ecosystem Approach of Fisheries. Precision was given that data algorithms are not very sophisticated, mainly to estimate whether vessels are steaming or fishing.

It was noted that an international database for stock assessment is possible with D4Science. On the same note, a system like OBIS puts trawl surveys on line plus the science work behind, and D4Science could really add more to the current status. Also valuable pilot studies could be tried with Ocean Tracking Network.

The VMEs (Vulnerable Marine Ecosystem) could benefit from a D4Science VRE as they exploit remote sensing data, and heterogeneous sources of data. Predictive distribution maps, like those provided by AquaMaps, are a fundamental part of the tools used for the management of VMEs. Hence the development of a VME VRE would immediately benefit from the AquaMaps VRE (in addition, some of the targeted ecosystem are seamounts that SeaLifeBase has documented intensively in collaboration with the CoML/CENSEAM project).

Some of them would need the digital library functionalities (IOC/IODE, ReefBase, ...) and could benefit directly from the INSPIRE and DRIVER projects.

On a slight negative side, various communications from developing countries regional bodies indicated that after 10 years or more, only slow progress could be achieved to integrate new technological computing capacities. This negative appreciation can however be turned into a positive action to be undertaken under the lead of FAO: by promoting the tools developed under D4Science, many duplicate efforts (as obvious from listening to all these presentations) could be avoided in future, and the infrastructure could constitute a safe repository preventing loss of data.

One fundamental issue that remain is the difficult political access to the data managed by the regional bodies, which is independent from the availability of tools or not.

D4Science-II

Marc Taconet opened the last session on “how D4Science-II can respond to users expectations”.

With a reference to a compiled synthesis of needs as expressed during the morning session (see Annex 4) he explained how current D4Science II objectives and work plan can accommodate expectations of user communities, distinguishing between the communities which are formally and non-formally involved in the project.

Objective 1 “enhancing the D4Science e-Infrastructure technological and operational capabilities”, will tackle all facets of interoperability, ie at data level (schemas), upper communication protocols (REST, SOAP, WSRF, ...), policies (security access, quality, flows, ...), functionality logic (software, operation semantics). The goals are to implement interoperability solutions that preserve as much as possible the autonomy of the participating data e-Infrastructures, and to transform VRE mechanism into a commodity that can be programmatically consumed. These activities will take into account user communities standards and will involve their developers, with priority to project front line members (Genesi-DR, AquaMaps, Inspire, Driver, FAO - FIGIS, Geonetwork, DocRep). Project second line partners (NAFO, ICES, NEAFC (Stats, VMS), OBIS) will be supported as resources permit, while agreements to involve other infrastructures (Catalogue of life, Biofresh, Barcode of life, Ocean Data Portal) might be established in passive or proactive way.

Objective 2 “setting-up a number of VREs for scientific application scenarios in the domain of the D4Science e-Infrastructure” will aim at the development of services / data sources for scenarios formally identified in project document: existing VREs will be improved (AquaMaps, ICIS, FCPPS) and new VREs created (VMS, Genesi-DR, Inspire, Driver). The developed services will be available for use by any VRE. Non formally involved agencies can express their interest to create a VRE where available application and data resources will be activated, and the project might allocate some resources for the necessary training.

Objective 3 “identifying appropriate models for successful outreach and sustainability” intends to respond to project’s long-term desired outcome for a broadly populated ecosystem serving a growing number of scientific communities. Activities include the participation to standardization bodies, focused dissemination and training activities to promote the knowledge ecosystem vision and gather feedback, and a specific focus on the sustainability of the D4Science infrastructure and the entire ecosystem. An

appropriate plan for future sustainability will be defined by establishing synergies with other on-going European initiatives, and taking into account feedback from newly involved communities.

Donatella Castelli then presented her synthesis on the topic 'Long-term sustainability and maintenance':

Access to software code: a good point is that an open access approach is followed, thus the entire gcube software code developed by the project is available for free. In D4Science1, it's not possible to access to inner part of the software but efforts will be made in the context of D4Sciencell to open this set of facilities to other communities.

Support to workflows: currently one can set-up a VRE for creating a specific workflow, but there is no workflow construction interfaces. Process management facilities with graphical offering on top of them were available in Diligent and these will be upgraded and available in D4Sciencell.

Support to user communities, and possible current participation: in a way, support to user communities is enshrined in D4Science philosophy since although D4Science is a cloud, services are tailored for data management. D4Science-II has got some funds to integrate new communities during its lifespan, and proposal can be made (the BioFresh Consortium was proposed since the meeting to use this facility, discussion is on-going). At present, a new user community can exploit the project outcomes by creating VREs on an existing VO, creating a new VO and associated VREs, or deploying the gcube software on their own infrastructure. as demonstrated with ESA.

