DANUBIUS-PP

Preparatory Phase for the pan-European Research Infrastructure DANUBIUS–RI "The International Centre for advanced studies on river-sea systems"

e-IRG interaction plan and report - final

Deliverable 8.13

European Commission

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739562

Project Full title	Preparatory Phase for the pan-European Research Infrastructure DANUBIUS–RI "The International Centre for advanced studies on river-sea systems"
Project Acronym	DANUBIUS-PP
Grant Agreement No.	739562
Coordinator	Dr. Adrian Stanica
Project start date and duration	1st December 2016, 36 months
Project website	www.danubius-pp.eu

Deliverable No.	D8.13	**Deliverable Date**	M36
Work Package No.		WP8	
Work Package Title		ICT e-Infrastructure and digital data cloud storage	
Responsible		RoEduNet	
Authors & Institutes Acronyms		ROMANIA ***Agenția de Administrare a Rețelei Naționale de Informatică pentru Educație și Cercetare (RoEduNet)*** Octavian RUSU, Paul GASNER, Ciprian PINZARU, Valeriu VRACIU, Gheorghe DINU	
Status:		**Final (F)**	X
		Draft (D)	
		Revised draft (RV)	
		Public (PU)	X

Dissemination level:	**Restricted to other program participants (PP)**	
	Restricted to a group specified by the consortium (RE)	
	Confidential, only for members of the consortium (CO)	

Executive summary / abstract

What is the focus of this Deliverable?

This report presents the e-IRG recommendations for European research electronic infrastructures, the interaction of DANUBIUS-PP with e-IRG, and the paths followed by DANUBIUS-PP for design, planning, deployment, development and operating the planned DANUBIUS-RI e-Infrastructure as support to accomplished the scientific research objectives and provide associated e-services.

What is next in the process to deliver the DANUBIUS-PP results?

By valorising the results of the other Work Packages related to digital data, the type of used simulation and modelling software and their requirements in terms of digital processing power, connectivity speed and data access mechanisms and security, DANUBIUS-RI will take the necessary actions to operate and manage its own e-Infrastructure, integrated and harmonised with the other existing or future ESFRIs.

What are the deliverable contents?

The report shortly describes the main roles and objectives of e-IRG, other e-Infrastructures groups/organizations related to Information and Communications Technologies, focused on the key recommendations to consider for the implementation of DANUBIUS-RI e-Infrastructure. The interaction of DANUBIUS-PP and e-IRG, the manner of following the e-IRG recommendations during the work of planning the DANUBIUS-RI e-Infrastructure and the e-services, and the main proposed rules for data management plan from the perspective of e-IRG recommendations are presented in this document.

Contents

1 Introduction

The DANUBIUS-RI Preparatory Phase (DANUBIUS-PP) develops the structures and processes to ensure that DANUBIUS-RI can deliver an integrated understanding of the functioning of river-sea systems. The overall objective of this DANUBIUS-PP is to support the further development of DANUBIUS-RI, to get the consortium ready for implementation as a pan-European distributed research infrastructure.

The e-Infrastructure has a crucial role across the components of DANUBIUS-RI and also worldwide. One of the main objectives to be achieved in the Preparatory Phase is to establish the necessary computing, storage and communication infrastructure and the simulation tools for a distributed RI. The technical aspects of processing power and the storage are subject of other ongoing studies, as well as the connectivity between the Data Centre, the Hub, the Nodes and Supersites, Technology Transfer Office and users.

In the DANUBIUS-RI context, it is imperative to interact with e-IRG in order to identify and implement recommendations and best practices towards development of DANUBIUS-RI e-Infrastructure and to fit the needs of research community. A draft report has been submitted describing the e-IRG recommendations and the contact established between e-IRG and DANUBIUS-PP. Following e-IRG recommendations, DANUBIUS-PP proposed an e-Infrastructure and the associated services, mainly described in the deliverables resulted from WP7 and WP8. How the e-IRG recommendations have been followed? How DANUBIUS-RI will interact with e-IRG? The next chapters will describe the e-IRG principles, and their transposition into the DANUBIUS-PP work.

2 About e-IRG

e-IRG (e-Infrastructure Reflection Group) is defined as an European level strategic body for integration of European e-Infrastructure and related services into a common vision for the future open e-Infrastructure.

According to the e-IRG website mission and objectives of the body are (http://e-irg.eu/mission):

"Mission and Vision

e-IRG is a strategic body to facilitate integration in the area of European e-Infrastructures and connected services, within and between member states, at the European level and globally.

