Mission Microbiomes AtlantECO
Data Sharing & Publication Best Practices

Definitions

"MMA" or "the mission" means Mission Microbiomes AtlantECO

"Partner(s)" means individual(s) from institutions that are party to the AtlantECO project Consortium Agreement and individual(s) involved in the production and exploitation of MMA Results.

"MMA Result(s)" means samples, derived samples, primary data and derived data resulting from Mission Microbiomes AtlantECO. MMA Results are jointly owned by the Partners.

Context

Mission Microbiomes AtlantECO (MMA) is an international study of the most fundamental fabric of the ocean — the ocean microbiome — aiming to understand its structure, functioning and connectivity in the Atlantic Ocean. It is an initiative of the European Union's research and innovation project AtlantECO, which aims at sharing capacity along and across the Atlantic, and providing knowledge-based resources to help design policies for the management and protection of Atlantic Ecosystem Services. The success and legacy of MMA on science and society relies on the fair, open and inclusive collaboration among its Partners, and the exploitation of jointly owned MMA Results. The adoption of the present Data Sharing and Publication Best Practices is therefore essential to achieve the scientific objectives of MMA and maximise their impacts.

Individual Commitment

We ask everyone who took part in the planning and/or execution of the field expedition and everyone who will contribute to the analysis and/or scientific work that will follow, to adopt the best practices described in this document. Anyone interested to collaborate in the future will also be asked to adopt the best practices. The list of individuals who adopt the best practices will be available publicly and will be used in the implementation of the best practices.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Table of Contents

Mission Microbiomes AtlantECO 1
- Definitions 1
- Context 1
- Individual Commitment 1
- Table of Contents 2

Publication Best Practices 3
- Early notification - exploitation of results 3
- Intellectual Property Rights (IPR) 5
- Access & Benefits Sharing (ABS) 5
- Open Access (OA) 6
- Co-authorship 6
- Acknowledgements 8

Data Sharing Best Practices 10
- Data Management Plan 10
- Early deposition 10
- Early notification - release of data 10
- Primary data 11
- Annotated data 12
- Aggregated data 13
- Data Hub 14

ANNEX I 15

Relevant Articles from the AtlantECO Grant Agreement 15
- ARTICLE 28.1 “Obligation to exploit the results” 15
- ARTICLE 29.1 “Obligation to disseminate results” 15
- ARTICLE 29.2 “Open access to publications” 15
- ARTICLE 29.3 “Open access to research data” 16

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author’s view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Publication Best Practices

Early notification - exploitation of results

Until 31 August 2029, date by which Partners have to exploit their MMA Results according to the AtlantECO project Grant Agreement (Annex I), Partners wishing to use MMA Results for the purpose of a planned communication or publication are required to send an summary/abstract of the proposed work/manuscript to all Partners (Figure 1), including a running title and a tentative list of co-workers/co-authors. An early notification of intent should be sent by email (Box 1) preferably before starting the proposed work, in which case Partners interested in contributing to the proposed work are encouraged to contact the proposing Partner directly. Otherwise, the notification must be sent no later than 45 calendar days before the work is first submitted to a journal or a conference, in which case the full text should be circulated to all Partners.

Any objection to the proposed work should be made in writing to the Coordinator (atlanteco-coordteam@googlegroups.com). If no objection is received within 30 calendar days after the early notification, the proposed work is deemed approved by all Partners. Once an objection has been raised, the involved Partners shall discuss how to resolve it on a timely basis, for example by amending the proposed publication. The Coordinator will mediate the discussion for a maximum period of 90 days, and the objecting Party shall not unreasonably continue the opposition if appropriate measures are taken following the discussion. If the objection cannot be resolved amicably, the decision shall go to vote by the Parties, following rules described in Section 6.1.3 of the AtlantECO Consortium Agreement, and any resulting dispute will be settled according to Section 11.8 of the AtlantECO Consortium Agreement.

In the case of a publication, the proposing Partner is required to inform all Partners (atlanteco-general@googlegroups.com, atlanteco-mma@googlegroups.com) once the manuscript is submitted, and once again when it is published. Following submission, and preferably before submission, the Coordinator is responsible for ensuring that obligations regarding open access, co-authorship and acknowledgements are fulfilled. These obligations are detailed in the following sections.