Costs: at this stage, costs are difficult to establish and depend on whether there will be new developments entailing the programming of new specific functionalities.

Sustainability issues: some communities are ready to elaborate a VRE (BioFresh, possibly CoL, ...), but their delegates explicitly do not want to spend time if the infrastructure does not have a long-term visibility. If the project ended with D4Science-II, any computer team could thus in essence acquire the code and deploy gCube nodes on their infrastructure under their responsibility. Discussions are on-going with EGEE regarding the foundational support which gcube maintenance requires. User Communities can also organize themselves to support an infrastructure which they would share.

To summarize, the adoption of the infrastructure by other communities, and its long-term sustainability look as a dynamic entanglement: there are no community willing to be involved if there is no long-term infrastructure; there is no long-term sustainability if there is no community using it. At some point, some communities have to take the risk to use it now, just making sure that they have a fall-back position allowing benefiting from their involvement in the development of D4Science.

Donatella Castelli gave a summary of her presentation to explain which ways could be (and could not be) explored. Frank Bisby suggested that there should be one instance of the infrastructure where students and "young computer scientists" could experiment themselves, providing a maintenance community, a reservoir of possible mediators between scientist using dedicated VRE and the infrastructure at technical level, and new tool creative force.

At the end of the session, sheets were distributed to participants for them to indicate their needs with respect to D4Science (see annex 5). Below is a summary of the essential and recurrent needs expressed by the participant:

- * Governance – business plan
- * GIS capabilities - interactive mapping tool
- * Opening existing capacities to students, eg use R-computing capacities, available tools / data
- * More analytical tools on-line (stock assessment)

- * Linking to / emulating resources with existing infrastructure equipped with web-services (OBIS- IOC – 4D4life)
- * Interest for building together next projects (ISTAM, 4D4life, Biofresh) for filling gaps
- * Best practices, standards, test beds for promoting sharing (VMS)
- * Accommodating data resources and related standards (WFC, SEAFDEC, CSRP)

These needs will be channelled on the D4Science II Advisory Board Meeting and the Outreach Strategy will involve the response of the needs of the participants.

At 16h30, Johannes Keizer concluded the meeting thanking warmly all the participants for their inputs to the meeting, on behalf of FAO and of the D4Science project management.

Annex 1: Agenda

| 25 November 2009 | |
|--|---|
| Day 1 - Fair for broad audience outreach | |
| 08:30 – 09:15 | <i>Registration</i> |
| 09:15 – 09:45 | <p>Opening remarks</p> <p><u>Anton Mangstl</u>, Director of the Knowledge Exchange and Capacity Building Division (Food and Agriculture Organization of the United Nations)</p> <p>Welcome and meeting objectives</p> <p><u>Johannes Keizer</u>, D4Science Outreach Manager (Food and Agriculture Organization of the United Nations)</p> |
| 09:45 – 10:30 | <p>Keynote speech on “<i>Digital Library Infrastructure Support for Virtual Collaboration in Science</i>”</p> <p><u>Peter Young</u>, Chief of the Asian Division (Library of Congress) - <i>Short biography</i></p> |
| 10:30 – 10:45 | <i>Break</i> |
| 10:45 – 11:30 | <p><i>Vision of the D4Science project, including D4Science-II</i></p> <p><u>Donatella Castelli</u>, D4Science scientific coordinator (National Research Council, ISTI-CNR)</p> |
| 11:30 – 12:15 | <p>Keynote speech on “<i>The Data Conservancy: A Digital Research and Curation Virtual Organization</i>”</p> <p><u>Sayed Choudhury</u>, Data Conservatory project, National Science Foundation DataNet project (Johns Hopkins University) - <i>Short biography</i></p> |
| 12:15 – 12:45 | <p>Presentation on “<i>Scientific Data e-Infrastructures in the European Capacities Programme</i>”</p> <p><u>Krystyna Marek</u>, Project Officer, GÉANT & e-Infrastructure Unit (European Commission)</p> |
| 12:45 – 13:45 | <i>Lunch</i> |
| 13:45 – 14:30 | <p>Keynote speech on “<i>Fishery science and policy-making: Connecting information and decision</i>”</p> <p><u>Serge Michel Garcia</u>, former director of the Fisheries Management Division (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations) - <i>Short biography</i></p> |

