**Work package 4**

**Deliverable:** 4.7 Report proposing new programme's data strategy and mechanisms for its implementation

**Lead organization:** Flanders Marine Institute (VLIZ), Task Leader: Jonas Lescroart

**Submitted by:** Andris Andrusaitis, Coordinator of BANOS CSA
andris.andrusaitis@bonuseeig.fi

**Submission date:** 30 November 2020

*The main outcomes*

- The report identifies the most relevant policy documents and projects concerning the management and sharing of research data in a European context, and its outcomes serve as a guidelines and recommendations for BANOS' strategy on open access and open date.
- In general, the data strategy should include a proactive role towards changing research cultures and treating data as a first-class citizen (changing research funding from being mainly citation driven to being open data driven). Soft-type action should be taken, such as promoting good practice, establishing better institutional reward systems for contributing researchers, etc.
- The data strategy should require that all data generated through potential BANOS projects fulfills the minimal requirements for FAIR and open data set by the Open Data Directive and Horizon Europe. The consortium should thereby try its best to provide the necessary support to bring all of its constituents up to that level.
- Similar to the Directive, BANOS' data strategy should handle the credo: "as open as possible, as closed as necessary", meaning that data should be open by default, yet allowing for data to be closed when there are valid reasons relating to publisher requirements, moratorium periods, privacy and security issues, etc. In light of this, the use of standard open licenses is recommended to indicate the status of the data.
- All Research data, and the associated metadata, should be deposited in an appropriate standard repository. It is recommended to provide a flexible list of recommended repositories from which researchers can choose. An impetus for this list is given in this report.
- Data should be accompanied by rich metadata. It is recommended to use appropriate and internationally accepted (machine-actionable) standards, protocols and dictionaries to document and exchange metadata.
- Mandatory Data Management Plans (DMPs) are put forward as the main implementation instrument for the BANOS programme. The consortium should agree on the template that is to be provided to the partners in order to aid them in setting up their DMP's.

Authors: Jonas Lescroart[1], Laurian Van Maldeghem[1], Gilbert Maudire[2], Katrina Exter[1], Floor ten Hoopen[3]

Affiliations: [1]Flanders Marine Institute (VLIZ), Belgium; [2]French Research Institute for Exploitation of the Sea (Ifremer), France, [3]Swedish Agency for Marine and Water Management (SwAM), Sweden

Contact: katrina.exter@vliz.be

## Description of task: (i.e. as in the Description of Work)

One of the major challenges of data driven science is to ensure the accessibility, interoperability and reusability of scientific data. Open data is essential to implement an open research policy - it has significant potential for reuse in new products and services, helps discovering new and innovative solutions to address societal challenges, fosters participation of end-users and increases transparency of science. Within Europe initiatives and legislations to promote open data and an open data policy are becoming mature, whether they focus on environmental data (INSPIRE), public sector information (PSI) or marine data (EMODnet). Each of them develop a specific set of guidelines, standards and protocols to improve accessibility and interoperability but serve the main overall objective i.e. the reusability of the data. As new technologies and sensors to study the marine environment develop, data becomes more complex and bigger. Processing of these data is more demanding and initiatives to pool infrastructures and resources to develop a common architecture, functionalities and collaborative tools are initiated under the European Open Science Cloud (EOSC). These developments will facilitate and improve the analysis of novel big data, but the principles to implement Findable, Accessible, Interoperable and Reusable (FAIR) data will remain crucial.

Based on the ongoing BONUS Art. 185 data policy and the lessons learned while creating the BONUS meta-data collection, following principles of dataset citations so researchers receive the appropriate credit and recognition for their work and using Digital Object Identifiers so data stays persistently accessible, we will develop a long term strategy related to Open Data policy and using open data infrastructures to reach these Open Data policies, framed within the overall EU framework programme's policies on Open Science.

To provide an information input for this task, the views and suggestions from all project participants will be collected and the Open Data policies, tools and infrastructures existing in Europe, in particular, around the Baltic and the North Seas will be collected. The output of this task will constitute a strategy to make the joint Baltic Sea and North Sea research and innovation programme data Findable, Accessible, Interoperable and Reusable (FAIR) using existing Open data policies, infrastructures and tools. Data strategies and instruments will be developed concurrently with development of the SRIA, so that they can be reflected in programme's main programmatic document. The final output of this task will be a report proposing new programme's data strategy and mechanisms for its implementation containing provisions to make the joint Baltic Sea and North Sea research and innovation programme data Findable, Accessible, Interoperable and Reusable (FAIR) using existing open data policies, infrastructures and tools. This report will be finalised by M24 and presented for SC approval in M25.

# BALTIC AND NORTH SEA COORDINATION AND SUPPORT ACTION (BANOS CSA)

## BANOS CSA D4.7

## Report proposing new programme's data strategy and mechanisms for its implementation

Delivery date: 30 November 2020

Jonas Lescroart[1], Laurian Van Maldeghem[1], Gilbert Maudire[2], Katrina Exter[1], Floor ten Hoopen[3]

[1]Flanders Marine Institute (VLIZ), Belgium; [2]French Research Institute for Exploitation of the Sea (Ifremer), France, [3]Swedish Agency for Marine and Water Management (SwAM), Sweden

Contact: katrina.exter@vliz.be

## Contents

## Executive summary

The Baltic and North Sea Coordination and Support Action (BANOS CSA) is preparing a framework for the joint Baltic Sea and North Sea Research and Innovation Programme (BANOS). Open data is an essential part of contemporary research and innovation, and as such, a competent strategy for open data will enlarge the societal impact of BANOS. The main objective of the task presented in this report (D4.7) is to facilitate the adoption of an open data policy for BANOS and to provide guidance towards the implementation of such a policy.

The report opens by briefly situating the task in its broader context, stating its purpose within BANOS CSA and pointing out the interrelation with other tasks. To ensure that everyone reading the report is on the same page, a number of general concepts and definitions are introduced, including open data, the FAIR Data Principles, and the concept of Data Management Plans (DMP).

To accurately assess the current best practices concerning the management and sharing of research data, the European policy framework for open data is examined through identification of the most relevant policy documents and open data projects. In this context, the Open Data Directive published by the EU in 2019 constitutes a key document. The strategies for open data of the countries represented in the BANOS CSA consortium were surveyed by means of a questionnaire, which was distributed amongst the consortium members and other national funding agencies. A good majority of organisations surveyed declares to have a research data policy in place or in the works, and explicitly supports the use of a DMP. Further, most institutions are already compliant with the Open Data Directive or are taking steps in that direction, and the consortium members are largely in favour of implementing the Directive's requirements for open and FAIR data in the BANOS programme. However, a number of reservations on the practical implementation of the Directive in the BANOS programme will need to be addressed during ensuing discussions with the consortium.

In anticipation of the use of DMPs as a central instrument in BANOS, an overview of essential and optional components for an effective DMP is presented in this report, with additional recommendations where appropriate. Once the BANOS SRIA is fully defined, this resource will enable the creation of DMP templates, which can be distributed to the programme's beneficiaries for implementation.

Lastly, the report offers a set of general recommendations regarding the adoption and implementation of an open data policy. These recommendations are drawn from the literature, such as project reports, and supplemented with suggestions obtained during an interview with EMODnet, a principal repository for marine data in Europe. In addition, a budding list of recommended repositories is provided, for the submission of project-generated data.

# Introduction

## Context

The Baltic and North Sea Coordination and Support Action (BANOS CSA) is preparing a framework for the joint Baltic Sea and North Sea Research and Innovation Programme (BANOS), to be ready in 2021. In line with the Europe 2020 strategy for a smart, sustainable and inclusive economy in Europe (European Commission 2010), knowledge and innovation are considered to be of central importance to the generation of blue growth. Broad access to the data that underpin research and innovation enhances multiple aspects of the research and innovation process. It helps to build on previous achievements, improving the quality of new results. It encourages collaboration and the avoidance of duplication, resulting in greater efficiency. It can speed up innovation by enabling faster uptake by the market, which translates to faster growth. Lastly, access to data makes the scientific process more transparent, potentially boosting the involvement of citizens and society (DG RTD 2017). Therefore, improving the openness and accessibility of research and innovation data will be an essential cross-cutting component to the future programme.

The purpose of BANOS CSA task 4.6 Developing strategies and instruments supporting open data is to formulate a general strategy towards open data that will serve as a guideline for the implementation of open data in the BANOS programme. To help develop a sound strategy, it is crucial to be aware of current best practices among the BANOS consortium members. To this end, a questionnaire was distributed to all currently involved organisations, and to potentially interested future funders and strategic partners. In addition, relevant European projects and policy documents are listed to provide an overview of best practices at the European level. The information gathered from the questionnaire will serve as a baseline for further discussions on the data management strategy among the consortium members. To ensure a proper implementation of the devised strategy, it is important to involve the beneficiaries of the programme from the beginning. Data Management Plans (DMPs) are a key tool in achieving this involvement, and this report outlines and reviews the components that are necessary to build a DMP template fitted to the BANOS programme.

Task 4.6 is part of BANOS CSA work package 4: Specific measures reinforcing future programme's lasting impact. The open data strategy is primarily a means to increase the impact of the future programme. If data generated throughout the programme are openly accessibly and up to the necessary standards for efficient re-use, more people can make use of the data, thus enlarging the reach and impact of the programme as a whole and increasing the value gained for the money spent.

Two other tasks within BANOS CSA WP4 share the theme of 'open' with task 4.6, with the same goal. Specifically, task 4.4 Strategies supporting firm establishing of 'open science' deals with open access to scientific literature, and task 4.5 Developing strategies and instruments stimulating innovation diffusion and 'open innovation' deals with knowledge sharing in the innovation process. The contents of these tasks were carefully attuned to one another in order to produce a trio of synergetic reports covering strategies for the sharing of research and innovation outcomes. The questionnaire mentioned earlier was designed to collect information for both task 4.4 and task 4.6.

The BANOS open data strategy does not start without a precedent. It builds on the BONUS Data Policy Statement (2012) and Metadatabase (2017), adding new insights from recent literature (e.g. Redkina 2019; Mons 2020) and seeking coherence with new policies (mainly the Open Data Directive, cf. infra).

## Open data: general concepts and definitions

Research data are defined as digital files other than scientific publications, which are collected or produced in the course of scientific research activities that are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results (Directive 2019/1024/EU). Examples of research data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

Open access is understood as the practice of providing online access to research outputs free of charge for the end user and without restrictions on use and re-use beyond the possibility to require acknowledgement of authorship (Directive 2019/1024/EU). Open access is a core strategy of the European Commission to improve knowledge transfer. It is an important feature of Horizon 2020 and will continue to be so in Horizon Europe.

Open data are digital data whose access and use are granted to all. They may be of public or private origin, produced in particular by a community, a public service, a research organisation, a citizen group or a private company. They are disseminated in a structured manner according to a method and an open license guaranteeing their free access and re-use by all, without technical, legal or financial restrictions (DG RTD 2017). Open data as a concept is part of a movement, cf. the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003), that recognises public information as a common good (e.g. as defined by Elinor Ostrom, Nobel Price of Economy) whose dissemination is in the general interest. The broader movement of open science and its implementation in the future BANOS programme is further discussed in BANOS CSA task 4.4 Strategies supporting firm establishing of open science. Open data now evolves toward public policies of publishing freely-accessible and freely-exploitable data. For example, in Europe the Open Data Directive (Directive 2019/1024/EU) provides a common legal framework for the re-use of public sector information, including publicly funded research data (see section "The European policy framework for open data").

The FAIR Data Principles are a set of guiding principles to make data findable, accessible, interoperable and reusable (Wilkinson et al. 2016). These principles provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital ecosystem. The concepts of open data and FAIR data are different, but complementary: open access is not a requirement for data to be FAIR, and open data do not necessarily adhere to the FAIR principles. The FAIR acronym stands for:

- Findable: metadata and data should be easy to find for both humans and computers;
- Accessible: once the user finds the required data, the user needs to know how they can be accessed: access conditions (data license, possibly authentication and authorisation) and access services;
- Interoperable: the data may be easily integrated with other data and/or may interoperate with applications or workflows for analysis, storage, and processing;
- Reusable: the metadata and data should be well described (e.g. by using commonly adopted vocabularies and formats) to optimise the re-use of data, which is the main objective of FAIR principles.

A Data Management Plan, or DMP, is a fundamental document that describes the research data life cycle. A DMP must demonstrate the application of the FAIR principles for the data that are collected, managed and processed. It should provide information about

- data types that will be managed and used;
- how they will be preserved after the end of the project;
- how they will be shared following the Open Data Directive and its principle "as open as possible, as closed as necessary".

Since 2017, a DMP is required by default for all European projects funded through the European Framework Programmes, and a template in free-form text is available (European Commission 2016). In addition, several tools exist that help structuring and writing DMPs, such as DMPonline (Digital Curation Centre 2010). In the section "Ingredients for a successful Data Management Plan" of this report, a list of potential DMP content is provided to guide the creation of a DMP template for the BANOS programme.

Data citation, i.e. proper recognition and reference to data sources, helps supporting data access, sharing and re-use. Data citation forms the basis for the reproducibility of scientific results. Data citation principles, articulated through the Joint Declaration on Data Citation Principles (Martone 2014), represent an advance in the state of the practice and a new consensus on citation.

The use of persistent identifiers such as the Digital Object Identifier (International DOI Foundation 2000) facilitates data citation by providing a unique reference or 'handle' for data to be cited. Through this unique reference, data may be retrieved and accessed. Assigning a permanent identifier to a dataset also requires that the dataset is described by associated metadata and enables long-term preservation and versioning of the data.