The mission of e-IRG is to support both coherent, innovative and strategic European e-Infrastructure policymaking and the development of convergent and sustainable e-Infrastructure services.

Objectives
Sustain e-IRG as the leading European advisory body on e-Infrastructures.

To reach this objective:

- *e-IRG will act as an external and neutral advisory body being counterpart of reference by:*
    - *producing strategic and policy reports, analyses and recommendations;*
    - *actively stimulating discussion with and between all stakeholder groups;*
    - *informing and influencing policy makers (regional, national, European and international) based on work carried out by e-IRG delegates;*
    - *obtaining a comprehensive understanding of the state of the art and progress in the field of e-Infrastructures;*
    - *promoting a user-driven and scientific excellence approach in e-Infrastructure policy making taking into account e-IRG Members' recommendations and constraints;*
    - *advancing the European landscape of e-Infrastructures taking into account national roadmaps;*
    - *proposing objectives taking into account the landscape;*
    - *proposing synergies and future cooperation between projects and/or initiatives;*
    - *monitoring the impact of e-IRG recommendations in the European research infrastructure landscape.*

- *e-IRG will continue its policy of independence and openness, providing opportunities for all stakeholders to debate their views, and offer perspectives, analyse and propose approaches and consolidate ways forward.*

Maintain and further develop e-IRG in its role as the facilitator and stimulator of European e-infrastructure collaboration

To reach this objective e-IRG will:

- *Facilitate the collaboration between all actors in the field of pan-European e-Infrastructures – networking, high-throughput and high-performance computing, and in particular data management and related processes and services.*
- *Stimulate work and strongly facilitate the international coordination of e-Infrastructure activities and the presentation of all different components to users as integrated e-Infrastructure services.*
- *Foster innovation in the data area, including stimulating the interaction between data producers (e.g. Research Infrastructures), e-Infrastructure providers and leading-edge users.*
- *Cooperate with ESFRI, ESFRI RI projects and other pan-European research infrastructures, as well as with the long tail of science in the effect possible."* [1]

There are other groups/organisations that provide recommendations to the European Commission (EC) in Research and Information and Communications Technology (ICT) areas [2]:

- Digital ERA forum was initiated by DG CONNECT (EC) for reporting about national developments.
- European Strategy Forum on Research Infrastructures (ESFRI) was initiated by member states and DG RTD to support a coherent strategy on research infrastructures in Europe facilitating the best development and usage of these infrastructures at the EU level.
- Horizon 2020 Advisory Group on Research Infrastructures including e-Infrastructures was initiated by the DG RTD and DG CONNECT to contribute to the implementation of Horizon 2020 the funding programme for research and innovation for the period 2014 – 2020.
- Research Infrastructures Programme Committee (or Management Committee - responsible for the Work Programmes) is initiated by DG RTD and DG CONNECT and it is responsible for next Work Programme and approving funding requests for RI and e-Infrastructures related projects.
- National Contact Points on Research Infrastructures (NCPs on RIs) was initiated be the member states being responsible for the dissemination of information about calls and support the Research Infrastructures Programme Committee.
- Organisation for Economic Co-operation and Development (OECD) - Global Science Forum (GSF) initiated by OECD provides action recommendations on major and high-priority science projects at international scale.

- EIROforum Information Technology initiated by EIROforum to coordinate the major IT issues facilitating cooperation on large projects in specific IT areas.
- The Carnegie group of G8+05 Science Advisers - Group of Senior Officials (GSO) on Global Research Infrastructures was initiated by G8+05 to find a common understanding in funding and management of large scale research infrastructures.
- Commission High Level Expert Group European Open Science Cloud (EOSC) initiated by EC as expert group to provide strategic advice to the Commission on the strategy for the European Open Science Cloud initiative.
- European Open Science Policy Platform initiated by EC to design and develop an Open Science Policy Agenda for Europe.

e-IRG provides guidance documents for research infrastructures as white papers. These documents offer recommendations, advices and best practices in various areas. New versions are issued regularly with more details and improvements for the next ESFRI Roadmap, an updated e-Infrastructure Commons concept.

3 DANUBIUS-PP and e-IRG

DANUBIUS-RI ICT infrastructure should be able to offer services to the ERIC members and to the public. The future DANUBIUS-RI should consider and benefit of the recommendations from e-IRG as published in the "Guide to e-Infrastructure Requirements for European Research Infrastructures"[3]. This document was issued, and it is used to help research infrastructures in the process of submitting proposals for the ESFRI Roadmap 2018.