The Coordinator and the communication partner of AtlantECO (Tara Ocean Foundation) are responsible for promoting publications using the AtlantECO website, social media, and press releases as appropriate. OpenAire is monitoring all publications that acknowledge project AtlantECO, including those that use MMA Results, and provides a bibliographic portal with linked access and bibliometrics (OpenAIRE AtlantECO).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

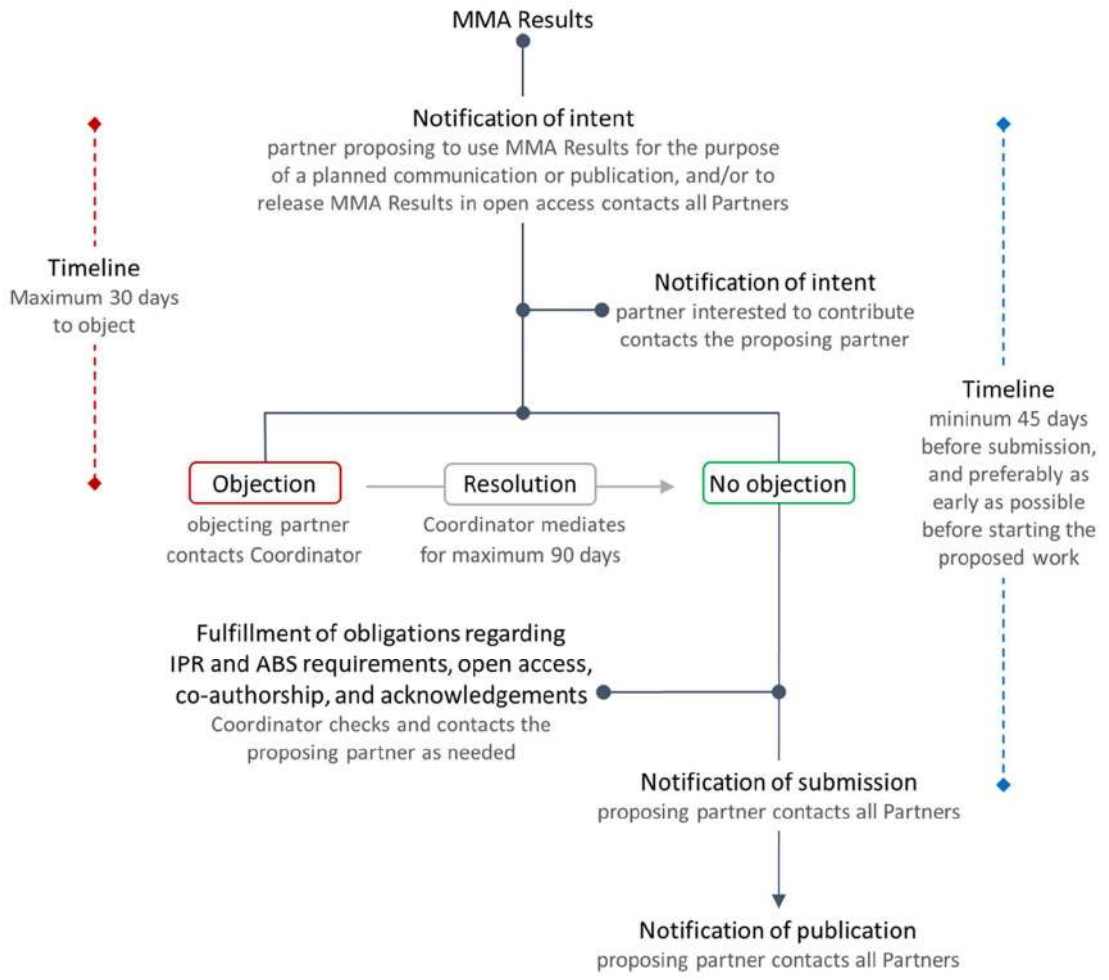

Figure 1. *Summary of the steps, timelines and decision process that follow the early notification for the release of data and the intent to use results for the purpose of a planned communication or publication.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Box 1. Email template for the early notification of intent to use MMA Results

To: atlanteco-general@googlegroups.com, atlanteco-mma@googlegroups.com

Dear Mission Microbiomes AtlantECO Partners,

We are proposing to use jointly owned results from Mission Microbiomes AtlantECO for the work described in the attached document, which we intend to submit for communication/publication in [insert journal or conference name] on [insert timeline]. If you are interested in contributing to the proposed work please contact us directly by return of email.

