

# Data Management Plan (DMP) for Arctic-BIODIVER

(Belmont/BioDiversa 2019–2021)

Updated version 6 July 2020

## I. DATA MANAGER(S)

<ul style="list-style-type: none"><li>● Who will be <b>responsible for managing the data</b>?</li></ul>	Project leaders and WP-leaders, i.e. Willem Goedkoop and Joseph Culp and NN part-time data engineer.
<ul style="list-style-type: none"><li>● Who will ensure that the data management plan is carried out? (<i>e.g. a Data Manager will take care of the DMP and coordinate the work of the data collectors/providers/users from each WP</i>)</li></ul>	Project leaders (see above).
<ul style="list-style-type: none"><li>● Will a <b>specialized and experienced</b> data expert be part of the project team?</li></ul>	We have a part time (5%) data engineer in our budget. Data experts are available for the circum-Arctic data at CAFF's Arctic Biodiversity Data Service (ABDS). As SLU is the national data host for freshwater data in Sweden, there is considerable database competence in house.

It is recommended to have one (or several) appointed Data Manager(s) to produce the DMP and overview the data management practices before, during, and after the project. This should ideally not be the Project Coordinator itself, and it is a great advantage to have people experienced in data management in the team.

It is recommended also to store the DMP as a project document, which can be updated whenever needed (for example: to update the information about upload of new data etc.).

## II. DATA IDENTIFICATION & DESCRIPTION

<ul style="list-style-type: none"><li>● What's the <b>purpose</b> of the research?</li></ul>	To build scenarios for changes in freshwater biodiversity in the Arctic regions and study ecosystem structure and function across latitudinal gradients.
<ul style="list-style-type: none"><li>● <b>What datasets</b> of long-term value do you expect that the project will collect, process, and produce?</li></ul>	Data on abiotic and biodiversity data compiled at ABDS is basis for scenario building and forms a base line for future change in Arctic regions. Latitudinal gradient studies will supply data that contribute to our understanding of how ecosystem structure and function will change.
<ul style="list-style-type: none"><li>● What is the data? How and in what <b>format</b> will the data be collected? Is it numerical data, image data, text sequences, or modelling data? (<i>e.g. microscopic images, video recordings of interviews, etc.</i>)</li></ul>	Data at ABDS are currently in ACCESS. New data from latitudinal gradient studies will have a different, yet undefined, format. These data will largely be species and e-DNA data from the

	different sites that have been sampled, as well as general information on the sites.
● <b>How much</b> data will be generated for this research?	100 gigabyte range.
● <b>How long</b> will the data be collected and how often will it change?	During 2019–2021, i.e. the duration of the project.
● Are you using data that <b>someone else produced</b> ? If so, where is it from?	Yes. Data compiled in the ABDS originate from multiple data sources (monitoring, research, industry) in participating countries.
● Will there be <b>other types of material</b> of long-term value produced? If so, what are your plans for ensuring these are also available over the long-term? (e.g. <i>samples of specimens collected on field; physical collections, software, curriculum materials, etc.</i> )	Some partners will store biological samples collected during the project in national collections, such as the Alaska Center for Conservation Science Herbarium, and the National Phycology Collection in Canada.

The DMP should consider, not only primary data, but also secondary data; not only peer-reviewed publications but also communications material, documents for stakeholders, etc. In addition, it should include physical objects such as specimens, etc.

Description of the data and datasets should be complete and detailed: type, flows, quantity, format, etc. This applies both to data collected/generated as to data (e.g. from other research projects) that will be (re)used.

“Long-term” means those datasets that, over time, will or may be of value to others within your research community and/or the wider research and innovation community.

### III. DATA ORGANISATION & EXCHANGE (internally, during the project)

● How do you intend to <b>manage the data</b> during the life of the project to ensure their long-term value is protected?	On an internet work space at SLU or using Github or Google Drive. See also below under “archiving of data”.
● What is your strategy for <b>organising your data</b> ? How do you organise your folders and name your files? What directory and file naming convention will be used?	Will be in our “quality handbook” that will be produced during the first year of the project. and decided along the way.
● <b>What data will be shared</b> among your colleagues/partners, when, and how?	All data produced within the project will be shared.
● <b>Where</b> will the data be held during the project?	At local hard drives and servers (back-ups) with partners and in the cloud during data production (e.g. Github and Google Cloud). Following data production and QA/QC, all data will be stored at a common work space (Share Point) and made available for analyses by all partners.
● <b>Who</b> has the right to <b>manage this data</b> ?	Addressed under next bullet point.