Metadata are information that describe data and hence can be defined as "data about data". In the case of environmental or ecological data, metadata document the "who, what, when, where, why, and how" of datasets. "Who" means which organization or institution collected the data, "what" the data types, "when" the dates the data were observed, "where" the locations of the observations, "why" the rationale for the data observation and "how" the methodology applied for obtaining the observation. Metadata are formatted using standards that are relevant for the data to be described, such as:

- The DataCite Metadata Schema (DataCite Metadata Working Group 2019) is a general, multidisciplinary standard. It is used when assigning a DOI, since it defines the minimum set of information that is mandatory for that purpose.
- The Dublin Core (Dublin Core Metadata Initiative 2020) is another general, multidisciplinary standard to describe digital resources and documents.
- The Darwin Core (Biodiversity Information Standards 2018), derived from the Dublin Core, and the Ecological Metadata Language (Jones et al. 2019) are metadata standards that are relevant to describe biological and ecological data sets.
- The ISO 19115 (ISO/TC 211 2014) is the standard to describe geospatial datasets. Using this standard is mandatory to be compliant with data management requirements of European environmental directives such as the INSPIRE Directive and the Marine Framework Strategy Directive.

It should be noted that ascending compatibilities exist between some these standards. For example, metadata expressed using ISO 19115 could be mapped onto the Dublin Core scheme and again onto the DataCite Metadata Scheme. The Metadata Standards Catalog maintained by the Research Data Alliance provides more information and recommendations about metadata (RDA Metadata Standards Catalog Working Group 2019).

A license for research data outlines the conditions for reusing a dataset. A data license is a legal document that describes the rights and the obligations of data users for a specific data set. Agreeing on the terms of the license is a prerequisite for being authorized to access or to download data sets. Adopting a data license is recommended even if data are made freely accessible (Open Data): the data license helps data providers to waive all their copyright and related rights in their works and to protect against abusive use or misinterpretation of provided data.

The open licenses such as Creative Commons CC0 (Creative Commons 2020) have been endorsed by the EU Commission for open data (European Commission 2019) and by several European Members States or associated States, for example the French Etalab 3.0 (Etalab 2019) and the UK Open Government Licence (The National Archives 2020). Some additional restrictions may be added to the Creative Common CC0, such as the "Attribution Restriction - CC-BY" (i.e. obligation of citation of the data source) that imposes users to give appropriate credits to the data providers. The EUDAT License Selector (EUDAT 2017) may help to select the most appropriate license.

The SEIS principles (European Environment Agency 2018) is a set of seven principles that dictates the collection, exchange and use of environmental data and information across Europe. The principles are advocated by the Shared Environmental Information System (SEIS), which in turn is promoted by the European Environmental Agency. SEIS aims to create an integrated, web-enabled, EU-wide environmental information system. The principles advise that information should be

- managed as close as possible to its source;

- collected once and shared with others for many purposes;
- readily available to easily fulfil reporting obligations;
- easily accessible to all users;
- accessible to enable comparisons at the appropriate geographical scale and the participation of citizens;
- fully available to the general public and at national level in the relevant national language(s), and
- supported through common, free, open software standards.

## Best practice

### The European policy framework for open data

The perspective of the European Commission on open data is that its re-use can maximise Europe's digital potential and contribute significantly to growth of the European economy. The Open Data Directive (Directive 2019/1024/EU) is the central piece of legislation in the European Union with respect to open data. It stems from the Public Sector Information (PSI) Directive (Directive 2003/98/EC) and its amendments (Directive 2013a/37/EU), which aimed to create a common data space. The Open Data Directive builds on the PSI Directive addressing the remaining barriers for a wide re-use of publicly funded information. Published in June 2019, Member States should transpose and enforce the Open Data Directive under National law by July 2021. The Directive dictates that publicly-funded research data belong to the public, and therefore should be open access by default. To accommodate the necessary exceptions to open access and comply with General Data Protection Regulations on data protection (Regulation (EU) 2016/679; IDRC 2019), the Directive refers to the credo "As open as possible, as closed as necessary". Concerns in relation to privacy, protection of personal data, confidentiality, national security, legitimate commercial interests such as trade secrets, and to intellectual property rights of third parties are all valid arguments for non-disclosure of public data. When licenses are issued to meet the required restrictions on re-use of data, they should come from standardised open licenses that place as few restrictions on the re-use as possible and impose objective, proportionate and non-discriminatory conditions for re-use of data. The Directive also specifies that data should be compliant with the FAIR Data Principles. Of further focus in the Open Data Directive is the promotion of artificial intelligence and real-time data applications. To this end, the Directive introduces the concept of high-value datasets for society (Art. 13). These are datasets with high potential to generate significant socio-economic or environmental benefits and innovative services and assist in generating revenues for a high number of users, including SMEs. In the EU, 26 countries have started work to identify these datasets (European Data Portal 2018), which cover various thematic categories (geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, mobility). To allow the full exploitation of their potential, these datasets will be made available free of charge, accompanied by an open license, in a machine-readable format and provided via an application programming interface (API) or as a bulk download.

Open access to research data covering all thematic areas in the currently concluding Horizon 2020 is enabled through the Open Research Data Pilot (DG RTD 2016). Projects are given the option to opt out at any stage when requirements for open data cannot be met. However, the restrictions to making research data openly available must be clearly justified and must refer to those also listed in the Open Data Directive such as the protection of individual data, protection of the environment and endangered species, ownership of non-European countries and other arguments as mentioned earlier. In line with the adoption of the Directive in the next Framework Program Horizon Europe, open access ("as open as possible, as closed as necessary") and the use of DMPs will be mandatory. Open data will be the default and research data management will be part of the evaluation of project proposals for Horizon Europe projects (Glinos 2020).

The European Open Science Cloud (EOSC), one of the 49 proposed European partnerships under Horizon Europe (EOSC Executive board 2020), aims to engage all relevant research and innovation stakeholders to jointly develop a virtually federated environment that provides open access to FAIR research data across Europe and

across scientific disciplines. Participation in EOSC is on a voluntary basis, but stakeholders are required to be exemplary in delivering open access and applying FAIR Principles, uphold EOSC standards and agree to the Rules of Participation (RoP). The EOSC Rules of Participation are a clear set of rules that define rights, obligations and accountability governing all EOSC transitions of digital resources by EOSC users, providers and operators. Differentiating rules will be applicable depending on the EOSC users' level of readiness with regard to infrastructures, specificities of scientific disciplines, etc. The RoP serve as a framework to harmonise the heterogeneous European research infrastructure landscape and guarantee a secure and cost-effective federated EOSC (EOSC RoP Working Group 2020). The requirement of stakeholders to be exemplary with regard to open access and FAIR translates to making research data FAIR and compatible. Therefore, working towards EOSC in practice constitutes making research data FAIR.

As mentioned before, Member States should enforce national policies on Open Data by July 2021. Analysis of the current standing on open data in Europe revealed that the level of policy development for open research, consisting of both publications and research data, differs between EU member states (European Data Portal 2018; Tsoukala et al. 2018). The policy landscape regarding open access to publications is described in more detail in BANOS CSA D4.5. It should be mentioned that a national law or policy is not necessarily stronger than a mandate put in place by an institution, funder or open access initiative, especially when funding depends on the latter. An example of such an initiative is Plan S (cOAlition S 2019), which aims to have publicly funded research published in compliant open access journals or platforms. National policies relating to open research data are generally less developed than those for open access to publications, with only a few countries having actual policies in place. However, the importance of open data is well on the radar of EU countries. Many are working on aspects of open research data (including data management, DMPs, FAIR data principles, etc.), either via national governments, funders or institutions. Varying levels of progress in relevant policies and practices are observed across Europe and coordination will be necessary to expedite processes and align policies and practices. As a step towards implementation of open research data, many EU countries reported on having 'current research information systems' (CRIS) in place which are specific platforms harvesting meta-data related to research output and other information of publicly funded research projects. In Flanders for example, the Flanders Research Information Space (FRIS) platform (Department EWI 2019) brings together all research data from publicly financed Flemish scientific research, for which open access is required since the enforcement of the Open Data Directive as amended to the W&I decree in 2019 (Vlaamse Regering 2019).

Apart from EU Member States, both Norway and the UK are represented in the BANOS CSA consortium. Norway and the UK are not directly subjected to the legal framework of the EU, but policies on open data in both countries are on the same level of development as those of many EU Member States and the content is broadly in line with international developments (European Data Portal 2018). In December 2017, the Norwegian Ministry of Education and Research issued the Norwegian national strategy on access to and sharing of research data (2018), stating that research data should be 'as open as possible and closed as necessary' and 'managed and curated to take full advantage of their potential' and with the goal of having all publicly-funded research openly available by 2024. Conformant to the national strategy, the Norwegian Research Council updated its policy 'Open access to research data' (2017) requiring that all research data collected through projects funded by them are open access, with allowance for exceptions where needed, and its management is compliant with the international FAIR principles. The UK government's commitment to public access of publicly funded research by accepting the Finch report (2012) has led to a well-established custom of open access to publications in the UK (Chan 2019). However, similar to many EU countries, policies on open data are not yet in place in the UK. This is primarily caused by the fragmentation of key stakeholders, with the seven research councils having different allocations of resources, roles and responsibilities. In 2016, the British Open Research Data Task Force was created to identify and investigate the possible future steps towards open research data, as published in their final report (Open Research Data Task Force 2018). The report builds on the Concordat on Open Research Data (2016), which is a set of principles on best practices regarding open data published by the scientific community itself.

Due to its aim to make science more efficient, reliable and responsive to societal challenges as well as its implementation in Horizon 2020, the Open Data Directive is a crosscutting component for all fields of research and innovation. In the context of BANOS CSA, several other EU Directives and policies are of relevance for open data:

- The Marine Strategy Framework Directive (MSFD) (2008/56/EC) aims to protect the marine ecosystem and biodiversity across Europe upon which many marine-related economic and social activities depend. Good environmental status (GES) of the European seas was set out to be achieved by 2020, which is unlikely to be fulfilled on time, and significant effort is still needed from the Member States. Article 19.3 of the directive imposes that the related data are made available to the public using the INSPIRE standards. The MSFD constitutes the environmental pillar of the EU Integrated Maritime Policy (IMP) (COM(2007) 575). The integrated maritime policy seeks to provide a holistic approach on the cross-coordination of different marine and maritime policies, and addresses issues that require the coordination of different sectors and actors, since these are not covered by a single sector-based policy.

- The Data Collection Framework (Regulation (EU) 2017/1004), adopted in 2008 and recast in 2017, requires Member States to provide access to high-quality fisheries data, which are collected and managed in the framework of multi-annual national programs, to support the formation of appropriate scientific advice for fisheries management and policy development. Beyond these purposes, access to data, their assembly on a sea-basin level and re-use of assembled data currently requires the consent of all data owners concerned. The scientific advice is used in support of formation of policy advice for the Common Fisheries Policy (Regulation (EU) No 1380/2013b) which aims to conserve fish stocks and reduce overfishing in order to provide EU citizens with a long-term stable, secure and healthy food supply.

- The INSPIRE Directive (2007/2/EC) aims to make spatial data and information publicly available and facilitate international policy-making activities regarding the environment through the establishment of a European infrastructure for geospatial information. The INSPIRE infrastructure, which entered into force in 2007, is based on the spatial data infrastructures established and operated by EU Member States. With regard to the sharing of datasets and services between public authorities, INSPIRE requires Member States to adopt a set of standards for metadata and data services. These must also be used in the framework of other European Environmental Directives such as the Marine Framework Strategy Directive.

- The Maritime Spatial Planning (2014/89/EU) aims to manage European waters in a more cooperative manner across borders and sectors to reduce conflicts, encourage investment, protect the environment and to ensure that activities take place in an efficient, safe and sustainable way. The Directive requires Member States bordering the same marine region to draw up coherent maritime spatial plans that map existing activities in their marine waters and identify most effective future spatial development. Article 10 of the Directive imposes Member States to organise and share the data and information required for maritime spatial plans.

- The Water Framework European Directive (2000/60/EC) aims to protect the inland surface waters, transitional waters, coastal waters and groundwater in qualitative as well as in quantitative terms. The key aims consist of expanding the scope of water protection to all waters and implementing water management based on river basin level to achieve 'good status'. The Directive imposes Member States to make related data available to all publics.

Lastly, global and regional policies, conventions and agendas are also of relevance to open marine data management and data distribution under the future BANOS programme, especially when non-European Countries are bordering the marine regions under consideration:

- The United Nations Convention on the Law of the Sea (UNCLOS) (Montego Bay, 1982) captures the notion that all ocean matters are closely interrelated and should therefore be addressed as a whole. It provides the legal framework for the conservation and sustainable use of oceans through rules governing all aspects of ocean space, such as delimitation, marine scientific research, economic activities, technology transfer and the settlement of disputes relating to ocean matters.

- The United Nations Decade of Ocean Science for Sustainable Development, 2021-2030 (IOC/BRO/2018/2), is a global agenda looking to empower interdisciplinary ocean research at all levels, to support the timely delivery of the data, information and knowledge needed to achieve a well-functioning ocean in support of all Sustainable Development Goals of the UN 2030 Agenda for Sustainable Development (United Nations 2015, A/RES/70/1). The Ocean Decade expresses clear provisions for marine open data management and reuse in its objectives and desired outcomes as presented in its Implementation Plan (IOC-UNESCO 2020). This Plan calls for an accessible ocean with open and equitable access to data, information and technology and innovation, and explicitly sets the objectives to develop interoperable, open access platforms to share data, and to enhance services for the facilitated use of ocean-related data including the social, cultural, environmental, and economic characteristics of the ocean.

- The Convention on Biological Diversity (CBD) (Rio de Janeiro, 1992), which entered into force in 1993, has the goal to globally protect biodiversity. Its main objectives are the conservation of biological diversity, the sustainable use of biodiversity components and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources, which is also known as the concept of Access and Benefit Sharing (ABS). The protocol on 'Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from Their Utilization', better known as the Nagoya Protocol (2011), is a supplementary agreement to the CBD which entered into force on 12 October 2014. It provides a clear, legally-binding framework for implementing ABS, or, in other words, how access to the genetic resources and associated traditional knowledge can be obtained and how the benefits arising from their use will be shared. It states that traceability of all biological samples and related analysis must be maintained as to easily identify their geographical provenance. The protocol has been signed by many countries and was adopted by the EU as part of its regulatory framework.