DANUBIUS-RI as a European distributed system consisting of Hub, Supersites and Nodes as main data sources in different scientific domains from a big variety of scientific equipment will lead to a complex data storage system hosted by the Data Centre. For the hardware solutions, the most critical parameters will be the scalability and the reliability of the equipment considering the difficulties in estimating the needed capacities and the long-time preservation of the data required by the scientific community in the river-sea interaction studies. The hardware challenges will be addressed by the Data Centre construction procedures.

To comply with the e-IRG recommendation on ICT research infrastructure, a close collaboration between the DANUBIUS-PP team responsible for the ICT infrastructure and the e-IRG has been recognized in the early stages of the project preparations. The first evaluation of the necessary e-Infrastructure for DANUBIUS-RI brings into attention the idea that data storage system complexity will be an important challenge of the whole project. To reach the proposed goals in data management, common European principles should be followed in construction process of the DANUBIUS-RI e-Infrastructure. In DANUBIUS-PP a special task is dedicated to the interface with e-IRG with the goal to follow the European guidelines for e-Infrastructures and build a valuable data repository for the research community and for the public. DANUBIUS-RI goals and its ICT proposed infrastructure has been presented to e-IRG members in November 2017.

The quantity of data to be generated in DANUBIUS-RI is difficult to be estimated for a specific time frame as well as to estimate the growth rate. These difficulties are justified by the numbers of scientific disciplines involved in the research infrastructure and the fact that each one addressed the data related problems differently over the years. There are discipline-specific approaches related to data generation, storage, findings and usage but it is clear that the DANUBIUS-RI repository should offer data not only for discipline specific approach but for inter-disciplinary approach too. To reach that goal each recommendation of the e-IRG will be considered.

3.1 Interaction with eIRG

The interaction between DANUBIUS-PP and e-IRG should be reflected in the key documents for the DANUBIUS-RI ICT components. DANUBIUS-PP e-Infrastructure preparation activities was introduced at the e-IRG meeting on November 27th, 2017 with a presentation about expected e-services and DANUBIUS-RI data policy main ideas. The e-IRG delegates followed the presentation, appreciated the interest of DANUBIUS-RI to provide services in accordance with the e-IRG vision, but the responsibility for these services rests with DANUBIUS-RI. Thus, the collaboration between e-IRG and DANUBIUS must be reflected in the essential documents describing the electronic services and data policy by considering the recommendations of the e-IRG and not through a bilateral collaboration, as the e-IRG is not directly involved in the projects to realize the European research infrastructures.

The interaction between DANUBIUS-PP and e-IRG has been seen in a form of reporting mechanisms from DANUBIUS-PP to e-IRG. e-IRG is an advisory body on e-Infrastructure and it is not involved in any specific project. It does not receive reports about evolution of any specific project, it's mission being to support coherent, innovative and strategic European e-Infrastructure policymaking and the development of convergent and sustainable e-Infrastructure services. This mission is accomplished by publishing strategic and policy reports, analyses and recommendations to be considered by each e-Infrastructure.

Direct reporting to e-IRG was not the appropriate approach, the right approach would be to plan the ICT infrastructure and services in accordance with e-IRG recommendations. To achieve that goal the following actions were considered:

1. Study the documents and recommendation of e-IRG for e-Infrastructures for European research infrastructures

2. Get information regarding available and recommended services for e-Infrastructures on different aspects like access policies, interoperability with other e-Infrastructures, planning of hardware and software solutions etc.

3. Use the e-IRG proposed evaluation procedures of e-Infrastructure for DANUBIUS research infrastructure.

3.2 Key recommendation of e-IRG and DANUBIUS-RI approach

The key recommendations of the "Guide to e-Infrastructure Requirements for European Research Infrastructures"[3] considered for DANUBIUS-RI e-Infrastructure are succinctly exposed below, accompanied by the status of well-known ICT infrastructures and the main directions to follow by DANUBIUS consortium:

a. Check existing e-Infrastructures and related services for their use before defining the ICT infrastructure for your research infrastructure.
b. Check with existing research infrastructures how they realised their ICT infrastructure, learn from existing similar RI.
c. Contact existing e-Infrastructures at national and European level and ESFRI Projects and Landmarks.
d. Work with other Projects and Landmarks to encourage the development of the e-Infrastructure commons.
e. Consideration about interoperability of services and data
f. Access to ICT resources
g. Take into account the FAIR principles for your research data
h. Define a data management plan that fits your Research Infrastructure

3.2.1 Existing infrastructures

DANUBIUS-RI will have to offer e-services for partners and the public. These services must be available online through a portal. According to the requirements of the DANUBIUS-RI members, a series of main services were identified (Deliverable 5.12) as follows:

- Preservation and access to stored data for all users, according to DANUBIUS-RI policy. This is the most important service involving the most important resources and could be elaborated into specific services like: long time preservation of all data, providing access to open data to the general public, quality control for all data on ingest flows, thematic data access aggregation of data and metadata from different distributed repositories.
- Software and hardware resources for data processing including data analysis and data mining tools; the necessary interoperability between the data systems are considered to be achieved by using international standards for data and metadata.
- HPC and cloud services for the members of DANUBIUS-RI were requested and, taking into account the similar infrastructures, a kernel of hardware resources should be available, but it should be integrated into the European HPC and Cloud e-Infrastructures to be able to offer scalable services for each type by using the available resources shared by different e-Infrastructures.
- Services for education and knowledge exchange will include services for academic environment (e.q. e-Learning) and other services such as subscription services for data product, thematic data access and coverage access services.
- Other e-services involving software engineering and e-Infrastructure on demand should be provided including but not limited to: software engineering for web appliances, specific tools for process automation, custom database creation, data analysis tools for data assembly, metadata registry operation (upgrade, new schema etc.), generate data collections on demand, etc.

Taking into account the services listed above, services provided by other European e-infrastructures could and should be used. In fact, DANUBIUS-RI e-services will only be possible using other services provided by other European e-Infrastructures. The e-infrastructures and related services in data communications, computing facilities and data storage are considered in the process of defining the DANUBIUS-RI e-Infrastructure.

Network and data communication requirements for DANUBIUS-RI services will take advantage of the GÉANT and NRENs (National Research and Education Networks) infrastructures. Connectivity services will be used to interconnect the collecting sites, processing and storage sites. Requirements in terms of data security for DANUBIUS-RI are analysed and could lead to a solution involving some enhanced GÉANT services like Layer 3 Virtual Private Network (L3VPN), which add security by isolating specific traffic between the sites from the usual IP traffic. Accessing GÉANT connectivity services is possible using facilities of NRENs that offer connectivity and services for research and education community inside a country. A study to identify the connectivity to GÉANT for all key players of DANUBIUS has been conducted; the results were included in Deliverable 8.5. In terms of connectivity, according to Deliverable 8.5, all DANUBIUS-RI components are connected to the GEANT network, connectivity issues being identified between GÉANT and the Romanian NREN: the existing connectivity was 20 Gbps heavily used by the Romanian RO-LCG community (RO-LCG is a consortium that represents Romania within the Worldwide Large Hadron Collider Computing Grid collaboration) with negative impact for other institutions especially when they have a large number of jobs from CERN. To cope with this, connectivity upgrade has been considered by the Romanian NREN. At time being, the bandwidth of existing link has been doubled and a project to secure a second DF (dark fibre) link

to GEANT is underway to provide seamless upgradability of the bandwidth on request with minimal further investments.

All services provided by GÉANT and the NRENs are analysed in the process of preparing the e-Infrastructure of DANUBIUS-RI and enhance the future provided services. There are some services already considered to be used for accessing the resources in DANUBIUS. The most important ones, also used by other European research infrastructures, are connectivity and monitoring services, community services and cloud brokerage services. Connectivity services are the basic service set for all activities within distributed components of DANUBIUS and with the Data Centre; monitoring services were used to gather the necessary information about the types and the provider of the networking connectivity for each DANUBIUS partner. Community services shall, and will be, used in by DANUBIUS

a. eduGAIN to provide authentication services to access the data repository and all other e-services offered by the Data Centre as service provider within eduGAIN federations.
b. eduPKI shall be used to obtain PKI certificates for all e-services.
c. GÉANT Cloud brokerage services were considered but other cloud services offered by other European infrastructures could be more suitable to be used.

More details about all available services considered to be used by DANUBIUS-RI were identified and described in Deliverable 8.8.

Computing capabilities services at the European level are provided by the High-Performance Computing (HPC) Infrastructure and Distributed Computing e-Infrastructures. The HPC Infrastructure is represented by Pan-European High-Performance Computing infrastructure and services (PRACE) that gives to academic and industry users access to the supercomputing facility. Access to PRACE resources is based on peer review. The project of Distributed Computing e-infrastructures is called European Grid Infrastructure (EGI) which provides access to high-throughput computing resources based on membership in a community-supported research group such as ATLAS, ALICE, LHCb, and others.