Partners are provided 30 days to object to the proposed work or to its publication. Objections are justified if:

- *the protection of the objecting partner's results or background would be adversely affected,*
- *the objecting partner's legitimate interests in relation to the results or background would be significantly harmed, or/and*
- *the proposed publication includes confidential Information of the objecting partner.*

Please submit any justified objections to the AtlantECO project coordinator (atlanteco-coordteam@googlegroups.com) together with proposed steps towards agreement within 30 days, so before [insert date].

If no objections are received, we assume that all partners agree with the publication of these results.

Best regards,

Intellectual Property Rights (IPR)

Partners are entitled to use any MMA Results for non-commercial research activities on a royalty-free basis, and to grant non-exclusive licences to third parties, without any right to sub-license. As a necessary protection measure, those intellectual property rights are subject to, and may be contested through the early notification process described above (Figure 1).

Access & Benefits Sharing (ABS)

MMA Results, especially Genetic Resources, fall under the United Nations Convention of Biological Diversity (CBD), more specifically under the Nagoya Protocol on Access and Benefit Sharing (ABS). Partners must comply with the conditions set forth by the coastal States prior to sampling, following the United Nations Convention on Law Of the Sea (UNCLOS) regulations.

The transfer of samples or derived samples (part of MMA Results) among Partners shall be formalised by a Material Transfer Agreement (MTA), especially when involving Genetic Resources. The aim of the MTA is to ensure that (1) the intended use is clearly communicated, (2) samples and derived samples are handled in a safe and responsible manner, especially if the intended use is destructive, and (3) obligations resulting from the CDB, ABS and UNCLOS regulations are communicated and respected.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

The AtlantECO Coordinator is responsible for providing all Partners with (1) links to State-specific regulations on the CBD Access and Benefit-Sharing Clearing-house, (2) all written consent documents obtained from the coastal States, as well as (3) a comprehensive list of State-specific requirements at the granularity of each sample collected on board SV Tara during the field expedition of Mission Microbiomes AtlantECO.

In case the exploitation of MMA Results involves non-commercial or commercial use of Genetic Resources that exceeds the scope of the State-specific requirements, Partners must contact the AtlantECO Coordinator and shall contact the relevant State authority(ies) in order to obtain the right for such additional use.

Open Access (OA)

Partners must ensure open access (i.e. free of charge online access for any user) to all communications and publications relating to the MMA Results. Data supporting any publication (i.e. in closed or open access in a peer reviewed journal, a preprint server or any public repository) must be available in open access before the date of publication.

Partners are encouraged to publish in peer-reviewed journals that offer gold open access, which often involves a publication charge. The AtlantECO project has a dedicated budget to support the publication of collaborative and multidisciplinary work in open access journals. Requests to use the AtlantECO open access budget should be made in writing to the Coordinator of project AtlantECO (atlanteco@szn.it). Partners are also encouraged to consider publishing in the European Commission's open peer-reviewed journal Open Research Europe, which offers gold open access, free of charge.

Partners are encouraged to deposit preprints (before peer-review) and postprints / pre-proofs (peer-reviewed, author accepted manuscript) in public repositories if green open access is the preferred option. We recommend self depositing preprints and postprints in the BioRxiv server, or in the European Commission's free repository Zenodo, as part of the AtlantECO community.

Co-authorship

We strongly encourage providing publishers with the ORCID of all authors. The composition and order of authors on a publication that uses MMA Results is as follows:

- ***First author(s)*** – We encourage early stage researchers to lead publications.
- ***Named authors list #1*** – includes anyone whose contribution to either the sampling effort, data generation, analysis and interpretation of MMA Results, or writing the manuscript was essential to publish the work. The composition and order of co-authors in this list is determined by the first author(s) and the Named authors list #2.
- ***Collective author "Mission Microbiomes AtlantECO"*** – is included as co-author when the publication is either (1) the first use of MMA Results from a large fraction (e.g. >50%) of samples collected from any protocol, from a large biogeographic region or from a MMA topical study, or (2) a collaboration involving a large proportion (e.g. >50%) of Partners are named authors. The composition of the collective author will vary from case to case by combining pre-determined lists of contributors (see below).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

- ***Named authors list #2*** – includes senior scientists who contributed particularly to supervising the work, developing its concept, and/or writing the paper.

The collective author "Mission Microbiomes AtlantECO" is used only if the publisher is able to index all Partners listed in the collective author as authors of the paper even though they do not appear as named author. This usually involves a note in the list of the authors affiliation, indicating that, for example, "the Mission Microbiomes AtlantECO Partners and their affiliations are listed at the end of this manuscript". The list usually appears before the acknowledgements and Partners are ordered alphabetically. Partners who contributed more specifically to the paper can also be named individually in the contributing authors lists #1 or #2, whichever is more appropriate. The publisher is responsible for deduplicating authors in the different lists so that they are not indexed more than once. In case the publisher cannot index individuals that are part of a collective author name, all individuals must appear as named authors in alphabetical order.