- The Regional Seas conventions constitute cooperation structures that bring together the EU Member States and neighbouring countries of a marine region and require them to work together with an aim to protect the marine environment. They also form an important aspect in the implementation of the United Nations Convention on the Law of the Sea. There are four European Regional Seas conventions: (1) the OSPAR Convention (1992) for the protection of the marine environment of the North-East Atlantic region, (2) the Helsinki Convention (1974) for the protection of the marine environment of the Baltic Sea region, (3) the Barcelona Convention (1975) for the protection of the marine and coastal region of the Mediterranean region and (4) the Bucharest Convention (1992) on the protection of the Black Sea region against environmental pollution.

## National strategies for open data

For the BANOS open data strategy to be successful in its future implementation, it should be endorsed by all members of the consortium. Preferably, the strategy should also be supported by the funding agencies that may become partners or participants of BANOS. To this end, we polled the consortium members and relevant national funding agencies on the current status of their R&I data management and data sharing policies, and on their disposition towards data management and sharing practises for the future BANOS programme.

A questionnaire was designed (Annex B: BANOS CSA T4.6 questionnaire on Open Data) and distributed in December 2019, addressed to (1) the fifteen member institutions of the BANOS CSA consortium, being the

leading R&I funders of twelve countries around the North and Baltic Seas; (2) associated parties, i.e. the observers (including Finland as the thirteenth country) and strategic partners to the programme with the exception of the European Commission (observer); and (3) other relevant national funding agencies in the countries represented in BANOS CSA.

The questionnaire was developed as an online, interactive form by the Flanders Marine Institute (VLIZ) using in-house software. It was composed of six header questions, each expanding into specific subsequent questions depending on the answer options selected by the respondent. The questions inquired about the respondent institution's policy for open access to scientific articles (further discussed in BANOS CSA report D4.5), current policy regarding research data management and data sharing, current compliance with the Open Data Directive (see section "The European policy framework for open data"), and most importantly, the institution's preferences and concerns regarding implementation of the Open Data Directive in the BANOS programme. The exact questions are included in Annex B: BANOS CSA T4.6 questionnaire on Open Data, and the results are discussed below. Please take note that respondents from the UK and Norway were presented with the same questions pertaining the Open Data Directive, but are not directly subjected to this EU policy.

Submissions to the questionnaire were processed and stored at the Flanders Marine Institute (VLIZ) and the data are available upon request, excluding personal details of the respondents. All answers and information provided by respondents are treated as institutionally-authorised submissions.

Of the 55 institutions that were approached for this survey, all twenty consortium members and associates submitted a complete response, along with fourteen other national funding agencies (total response rate of 62%). The national funding agencies of the countries represented in the consortium are listed in Annex 2 of the BANOS CSA report D2.2 Report on national funding landscape and modalities (ANR and BONUS EEIG 2019) and are listed in the map of the programme's stakeholders presented in the BANOS CSA report D3.2 (Lescroart et al. 2019). In this report, an overview of the institutions targeted for the questionnaire is provided in Annex C: List of institutions targeted for questionnaire with indication of which institutions responded to the questionnaire. For the discussion of the questionnaire results, 'consortium' refers to the BANOS CSA consortium members and associated parties, i.e. observers and strategic partners, and 'other national funding agencies' refers to those surveyed funding agencies that are not involved with the consortium.

In first instance, respondents were asked about the current practises maintained at their organisation regarding data management. Out of the 34 represented organisations, 20 have institutional guidelines in place to handle research data, 5 are in the process of developing such a policy, and 2 follow a higher-level policy (Figure 1).
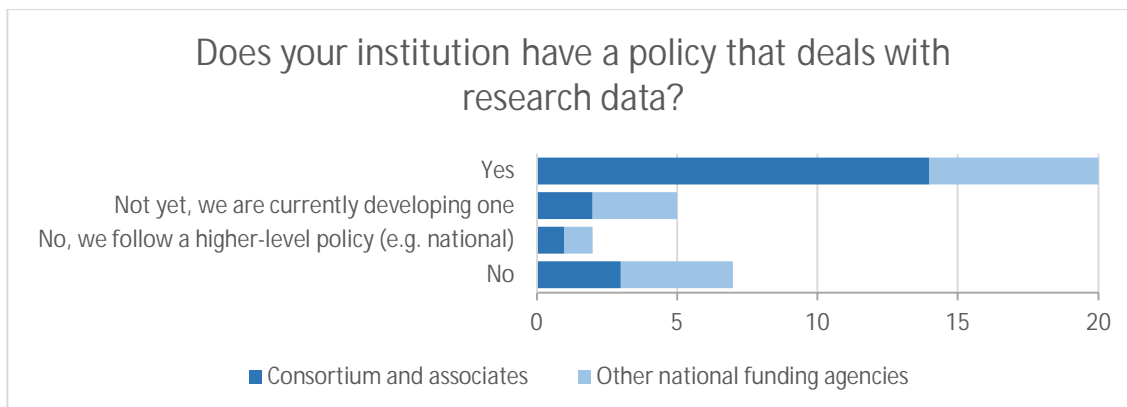


Figure 1. Results for Q2: "Does your institution have a policy that deals with research data?". Respondents could select one out of four answer options. The response from the BANOS consortium members, observers and strategic partners is shown in dark blue, the response from other funding agencies is stacked on top in lighter hue.

Regarding the 7 funding organisations without a clear research data policy, the prospect of adopting a policy is being discussed by some, and in practice the handling of research data is often determined through participation in national or international projects. Most of the established research data policies have explicit provisions for data sharing or open data, and a good majority explicitly supports the use of a Data Management

Plan (Figure 2). The evidently widespread use of DMPs indicates that the DMP is regarded by funding agencies as a viable instrument to implement data regulation.



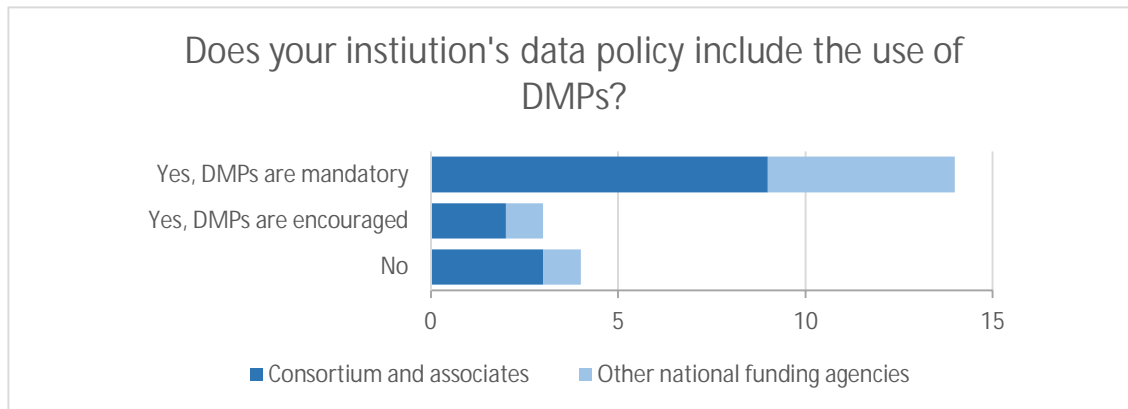Figure 2. Results for Q2.2, Q2.5 and Q2.9: "Does/will your institution's data policy include the use of DMPs?". Respondents could select one out of three answer options. The response from the BANOS consortium members, observers and strategic partners is shown in dark blue, the response from other funding agencies is stacked on top in lighter hue.

As discussed in the preceding section "The European policy framework for open data", the future BANOS programme will have to comply with the Open Data Directive (Directive 2019/1024/EU), published in June 2019. Given that this is a rather recent development, many European Member States along with their respective national funding agencies are likely to be in a process of transition, reshaping national data policies to meet the requirements of the Directive. Since the state of progress of the individual members of the BANOS consortium will influence the ease of implementation in the programme as a whole, it is useful to assess the current state of compliance among European funding agencies (Figure 3). Among the BANOS consortium members and associates (20 institutions, see Annex C: List of institutions targeted for questionnaire), 8 report to be compliant with the Directive (in particular: AKA, Cefas, HELCOM, Ifremer, JPI Oceans, NWO, OSPAR and Jülich) and 10 indicate that steps are currently being taken. Such steps include "the Directive is under discussion" (ICES, LMT, SwAM), "we are awaiting higher-level decisions" (ANR, IFD, NCBR, RCN, SEDA), and "we are actively drafting a new policy" (ETAG, VLIZ). BONUS and Formas stated that they are not compliant, and ICES added the comment that they as an organisation do not act as a primary generator of research data, but rather their network of institutes do. Overall, the results imply that funding agencies are at least aware of the Open Data Directive and moving towards implementation, and in fact this has already been accomplished by many. The consortium members whose current research data policy does not yet satisfy the requirements of the Directive were asked whether there are any aspects of the Directive that they would rather not have to comply with, for any reason at all (financial, technical, administrative…). Out of 12, 6 foresee no problems with any of the requirements, 5 find it too soon to tell, and one consortium member notes that they would rather not have to comply to the principle of 'open by default'. For an overview of requirements in the Directive, see Q3.2 and Q3.3 in Annex B: BANOS CSA T4.6 questionnaire on Open Data.

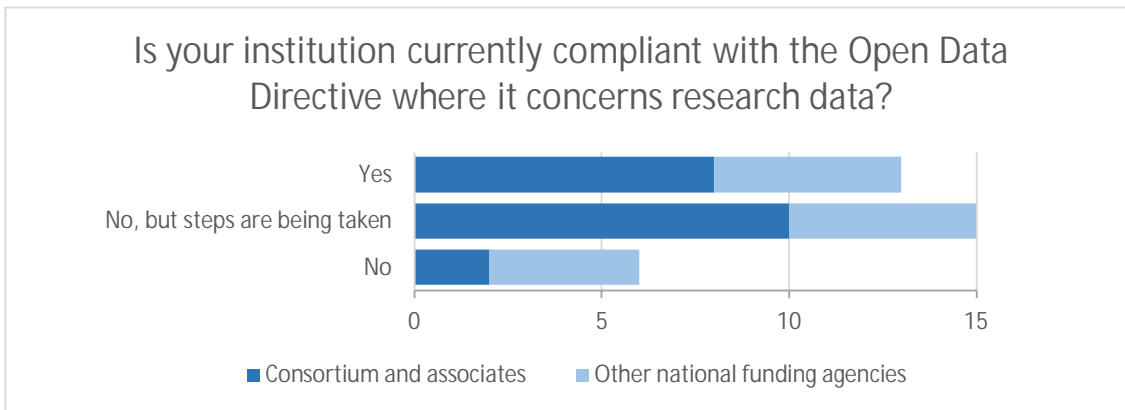**Is your institution currently compliant with the Open Data Directive where it concerns research data?**

Figure 3. Results for Q3: "Is your institution currently compliant with the Open Data Directive where it concerns research data?". Respondents could select one out of three answer options. The response from the BANOS consortium members, observers and strategic partners is shown in dark blue, the response from other funding agencies is stacked on top in lighter hue.

The respondents were further given the opportunity to share where they currently go or intend to go beyond the minimum stipulations of the Directive, i.e. where their policy is more progressive towards data sharing practices than is required by the EU (Table 1). Such efforts can be seen as a reflection of the funding institution's data sharing priorities, and in that light, the measures that are shared among the most consortium members might be interpreted as specific accents to be taken into account for the BANOS programme. However, since the measures could have been established as a consequence of technical opportunities or to meet national or other requirements, rather than as clear-cut institutional priorities, the identification of priorities for the BANOS programme is better addressed by Q5. These considerations aside, the results do at least indicate the current state-of-play among the consortium where it concerns data sharing.

The responses from the consortium show that a slight majority (11 out of 20) does not currently go beyond the Directive's minimum regulations. Among those that do, the majority of the measures are the provision of data without the need for formal access requests, the use of the most open licenses available, and the sharing of all data via APIs and as machine-readable, rather than only the data belonging to high-value datasets. Among the other national funding agencies, a large fraction (10 out of 14) does not go beyond the Directive's requirements. For those that do, the most common measures are the use of the most open licenses available, the provision of all data free of charge without exceptions, and the provision of all data in a machine-readable format, not only data belonging to high-value data sets.

Table 1. Results for Q4: "Are there any ways in which your institution goes beyond the minimum regulations outlined in the Open Data Directive, or intends to do so?". Respondents were free to select multiple answer options. Complete descriptions for those options are listed under Q4 in Annex B: BANOS CSA T4.6 questionnaire on Open Data. Some made additional comments: Cefas remarks that they use the UK Open Government Licence for all publicly funded data, HELCOM remarks that they deal mostly with public monitoring data, which they don't consider to be research data, and Ifremer remarks that while they support the use of open licenses, some restrictions could be applied to certain data sets dealing with e.g. environmental protection or International Regulations for the Ocean.

| | AKA | ANR | BONUS EEIG | Cefas | ETAG | VLIZ | HELCOM | ICES | Ifremer | IFD | JPI Oceans | NWO | OSPAR | Jülich | Formas | LMT | SwAM | NCBR | RCN | SEDA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| none | x | x | x | | x | | | | | x | x | | | x | x | x | | x | x | | 11 |
| no formal application necessary | | | | x | | | | x | | | | | x | | | | x | | | x | 5 |
| use of most open licenses | | | | | | | x | | x | | | x | x | | | | x | | | | 5 |
| workflows, scripts... as open and FAIR | | | | | | x | x | | | | | x | | | | | | | | | 3 |
| publication of all the collected data | | | | | | | x | | x | | | | | | | | | | | | 2 |
| all data free of charge | | | | x | | | x | x | | | | | | | | | | | | | 3 |
| all data via APIs | | | | x | | x | x | x | x | | | | | | | | | | | | 5 |
| all data machine-readable | | | | x | | x | x | x | x | | | | | | | | | | | | 5 |
| all data as bulk download | | | | x | | | x | x | x | | | | | | | | | | | | 4 |

With the institutional views and practices mapped, respondents were asked if they, representing their institution, hold the opinion that the BANOS programme should be compliant with the Open Data Directive where it concerns research data (Figure 4).
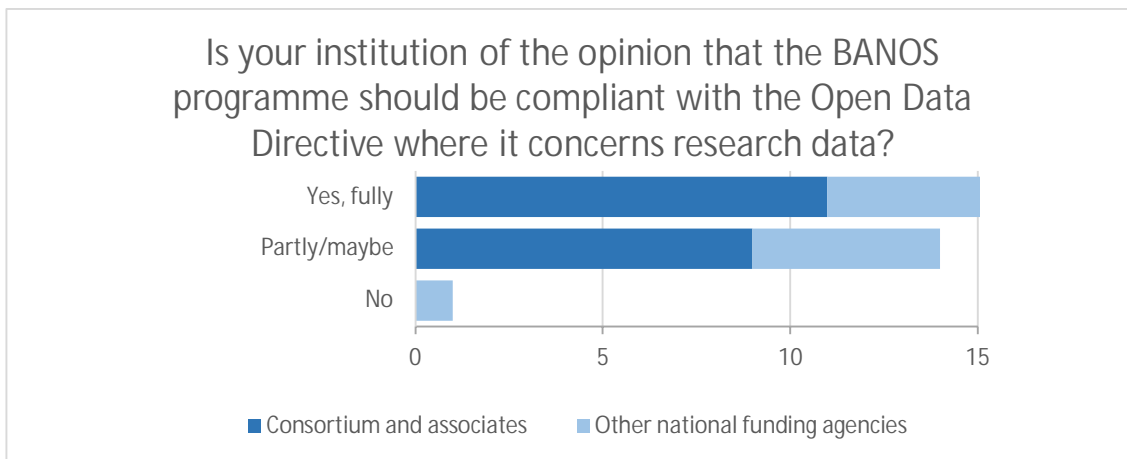


Figure 4. Results for Q5: "Is your institution of the opinion that the BANOS programme should be compliant with the Open Data Directive where it concerns research data?". The respondents could select one out of three answer options. The response from the BANOS consortium members, observers and strategic partners is shown in dark blue, the response from other funding agencies is stacked on top in lighter hue.