DANUBIUS-RI should operate its own Data Centre according to the ESFRI application with two components: long term storage system for data repository and computing resources for the modelling node. The estimations of the storage capacities and computing needs for repository operation concluded to reasonable values. On the other hand, for the computing power requested by the modelling node, there are important variations. To solve this issue the most appropriate approach would be to have a relatively small computing power installed in the Data Centre (around 1000 CPU cores) connected with the European infrastructures to be able to benefit on their services when needed. All computing services offers are considered for DANUBIUS-RI to complement the computing resources from DANUBIUS-RI Data Centre.

Data storage services at the European level are provided by the European e-infrastructure of integrated data services and resources (EUDAT) infrastructure to store vital information in a secure environment. The EUDAT it is a collaboration project consisting from multiple nodes that provide services for: upload and retrieval, identification and description, movement, replication and data integrity.

The long-term data repository for DANYUBIUS-RI has been the main reason for building own Data Centre. In the scientific fields of interest of the research infrastructure, besides the accuracy of the measurements in order to obtain correct data, an important aspect is the continuity of the data obtained from the measurements for long term, decades or even more. Ensuring the integrity and availability of the measured data over a very long term was the main concern of the team who submitted the ESFRI application. Within WP5 and WP8 a new requirement regarding the data warehouse has been identified; the main repository should be interconnected with the existing repositories of the partners so that, in fact, a distributed repository should be realized to include existing time series data and, in the same time, keeping the ownership, privacy and access conditions to that existing data.

The European Open Science Cloud (EOSC) initiative is appreciated and considered by the DANUBIUS-PP partners; the results of the pilot project will be analysed in the context of planning the DANUBIUS-RI e-Infrastructure.

3.2.2 Similar infrastructures

The DANUBIUS-PP consortium has 29 members and some of them already use their own ICT infrastructure with computing and storage facilities and specific services for data storage and analysis in the river-sea interaction domain. DANUBIUS-RI will benefit of the experience and expertise of the partners in all areas including e-Infrastructure. The DANUBIUS-RI Data Centre and associated services will be tailored to take advantages of the existing partners' resources and expertise.

In order to check how other similar research infrastructures have built their own ICT infrastructure, we've studied 23 European research infrastructures operating within the field of environment and life science, by gathering public information on the three pillars: data network and communications, computing power, and data storage facilities. The most relevant distributed ICT infrastructures have been found on Deltares, The European Marine Biological Resource Centre (EMBRC-ERIC), Infrastructure for the study of tectonics and Earth surface dynamics (EPOS-ERIC), EURO-ARGO ERIC, In-service Aircraft for a Global Observing System (IAGOS), LIFEWATCH-ERIC, Svalbard Integrated Arctic Earth Observing System (SIOS), ELIXIR (Deliverable D5.12).

All mentioned research infrastructures have built their own ICT infrastructures, according to their needs, and provide services to their research communities, collaborators, and/or general public. The common point is the deployment over GÉANT network, as expected; on another hand, the amount of installed/used resources in terms of compute power and storage are spread over a large domain. Where were possible, the data centres of partners have put in common resources, dedicated or shared with other programs/projects, yet most of ESFRIs have built their own data centre and used the partners' resources as extensions. Also, the resources needs are increasing in time and each e-Infrastructure owner has estimated the necessary for the next years.

The requests from DANUBIUS community and the analysis of the existing ICT infrastructures of similar organisations helped to lay down electronic services to be provided by DANUBIUS-RI, as described in Deliverable 5.12, and consistently estimate the necessary compute power, storage system and data network.

In a few figures, the main identified capacities of computing power and storage are presented in the table bellow.

ESFRI Data Centre	CPU & RAM	Storage
Deltares	1050 cores	1.4 PB
EMBRC	1864 CPU, 2 TB RAM	800 TB
EURO-ARGO	>11000 cores	500 TB + 550 TB
IAGOS	>1900 cores	>6 PB
LIFEWATCH	39 nodes, 192 GB RAM per node	600 TB
ELIXIR	>10000 cores, >100 TB RAM	>5 PB
	>49000 cores, 150 TB RAM, 32 GPU nodes	n/a
	16048 CPU cores, 92 TB RAM	3 PB
	10000 cores	2 PB
	7624 cores, 37 TB RAM	3.7 PB
	>6000 cores	6 PB

3.2.3 Services from existing infrastructures

The e-Infrastructure of DANUBIUS-RI will be supported by the Romanian NREN to provide high speed connectivity for the Data Centre and the Hub. Providing connectivity (high speed link and high reliability) for the Hub is a challenge considering the location of the Hub in the Danube Delta and various options are considered. Close collaboration with the Romanian NREN provides a better understanding of the services in the data communications field and in the range and types of services provided by the European network GÉANT through the NRENs. Discussions have been already started with GÉANT representatives to prepare the future DANUBIUS network and services for the members, for the research community and for the public.