<ul style="list-style-type: none"> <li>Who will have <b>access</b>?</li> </ul>	<p>All project participants will have access to the data generated during the latitudinal gradient studies (reading), but changes to the final data base will be restricted to the project leads (editing). The data from circum-Arctic lakes and rivers at ABDS will be accessible for all partners (reading), but managed by CAFF-staff (editing).</p>
<ul style="list-style-type: none"> <li>How do you <b>exchange files</b> (and other information) with your collaborators?</li> </ul>	<p>Files will be shared on a common work space on the internet. SLU can provide a log-in protected common work space on a file-server.</p>
<ul style="list-style-type: none"> <li>How do you take care of <b>consistency and quality of the data</b>?</li> </ul>	<p>Data will be produced according to the guidelines in the project's quality handbook that will be published on the project's web page in August 2020.</p>

Be specific and complete in explaining WHEN exactly the data will be available (e.g. at any moment of the project) and WHERE (e. g. name of repository or website).

**IV. DATA STORAGE AND BACK-UP (during the project)** What is your **data storage and backup** strategy?

<ul style="list-style-type: none"> <li>How much <b>data storage</b> you need for the project and what is the <b>estimated increase per month</b>?</li> </ul>	<p>Approximately 1 Gb. Data increase per month not estimated.</p>
<ul style="list-style-type: none"> <li>How frequently do you do your <b>backups</b>?</li> </ul>	<p>Back ups will be made on local hard drives, as well as in "the cloud" (e.g. Github), by each of the partners. Data storage capacity and back-up handling can be purchased at SLU:s servers. Local data producers (i.e. project partners) are responsible for the secure handling of data files during the project and have own routines for backing up data.</p>
<ul style="list-style-type: none"> <li>At how many <b>independent locations</b>?</li> </ul>	<p>Each of the project partners will have own data files and storage(?) during the phase of data production. These data will be merged, however, when analyses have been completed. After that, all partners will have access to all data for the specific research questions they address in their WPs. These files will be made accessible at a common work space for the project.</p>
<ul style="list-style-type: none"> <li>What are your local storage and backup <b>procedures</b>? Will this data require <b>secure storage</b>?</li> </ul>	<p>A log-in protected common work space (Microsoft Share Point) at SLU will be made available during the project. Here joint work files will be stored, updated, and securely backed up.</p>

## V. DATA SHARING, STANDARDS<sup>1</sup> & METADATA (with externals)

<ul style="list-style-type: none"> <li>Are you using a <b>file format</b> that is standard to your field? If not, how will you document the alternative you are using?</li> </ul>	Working files containing raw data will be primarily in Excel, as “comma-separated values files, and/or GIS data in geodatabases.
<ul style="list-style-type: none"> <li>Which <b>methodology</b> and <b>standards</b> will be applied? (<i>e.g. NetCDF (Network Common Data Format) files (for multidimensional spatio-temporal data)</i>)</li> </ul>	Not decided yet. This is something for the consortium to look into.
<ul style="list-style-type: none"> <li>What tools or software are required to <b>read or view the data</b>?</li> </ul>	MS Excel for data from latitudinal gradients studies and MS ACCESS for biodiversity data stored at ABDS.ARC-GIS will be used for geospatial data.
<ul style="list-style-type: none"> <li>What <b>supporting documentation</b> will you be creating and make publicly available in order to make the data understandable by other researchers and support the longer-term re-use of the data?</li> </ul>	References to existing standards and/or scientific publications describing the methods used will be included in data files. Also our quality handbook of methods will be of help to clearly define these variables and their methods of quantification. We will look into following DataOne's ‘Knowledge Network for Biocomplexity (KNB)’ requirements for supporting documentation and metadata
<ul style="list-style-type: none"> <li>Are you using <b>metadata</b> that is standard to your field? How will the metadata be managed and stored?</li> </ul>	Metadata exists for the circum-Arctic sites that currently are stored in the ABDS at CAFF. Metadata will be compiled for the sites that will be part of our latitudinal studies. We will look further into the choice of metadata standards.