Of the consortium members and associates, 11 agree that the BANOS programme should be fully compliant with the Directive, none are opposed, and 9 indicate partial or provisional agreement, stating the following reservations (Q5.1 and Q5.2 in the questionnaire):

- BONUS EEIG remarks that the level of compliance might well depend on the source of funding. If EU funding will be obtained, then complying with the EU Directive is compulsory. Since the Directive is compulsory anyway for all member states from 17 July 2021 onward, even if only national funding agencies are involved, the minimum requirements of the Directive should be fulfilled.
- ICES points out that technical directives are always difficult to implement, and can leave behind those partners that do not have the knowledge or means to be fully compliant. As such, BANOS should make the best efforts to bring all of its constituents up to a level with an aim for full compliance, but not as an absolute.
- ETAG advises to follow the requirements of the scientific publishers, which could mean that not all data can or must be open immediately.
- SEDA represents the National Contact Point and does not store nor generate original research data. Open data regulation is implemented by the Latvian Ministry of Science and Education. Therefore, this Ministry was also contacted with the questionnaire. They are of the opinion that the BANOS programme should be fully compliant with the Open Data Directive where it concerns research data.
- At the time of the questionnaire, some consortium members are hesitant to agree to a full implementation of the Directive in the BANOS programme because of potential discordance with institutional policies. LMT is currently not compliant with the Directive. Ifremer maintains a moratorium for very recent research data. They also have extra access regulations in place for certain data on natural resources, fishery and aquaculture, although they report to be compliant for most other data sets (marine physics, chemistry, biodiversity...). Cefas currently only mints DOI's for selected datasets of the highest scientific value and reports to have a large backlog of legacy data which is not yet available. However, they intend to publish more legacy data alongside all current publicly-funded project data, and to increase their rate of DOI publication for datasets.
- RCN wants to be sure that an implementation of the Directive would agree with the national strategy on access and sharing of research data developed by the Norwegian government. The main principles of the strategy are the same as those pushed in the Directive i.e. 'as open as possible, as closed as necessary', with legitimate considerations to restricted access being security, protection of personal privacy, intellectual property rights, trade secrets and others.
- In general, NCBR agrees with full compliance. However, in the early phase of the discussion, they cannot yet commit to this decision.

The consortium members who stated that BANOS should be fully compliant with the Directive were then asked if the programme should make extra efforts for research data sharing, on top of what the Directive dictates as minimum requirements (Table 2).

Table 2. Results for Q5.1: "Are there aspects where the BANOS programme should go beyond the Open Data Directive?". Respondents were free to select multiple answer options. Complete descriptions for those options are listed under Q5.1 in Annex B: BANOS CSA T4.6 questionnaire on Open Data. With the option 'other', VLIZ elaborated that this question should be evaluated on a case-by-case basis. *As communicated by the Latvian Ministry of Science and Education.

| | AKA | ANR | SEDA* | VLIZ | HELCOM | IFD | JPI Oceans | NWO | OSPAR | Jülich | Formas | SwAM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| none | x | x | | | | x | | | | x | x | | 5 |
| no formal application necessary | | | | | | | x | | x | | | x | 3 |
| use of most open licenses | | | x | | x | | | x | x | | | x | 5 |
| workflows, scripts... as open and FAIR | | | x | | x | | | x | x | | | | 4 |
| publication of all the collected data | | | x | | x | | | | | | | | 2 |
| all data free of charge | | | x | | x | | | | | | | | 1 |
| all data via APIs | | | x | | | | | | | | | | 1 |
| all data machine-readable | | | x | | | | | | | | | | 1 |
| all data as bulk download | | | x | | x | | | | | | | x | 3 |
| Other | | | | x | | | | | | | | | 1 |

A good portion is in favour of not including any systematic extra measures for open data. For those that would like to see extra measures in place, the most supported measures are the default use of the most open licenses available and the application of 'open by default' and the FAIR principles to workflows, scripts and processes leading up to the data, and not just to the data itself.

The questionnaire concluded with an opportunity to add any further comments, recommendations, or sources, such as initiatives or policies that should be taken into account for the BANOS strategies regarding open data and open access (Q6):

- BONUS EEIG notes that complying with open access requirements could be problematic in case of project-based funding. As a general rule, no costs are allowed after a project ends (with some exceptions). However, a substantial part of the outcomes (e.g. articles) will be published after the end of the project. To BONUS EEIG, an important question is how the open access costs will be covered in these cases, and what the relevant time frame should be. Should project participants be able to request an embargo on the research data after the end of the project, whilst preparing a publication linked to the data? Should there be an obligation to make all data available that is generated during the project, even when those data are not tied to publications in scientific journals? They provide an example from the BONUS programme: the BONUS Viable Ecosystem projects (7 projects between 2014-2017) have published 410 peer-reviewed articles up to date, of which 161 (and more to come) were published after the formal end of the projects. Bibliometric analysis of the BONUS+ publications show that 70% were published in the five years after the project ended (Snoeijs-Leijonmalm et al. 2017).
- NCBR remarks that there are many questions that need to be addressed before anyone can make final commitments to the open access and open data strategy. Such questions pertain the implementation of the Open Data Directive, financial issues of the implementation, institutional issues, infrastructural matters, technical difficulties, legal risk (copyrights and related rights) and the opinion of the beneficiaries.

Others used the opportunity to highlight relevant institutional polices and to bring other initiatives to the attention. This included references to: Plan S (cf. supra); the San Francisco Declaration on Research Assessment (DORA); EMODnet (cf. infra); the Defra Data Services platform; the UK Marine Environmental Data and Information Network (MEDIN); the Cefas Data Hub; the International Oceanographic Data and Information Exchange (IODE) network of National Oceanographic Data Centres; the Core Trust Seal delivered jointly by the Research Data Alliance (RDA, cf. supra) and the World Data System (WDS); and the Strategy for Environmental Data Management (2016) adopted by several Swedish agencies.

The conclusion of the questionnaire is that the consortium members are largely in favour of implementing the requirements for open and FAIR data in the BANOS programme, as imposed by the EU Open Data Directive. However, a number of reservations on the practical implementation of the Directive in the BANOS programme will need to be addressed during ensuing discussions with the consortium. This questionnaire has mapped the initial opinions and current policies of the consortium members and other national funding agencies where it concerns open access and data sharing, and as such, the results provide a starting point for a formal discussion on the strategy that will be maintained for the BANOS programme.

## Open data projects

Marine-specific data management is well supported in Europe and relies on three main components: the Pan-European Infrastructure for Ocean and Marine Data Management (SeaDatanet), the European Marine Observation and Data network (EMODnet) and the Copernicus Marine Environment Monitoring Service (CMEMS). Their relative positions in the marine data landscape are illustrated in an infographic prepared by EuroGOOS (Annex A: "European marine data landscape (infographic)"). This section gives a short description of the three main and other marine data systems including the most relevant European projects supporting open and FAIR data.

The Pan-European Infrastructure for Ocean and Marine Data Management (SeaDataNet) is a research infrastructure for marine data management, which networks National Oceanographic Data Centres at a pan-European scale (35 countries, 110 data centres). It provides harmonised data collections, services and guidelines for marine data management such as common data vocabularies, metadata profiles, file formats descriptions, DOI assignment to datasets and other tools. It works closely with EuroGOOS, CMEMS, Euro-Argo, ICES, and EurOBIS, is a major driver of EMODnet, and is a principal initiator of the international Ocean Data Interoperability Platform (ODIP). SeaDataNet is currently supported through the SeaDataCloud project (DG-Research-H2020), working together with EUDAT and considering a connection to EOSC.

URL: www.seadatanet.org

The European Marine Observation and Data network (EMODnet), supported by DG MARE, provides data and integrated data products related to bathymetry, benthic habitats, seafloor geology, marine physics, chemistry and biology, and statistics on marine human activities. With more than 150 organisations working together, EMODnet is making excellent progress in developing added-value services and products, which are attracting users from government, research and industry.

URL: www.emodnet.eu

The Copernicus Marine Environment Monitoring Service (CMEMS) provides real-time and in delayed-mode ocean products (in-situ, satellite observations, forecast models), ocean monitoring indicators, and ocean state reports. It is the marine component of Copernicus, the EU's Earth Observation Programme coordinated by DG Grow. CMEMS's focus is on meteorological and physical oceanographic parameters, with recent expansion into bio-geochemical parameters, to support ecosystem modelling. Next to providing near real-time datasets,

CMEMS has a large interest in archived data collections, which are used as additional input for its forecasting models.

URL: www.marine.copernicus.eu

The European Open Science Cloud (EOSC), as discussed in the section "The European policy framework for open data", is the main European initiative on open and FAIR data and was proposed by the European Commission in 2016. The EOSC architecture is being built in stages, progressively moving towards a federation of infrastructures at the European level, through several initiatives and projects.

URL: www.eosc-portal.eu

> The regional EOSC projects (DG-Research-H2020): EOSC-Pillar (Austria, Belgium, France, Germany, Italy), EOSC-Nordic (Scandinavian and Baltic Countries), etc., aim to coordinate national initiatives for Open Science, ensuring their contribution to and their readiness for the implementation of EOSC.
>
> URL: www.eosc-pillar.eu and www.eosc-nordic.eu
>
> The EOSC-Hub (DG-Research- H2020) is an early adopter program for thematic communities to test and to adapt the services provided by several Open Science service providers such as OpenAire, Geant, OCRE, EUDAT, etc.
>
> URL: www.eosc-hub.eu
>
> Blue Cloud aims to build and demonstrate the pilot Blue Cloud as a thematic EOSC cloud to support research for improving the understanding and management of the many aspects of ocean sustainability, ranging from sustainable fisheries to ecosystem health to pollution.
>
> URL: www.blue-cloud.org
>
> EOSC-Life brings together research infrastructures in the life sciences. The project will publish FAIR data and a catalogue of services provided by participating research infrastructures for the management, storage and reuse of data in EOSC. The marine component is represented by the European Marine Biological Research Centre (EMBRC-ERIC).
>
> URL: www.eosc-life.eu and www.embrc.eu

The community of Environmental Research Infrastructures (ENVRI) is a network dedicated to earth observation and environmental research infrastructures, including e-infrastructures in support of data solutions. Currently the community is supported through the ENVRI-FAIR project (DG-Research- H2020), promoting the adoption of open data and the FAIR principles to connect the community to EOSC, and proposing guidelines and tools towards the implementation of those principles.

URL: www.envri.eu

The Ocean Data Interoperability Platform (ODIP) is a collaborative platform between a consortium of international partners from the EU, USA and Australia. Its objective is the development of a common approach to marine data management to facilitate interoperability of ocean and marine data management infrastructures and hence harmonise the diverse regional systems and stimulate effective data sharing. Products and services developed by ODIP are actively promoted at the international level through IOC/IODE, who is also involved in its implementation and operation, the Research Data Alliance and GEOSS.

URL: www.odip.eu

The Marine Information System for Europe (WISE Marine), which represents the marine branch of the Water Information System for Europe (WISE), is a data infrastructure and portal to share marine environmental information at a European level such as data reported by Member States under the Marine Strategy Framework Directive (cf. supra). WISE Marine fits in the landscape of existing Directive reporting and associated portals, European Data infrastructures, and Regional Sea Conventions – all of which are heavily dependent on data and information coming from EU member states.

URL: www.water.europa.eu/marine

The ICES Data Portal is a data system, governed by the International Council for the Exploration of the Sea (ICES), which holds dataset collections related to the marine environment. The datasets are organised according to specific themes and can be accessed either directly or through a programming interface provided by web services. They contribute to advancing and sharing scientific understanding of marine ecosystems to meet conservation, management and sustainability goals.

URL: www.ecosystemdata.ices.dk

PANGAEA represents an information system for archiving, publishing and distributing georeferenced datasets from earth-system scientific research including oceans and fisheries. It operates as an Open Access repository, providing freely available data that can be used under the mentioned license terms. As a member of the World Data System, which is an interdisciplinary body of the International Science Council, PANGAEA contributes to promoting long-term, universal and equitable access to quality-assured scientific data and services.

URL: www.pangaea.de

The European Union Open Data Portal (EU ODP) is a standardized catalogue that provides open access to data published by EU institutions and bodies. As such it increases the transparency of EU institutions on open data. Besides this, the EU ODP also offers a range of web services for reusing those data, a query editor and tips on how to make best use of the open data portal.