Predetermined lists of collective co-authors are provided on the MMA community page at Zenodo (https://zenodo.org/communities/mission-microbiomes-atlanteco/), together with up-to-date affiliations and ORCIDs. The lists correspond to ten categories that can be combined according to the scope of the publication:

1. *Co-authors who contributed to the scientific coordination & logistics of MMA: 20 persons*
2. *Co-authors who contributed to the analysis of samples or sensor data from MMA: 34 persons*
3. *Co-authors who contributed to the topical study Amazon river plume: x persons*
4. *Co-authors who contributed to the topical study Trinidad seamounts biodiscovery: x persons*
5. *Co-authors who contributed to the topical study Coccolithophore bloom: x persons*
6. *Co-authors who contributed to the topical study Weddell Sea & iceberg: x persons*
7. *Co-authors who contributed to the topical study Sub-mesoscale: x persons*
8. *Co-authors who contributed to the topical study Benguela upwelling region: x persons*
9. *Co-authors who contributed to the topical study Senegal upwelling region: x persons*
10. *Co-authors who contributed to the topical study African river plumes: x persons*

Here, we provide examples of cases that require a collective co-author, and how lists can be combined:

- A publication that uses MMA Results from a specific set of protocols, covering a large biogeographic region (e.g. the South Atlantic) would include list #1 only.
- A publication that uses MMA Results from a broad set of protocols, covering a large biogeographic region (e.g. the west coast of Africa) would include lists #1 and #2
- A publication that uses MMA Results from a broad set of protocols, covering the Benguela upwelling region only would include lists #1, #2 and #8.

The proposed use (or not) of a collective author should be communicated to the Partners as part of the notification of intent to publish (Figure 1, Box 1), so that it can be agreed early on by all Partners,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

and individuals who are part of the collective author can engage as co-author. We recommend that you contact the coordinator (atlanteco-coordteam@googlegroups.com) before sending the early notification in order to discuss the possible co-authorship options.

Acknowledgements

We provide three acknowledgements that must be included in all publications that use MMA Results. These should of course be combined with additional, more specific acknowledgements, but we kindly ask to start with Acknowledgement #1. The list of those who participated in Mission Microbiomes AtlantECO in various capacities is given in Box 2. The first author(s) may decide at their own discretion to include any of the proposed lists in the acknowledgement section of the publication, for example when the collective author "Mission Microbiomes AtlantECO" is not used.

Acknowledgement #1 - "We wish to thank the Tara Ocean Foundation, the SV Tara crew and all those who participate in Mission Microbiomes AtlantECO and adopt its Data Sharing & Publication Best Practices (https://zenodo.org/communities/mission-microbiomes-atlanteco/)."

Acknowledgement #2 - "This publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923 (project AtlantECO). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein."

Acknowledgement #3 - "We are keen to thank the commitment of the following institutions for their financial and scientific support that made Mission Microbiomes AtlantECO possible: Stazione Zoologica Anton Dohrn, European Bioinformatics Institute (EMBL-EBI), Centre national de la recherche scientifique (CNRS), Centre National de Séquençage (CNS, Genoscope), agnès b., BIC, Capgemini Engineering, Fondation Groupe EDF, Compagnie Nationale du Rhône, L'Oréal, Biotherm, Région Bretagne, Lorient Agglomération, Billerudkorsnas, Havas Paris, Fondation Rothschild, Office Français de la Biodiversité, AmerisourceBergen, Philgood Foundation, UNESCO-IOC, Etienne Bourgois."

Box 2. Mission Microbiomes AtlantECO participants

We acknowledge the contribution of individuals involved in the logistics & scientific coordination of Mission Microbiomes AtlantECO: *Aliénor Bourdais, Chris Bowler, Clementine Moulin, Colomban de Vargas, Daniele Iudicone, Elisabeth Catafort, Emmanuel Boss, Emmanuelle Petit, Eric Mayeux, Etienne Bourgois, Fabien Lombard, Juliette Schramm, Lionel Guidi, Miguel Moll, Patrick Wincker, Romain Troublé, Sabrina Speich, Sophie Sanchez, Stéphane Pesant, Thomas Linkowski*