As described above in paragraph 3.2.1, the data network is available for DANUBIUS-RI across the pan-European network GÉANT and the European NRENs. Also, some storage services are available at partners' facilities and will be integrated as distributed storage services. The existing storage capacities are not sufficiently by far and the amount and the heterogeneity of data kind requires dedicated storage services, starting at data gathering stage and going to next steps like sorting, checking, preservation, etc. up to providing data to the community.

3.2.4 Collaborate in other projects to encourage the development of the e-Infrastructure commons

DANUBIUS-RI will be distributed around Europe with its Hub, Supersites, Nodes, Data Centre and Technology Transfer Office and will have to develop common strategies for data management, data storage and access in different research fields. Special attention is paid to the study of the storage of scientific data, their format and associated metadata to ensure interoperability with other data storage systems in the scientific area of interest. Methods and procedures are also being explored to ensure the availability of data for a very long time given the specificities of the scientific activity in the field of the project.

An important opportunity for the DANUBIUS community was the inclusion in the ENVRI-FAIR (ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research) project of several members of the DANUBIUS-PP, those who have attributions in the field of data processing, storage and access. Considering the goal of the project, all participating RIs should built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC and the status of DANUBIUS-RI, the advantages of being involved in this project are very important: the DANUBIUS services can be designed and implemented according to FAIR principles from the beginning. ENVRI-FAIR is considering the following ways to reach its main goal:

- Common policies and standards, aligned with European policies, for data;
- Implement transparent and auditable data services compliant with FAIR principles;
- Define the catalogue of services for each RIs involved;
- Thematic data services and tools to be exposed under EOSC catalogue of services via EOSC-hub.

This is a great opportunity for DANUBIUS to work to an ICT synergy with other Projects and Landmarks to encourage the development of the e-Infrastructure commons.

3.2.5 Interoperability of services and data.

Following the recommendation in the Guide to e-Infrastructure requirements for European Research Infrastructures [3] is the first step to provide good interoperability of data and services with other research infrastructures at the European level. Considering the types, sources of data and different data measurement systems, as well as the amount of data to be acquired, stored and managed by DANUBIUS-RI providing interoperability of data is a long-term job that is susceptible to changes and adjustments according to the evolution of standards in scientific interest and industry standards fields. In the service area DANUBIUS-RI

will use services already developed and used in other research infrastructure as well as good practices for storage, manage and access the scientific data.

3.2.6 Access to ICT resources.

The most important ICT resources (computing and data storage equipment) will be concentrated in the Data Centre. The stored data will be obtained from the measurements carried out in various DANUBIUS-RI research projects as well as historical data. Four types of data are envisaged: observational, experimental, and computational (simulations/modelling results), and derivative (data obtained by applying transformations - aggregation - of existing data). The Data Centre and the associated data management software will take into consideration the flow of the data on the input side to provide quality assurance and quality control of the stored information and, also, security mechanism for the data transfer procedures. Access policies and methods to the stored data will be elaborated according to open access policies. Availability and reliability of the ICT resources will be considered as key requirements in the process of building the Data Centre by planning redundant storage systems and network links. Distributed storage services will be built by collaborating with project members able to offer these types of ICT services by sharing the storage space, providing backup mechanisms for disaster recovery.

3.2.7 FAIR principles for research data

FAIR stands for Findable, Accessible, Interoperable and Re-usable. The data repository of DANUBIUS-RI will provide access to observational, computational, experimental and aggregated data. The variety of the data types put a big challenge on the data management procedures to fulfil the **findable** principle the work is underway to identify the standards in the field in order to use the most appropriate sets of metadata for persistency, data description and data identification. DANUBIUS-RI data will be **accessible** using common internet protocols able to provide security mechanisms, authentication and authorisation for transactions between the client and the server. A set of open protocols are considered not only for accessing the data but for storing and identifying the reliable data sources as well. DANUBIUS-RI involves multiple research domains and data will be gathered from different sources and equipment types. **Interoperability** of the stored data is a key element of DANUBIUS-RI: the whole community should take advantage of the data originating in different research fields. Policies for data storage and associated metadata will align to the open standards and will use a common vocabulary for knowledge representations. Proprietary data from different equipment on filed should be translated into a common representation to provide accessibility and interoperability, each of these procedures will be developed case by case. **Re-usability** of the data will be guaranteed by using well known standards for the representation of the data and the associated metadata and by including the clear usage licence. These principles will be used also for historical data, where available, bringing historical valuable information in a common format to the community.