It is essential to associate datasets with metadata so that other researchers can understand how the data was collected and under which conditions it can be reused.

## VI. DATA RESTRICTIONS

<ul style="list-style-type: none"> <li><b>How open</b> will the data and outputs be?</li> </ul>	Data will be made publicly available. Open access.
<ul style="list-style-type: none"> <li>Will you be dealing with <b>sensitive/personal/restricted data</b> and <b>why</b>? (<i>e.g. spatial/temporal information about endangered species; personal information from interviews;...</i>)</li> </ul>	<p>In European countries we will, where appropriate, follow the European General Data Protection Regulation (2016/679). <a href="https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679">https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679</a></p> <p>Information on distribution of endangered species will follow the methods used by national species data services (or similar organizations) in participating countries.</p> <p>No interview data will be produced. Land owner data will not be stored.</p>

<ul style="list-style-type: none"> <li>Do you expect there will be any <b>restrictions</b> on how the data can be accessed or reused (e.g. due to Intellectual Property Rights (IPRs))?</li> </ul>	<p>There will be no such restrictions for the data from latitudinal field studies produced within our project. There may be some availability restrictions to the compiled data on Arctic biodiversity that currently are stored at ABDS, as part of this data set originates from industry and other scientific projects. Data agreements for the purposes outlined in Arctic-BIODIVER will be made with data owners.</p>
<ul style="list-style-type: none"> <li>Does sharing the data raise <b>privacy, ethical, legal, or other confidentiality</b><sup>2</sup> concerns? (please note that there may be country-dependent rules what comes to these issues.</li> </ul>	<p>At this stage, we cannot see that sharing the data caused any such conflicts. We will, however, have an active, ongoing dialogue about this during the course of the project.</p>
<ul style="list-style-type: none"> <li>Do you have a plan to protect or anonymize data, if needed?</li> </ul>	<p>Data on sites with protected species of aquatic flora and fauna may need to be protected, as does information on land owners. Data will be anonymized where this is judged necessary for the protection of this species.</p>

Data should be as open as possible, though with restricted or closed access where appropriate and necessary; for example, if there are sensitive data involving human subjects. Depending on the nature of the research, the degrees of data openness may vary, extending from fully open to strictly confidential data.

The reason for restricting the access or use of some data should be explained and justified. As for personal data, make sure it complies with the EU General Data Protection Regulation<sup>3</sup> (GDPR).

## VII. DATA PUBLISHING & LICENSING

<ul style="list-style-type: none"> <li>Where and how will the data be published?</li> </ul>	<p>All data generated within the project will be published after the project ends in late 2021 or when scientific publications have been published. The DataOne's Knowledge Network for Biocomplexity and UAA Alaska Center for Conservation Science data portal is one suitable forum for this. Data of suitable format (e.g. species data, water chemistry) will also be stored in publicly available national monitoring data bases.</p>
<ul style="list-style-type: none"> <li>Under which <b>licences</b> will the data be published? Have <b>possible licensing</b> issues been considered?</li> </ul>	<p>This is an issue that still needs to be addressed by the consortium.</p>
<ul style="list-style-type: none"> <li>Will the research be published in a journal that requires the <b>underlying data</b> to accompany articles?</li> </ul>	<p>Data that are part of scientific publications will be made available in the supplementary materials of these publications.</p>
<ul style="list-style-type: none"> <li>Will there be any <b>embargoes</b> on the data? If yes, explain why, until when, and what happens when the embargo is over.</li> </ul>	<p>We do not see any such embargoes at present.</p>

<ul style="list-style-type: none"> <li>● What are your intended <b>Open Access publications practices</b> (Green, Golden, Hybrid<sup>45</sup>)? Data should be made available as soon as the results of the research have been published.</li> </ul>	<p>Thesis (MSc and PhD) will be published in self-archiving systems at our universities (“Green”). Scientific papers will be published in journals with open-access publishing (“Golden”) and/or Hybrid.</p>
--	--