We acknowledge the contribution of scientists involved in the integrated analysis of samples or sensor data from Mission Microbiomes AtlantECO: *Alexandra Ter Halle, Céline Dimier, Daniela Banaru, Darshika Manral, Emmanuel Boss, Fabien Lombard, Flora Vincent, Florian Kokoszka, Georg Pohnert, Gilles Reverdin, Hugo Sarmento, Ian Probert, Jean-François Ghiglione, Jochen Horstmann, Josep Gasol, Lars-Eric Heimbürger, Linda Amaral Zettler, Lionel Guidi, Mak Saito, Matt Sullivan, Michel Flores, Mireille Pujo Pay, Naiara Rodriguez-Ezpeleta, Nicolas Metzl, Pedro Oliviera, Paulo Calil, Rainer Pepperkok, Rémi Laxenaire, Rubens Lopes, Seth John, Shaun Dayzel, Stéphane Pesant, Thulani Makhalanyane, Ulisse Cardini*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

We acknowledge the contribution of scientists who participated in the fieldwork on board SV Tara: *Alessandra Gomes, Alessandro Tagliabue, Alison Chase, Andrea Freire, Andrea Green Koettker, Ange Bouramanding Diedhiou, Antonella Ruggiero, Céline Dimier, Charlotte Begouen, Chris Bowler, Clara Trellu, Constanze Kuhlisch, Cora Hörstmann, Damaria Ardène Boussiengue, David Leistenschneider, Douglas Couet, Edouard Lavergne, Emma Rocke, Emmanuel Boss, Eric Pelletier, Erica Caroline Becker, Flora Vincent, Giancarlo Bachi, Gleice Sousa Santos, Guillaume Bourdin, Helena Cruz de Carvalho, Isabelle Calves, Jean-François Ghiglione, Josep Erta, Karine Lebaron, Léa Olivier, Lee Karp-Boss, Leïla Meitertzheim, Mancha Mabaso, Mathilde Bourreau, Michel Flores, Miléna Cerda, Morgane Ratin, Nastassia Patin, Ndamonoghenda Mateus, Nicole Dames, Noé Poffa, Paula Huber, Pedro Junger, Rémi Laxenaire, Ricardo Silva, Roland Ngomo, Samuel Chaffron, Stéphane Pesant, Suzana Nicolaus, Thomas Leeuw, Thomas Linkowski, Thulani Makhalanyane, Xiomara Garcia Diaz*

We acknowledge the contribution of the crew, media and artists who participated in the fieldwork on board SV Tara: *Alain Prieur, Antoine Bertin, Arthur Larie, Baptiste Regnier, Carole Pire, David Monmarché, Edson Macalini, François Aurat, Guilia Grossman, Guillaume Tauran, Hugo Viel, Irene Kopelman, Julie Nedelec-Andrade, Kimerudi Motswai, Lara Tabet, Laurent Rogniaux, Léo Boulon, Leslie Moquin, Loïc Caudan, Louis Wilmotte, Maeva Bardy, Manon Lanjouère, Martin Hertau, Martin Leroux, Mathieu Oriot, Maxime Horlaville, Nicolas Bin, Nicolas Michel, Samuel Audrain, Sophie Bin, Vincent Jolly, Yves Tourmon*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Data Sharing Best Practices

Data Management Plan

The AtlantECO project, and hence its flagship Mission Microbiomes AtlantECO, adopt the FAIR Guiding Principles for scientific data management and stewardship, making data Findable, Accessible, Interoperable and Reusable. The Data Management Plan (DMP) of the AtlantECO project sets out best practices aimed at sharing and integrating six types of data that are generated through the project's augmented observation activities, i.e. provenance, environmental, genomics, proteomics, metabolomics and imaging data. The best practices also address the workflows by which data are curated, analysed and archived throughout their life cycle as primary data, annotated data and aggregated data. The present document details the best practices set out in the DMP.

Early deposition

The AtlantECO project requires the early deposition of primary, annotated and aggregated data in the selected permanent archives following their analytical and technical validation, which must be completed within three months after the analysis is performed. Data shall be deposited in pre-publication, closed access mode, unless open access is otherwise agreed following the early notification process (Figure 1)

Early notification - release of data

As a guiding principle, jointly owned data (part of MMA Results) shall be made available in open access as soon as possible and no later than 31 August 2029, date by which Partners have to exploit their Results according to ARTICLE 28.1 of the AtlantECO project Grant Agreement (Annex I). The release of jointly owned data in open access before that deadline requires the approval of Partners (Figure 1), following the rules described in ARTICLE 29.1 of the AtlantECO Grant Agreement (Annex I). A notification of intent must be sent by email (Box 3) as early as possible, and no later than 45 calendar days before the data is released in open access.