These principles are considered in the preparatory phase of DANUBIUS-RI, possible solution being analysed for each one. The open standards and/or well-known standards for data will be identified in WP7.

The participation of DANUBIUS in the ENVRI-FAIR project is a very good opportunity to benefit, before building the data repository and associated services, from the experience of other RIs in FAIR principles implementation.

3.2.8 Data management plan

Data management plan and data management policy for DANUBIUS-RI will follow the FAIR principle and will focus on long-time preservation and high availability of stored information. DANUBIUS Data Policy has been drafted in WP 7 taking into account the European legal framework related to environmental data, information and databases. DANUBIUS Data policy aims to open exchange of data, metadata and elaborated data products being made available with minimum time delay and at no-cost. The most important regulatory documents used in drafting the Data Policy were: INSPIRE Directive, Database Directive, Software Directive, COMMISSION RECOMMENDATION of 17.7.2012 on access to and preservation of scientific information C(2012) 4890 and Aarhus Convention.

Following relevant EU legal framework documents, such as INSPIRE Directive (sharing of the spatial information among public sector organizations and access to the spatial data), Database Directive (protection of the databases), Software Directive (protection for computer programs), Commission Recommendation of 17.7.2012 on access to and preservation of scientific information C(2012) 4890 or Aarhus Convention (access to environmental data), proposed data policy governs and moderates the desired openness of data, metadata and elaborated data products of DANUBIUS-RI.

- ***Users Access to DANUBIUS RI Data***

DANUBIUS-RI requires collection, processing and availability of high quality DANUBIUS data including DANUBIUS metadata. DANUBIUS-RI data users may access DANUBIUS-RI data under the conditions set forth by the data access policy and DANUBIUS-RI future Terms of Access, including possible payment of fees.

DANUBIUS-RI data, data products and services are governed by Open Access principles, defined by the Directorate-General for Research and Innovation issued document “European Charter for Access to Research Infrastructures”. Proposed data policy’s components conform to mentioned document and address its various aspects:

- non-discrimination in granting access of users from academia, business, industry and public services; procedures to approve/accept restrictions and embargo conditions will be handled by a dedicated DANUBIUS ERIC committee;
- adhere to the standard codes of conduct and ethical behaviour in scientific research and to research integrity;
- users should acknowledge the contribution of the DANUBIUS RI in any output (publication, patent, data, etc.) deriving from research conducted within its realms;
- research data are appropriately maintained, archived for a reasonable period, and available for review and re-use;

- comply with national and international law and agreements, particularly, but not only, in areas such as intellectual property rights and the protection of privacy, ethical considerations as well as safety, security and public order regulations;
- DANUBIUS RI will provide its users with instructions for the effective and efficient access to a research infrastructure.

- ***Data Processing and Archiving***

DANUBIUS-RI will act as curator for data, raw data, meta-data, processed data and results. DANUBIUS-RI reserves the right to delegate this task to local scientific partners, as long as an agreement can be found with them that does not alter the DANUBIUS-RI data policy.

Data management and archiving will be the responsibility of the Data Centre and the DANUBIUS ERIC will make sure that the DANUBIUS data is available to users. All DANUBIUS data will be archived in the DANUBIUS Supersites and the DANUBIUS Data Centre. DANUBIUS data will be made available through the Data Centre as soon as possible after further Quality Assurance/Quality Control (QA/QC) procedures have been applied. DANUBIUS-ERIC and the DANUBIUS Data Centre reserve the right to remove any material placed in the Data Centre servers should the said material be found to be erroneous or contain malicious content or if the external data providers have not respected the DANUBIUS data policy or author's rights.

- ***DANUBIUS RI Data Attribution and Citation***

DANUBIUS data users must acknowledge the persons and organizations, which have originally generated the DANUBIUS data or processed the DANUBIUS data. An identifier with the information of the data providers/authors and such will accompany every DANUBIUS dataset. In DANUBIUS-RI access policy, the acceptance of the "Terms of Access" is mandatory for all researchers, to credit DANUBIUS-RI for access: DANUBIUS-RI is therefore acknowledged in the paper as access provider whenever DANUBIUS-RI provided access to the authors of the paper. A tracking solution for use of the DANUBIUS data in research papers should be made available at DANUBIUS Data Centre.

- ***Services***

DANUBIUS-RI will provide data services to users. Subscribing to the principles enounced in the European Charter for Access to Research Infrastructures, DANUBIUS-RI aims at minimizing as much as is possible fees demanded for data access from research and student users. However, and within the limits of what is authorized by European and local regulations, DANUBIUS-RI reserves the right to charge fees to users.