### VIII. DATA ARCHIVING (after the project ends)

<ul style="list-style-type: none"> <li>● How will the data be managed <b>after the project ends</b> to ensure their long-term availability?</li> </ul>	<p>Long-term storage of the data from sites across the Arctic region will be done at ADBS (CAFF), where the data currently are stored. Data from latitudinal field studies will be ?</p> <p>Newly produced data from fields studies across latitudinal gradients will be archived at SLU, the lead university for this project. Archiving will be done in accordance with the principle of public access to official documents in Sweden.  <a href="https://www.slu.se/en/subweb/library/publish-and-analyse/Archiving-and-publishing-research-data/archive-data/">https://www.slu.se/en/subweb/library/publish-and-analyse/Archiving-and-publishing-research-data/archive-data/</a></p>
<ul style="list-style-type: none"> <li>● Will the data be published with a <b>Digital Object Identifier (DOI)</b> and/or be placed in a <b>recognized long-term repository</b> or data centre, and when will this take place?</li> </ul>	<p>DOI will be used for the data base, as well as for data sets that have been used for specific analyses. This can, for example, be done at the Swedish National Data Service (SND).</p>
<ul style="list-style-type: none"> <li>● How will you be <b>archiving the data</b>? Will you be storing it in an archive or repository for long-term access? If not, how will you preserve access to the data?</li> </ul>	<p>Long-term storage of data will be done at SLU, as the project is led by SLU. As such, the university will also guarantee long-term availability of the data as part of its official responsibilities. Other participating universities may also archive the data according to their routines at local servers.</p>
<ul style="list-style-type: none"> <li>● How will you prepare data for <b>preservation or sharing</b>? Will the data need to be anonymized or converted to more stable file formats?</li> </ul>	<p>Long-term storage (archiving) of the data will be in archive-supporting format such as text- and/or CSV-format.</p>
<ul style="list-style-type: none"> <li>● Are <b>software or tools</b> needed to use the data? Will these be archived?</li> </ul>	<p>Excel, ACCESS, ARC-GIS and specific statistical software (e.g. R, JMP) will be used during the data analysis phase. Textfiles, for which no specific software is needed, will be used for archiving data.</p>
<ul style="list-style-type: none"> <li>● <b>How long</b> should the data be retained? 3-5 years, 10 years, or forever?</li> </ul>	<p>Data should be archived. Data produced within the project and those delivered to SLU from project partners, will be archived at SLU. Our ambition is that these data will be retained for an indefinite future in accordance with Swedish law.</p>

sBe specific and complete in explaining WHEN and FOR HOW LONG exactly the data will be available (e.g. after project ends, during 5 years,...) and WHERE (e. g. name of repository or website). Vague statements like “could be/may be/we plan to...” should be avoided and replaced with more accurate terms.

The DMP should include the full list of data storages/repositories/catalogues/websites (with working hyperlinks). It is highly recommended to use Digital Object Identifiers (DOIs).

We encourage projects to store the data for as long as it possible: 5-10 years storage is seen as minimum period of time.

<sup>2</sup>For more information, please visit: <https://libraries.mit.edu/data-management/share/confidentiality/>

<sup>3</sup>See [https://ec.europa.eu/info/law/law-topic/data-protection\\_en](https://ec.europa.eu/info/law/law-topic/data-protection_en); but also: <https://www.scienceurope.org/legislation/activities/data-protection>.

<sup>4</sup>Licensing and publishing references and tools can be found under **4.4 Licensing, citing & publishing** of the guidance document. <sup>5</sup>See **Annex I – Glossary** for a description of the different OA publication types.

## VIII. COSTS

<ul style="list-style-type: none"> <li>● What are the estimated costs for managing your data and other materials during/after the project?</li> </ul>	<p>There is funding for a part-time (5% of full time, or 35000 SEK) data engineer in the project.</p>
<ul style="list-style-type: none"> <li>● How have you accounted for the costs to ensure long-term availability? (e.g. <i>cloud hosting costs at www.example.com will be of 15,000€/year, budget included in the project</i>).</li> </ul>	<p>Such costs have not been included in the budget of the project. However, SLU will guarantee long-term availability of the data as part of its official responsibilities. At present this service can be used, as there is no cost for this.</p>