URL: www.data.europa.eu/euodp

The Open Access Infrastructure for Research in Europe (OpenAIRE) is an initiative, supported by the European commission, to advance Open Science in Europe. As an innovative digital infrastructure, it provides a catalogue on research projects and data as well as a variety of Open Science services such as toolkits and workshops. Besides this, OpenAIRE also represents a well-established human network of National Open Access Desks (NOADS), which are local Open Science experts who work to develop policies and help researchers adopt Open Science practices. The most recent phase, OpenAIRE Advance, continues the mission of OpenAIRE namely sustaining the current infrastructure, advancing the NOADs, incorporating citizen science initiatives through the Open Innovation programme and also cooperates with EOSC-HUB to prepare for EOSC.

URL: www.openaire.eu

The EU-funded project Policy RECommendations for Open Access to Research Data in Europe (RECODE) aimed to identify barriers to open access and dissemination of research data in Europe. The project therefore leveraged existing networks, communities and projects working around Open Access to research data, which are fragmented by boundaries such as disciplines, geography and stakeholder categories. And the findings have led to the formation of a series of policy recommendations on open access to research data targeted at key stakeholders, namely research funders, research institutions, data managers, and publishers.

URL:
http://www.gsrt.gr/EOX/files/Policy%20recommendations%20for%20open%20access%20to%20research%20data.pdf

# Ingredients for a successful Data Management Plan

Typically, a template of a Data Management Plan (DMP) is compiled by the funder and then provided to the (aspiring) beneficiaries i.e. the project participants or applicants, who use the template to draft a full DMP. The funder can enforce its data policy by including mandatory sections and provisions in the DMP template. It is highly recommended that the future BANOS programme should create such a DMP template for use by all the programme's beneficiaries. This template will assist grant beneficiaries in creating a DMP that complies with all the minimum requirements for data management and data sharing as set by the funding programme. There are several reasons to abstain from presenting a ready-to-go DMP template, concluded from the work presented in this report. Most importantly, the BANOS CSA consortium first needs to make fundamental decisions regarding the programme's desired policy for data management and data sharing. The creation of a finalised DMP is dependent on a fully-defined programme structure and policy. In addition, the target groups of the DMP template need to be known, when these are defined in tandem with the call topics. These in turn depend on the final BANOS SRIA and its concrete implementation. Different DMP templates may be necessary for distinct call topics. Therefore, rather than a finished DMP template, this report presents a list of the typical components that go into the making of a DMP, with additional advice and a presentation of the options where relevant. This resource will aid in the swift compilation of a DMP template once the requirements outlined above are met.

Multiple existing DMP templates and other sources were consulted to compile this list of components. The structure in which these components are presented here follows the FAIR principles. The same structure is used in the DMP template of H2020 (European Commission 2016), which is one of the two major sources for the compilation of this list. The other major source is a report from Science Europe titled Practical Guide to the International Alignment of Research Data Management (Science Europe 2018), which proposes a set of Core Requirements that any DMP should fulfil. Where possible, reference to the corresponding Core Requirement (CR1-6) is made in our component list. The components also are accompanied by specific recommendations where relevant, as highlighted in grey. Additional DMP templates were scanned to make sure that no crucial components were missed. These are the DMP templates of the Dutch Research Council (NWO 2019), Research Foundation – Flanders (FWO 2018) and the Flanders Marine Institute (VLIZ, unpublished). Note that although the components listed here are quite extensive, they are not exhaustive. On a similar note, the list is meant as a broad overview of options and it is certainly not advisable to include all of the presented components in the future BANOS DMP template.

## General information

- Name of applicant.
- Project number, title, acronym, lead, participants, abstract.
- Funding programme.
- DMP title, version.
- DMP authors, affiliations, contact details.
- Person (name, contact details) bearing the end responsibility for the coordination of data management in the project, and updating and implementing the DMP.
- Reference to other procedures for data management that are being used, if any (e.g. national/funder/sectorial/departmental).

## Data summary

- Types of data, the kind of content: numeric or textual databases, images, audio, video, geospatial... concerning experimental results, environmental variables, biodiversity measures, genetic data, user-submitted data [CR1b].
- Formats of data, i.e. the way in which the data are encoded for storage as often reflected by the filename extension: CSV, TXT, RDF [CR1b].
- Justification of the choice of format: staff expertise, standards accepted by target repositories, widespread usage within the research community, restrictions/compatibility with software or equipment, preference for open formats [CR1b].
- Expected size of the data (e.g. in bytes, number of files, number of observations...) [CR1b].
- Re-use of existing data, reference to their source, restrictions put in place by third parties to the dissemination or exploitation of data that will be re-used.
- Purpose of data creation, relation to the project objectives, data utility i.e. user groups that can benefit from the data.
- Geographical origin of the new data.
- Brief description of the methodology for collecting or generating the new data [CR1a].
- Software needed to produce and/or process the data [CR1a].
- Person or institution responsible for data capture.
- Reasons for considering but subsequently discarding the re-use of any existing data sources [CR1a].

---

- ➤ Give preference to expression of the size in bytes, as it unambiguously reflects the required storage space. However, other field-specific size indicators are welcome additions as these can be more human-readable and more easily interpretable.
- ➤ Give preference to open, community-common standardised data formats. These facilitate the sharing and long-term re-use of data.
- ➤ Follow the preferred data formats of the target repositories [CR1b].

## To be Findable, and provisions for metadata

---

- Discoverability: location of the metadata: can be in the data files themselves, in separate files, or only in the catalogue entry.
- Discoverability: location of the data: name of the repository, indexation in a catalogue [CR5a, CR5b].
- Identifiability of data i.e. the standard identification mechanism. Assignment of a unique persistent identifier such as a Digital Object Identifier (DOI) to the data set.
- Approach towards data discovery keywords.
- Approach for clear versioning (i.e. keep track of updates and changes to the data), data provenance (i.e. logging data sources and ownership) and folder structures [CR1a, CR2a].
- Specification of the standards used in the metadata creation (e.g. DDI, TEI, EML, MARC, CMDI). If no standards are available for a given discipline, description of how and what metadata will be provided [CR2a].
- Provision of documentation that enables understanding and re-use of the data: methodology, procedural information, definitions of variables, units of measurements etc. Description of where this information will be recorded e.g. in a database, a 'readme' text file, file headers, code books or lab/field notebooks [CR2a].
- Person bearing the responsibility for proper data documentation and provision of adequate metadata [CR3b].

---

- ➤ Always create a metadata record in a public, community-common archive [CR2a, CR5b].
- ➤ Use community (meta)data standards where these are in place [CR2a].
- ➤ Persistent identifiers allow reliable and efficient location of and referral to the data. They also help to track citations and other metrics for re-use.

- Accessibility of the data: method of access (e.g. direct download, handling of data requests...), use of a secure data service, storage in a repository. If no established repository is proposed, demonstration that the data can be curated effectively beyond the lifetime of the grant [CR5b].
- Person responsible for the preservation, archiving and sharing of the data e.g. by submitting the data to a trusted repository.
- Documentation of the data: location of metadata, code, (reference to) software tools and additional documentation required to access, understand and re-use the data.
- Owner of the data, with the rights to control access [CR4b].
- Restrictions to data access, outlining the rationale for restricted access and the data it applies to, who will have access and how access will be provided, what actions will be taken to overcome or minimize restrictions [CR4b, CR5a].
- Restrictions on the re-use of third-party data [CR4b].
- Intellectual property rights: (potentially) affected rights, how they will be dealt with [CR4b].
- Minimum period of data retention after the project. Potential reasons for not preserving complete or partial data (legal or contractual restrictions, physical preservation issues...) [CR5a].
- Selection of data for long-term preservation, rationale behind selection, description of selected data. Indication of data that needs to be destroyed or retained specifically for contractual, legal or regulatory purposes [CR5b].
- Timing of data release and embargo on data for exclusive use: duration, justification e.g. intention to publish, protection of intellectual property, seeking patents [CR5a].

---

➢ Data access should be free of charge.
➢ Choose only trusted, long-term repositories. Criteria for the selection of trustworthy repositories are proposed by Science Europe in their report Practical Guide to the International Alignment of Research Data Management (2018).
➢ The Open Data Directive states that data must be immediately accessible, which would improve the efficiency of the handling of data requests.
➢ The owner of the data can be the funder, the project leader, the institution generating the data... This needs to be specified in the grant agreement/rules of participation of the future programme. For multi-partner projects and multiple data owners, cover the rights of control to data access in the project consortium agreement [CR4b].
➢ As a common measure, funders require the participants to retain the data for 5 years after the project ends. In principle however, research and innovation actors should strive to retain data 'indefinitely'.
➢ Preferably, no embargo is imposed on the data and in any case, research data should be made available as soon as possible after the end of a project. If necessary, e.g. to give time to publish or to seek patents, the commonly-accepted period of embargo is 6 or 12 months.
➢ Preferably, all research data outlined in the DMP of a project is made available. However, depending on the policy decided by the funders and taking into account higher-level policies, researchers could limit themselves to make only data available that is linked to publications.
➢ Consider the sustainability of software needed for accessing the data.

## To be Interoperable

---

- Usage of standard data formats and methodologies.
- Specification of standard vocabularies for all data and metadata. Where non-standard vocabularies are used, provision of mapping to more common ontologies.

- Control and documentation of the consistency and quality of data collection e.g. calibration, repeated measurements or samples, standardised data capture, data entry validation, peer review of data, representation with controlled variables [CR2b].

## To be Reusable

- Licencing of the data: choice of license, restrictions and conditions for re-use and if applicable, justification for setting restrictions and conditions for re-use.
- Potential for intellectual property claims: data of relevance, ensuing restrictions.
- Indication of the potential future research uses/users of the data.
- Person responsible for overseeing the quality of the data.

> - Consistent, well-ordered research data is easier to find, understand and re-use [CR2a].
> - Data should be licensed as such to permit the widest re-use possible. For this purpose, usage of the licenses CC0 and CC BY 4.0 is endorsed by the European Commission (2011/833/EU). See the section "Open data: general concepts and definitions" for more on the licencing of research data.
> - The licenses of choice should be standard and widely used.
> - The choice of license(s) should be defined in the project consortium agreement.

## Data storage and backup

- Persons responsible for performing regular data backups during the project [CR3b].
- Persons responsible for proper data storage [CR3b].
- Estimation of the costs of long-term preservation of the data. Estimations could cover
    o the project duration;
    o an extra five years after the project;
    o indefinite storage.
- Description of the potential value of long-term preservation of the data.
- Estimation of the costs for data management, for during and after the lifetime of the project.
- Availability of sufficient storage and backup capacity. Location and provider(s) of storage and backup capacity. If the available capacity will be insufficient, explanation of how the issue will be dealt with. Description of the data backup and recovery process [CR3a, CR3b].
- Measures for data security and transfer of sensitive data i.e. prevention of access or modification by unauthorised users [CR3b].
- Description of the main risks and appropriate risk management for projects involving sensitive data (e.g. personal data, politically sensitive information, trade secrets) [CR3b].
- Reference to institutional data protection policies that might be in place [CR3b].

> - Responsibilities should be assigned to individual people rather than institutional entities.
> - Reserve a dedicated fraction of data management and preservation costs in the project grant. In general, the recommended fraction is 5% of the total grant (Mons 2020).
> - Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

- Ethical issues concerning the creation and/or use of the data (e.g. trials involving humans or animals, dual use). Impact of ethical issues on data storage, transfer, access rights and preservation [CR4c].
- Usage of personal data. Description of the kind of personal data. Steps foreseen to handle, store and distribute personal data in accordance with personal data protection laws, notably GDPR (Regulation (EU) 2016/679) [CR4a].
- Compliance to the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity (2011).
- Requirement of any other types of permits for data collection and re-used data during the project.

➢ If there are ethical issues concerning the creation and/or use of data, request a formal approval by a relevant ethical review committee [CR4c].
➢ Ethical issues do not need to be covered in the DMP if the project application requires a dedicated ethical section elsewhere.
➢ Gain informed consent for preservation and/or sharing of personal data.
➢ Consider the following procedures for personal data [CR4a]:
  - Anonymisation: truly anonymous data are no longer considered personal data
  - Pseudonymisation: reversible anonymisation
  - Encryption: pseudonymisation where an encryption key is needed to reverse the anonymisation. The key should be stored separately from the data (e.g. with a trusted third party such as the funder)

# Recommendations

The following overview of recommendations is non-exhaustive and was established through consultation and summary of the recommendations from several sources, notably those proposed by the RECODE project and the JERICHO-NEXT project (Finn et al. 2014; Tsoukala et al. 2015; European Data Portal 2016; Gorringe et al. 2017). In addition, an interview was conducted with the EMODnet Secretariat to add to these recommendations from the viewpoint of a leading European data repository (Lescroart 2020).

Outline clear roles and responsibilities for each stakeholder and promote open research.

Open science increases the quality and impact of research, making it more reliable and more responsive to societal needs and is expected to contribute significantly to innovative development. However, an obstacle to achieving this are the current incentive structures of academic research, which often fail to recognize, value and reward efforts to open up scientific research. A key aspect in fostering the necessary change in research culture is the development of explicit policies on open research data with a clear description of roles and responsibilities for each stakeholder. Policies should set open access as the default and follow the credo "open if possible, restricted if necessary" to accommodate the possibility for closed data when ethical, confidentiality and security issues are of key concern. For research data produced by projects, they should require DMPs, use of open licenses and machine-actionable schemes to enable proliferation of reusable research data. Funders should also take the lead in collaborating with all relevant stakeholders and networks (researchers, research institutions policy makers, data managers and publishers) while establishing these policies, as this is important in developing aligned policies and strategies.

Besides policy development, funding initiatives should also take a proactive role in changing research cultures through soft-type activities such as promoting good practices and, in view of providing incentives for researchers, establishing better institutional reward systems for contributing researchers. Such support could include access to technologies, skills and infrastructure related to data storage and management. Another

element that can contribute to a faster and deeper change in research culture are new generation metrics. In combination with conventional indicators, these metrics can both monitor development and measure performance towards openness. As a way to provide robust data on researcher's participation in open research, funders should look into the development and inclusion of such metrics in their policies.