| | |
|----------------------|---|
| <p>14:30 – 15:00</p> | <p>Community fair – Introduction of the D4Science communities, their challenges and the synergies</p> <p>Chairperson: <u>Marc Taconet</u>, Senior Fisheries Information Officer (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations)</p> <p><u>Nicolas Bailly</u>, FishBase Project Manager (The WorldFish Center)</p> <p><u>Luigi Fusco</u>, Earth Observation Programme Directorate (European Space Agency)</p> <p><u>Richard Grainger</u>, Chief of the Fisheries and Aquaculture Information and Statistics Service, (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations)</p> |
| <p>15:00 – 16:15</p> | <p>Community fair – Showcasing “Building dynamic research environments using D4Science technology”</p> <p>Chairperson: <u>Marc Taconet</u>, Senior Fisheries Information Officer (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations)</p> <p>Setting up a Virtual Research Environment (VRE)</p> <p><u>Veronica Guidetti</u>, Earth Observation Programme Directorate (European Space Agency)</p> <p>Species distribution modelling using the AquaMaps VRE</p> <p><u>Kristin Kaschner</u>, FAO Project consultant (University of Freiburg, Faculty of Biology)</p> <p>Generating reports from distributed heterogeneous data sources</p> <p><u>Anton Ellenbroek</u>, Work Package leader for D4Science Communities VREs Definition, Validation and Exploitation (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations)</p> |
| <p>16:15 – 16:30</p> | <p><i>Break</i></p> |
| <p>16:30 – 18:00</p> | <p>Panel discussion “How to exploit projects’ synergies”</p> <p>Chairperson: <u>Carlos Morais-Pires</u>, Head of Sector Scientific Data Infrastructure (Information Society and Media Directorate-General, European Commission)</p> <p><u>Frank Bisby</u>, (Distributed Dynamic Diversity Databases for Life, 4D4Life)</p> <p><u>Sayeed Choudhury</u>, (Johns Hopkins University, The Data Conservatory project, NSF DataNet project)</p> <p><u>Marc Goovaerts</u>, (Hasselt University, IOC/IODE Expert) - <i>Video</i></p> |

| | |
|---------------|---|
| | <p>Sujeevan Ratnasingham, (Barcode of Life Initiative, Canadian Centre for DNA Barcoding)</p> <p>Hendrik Segers, (BioFresh)</p> |
| 18:00 – 18:15 | Closing remarks |
| 19:30 | <i>Dinner</i> |

| 26 November 2009 | |
|--|---|
| Day 2 - Working session for D4Science members, partners and interested parties | |
| 09:15 – 09:30 | Opening remarks |
| 09:30 – 10:30 | <p>D4Science technical features and opportunities of the grid infrastructure for large scale data management</p> <p>Pasquale Pagano, D4Science technical director (National Research Council, ISTI-CNR)</p> |
| 10:30 – 11:00 | Community fair - Discussion on D4Science technical features |
| 11:00 – 11:15 | <i>Break</i> |
| 11:15 – 12:55 | <p>Introduction of user needs: Fishery statistics, Vessels monitoring, Biodiversity, Aquatic science digital libraries, Environmental monitoring, Integrated reporting - Detailed information</p> <p>Chairperson: Rainer Froese, Leader of the Fishbase Project (IFM-GEOMAR, Leibniz Institute of Marine Sciences)</p> <p>Mette Bertelsen (International Council for the Exploration of the Sea, ICES) - Part 1 - Part 2</p> <p>Bob Branton (Ocean Biogeographic Information System, OBIS)</p> <p>Hamady Diop (Research and Information Systems Sub-Regional Commission of Fisheries, CSRP)</p> <p>Ricardo Federizon (Northwest Atlantic Fisheries Organization,NAFO)</p> <p>Marc Goovaerts, (Hasselt University, IOC/IODE Expert)</p> <p>Dominique Greboval (FAO) - Part 1 - Part 2</p> <p>Jerome Guitton (Agrocampus)</p> |

| | |
|---------------|--|
| | <p>Kjartan Hoydal (North East Atlantic Fisheries Commission, NEAFC)</p> <p>Moi Khim Tan (WorldFish Center, WFC)</p> <p>Pouchamarn Wangsanga (Southeast Asian Fisheries Development Center, SEAFDEC)</p> |
| 12:55 – 14:00 | Lunch |
| 14:00 – 15:25 | Continuation of the morning sessions (Introduction of user needs) |
| 15:25 – 16:10 | <p>D4Science-II</p> <p>Wrap-up of the previous sessions: How D4Science-II can take expectations into account</p> <p>Marc Taconet, Senior Fisheries Information Officer (Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations)</p> <p>Long-term sustainability and maintenance</p> <p>Donatella Castellj, D4Science scientific coordinator (National Research Council, ISTI-CNR)</p> |
| 16:10 – 16:30 | Closing remarks |