As a best practice, Partners agree with the following rules that support the early production of provenance and environmental context data:

- Provenance (meta)data will be deposited in open access as soon as they are available, and will be augmented periodically with additional provenance and environmental context.
- All primary environmental data from Mission Microbiomes AtlantECO will be described in a single publication submitted to a data journal, and granted open access at the time of publication.
- Parts of the primary environmental data may be used to generate environmental context that will be made available in open access at BioSamples together with sample provenance (meta)data as early as possible, before the primary environmental data are published in open access.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Box 3. Email template for the early notification of intent to release MMA Results (data)

To: atlanteco-general@googlegroups.com, atlanteco-mma@googlegroups.com

Dear Mission Microbiomes AtlantECO Partners,

We are proposing to share the jointly owned data from Mission Microbiomes AtlantECO described in the attached document, which we intend to make available in Open Access in [insert data archive name] on [insert timeline].

Partners are provided 30 days to object to the release of these results. Objections are justified if:

- *the protection of the objecting partner's results or background would be adversely affected,*
- *the objecting partner's legitimate interests in relation to the results or background would be significantly harmed, or/and*
- *the proposed publication includes confidential Information of the objecting partner.*

Please submit any justified objections to the AtlantECO project coordinator (atlanteco-coordteam@googlegroups.com) together with proposed steps towards agreement within 30 days, so before [insert date].

If no objections are received, we assume that all partners agree with the release of these jointly owned data in open access.

Primary data

Specific archives were selected for the deposition and preservation of six types of data generated by the AtlantECO project (Figure 2). Once deposited in the selected archives, primary data will be curated according to community standards and assigned unique accession numbers.

Permanent archives selected for the different data types:

- EMODnet for environmental data (https://emodnet.eu/en)
- ENA for genomics data (http://www.ebi.ac.uk/ena)
- PRIDE for proteomics data (https://www.ebi.ac.uk/pride/)
- Metabolights for metabolomics data (https://www.ebi.ac.uk/metabolights/)
- BioImage Archive for imaging data (https://www.ebi.ac.uk/bioimage-archive/)

Partners shall have access to primary data available in both open and closed access modes (Figure 3). Partners shall use only [primary data] curated by the selected archives, and shall include the corresponding accession numbers, md5 checksum and provenance metadata throughout their work and publications in order to ensure traceability. Partners shall not exchange versions of the primary data that differ from those curated by the selected archives, thus ensuring that everyone works with the same data and metadata.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