Adopt a comprehensive approach in funding the implementation of open access to and preservation of research data.

For the implementation of mandatory policies for open access to research data, funders should fund and/or oversee that appropriate infrastructure and processes are in place. In other words, they should provide collaborative and scalable infrastructures and services for access to and long-term preservation of research data, take innovative actions that boost data re-use in the research and innovation sector and contribute to the development of skills among researchers and information specialists. Funders, especially major public research funders of nation-wide or transnational scale, are encouraged to mobilise complementary funding instruments, where available, to support the implementation of comprehensive policies for open access to research data. To that end, funders are encouraged to evaluate the services for research data access and preservation in their country along with the needs of the research community.

Promote data management as a distinct activity within the research process.

Data management plans have emerged as an important tool in achieving openness of research data. To acknowledge research data management and sharing as a distinct and required activity, funders should require a DMP for research data produced with project funding. Further, they should provide relevant guidance on the points that a DMP should address, i.e. by making DMP templates available to beneficiaries. Relevant resources should be allocated for the inclusion of data management in the project grant. Investing 5% of a project's budget into data management offers an excellent return on investment, as it can encourage institutional employment of data stewards. Data stewards support researchers with data-specialised tasks, lending greater speed and efficiency to the job. They can supervise data curation, storage and preparation for re-use, and advise researchers on experimental design and data capture to allow consistent downstream data processing. With a significant cut in complex and time-consuming tasks, researcher are better able to focus on high-quality research, leading to a boost in effective research capacity (Mons 2020).

DMP as two-step process.

It is recommended to require a draft DMP with project applications and, if the project is selected, a full DMP 4-6 months after the start of the project (H2020, NWO, FWO). This forces applicants to think ahead about the management of data that will be collected during their project, and gives them time to make necessary adjustments once the project is underway. Although DMPs (in principle) outline all the data that will be collected during the project, that does not necessarily mean that all data is used for publications. Depending on the policy decided by the funders, researchers could limit themselves to make only data available that is linked to publications.

Ensure that research data generated through BANOS projects is compliant with the requirements for open research data set by Horizon Europe.

The foremost recommendation for research data generated from projects within Europe is that it should be compliant with the requirements for open research data set by H2020 and Horizon Europe (DG RTD 2017; European Research Council 2017). These state that research data must be deposited in a research data repository as soon as possible, together with other collected data including associated metadata, and that this should be linked to publications and underlying data through use of persistent identifiers and data citations; that measures must be taken to enable third parties to access, mine, exploit, reproduce and disseminate this research data, free of charge; and that information should also be provided about the tools available to validate

the results, and if possible these tools themselves should also be provided. Therefore, the recommendations below were identified in relation to the use of data repositories, persistent identifiers, data citation and licensing.

Use existing repositories to submit research data.

As a funding programme, developing a new repository from scratch has the advantage that users and re-use can be monitored closely, but it also requires a significant amount of resources. A user base would need to be set up from scratch and, if data were to be made available in other repositories in an attempt to increase outreach, data would exist in duplications, which impedes unambiguous provenance and versioning. Taking these disadvantageous into account, it is recommended to have (meta)data submitted to existing repositories collected. An added benefit of using existing repositories is that these will enforce their standards on the data and metadata, which increases interoperability and quality.

For large-scale programmes such as the BANOS programme, it may be beneficial to set up fixed data flows with certain repositories (e.g. EMODnet, as communicated by the EMODnet Secretariat), rather than have project-generated data submitted through the regular ingestion portals of those repositories. In the case of EMODnet, if such a setup would be desired, BANOS should reach out to the separate EMODnet portals to decide on data flows at the portal level.

Given the various data that will be generated through BANOS projects, it is recommended to provide a flexible list of repositories from which participants can choose and to which they may refer in their DMP, rather than imposing a fixed and limited number of repositories ahead of the projects. An impetus for this list is given in Annex D: Impetus for list on recommended data repositories based on a list compiled in the ASSEMBLE Plus project (D4.1, 2018). Not all fields relevant to BANOS are represented yet, as user feedback will be needed to come to agreement on which repositories to use. It is also important to realise that many institutions already have their own go-to repositories, which could very well prove to be suitable repositories for specific types of data and which they will be reluctant to discard in favour of other repositories.

Adopt a metadata driven approach for making data findable.

In order to ensure interoperability of metadata and data collected through BANOS projects, it is recommended to collect rich metadata and use appropriate and internationally accepted standards, protocols and dictionaries (such as the BODC vocabulary, the INSPIRE standards, etc.) to document and exchange metadata. To increase findability of data, it is further recommended to make the metadata available and searchable through a catalogue with a web-based interface (cf. the repositories listed in Annex D: Impetus for list on recommended data repositories).

Besides collection of metadata, the use of globally resolvable persistent identifiers, such as DOIs or other local persistent identifiers is recommended in order to provide permanent links to and between publications and underlying data. Some repositories (e.g. EMODnet) may agree to mint persistent identifiers for project datasets, given a fixed data flow. Further, proper data citation is required to acknowledge contributors and data sources. Both these methods will contribute to the findability of data.

Use standardised open licenses to demonstrate data availability and address legal obligations.

Open licenses, such as the Creative Commons attribution licenses, outline how information should be accessed, shared and re-used. They provide a useful way to enable third parties to access, mine, exploit, reproduce and disseminate research data, free of charge, and also address legal and ethical issues. Therefore, the use of open licenses in research data management should be required. The EUDAT B2SHARE tool (EUDAT 2017) includes a built-in license wizard that facilitates the selection of adequate licenses for research data.

## Conclusion

Open data is an essential part of contemporary research and innovation and has the potential to significantly enlarge the societal impact of R&I projects. This is firmly understood by the EU and reflected as such in the recently published Open Data Directive and the upcoming Framework Programme Horizon Europe, and echoed by many other policies and initiatives throughout Europe and beyond. BANOS should therefore strive to implement the necessary provisions for the proper management and sharing of research and innovation data. The survey conducted during this study gathers the state of affairs and opinions of all the BANOS CSA consortium members on the matter of open data and, together with the expert recommendations included in this report, is to guide the consortium's decisions regarding the data policy of the BANOS programme. A central instrument for the practical implementation of an open data policy should be the use of Data Management Plans, and to that end BANOS should make DMP templates available to its beneficiaries. These templates can be compiled from the list of components presented in this report. Further, project-generated data should be deposited in high-quality repositories, and an impetus for a list of such repositories is given in annex.

## Consulted sources

ANR, BONUS EEIG. 2019. Report on national funding landscape and modalities. BANOS CSA/D2.2.

ASSEMBLE Plus. 2018. D4.1 ASSEMBLE Plus Data Management Plan. http://www.assembleplus.eu/results/wpreporting.

Biodiversity Information Standards. 2018. Darwin Core. https://dwc.tdwg.org/.

Chan TT. 2019. Open Research Policies in the United Kingdom - Open Science Monitor Case Study. https://ec.europa.eu/info/sites/info/files/research_and_innovation/open_research_in_the_uk_0.pdf.

cOAlition S. 2019. Accelerating the transition to full and immediate Open Access to scientific publications. Brussels. https://www.coalition-s.org/wp-content/uploads/PlanS_Principles_and_Implementation_310519.pdf.

Creative Commons. 2020. CC0 1.0 Universal. https://creativecommons.org/share-your-work/public-domain/cc0/.

DataCite Metadata Working Group. 2019. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. doi:https://doi.org/10.14454/f2wp-s162. https://schema.datacite.org.

Department EWI. 2019. FRIS Research Portal. Flem Gov. https://researchportal.be/en/.

DG RTD. 2016. H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

DG RTD. 2017. H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Digital Curation Centre. 2010. DMPonline. https://dmponline.dcc.ac.uk/.

Dublin Core Metadata Initiative. 2020. DCMI Metadata Terms. https://www.dublincore.org/.

EOSC Executive board. 2020. Draft proposal for a European Partnership under Horizon Europe European Open Science Cloud (EOSC) Partnership. https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-open-science-cloud-eosc.pdf.

EOSC RoP Working Group. 2020. Rules of Participation Working Group. https://www.eoscsecretariat.eu/working-groups/rules-participation-working-group.

Etalab. 2019. Etalab. https://www.etalab.gouv.fr/.

EUDAT. 2017. License Selector. https://www.eudat.eu/services/userdoc/license-selector.

EuroGOOS AISBL. 2017. Copernicus Marine Service and marine sea-borne observation infrastructure. http://eurogoos.eu/download/project_deliverables/Copernicus-2017-09-MarineService_insitu_obs_Sept.2017.pdf.

European Commission. 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Off J Eur Union. 43(L 327). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000L0060.

European Commission. 2007. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: An Integrated Maritime Policy for the European Union. COM/2007/0575 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52007DC0575.

European Commission. 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Off J Eur Union. 51(L 164). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008L0056.

European Commission. 2010. EUROPE 2020: A European strategy for smart, sustainable and inclusive growth. (COM(2010) 2020 final):37. https://www.eea.europa.eu/policy-documents/com-2010-2020-europe-2020.

European Commission. 2011. Commission Decision of 12 December 2011 on the reuse of Commission documents (2011/833/EU). Off J Eur Union. 54(L 330):39–42. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32011D0833.

European Commission. 2014. Directive 2014/89/EU of the European Parliament and of the Council of 23 July 2014 establishing a framework for maritime spatial planning. Off J Eur Union. 57(L 257). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0089.

European Commission. 2016. Horizon 2020 FAIR Data Management Plan (DMP) Template. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template.

European Commission. 2019. Commission Decision of 22.2.2019 adopting Creative Commons as an open licence under the European Commission's reuse policy. https://ec.europa.eu/transparency/regdoc/rep/3/2019/EN/C-2019-1655-F1-EN-MAIN-PART-1.PDF.

European Data Portal. 2016. Open Data Goldbook for Data Managers and Data Holders. https://www.europeandataportal.eu/sites/default/files/goldbook.pdf.

European Data Portal. 2018. Open Data Maturity in Europe. https://www.europeandataportal.eu/en/news/open-data-maturity-europe-2018.

European Environment Agency. 2018. Shared Environmental Information System. https://www.eea.europa.eu/about-us/what/shared-environmental-information-system-1.

European Parliament, Council of the European Union. 2003. Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. Off J Eur Union. 46(L 345):90–96. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2003.345.01.0090.01.ENG.

European Parliament, Council of the European Union. 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Off J Eur Union. 50(L 108):1–14. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002.

European Parliament, Council of the European Union. 2013a. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. Off J Eur Union. 56(L 175):1–8. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32013L0037.

European Parliament, Council of the European Union. 2013b. Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC. Off J Eur Union. 56(L 354):22–61. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32013R1380.

European Parliament, Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da. Off J Eur Union. 59(L 119):1–88. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504.

European Parliament, Council of the European Union. 2017. Regulation (EU) 2017/1004 of the European Parliament and of the Council of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common . Off J Eur Union. 60(L 157):1–21.

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R1004.

European Parliament, Council of the European Union. 2019. Directive 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). Off J Eur Union. 62(L 172):56–83. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024.

European Research Council. 2017. Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020. http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf.

Finn R, Wadhwa K, Taylor MJ, Sveinsdottir T, Noorman M, Sondervan J. 2014. Deliverable D3.1: Legal and ethical issues in open access and data dissemination and preservation. https://www.trilateralresearch.com/wp-content/uploads/2018/09/D3.1-legal-and-ethical-issues-FINAL.pdf.

Flanders Marine Institute. VLIZ Data Management Plan: Information and Guidelines for Researchers.

FWO. 2018. DMP template. https://www.fwo.be/en/the-fwo/organisation/data-management-plan/.

Glinos K. 2020. Making Open Science the new normal. In: Keynote presentation slides from the 15th International Digital Curation Conference. Dublin, Ireland: Zenodo. p. 35. https://zenodo.org/record/3688670#.X2SH8GgzY2y.

Gorringe P, Novellino A, Fernandez P. 2017. D5.1 Recommendation on a free and open data access policy. http://eurogoos.eu/download/project_deliverables/JERICO-NEXT-2017-D5.1-Recommendation-on-a-free-and-open-data-access-policy-2017.pdf.

Higher Education Funding Council for England, Research Councils UK, Universities UK, Wellcome. 2016. Concordat on Open Research Data. https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/.

IDRC. 2019. The State of Open Data: Histories and Horizons. Davies T, Walker SB, Rubinstein M, Perini F, editors. Cape Town and Ottawa: African Minds and International Development Research Centre. https://www.idrc.ca/en/book/state-open-data-histories-and-horizons.

International DOI Foundation. 2000. Digital Object Identifier System. https://www.doi.org/.

IOC-UNESCO. 2018. The United Nations Decade of Ocean Science for Sustainable Development, 2021-2030. IOC/BRO/2018/2. https://unesdoc.unesco.org/ark:/48223/pf0000261962.

IOC-UNESCO. 2020. Implementation Plan Version 2.0. https://www.oceandecade.org/resource/108/Version-20-of-the-Ocean-Decade-Implementation-Plan-.

ISO/TC 211. 2014. ISO 19115-1:2014 Geographic information — Metadata — Part 1: Fundamentals. https://www.iso.org/standard/53798.html.

Jones M, O'Brien M, Mecum B, Boettiger C, Schildhauer M, Maier M, Whiteaker T, Earl S, Chong S. 2019. Ecological Metadata Language version 2.2.0. doi:10.5063/f11834t2. https://eml.ecoinformatics.org.

Lescroart J. 2020. Interview with EMODnet Secretariat in context of BANOS CSA T4.6.

Lescroart J, De Raedemaecker F, Pirlet H, Lescrauwaet A-K, Sirola M. 2019. A holistic map of the programme's stakeholders. BANOS CSA/D3.2.

Martone M (ed). 2014. Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. doi:https://doi.org/10.25490/a97f-egyk. https://www.force11.org/datacitationprinciples.