Annex 2: List of participants

Pedro Andrade, European Organization for Nuclear Research
Nicolas Bailly, International Center for Living Aquatic Resources Management
Claudio Baldassarre, Food and Agriculture Organization of the United Nations
Pedro Barros, Food and Agriculture Organization of the United Nations
Mette Bertelsen, International Council for the Exploration of the Sea
Frank Bisby, Distributed Dynamic Diversity Databases for Life
Maurizio Bonafede, European Space Agency
Bob Branton, Dalhousie University
Fabrice Brito, TERRADUE
Pilar Cabestany, Food and Agriculture Organization of the United Nations
Florida Estrella Cainglet, European Organization for Nuclear Research
Francesco Calderini, Food and Agriculture Organization of the United Nations
Leonardo Candela, Consiglio Nazionale delle Ricerche
Caterina Caracciolo, Food and Agriculture Organization of the United Nations
Fabio Carocci, Food and Agriculture Organization of the United Nations
Donatella Castelli, Consiglio Nazionale delle Ricerche
Sayed Choudhury, Johns Hopkins University
Valerio Crespi, Food and Agriculture Organization of the United Nations
Elena Di Paola, Food and Agriculture Organization of the United Nations
Beniamino Dimartino, Second University of Naples
Hamady Diop, Research and Information Systems Sub-Regional Commission of Fisheries
Andor Dirner, 4D SOFT
Carole Durussel, International Union for Conservation of Nature
Anton Ellenbroek, Food and Agriculture Organization of the United Nations
Tina Farmer, Food and Agriculture Organization of the United Nations
Ricardo Federizon, Northwest Atlantic Fisheries Organization
Riccardo Fortuna, Food and Agriculture Organization of the United Nations
Rainer Froese, Fishbase Information & Research Group Inc.
Luigi Fusco, European Space Agency
Fabrizio Gagliardi, Microsoft Corporation
Giulio Galiero, Engineering - Ingegneria Informatica – SpA
Serge Garcia, consultant keynote speaker
Riccardo Gargana, European Space Agency
Aureliano Gentile, Food and Agriculture Organization of the United Nations
Pedro Goncalves, TERRADUE
Marc Goovaerts, Hasselt University
Colin Graham, British Geological Survey
Richard Grainger, Food and Agriculture Organization of the United Nations

Dominique Greboval, Food and Agriculture Organization of the United Nations
Armand Gribling, Food and Agriculture Organization of the United Nations
Veronica Guidetti, European Space Agency
Jerome Guitton, Rennes AgroCampus
Graham Higley, Encyclopedia of Life
Kjartan Hoydal, North East Atlantic Fisheries Commission
Yannis Ioannidis, National and Kapodistrian University of Athens
Alhaji Jallow, Food and Agriculture Organization of the United Nations
Yves Jaques, Food and Agriculture Organization of the United Nations
Kris Jelinek, Food and Agriculture Organization of the United Nations
George Kakaletris, National and Kapodistrian University of Athens
Stefka Kaloyanova, Food and Agriculture Organization of the United Nations
Kristin Kaschner, Food and Agriculture Organization of the United Nations
Johannes Keizer, Food and Agriculture Organization of the United Nations
Manisha Laloo, GridTalk
Erwin Laure, Royal Technical University at Stockholm
Anton Mangstl, Food and Agriculture Organization of the United Nations
Andrea Manieri, Engineering - Ingegneria Informatica – SpA
Krystyna Marek, European Commission
Luca Mellano, CHELYS srl
Jessica Michel Assoumou, European Research Consortium for Informatics and Mathematics
Diego Milano, Universität Basel
Carlos Morais-Pires, European Commission
Ahsan Morshed, Food and Agriculture Organization of the United Nations
Steven Newhouse, European Organization for Nuclear Research
Alexander Okhanov, Permanent representation of the Russian Federation to FAO
Neasan O'Neill, GridTalk
Pasquale Pagano, Consiglio Nazionale delle Ricerche
Rocco Carmine Pellegrini, Agenzia Spaziale Italiana
Richard Pepe, Food and Agriculture Organization of the United Nations
Axel Poigné, Fraunhofer Institute for Intelligent Analysis and Information Systems
Kanchanna Ramasamy Balraj, Engineering - Ingegneria Informatica – SpA
Sujeevan Ratnasingham, Barcode of Life Initiative, Canadian Centre for DNA Barcoding
Kathleen Reyes, International Center for Living Aquatic Resources Management
Paolo Roccetti, Engineering - Ingegneria Informatica – SpA
Philippe Rohou, European Research Consortium for Informatics and Mathematics
Jessica Sanders, Food and Agriculture Organization of the United Nations
Hendrik Segers, BioFresh
Fabio Simeoni, University of Strathclyde

Marc Taconet, Food and Agriculture Organization of the United Nations

Moi Khim Tan, International Center for Living Aquatic Resources Management

Merete Tandstad, Food and Agriculture Organization of the United Nations

Erik Van Ingen, Food and Agriculture Organization of the United Nations

Virginie Viollier, Food and Agriculture Organization of the United Nations

Pouchamarn Wongsanga, Southeast Asian Fisheries Development Center

Peter Young, Library of Congress