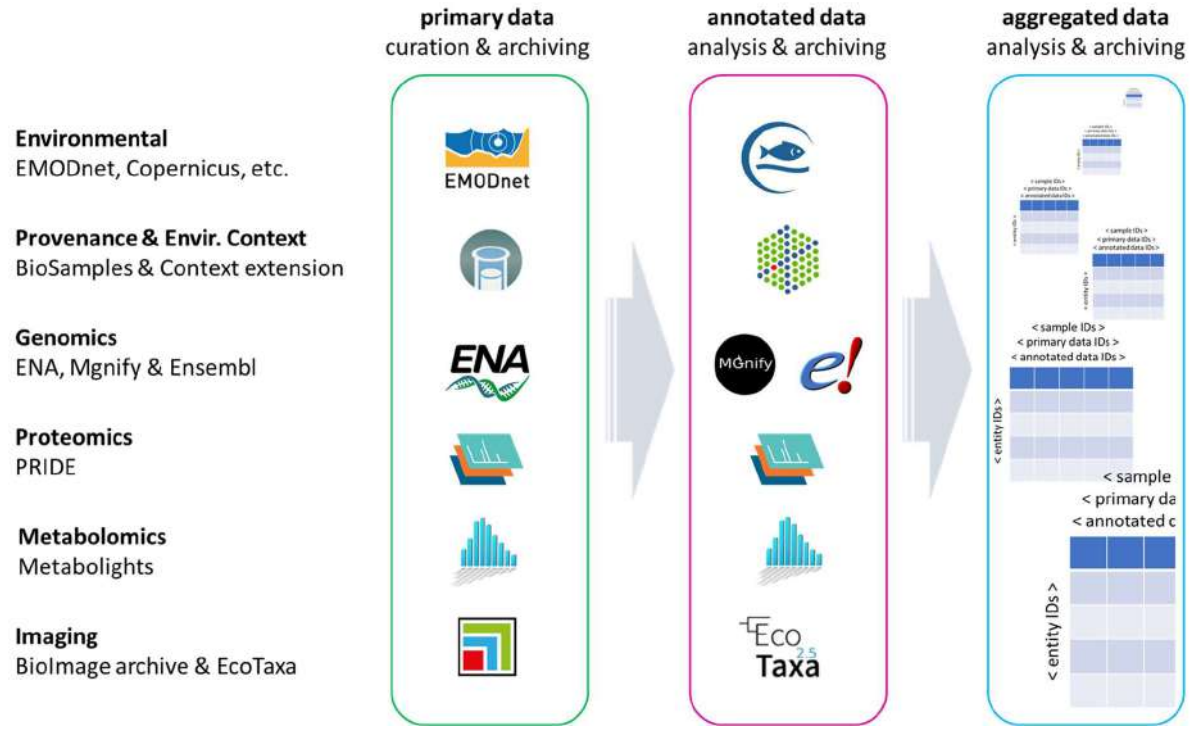

Figure 2. *Schematic diagram of data infrastructures & services that were selected to analyse, curate & archive primary, annotated & aggregated data from six data types.*

Annotated data

Specific analysis pipelines were selected for first-hand analysis and annotation of six types of data generated by the AtlantECO project (Figure 2). By depositing primary data in closed or open access mode in the selected archives, Partners agree to making them available to the respective analysis pipelines (Figure 3, blue workflow). Following the analysis, annotated data are curated according to community standards, assigned unique accession numbers, and archived in closed or open access mode, matching the access mode of the primary data used in the analysis.

Analysis pipelines selected for the different data types:

- BioSamples for provenance and environmental context (https://www.ebi.ac.uk/biosamples/)
- MGnify and Ensembl for genomics data (https://www.ebi.ac.uk/metagenomics/)
- PRIDE for proteomics data (https://www.ebi.ac.uk/pride/)
- Metabolights for metabolomics data (https://www.ebi.ac.uk/metabolights/)
- EcoTaxa for imaging data (http://ecotaxa.obs-vlfr.fr/)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Partners shall have access to annotated data available in both open and closed access modes (Figure 3). Partners are welcome to use other pipelines to analyse primary data provided that the resulting annotated data are made available to all Partners by depositing the annotated data in the selected archive in closed or open access mode, matching the access mode of the primary data used in the analysis. Partners shall use only annotated data files available in the data hub, and shall include the corresponding accession numbers, md5 checksum and provenance metadata throughout their work and publications in order to ensure traceability. Partners shall not exchange versions of the annotated data that differ from those available in the data hub, thus ensuring that everyone works with the same data and metadata.