In accordance with European Framework for State Aid for Research and Development and Innovation, private sector and public authority users may be required to pay fees compensating the full costs of offering the said services, plus a reasonable margin.

- ***Privacy***

DANUBIUS-RI will take into account the GDPR (Regulation EU 2016/679) on the protection of natural persons with regard to the processing of personal data and on the movement of such data. Statistics based on data profile will be reported in activity reports in an anonymized and aggregated form, in and out of the context of the DANUBIUS-RI KPIs (Key Performance Indicators)

- ***Intellectual Property Rights (IPR)***

DANUBIUS-ERIC keeps the property of the DANUBIUS data and any related tools and all rights related thereto. Any data, tools, models or databases created at the Data Centre will belong to the DANUBIUS-ERIC. DANUBIUS ERIC and all entities generating DANUBIUS Data shall acknowledge that, subject to any relevant Intellectual Property Right, they seek to promote the non-exclusive transfer of available information technology on mutually agreed terms to research institutions, especially in conjunction with training and capacity development programs.

- ***Licensing***

Except where otherwise agreed, DANUBIUS data is licensed under a Creative Commons Attribution 4.0 International License. For an effective rights management, the DANUBIUS ERIC shall establish and maintain a DANUBIUS data user license. DANUBIUS ERIC is aiming to grant one common DANUBIUS data user license for all the levels of DANUBIUS data.

To ensure the widest dissemination and publicity for DANUBIUS managed data, data products, software and available services, it is essential that metadata are easily and freely accessible at any time, with as few restrictions as possible.

- ***Data Policy Management***

The responsibility of the implementation and monitoring of the DANUBIUS Data Policy is that of the service providers.

Failures regarding the use or the implementation of the DANUBIUS Data Policy shall be reported to the DANUBIUS HUB, which will inform an ad hoc committee, whose statutes, missions and composition will be defined by the DANUBIUS-ERIC General Assembly. That committee will then take a decision accordingly.

- ***Liability***

DANUBIUS-ERIC users register - and in doing so - agree to relieve DANUBIUS-ERIC of any liability for any use of the DANUBIUS-ERIC data, data products, software and available services. DANUBIUS-ERIC is not liable for any misuse of data, data products, software and available services or associated metadata

- ***Policy Review***

Data-policy document is subject to revision according to changes in the law, the needs and strategy changes of DANUBIUS. The DANUBIUS-ERIC General Assembly will approve DANUBIUS Data Policy reviews.

4 Conclusions

DANUBIUS-RI e-Infrastructure has to offer services to the consortium members and to the public, and, as future research infrastructure has to consider the recommendations of e-IRG as published in White Papers to integrate into e-Infrastructure Commons guidelines of the next e-IRG Roadmap.

There are multiple faces of DANUBIUS-RI ICT infrastructure to consider in regards of planning, deployment and management. The e-IRG recommendations are the sum of experience and expertise provided by the EU States and will tailor the actions necessary to identify the needs of DANUBIUS-RI e-Infrastructure and the best solutions to achieve the goals of the research infrastructure and research community in a coherent manner.

Some of the DANUBIUS-RI ICT infrastructure components need can be accomplished by already existing e-Infrastructure (e.g. data communication via NRENs and GÉANT), other ones will be built up. The e-IRG Roadmap and recommendations helps to identify the current e-Infrastructure systems and with the common vision for higher degree of harmonisation and integration at European level.

A great opportunity for DANUBIUS community in order to harmonize the development of ICT infrastructure with other projects and landmarks is the participation of several members of DANUBIUS-PP in the ENVRI-FAIR (ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research) project. All participating RIs aim to design and implement the data services according to FAIR principles to improve the efficiency and productivity or research activities and support the innovation.

In this context, as described in this report, DANUBIUS-RI e-Infrastructure is designed, will be deployed, operated, managed upon the recommendation from e-IRG and will be integrated into the future e-IRG Roadmaps.

References

[1] e-IRG Mission. Retrieved from http://e-irg.eu/mission.

[2] E-IRG related policy groups. Retrieved from http://e-irg.eu/related-policy-groups

[3] Guide to e-Infrastructure requirements for European Research Infrastructures, ISBN 978-90-823661-5-0, 2017-03-01, http://e-irg.eu/catalogue/eirg-1004

DANUBIUS-PP

Preparatory Phase for the pan-European Research Infrastructure DANUBIUS–RI "The International Centre for advanced studies on river-sea systems"

European Commission

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739562