Max-Planek-Gesellschaft. 2003. Berlin Declaration. Open Access Max-Planck-Gesellschaft. https://openaccess.mpg.de/Berlin-Declaration.

Mons B. 2020. Invest 5% of research funds in ensuring data are reusable. Nat World View.:491. doi:10.1038/d41586-020-00505-7. https://www.nature.com/articles/d41586-020-00505-7.

Norwegian Ministry of Education and Research. 2018. National strategy on access to and sharing of research data (summary). https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/.

NWO. 2019. DMP template.

Open Research Data Task Force. 2018. Realising the potential: final report of the Open Research Data Task Force. https://www.gov.uk/government/publications/open-research-data-task-force-final-report.

RDA Metadata Standards Catalog Working Group. 2019. RDA Metadata Standards Catalog. https://rdamsc.bath.ac.uk/.

Redkina NS. 2019. Current Trends in Research Data Management. Sci Tech Inf Process. 46(2):53–58. doi:10.3103/S0147688219020035. https://doi.org/10.3103/S0147688219020035.

Science Europe. 2018. Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management.

Secretariat of the Convention on Biological Diversity. 2011. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity. Montreal: United Nations Environmental Programme. https://www.cbd.int/abs/text/.

SEPA, SwAM, County Administrative Boards of Sweden, Swedish Regional Water Authorities. 2016. Strategy for Environmental Data Management. https://www.naturvardsverket.se/upload/sa-mar-miljon/oppna-data/miljodatastrategi/strategy-for-environmental-data-management-161107-ver-1.02.pdf.

Snoeijs-Leijonmalm P, Barnard S, Elliott M, Andrusaitis A, Kononen K, Sirola M. 2017. Towards better integration of environmental science in society: Lessons from BONUS, the joint Baltic Sea environmental research and development programme. Environ Sci Policy. doi:10.1016/j.envsci.2017.10.004.

The National Archives. 2020. UK Open Government License for public sector information. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.

The Research Council of Norway. 2017. Open Access to Research Data, Revised 2017. Oslo, Norway. https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf.

Tsoukala V, Adoojan M, Dechamp J-F. 2018. Access to and Preservation of Scientific Information in Europe - Report on the implementation of Commission Recommendation C(2012) 4890 final. Luxembourg. https://op.europa.eu/s/ofxa.

Tsoukala V, Angelaki M, Kalaitzi V, Wessels B, Price L, Taylor MJ, Smallwood R, Linde P, Sondervan J, Reilly S, et al. 2015. Deliverable D5.1: Policy guidelines for open access and data dissemination and preservation. https://www.trilateralresearch.com/wp-content/uploads/2018/09/RECODE-D5.1-POLICY-RECOMMENDATIONS-_FINAL.pdf.

United Nations. 1982. United Nations Convention on the Law of the Sea. In: Third United Nations Conference on the Law of the Sea. Montego Bay, Jamaica: United Nations. https://treaties.un.org/pages/ViewDetailsIII.aspx?src=TREATY&mtdsg_no=XXI-6&chapter=21&Temp=mtdsg3&clang=_en.

United Nations. 1992. Convention on Biological Diversity. https://www.cbd.int/doc/legal/cbd-en.pdf.

United Nations. 2015. Resolution adopted by the General Assembly on 25 September 2015, 70/1. Transforming our world: the 2030 Agenda for Sustainable Development. A/RES/70/1. https://sustainabledevelopment.un.org/post2015/transformingourworld.

Vlaamse Regering. 2019. Decreet houdende wijziging van diverse decreten met betrekking tot het economisch, wetenschaps- en innovatiebeleid. Belgisch Staatsbl. https://codex.vlaanderen.be/Zoeken/Document.aspx?DID=1030522&param=inhoud&AID=1256828.

Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 3(160018):9. doi:10.1038/sdata.2016.18. https://www.nature.com/articles/sdata201618.

Working Group on Expanding Access to Published Research Findings. 2012. Accessibility, sustainability, excellence: how to expand access to research publications. https://www.sconul.ac.uk/sites/default/files/documents/finch-report-final.pdf.

## Annexes

## Annex A: European marine data landscape (infographic)

This infographic displays SeaDataNet, EMODnet and Copernicus Marine Service as the main components of the marine data landscape. It was published by Copernicus Marine Service, produced by EuroGOOS and designed by Visuality (EuroGOOS AISBL 2017).
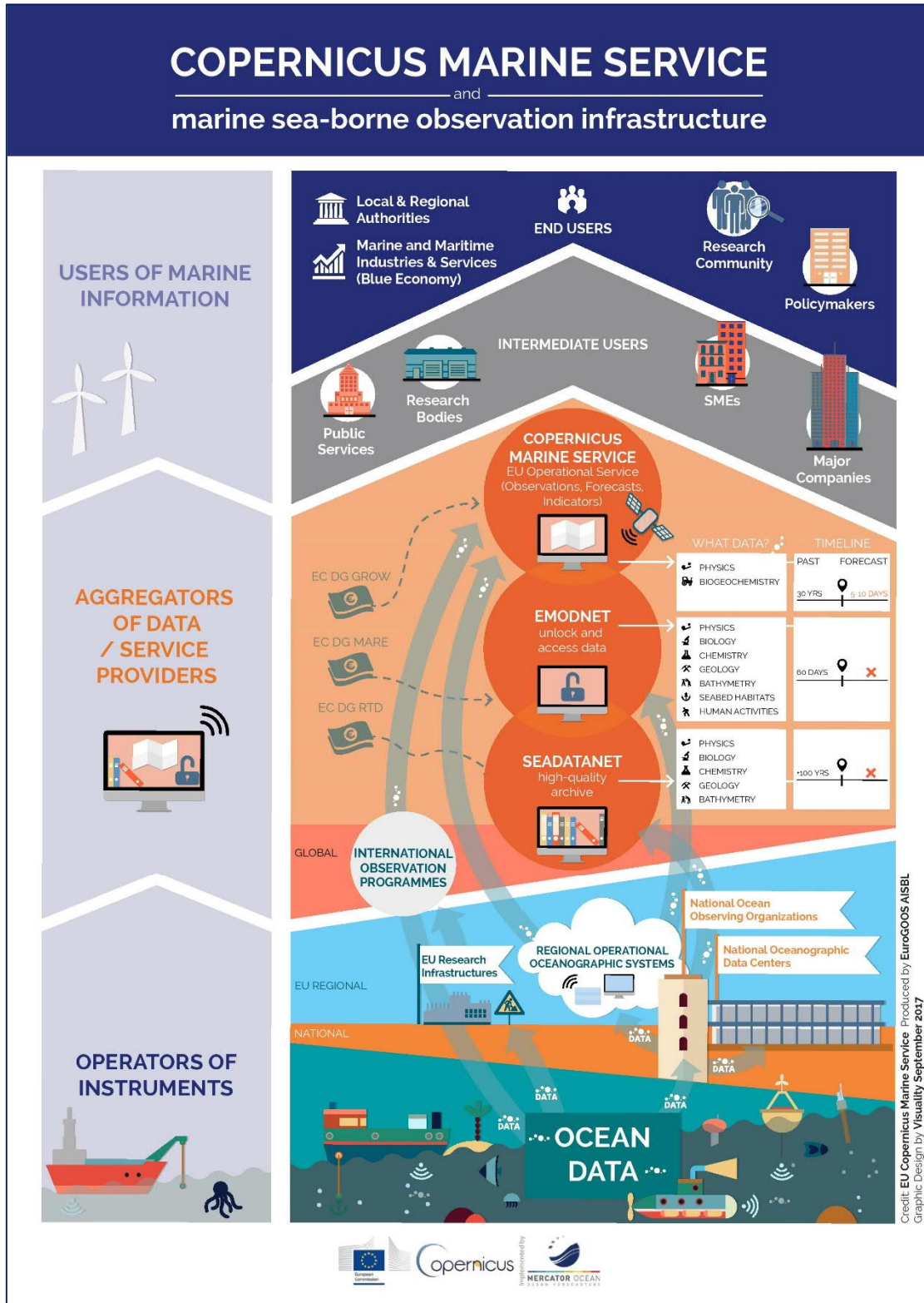
# BANOS CSA T4.6 questionnaire on Open Data

## 1 Context

This questionnaire is conducted as part of the Baltic and North Sea Coordination and Support Action (BANOS CSA). It seeks to collect the status and attitude towards open access and open research data of the research and innovation funders across the Baltic Sea and North Sea. Your responses will be taken into account for the development of the BANOS open access and open data strategies. This questionnaire is being distributed among the BANOS CSA Consortium Members and the research and innovation funding agencies of the countries represented in BANOS CSA.

## 2 Guidelines

- Please fill out the questionnaire by 22 January 2020 at the latest.
- Only one response per institution should be submitted. This submission is expected to reflect an institutionally-authorised response.
- For an overview of this survey, please find here a printable PDF version. In case you need to consult other persons or resources within your institution, we would suggest that you review the PDF version before you fill in this survey online, as only the online version can be used to submit your final answers.
- Please note that you will be able to go back and make changes to your answers before submission. However, your answers will not be saved when the questionnaire is closed.
- Please make sure to hit 'Submit' at the end of the survey to send your answers.
- After submitting the survey, a report of your answers will be emailed to you automatically.

## 3 Structure of the questionnaire

This questionnaire is structured in two parts:

> Part A – Respondent details
> Part B – Six main questions, about Open Access (1), Research Data Management (1), the Open Data Directive (3), and further suggestions (1)

## 4 Introductory information

For the purpose of this survey, please consider the following information:

Open access is understood as the practice of providing online access to research outputs free of charge for the end user and without restrictions on use and re-use beyond the possibility to require acknowledgement of authorship[1]. Open access is a core strategy of the European Commission to improve knowledge transfer. It is an important feature of Horizon 2020 and will most likely continue to be so in Horizon Europe.

Research data are defined as documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results[1]. Examples of research data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

Open access to research data refers to the right to access and re-use research data, under the specific terms and conditions set out by the data holder (e.g. proper acknowledgment of the data creator). Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge[2]. However, concerns in relation to privacy, protection of personal data, confidentiality, national security, legitimate commercial interests such as trade secrets, and to intellectual property rights of third parties, should be duly taken into account[1].

The Open Data Directive (Directive 2019/1024/EU)[1] is a recast of the PSI Directive[3] and its amendments[4]. It aims to enable full exploitation of information held by the public sector, including data gathered through publicly-funded research. Member States should transpose and enforce the Open Data Directive into national law by 17 July 2021. Of immediate relevance to this questionnaire are the following excerpts:

"Member States shall support the availability of research data [...] following the principle of 'open by default' and compatible with the FAIR principles. [...] Those open access policies shall be addressed to research performing organisations and research funding organisations." (Art. 10)

"The Commission shall adopt implementing acts laying down a list of specific high-value datasets [...] which shall be:

1. available free of charge [some exceptions apply];
2. machine readable;
3. provided via APIs; and
4. provided as a bulk download, where relevant." (Art. 14)

I.e. publicly funded research data should be FAIR and open by default, and data that belong to a specific selection of thematic categories, associated with important benefits for society, are subjected to additional requirements. Current categories are: geospatial; earth observation and environment; meteorological; statistics; companies and company ownership; mobility.

The FAIR Data Principles[5] are guiding principles for data management and stewardship to make data Findable, Accessible, Interoperable, Reusable.

---

## THE FAIR DATA PRINCIPLES

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Wilkinson et al. 2016

---

1. Directive 2019/1024/EU of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast) (OJ L 172, 26.6.2019, p. 56). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024.

2. DG RTD. 2017. H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020.
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

3. Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information (OJ L 345, 31.12.2003, p. 90).

4. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information (OJ L 175, 27.6.2013, p. 1).

5. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 3(160018):9. doi:10.1038/sdata.2016.18. https://www.nature.com/articles/sdata201618.

# 5    Disclaimer

Data and information provided are treated as institutionally-authorised submissions.

Personal information (i.e. name, function and email of respondent) will be kept confidential by the Flanders Marine Institute, Formas and BONUS EEIG. Personal information may be used by the Flanders Marine Institute, Formas and BONUS EEIG to follow up with the respondent in case further questions arise regarding the responses provided in the questionnaire. The respondent holds the right to request access, correction, deletion, and restriction of all processing of your personal information.

Information provided by the respondent in this questionnaire, with exception of the aforementioned personal information, will be shared with the BANOS CSA Consortium Members and may be publicly distributed as part of a deliverable or other form of publication by BANOS CSA.

# 6    Contact

Please send an email to Jonas Lescroart (jonas.lescroart@vliz.be) for further information and/or in case you have questions related to your submission.

## 7    Part A – Respondent details

Q1. Institution:

Q2. Full name of respondent:

Q3. Function of respondent:

Q4. Email:

## 8    Part B – Open Access, Research Data Management and the Open Data Directive

Q1. Does your institution have concrete plans to implement requirements stipulating immediate open access for scientific articles?

    OA1.              Such requirements are already in place

        Q1.1.   Please elaborate.

            OA1.1.         [Open field]

    OA2.               Yes, requirements in place from 2021 onwards (congruent with Plan S)

        Q1.2.   Please elaborate.

            OA2.1.         [Open field]

    OA3.               Yes, but timing is uncertain or projected to be later than 2021

        Q1.3.   Please elaborate.

            OA3.1.         [Open field]

    OA4.          No

        Q1.4.   Please elaborate.

            OA4.1.         [Open field]

Q2. Does your institution have a policy that deals with research data?

    A3. Yes

        Q2.1.   Does it cover aspects of data sharing/open data?

            A3.1.         Yes

            A3.2.         No

        Q2.2.   Does it include the use of Data Management Plans (DMPs) by researchers and/or beneficiaries?

            A3.3.         Yes, DMPs are mandatory

            A3.4.         Yes, DMPs are encouraged

            A3.5.         No

        Q2.3.   If possible, please provide a link to that policy

            A3.6.         [Open field]

    A4. Not yet, we are currently developing one

        Q2.4.   Will it cover aspects of data sharing/open data?

            A4.1.         Yes

            A4.2.         No

        Q2.5.   Will it include the use of Data Management Plans (DMPs) by researchers and/or beneficiaries?