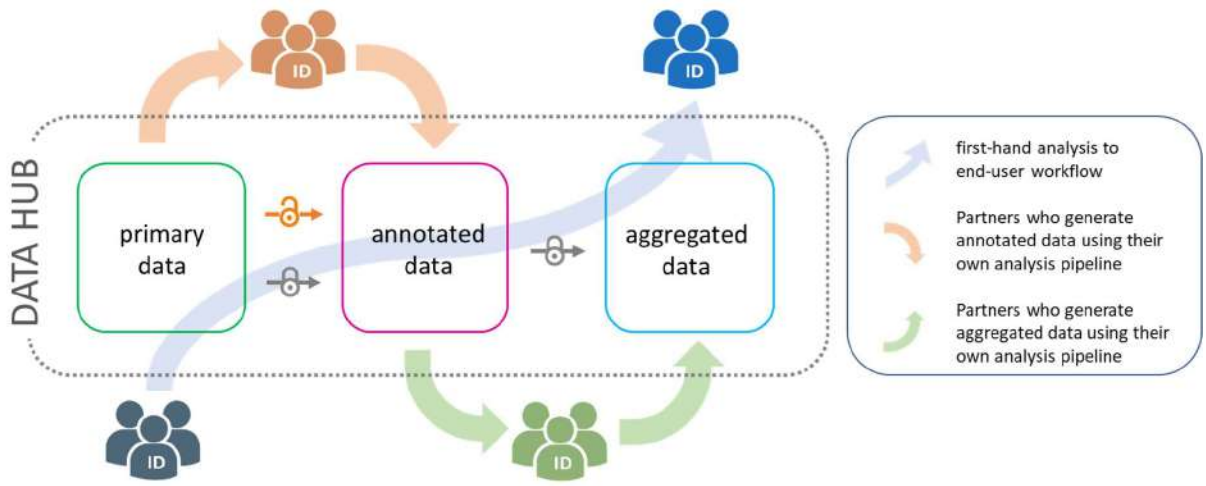

Figure 3. *Schematic diagram of the Data Hub showing how data generated by a partner are first deposited in the data archives in open and closed access before being shared with other partners. Three complementary workflows are depicted: (blue) Following their deposition in the Data Hub, primary data are automatically annotated, aggregated and made available to the Partners; (orange) Partners can also access primary data, use their own pipelines to perform analyses, and deposit the resulting annotated data back in the Data Hub, which are then picked-up by the first workflow; (green) Partners can also access annotated data, use their own pipelines to perform analyses, and deposit the resulting aggregated data back in the Data Hub, which are then made available to the Partners.*

Aggregated data

Annotated data in open and closed access mode from each data type will be aggregated into tables where columns have distinct provenance (i.e. distinct accession numbers for either samples, primary data or derived data) and rows correspond to distinct entities of a reference catalogue or library (Figure 2). Each table will only aggregate annotated data that share a common reference catalogue, and may be replicated in parts for publication purposes, so there will likely be several tables for each data type. We expect at least one table for the following types of annotated data: provenance & environmental parameters, taxonomy, functional & metabolic potential, proteins & peptides and metabolites. These aggregated data tables will be curated using community standards (e.g. the Biological Observation Matrix (BIOM) format) and will be citable with a unique identifier. The functionalities required to produce and archive aggregated data do not exist, and will be developed by EMBL-EBI.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Partners shall have access to aggregated data available in both open and closed access modes (Figure 3). Partners are welcome to use other pipelines to aggregate annotated data provided that the resulting aggregated data are made available to all Partners by depositing the aggregated data in the appropriate archive in closed or open access mode, matching the access mode of the annotated data used in the analysis. Partners shall use only aggregated data files available in the data hub, and shall include the corresponding accession numbers, md5 checksum and provenance metadata throughout their work and publications in order to ensure traceability. Partners shall not exchange versions of the aggregated data that differ from those available in the data hub, thus ensuring that everyone works with the same data and metadata.

Data Hub

Access to jointly owned data (primary, annotated and aggregated) in closed and open access modes will be provided to all Partners via a dedicated AtlantECO Data Hub, hosted by EMBL-EBI with a specific authentication system (Figure 3). There are several aspects of the proposed Data Hub that require development, notably the use of BioSamples IDs across the different data types, the use of a common authentication system, and the development of a data aggregation service. Until the Data Hub is operational, jointly owned data in closed and open access mode will be provided by EMBL-EBI on an FTP server with a unique AtlantECO login and password.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

ANNEX I

Relevant Articles from the AtlantECO Grant Agreement

ARTICLE 28.1 "Obligation to exploit the results"

Each beneficiary must — up to four years after the period set out in Article 3 — take measures aiming to ensure 'exploitation' of its results (either directly or indirectly, in particular through transfer or licensing; see Article 30) by:

a. using them in further research activities (outside the action);

b. developing, creating or marketing a product or process;

c. creating and providing a service, or

d. using them in standardisation activities.

This does not change the security obligations in Article 37, which still apply.

ARTICLE 29.1 "Obligation to disseminate results"

Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — 'disseminate' its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium). This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of — unless agreed otherwise — at least 45 days, together with sufficient information on the results it will disseminate. Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests.

ARTICLE 29.2 "Open access to publications"

Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results. In particular, it must:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

a. as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;

b. ensure open access to the deposited publication — via the repository — at the latest:

   i. (i) on publication, if an electronic version is available for free via the publisher, or

   ii. (ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.

c. ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication. The bibliographic metadata must be in a standard format and must include all of the following:

   i. the terms "European Union (EU)" and "Horizon 2020";

   ii. the name of the action, acronym and grant number;

   iii. the publication date, and length of embargo period if applicable, and

   iv. a persistent identifier.

d. deposit at the same time the research data needed to validate the results presented in the deposited scientific publication.

ARTICLE 29.3 "Open access to research data"

Regarding the digital research data generated in the action ('data'), the beneficiaries must:

a. deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

   i. the data, including associated metadata, needed to validate the results presented in scientific publications, as soon as possible;

   ii. not applicable;

   iii. other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

b. provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39,3 all of which still apply.

European Commission infographics on Open Access to research data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.