            A4.3.         Yes, DMPs will be mandatory

            A4.4.         Yes, DMPs will be encouraged

            A4.5.         No

            A4.6.         Too soon to tell

        Q2.6.   If possible, please provide a link to that policy

            A4.7.         [Open field]

    A5. No, we follow a higher-level policy (e.g. national)

        Q2.7.   At what level is that policy enforced?

            A5.1.         [Open field]

        Q2.8.   Does it cover aspects of data sharing/open data?

            A5.2.         Yes

            A5.3.         No

        Q2.9.   Does it include the use of DMPs by researchers and/or beneficiaries?

            A5.4.         Yes, DMPs are mandatory

            A5.5.         Yes, DMPs are encouraged

            A5.6.         No

        Q2.10. If possible, please provide a link to that policy

|        |       | A5.7. | [Open field] |
|---|---|---|---|

A6. No

    Q2.11. Why not?

        A6.1.        Not necessary

        A6.2.        Not considered

        A6.3.        Under discussion

        A6.4.        Other + [open field]


Q3. Is your institution currently compliant with the Open Data Directive where it concerns research data?

    A7. Yes

    A8. No, but steps are being taken (at a minimum, the Directive has entered discussion)

        Q3.1.    What is the current status?

            A8.1.        The Directive has entered discussion

            A8.2.        Awaiting higher-level decisions (e.g. national)

            A8.3.        New policy is being drafted

            A8.4.        Other + [Open field]

        Q3.2.    Are there any of the following aspects of the Directive that your institution would rather not have to comply with when it comes to publicly-funded research data (for any reason; considerations might be financial, technical, administrative...)?

            A8.5.        No

            A8.6.        Too soon to tell

            A8.7.        Principle of 'open by default'*

            A8.8.        Requirement to be compatible with FAIR principles

             A8.9.        High-value datasets: to provide via APIs

            A8.10.       High-value datasets: to provide free of charge

            A8.11.       High-value datasets: must be machine-readable/interoperable

            A8.12.       High-value datasets: to provide as bulk download

            A8.13.       Other... + [Open field with placeholder text 'pertaining to research data']

    A9. No

        Q3.3.    Are there any of the following aspects of the Directive that your institution would rather not have to comply with when it comes to publicly-funded research data (for any reason; considerations might be financial, technical, administrative...)?

            A9.1.        No

            A9.2.        Too soon to tell

            A9.3.        Principle of 'open by default'*

            A9.4.        Requirement to be compatible with FAIR principles

            A9.5.        High-value datasets: to provide via APIs

            A9.6.        High-value datasets: to provide free of charge

            A9.7.        High-value datasets: must be machine-readable/interoperable

            A9.8.        High-value datasets: to provide as bulk download

            A9.9.        Other... + [Open field with placeholder text 'pertaining to research data']

* Taking into account concerns relating to intellectual propery rights, personal data protection and confidentiality, security and legitimate commercial interests.


Q4. Are there any ways in which your institution goes beyond the minimum regulations outlined in the Open Data Directive, or intends to do so?

        A10.        None

        A11.        No formal application is necessary to access the data

        A12.        Default use of the most open licenses available (CC BY, CC0 or comparable), even when a stricter license would be reasonable under the Open Data Directive

A13.  Extension of the 'open by default' and FAIR principles to the workflows, scripts… used to process research data

A14.  Publication of all data collected during research, not just the data necessary to validate published research results

A15.  Extension of high-value dataset requirements to other data: to provide free of charge, even when a charge could be incurred under the Open Data Directive

A16.  Extension of high-value dataset requirements to other data: to provide via APIs

A17.  Extension of high-value dataset requirements to other data: must be machine-readable/interoperable

A18.  Extension of high-value dataset requirements to other data: to provide as bulk download

A19.  Other + [Open field]

Q5. Is your institution, as a BANOS CSA Consortium Member and/or as a potential participant of the future BANOS programme, of the opinion that the BANOS programme should be compliant with the Open Data Directive where it concerns research data?

 A20.  Yes, fully

  Q5.1. Are there aspects where the BANOS programme should go beyond the Open Data Directive?

   A20.1.  None

   A20.2.  No formal application is necessary to access the data

   A20.3.  Default use of the most open licenses available (CC BY, CC0 or comparable), even when a stricter license would be reasonable under the Open Data Directive

   A20.4.  Extension of the 'open by default' and FAIR principles to the workflows, scripts… used to process research data

   A20.5.  Publication of all data collected during research, not just the data necessary to validate published research results

   A20.6.  Extension of high-value dataset requirements to other data: to provide free of charge, even when a charge could be incurred under the Open Data Directive

   A20.7.  Extension of high-value dataset requirements to other data: to provide via APIs

   A20.8.  Extension of high-value dataset requirements to other data: must be machine-readable/interoperable

   A20.9.  Extension of high-value dataset requirements to other data: to provide as bulk download

   A20.10.  Other + [Open field]

 A21.  Partly/maybe

  Q5.2. Please elaborate.

   A21.1.  [Open field]

 A22.  No

  Q5.3. Why not? Please elaborate.

   A22.1.  [Open field]

Q6. Please add any further comments, recommendations, or sources, such as initiatives or policies, that should be taken into account for the BANOS strategies regarding open data and open access.

 A23.  [Open field]

# Annex C: List of institutions targeted for questionnaire

Institutions that responded to the questionnaire are displayed in bold. For short descriptions of the national funding agencies, see Annex 3 of the BANOS CSA report D2.2 (ANR and BONUS EEIG 2019).

| Country | Acronym | Institution | Relation to BANOS CSA |
|---------|---------|-------------|------------------------|
| Belgium | BELSPO | Belgian Science Policy Office | National funding agency |
| Belgium | VLAIO | Flanders Innovation & Enterpreneurship | National funding agency |
| Belgium | VLIZ | Flanders Marine Institute | Consortium member |
| Belgium | F.R.S.–FNRS | National Fund for Scientific Research | National funding agency |
| Belgium | FWO | Research Foundation - Flanders | National funding agency |
| Denmark | IFD | Innovation Fund Denmark | Consortium member |
| Estonia | EAS | Enterprise Estonia | National funding agency |
| Estonia | KIK | Environment Investment Centre | National funding agency |
| Estonia | ETAG | Estonian Research Council | Consortium member |
| Estonia | | Ministry of Rural Affairs | National funding agency |
| European | JPI Oceans | Joint Programming Initiative Healthy and Productive Seas and Oceans | Strategic partner |
| Finland | AKA | Academy of Finland | Observer |
| Finland | MMM | Ministry of Agriculture and Forestry | National funding agency |
| France | ADEME | Agency for Environment and Energy Management | National funding agency |
| France | AFD | French Development Agency | National funding agency |
| France | Ifremer | French Research Institute for Exploitation of the Sea | Consortium member |
| France | ANR | National Research Agency | Consortium member |
| Germany | BMWi | Federal Ministry for Economic Affairs and Energy | National funding agency |
| Germany | BMBF | Federal Ministry of Education and Research | National funding agency |
| Germany | BMEL | Federal Ministry of Food and Agriculture | National funding agency |
| Germany | Jülich | Research Centre Jülich | Consortium member |
| Latvia | LIAA | Investment and Development Agency of Latvia | National funding agency |
| Latvia | | Ministry of Education and Science | National funding agency |
| Latvia | SEDA | State Education Development Agency | Consortium member |
| Lithuania | MITA | Agency for Science, Innovation and Technology | National funding agency |
| Lithuania | ŽŪM | Ministry of Agriculture | National funding agency |
| Lithuania | LMT | Research Council of Lithuania | Consortium member |
| Macroregional | HELCOM | Baltic Marine Environment Protection Commission - Helsinki Commission | Strategic partner |
| Macroregional | BONUS EEIG | Baltic Organisations Network for Funding Science | Consortium member |
| Macroregional | OSPAR | Convention for the Protection of the Marine Environment of the North-East Atlantic | Strategic partner |
| Macroregional | ICES | International Council for the Exploration of the Sea | Strategic partner |
| Netherlands | RVO | Netherlands Enterprise Agency | National funding agency |

| | | | |
|---|---|---|---|
| Netherlands | NWO | Netherlands Organisation for Scientific Research | Consortium member |
| Norway | RCN | Research Council of Norway | Consortium member |
| Poland | NCBR | National Centre for Research and Development | Consortium member |
| Poland | NCN | National Science Centre | National funding agency |
| Sweden | VINNOVA | Sweden's Innovation Agency | National funding agency |
| Sweden | SwAM | Swedish Agency for Marine and Water Management | Consortium member |
| Sweden | SEPA | Swedish Environmental Protection Agency | National funding agency |
| Sweden | Mistra | Swedish Foundation for Strategic Environmental Research | National funding agency |
| Sweden | VR | Swedish Research Council | National funding agency |
| Sweden | Formas | Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning | Consortium member |
| UK | BBSRC | Biotechnology and Biological Sciences Research Council | National funding agency |
| UK | | British Academy | National funding agency |
| UK | Cefas | Centre for Environment, Fisheries and Aquaculture Science | Consortium member |
| UK | BEIS | Department for Business, Energy & Industrial Strategy | National funding agency |
| UK | DEFRA | Department for Environment, Food & Rural Affairs | National funding agency |
| UK | DFID | Department for International Development | National funding agency |
| UK | EPSRC | Engineering and Physical Sciences Research Council | National funding agency |
| UK | | Innovate UK | National funding agency |
| UK | Invest NI | Invest Northern Ireland | National funding agency |
| UK | NERC | Natural Environment Research Council | National funding agency |
| UK | STFC | Science and Technology Facilities Council | National funding agency |
| UK | Marine Scotland | Scottish Government, Marine Scotland Directorate | National funding agency |
| UK | UKRI | UK Research and Innovation | National funding agency |

## Annex D: Impetus for list on recommended data repositories

Potential data repositories for redistribution of the thematic data types based on ASSEMBLE Plus (D4.1, 2018).

| | Thematic datatype category | Research domains | Potential data repositories for redistribution | URLs |
|---|---|---|---|---|
| Biology | Biodiversity | biogeographic research, species traits, taxonomy | Global Biodiversity Information Facility (GBIF) | http://www.gbif.org |
| | | | Biology Portal of the European Marine Observation and Data Network (EMODnet Biology) | www.emodnet.eu/biology |
| | | | World Register of Marine Species (WoRMS) | www.marinespecies.org |
| | | | Morphobank.org | www.morphobank.org |
| | Genomic | sequencing (DNA, RNA), annotation of features, protein structural information, gene expression profiles, alignment data, chromosomal mapping, phylogenetic trees, Single | European Nucleotide Archive (ENA) | https://www.ebi.ac.uk/ena |
| | | | GenBank | https://www.ncbi.nlm.nih.gov/genbank/ |
| | | | dbSNP | https://www.ncbi.nlm.nih.gov/snp |
| | | | European Variation Archive (EVA) | https://www.ebi.ac.uk/eva/ |
| | | Nucleotide Polymorphisms (SNPs), functional genomics, metabolomics, proteomics, environmental DNA | dbVar | https://www.ncbi.nlm.nih.gov/dbvar |
| | | | Database of Genomic Variants Archive (DGVa) | https://www.ebi.ac.uk/ dgva |
| | | | EBI Metagenomics | https://www.ebi.ac.uk/metagenomics/ |

| | | | | NCBI Trace Archive | https://www.ncbi.nlm.nih.gov/ Traces/home/ |
|---|---|---|---|---|---|
| | | | | NCBI Sequence Read Archive (SRA) | https://www.ncbi.nlm.nih.gov/sra |
| | | | | NCBI Reference Sequence Database (RefSeq) | https://www.ncbi.nlm.nih.gov/refseq/ |
| | | | | Entrez Nucleotide | https://www.ncbi.nlm.nih.gov/nucleotide |
| | | | | UniGene | https://www.ncbi.nlm.nih.gov/unigene |
| | | | | HomoloGene | https://www.ncbi.nlm.nih.gov/homologene |
| | | | | Protein Information Resource (PIR) | https:// pir.georgetown.edu/ |
| | | | | Protein Data Bank (PDB) | https://www.wwpdb.org |
| | | | | Protein Circular Dichroism Data Bank (PCDDB) | pcddb.cryst.bbk.ac.uk/ home.php |
| | | | | ArrayExpress | https://www.ebi.ac.uk/arrayexpress/ |
| | | | | Gene Expression Omnibus (GEO) | https://www.ncbi.nlm.nih.gov/geo/ |
| | | | | GenomeRNAi | www.genomernai.org |
| | | | | dbGAP | https://www.ncbi.nlm.nih.gov/gap |
| | | | | The European Genome-phenome Archive (EGA) | https://www.ebi.ac.uk/ega/ |
| | | | | Database of Interacting Proteins (DIP) | http://dip.doe-mbi.ucla.edu/dip/Main.cgi |

| | | | IntAct | https://www.ebi.ac.uk/ intact/ |
|---|---|---|---|---|
| | | | Biological General Repository for Interaction Datasets | https:// thebiogrid.org |
| | Imaging | zooscan images, flowcam images, flow cytometry images, VPRvideos | EcoTaxa | ecotaxa.obs-vlfr.fr |
| | | | FlowRepository | flowrepository.org |
| | Biogeochemical | biochemical pathways, nutrients | Data Exchange Portal of MPI | https://www.bgc-jena.mpg.de/geodb/projects/Home.php |
| | Experimental | data resulting from lab experiments | | |
| Oceanographical | Physical | seawater temperature, salinity, ocean currents, waves, geospatial | The Physical Portal of the European Marine Observation and Data Network (EMODnet Physics) | www.emodnet-physics.eu/map/ |
| | Chemical | pollution, heavy metals | The Chemical Portal of the European Marine Observation and Data Network (EMODnet Chemistry) | www.emodnet-chemistry.eu/welcome |

| | | | | |
|---|---|---|---|---|
| Climatological | Modelling | climate models | BioModels Database | https://www.ebi.ac.uk/ biomodels-main/ |
| | | | Kinetic Models of Biological Systems (KiMoSys) | https://kimosys.org |
| Technical | Instrumental | experimental instruments, laboratory set-ups | | |