D1.2 Data Management Plan v1.0

Irene Andreassen

Sagnik Sengupta

ArcSolution

Arctic pollution in a One Health perspective
from complex challenges to sustainable solutions

Funded by
the European Union

This deliverable is a part of the ArcSolution project.

Deliverable number and info: D1.2 – based on Horizon Europe DMP template
Dissemination level: Public

Funded by the European Union under the Horizon Europe
Program, Grant No. 101135051 — ArcSolution

www.arcsolution.eu

Title: Data management plan V1.0
Author(s): Sagnik Sengupta, Irene Andreassen
Publication date: 31.01.2024
Page number: 16

Content

1 Summary

This deliverable presents the Data Management Plan (DMP) for the project, providing a comprehensive framework for data collection, processing, sharing, and storage. It ensures compliance with ethical standards and actively involves local communities to foster understanding of Arctic environmental health through a One Health approach. The DMP follows the Horizon Europe template, with sections 3 to 5 addressing the guideline questions outlined in the template.

2 Data Summary

A comprehensive mapping of all datasets and data types in ArcSolution has been carried out to form the foundation of this data management plan. This mapping is a dynamic process, as the need for additional data or the relevance of existing datasets may change over the course of the project. Accordingly, the mapping will be reviewed and updated regularly throughout the project's duration.

The table below provides a summary of the mapping results.

Table 1: Summary of Data Sets Generated by Each Work Package (WP)

WP no.	Title	Types of data generated	File formats
1	**Management and administration** (no scientific data produced in WP1)	**Administrative and financial records:** Documentation of project management, financial reporting, and risk management.	**.doc, .pdf, .xls**
		Progress reports: Biannual updates and detailed reports for the European Commission	**.doc, .pdf**
		Data management plan (DMP): Regularly updated plans addressing data handling, preservation, accessibility, and governance for Indigenous and local data.	**.doc, .pdf**
		Collaborative outputs: Records of international partnerships, contact lists, and documentation from collaborations with external projects and networks.	**.doc, .pdf, .xls**
2	**The ArcSolution One Health Framework**	**Dietary and cultural data:** Information on key dietary items and local/traditional knowledge about nature, wildlife, and human interactions.	**.doc, .pdf, .xls**
		Community engagement data: Inputs from surveys, focus groups interviews, unstructured interviews, co-production community workshop, as well as workshop evaluations.	**.doc, .pdf, .xls**
		Integrated datasets: Compilations of environmental, wildlife, and human data	**.doc, .pdf, .xls, model outputs**

		from WP3 and WP4, and synthesis across all WPs.	
		Gap analysis and research insights: Data and knowledge gap analyses and descriptions of citizen science projects.	**.doc, .pdf, .xls**
3	**Arctic Pollution in a climate change context**	**Environmental Monitoring:** Non-target screening of environmental samples (air, water, snow) and wastewater treatment in Longyearbyen.	**.doc, .pdf, .xls**
		Pollutant Transport and Modelling: Modelling of long-range transport of pollutants via air and ocean, along with climate modelling.	**.doc, .pdf, .xls Model outputs Model scripts**
		Biota and food web: Stable isotopes and fatty acid analyses of biota samples and food items, biota samples analysed for priority pollutants, and food web characterization.	**.doc, .pdf, .xls Model outputs Model scripts**
		Human Health Data: Human tissue samples from Longyearbyen, along with medical and diet questionnaires.	**.doc, .pdf, .xls**
		Citizen Science Contributions: Citizen science litter identification and hunters' records.	**.doc, .pdf, .xls**
		Cohort Datasets: Secondary data from existing cohorts (Finland, Greenland, Tromsø, and Faroe Islands).	**.doc, .pdf, .xls**
4	**Impacts under different scenarios**	**Toxicity and Biological Effects**: Reviews on pollutant toxicity, effect-directed analyses, predictions/modelling of effects, and zebrafish toxicity testing.	**.doc, .pdf, .xls**
		Health and Biomarkers: Data on metabolic health, thyroid function, brown fat activity, plasma biomarkers (clinical, reproductive, and endocrine), and inflammatory markers.	**.doc, .pdf, .xls**
		Wildlife Studies: Tissue and autopsy samples, gut flora genome sequencing.	**.doc, .pdf, .xls**
		Environmental and Lifestyle Data: Environmental exposure, demographic and lifestyle information, and dietary data from study locations.	**.doc, .pdf, .xls**
		Citizen Science: Contributions on the impacts of plastic pollution on wildlife.	**.doc, .pdf, .xls**
		Risk Assessment and Integration: Socio-economic consequences, risk assessment strategies, and compiled datasets on human exposure and cross-task inputs.	**.doc, .pdf, .xls**
5		**Local Stakeholder and Expert Input**: Discussions with local authorities,	**.doc, .pdf, .xls**

	Solutions - Resilience, adaptation, mitigation	stakeholders, and experts, along with analysis of proposed technologies and chemical management strategies.	
		Mitigation and Remediation Strategies: Data on mitigation strategies, including efficacy indicators for technological solutions and testing/validation for circular approaches.	**.doc, .pdf, .xls Patents**
		Toxicological and Quality Assessments: Toxicological assessments for chemicals of emerging Arctic concern (CEACs) and reuse materials, alongside product quality data for circular solutions.	**.doc, .pdf, .xls Patents**
		Pilot Study Data: Metrics from locally-adapted pilot studies in areas such as food safety, water supply, sewage treatment, waste disposal, and indoor environment improvements.	**.doc, .pdf, .xls Patents**
		Citizen Science and Regulatory Data: Citizen science data on waste handling and water circulation, along with existing regulatory standards and thresholds.	**.doc, .pdf, .xls Patents**
		Ecosystem and Risk Evaluation: Data on ecosystem services evaluation.	**.doc, .pdf, .xls**
6	**Exploitation, Dissemination and Communication**	**Citizen Science Results**: Conclusions and results from citizen science projects.	**.doc, .pdf, .xls**
		Stakeholder Feedback: Evaluation data and discussions on the continuation of stakeholder engagement efforts.	**.doc, .pdf, .xls**
		Educational Materials: Development of toolboxes for teaching and teaching materials for summer schools.	**.doc, .pdf, .xls**
		Dissemination Outputs: Policy briefs and peer-reviewed publications to communicate findings and recommendations.	**.doc, .pdf, .xls**
		Communication and Outreach: Data from the project's website, social media activities, as well as materials like logos, templates, and other promotional tools.	**.doc, .pdf, .xls**

2.1 Purpose of Data Generation and Reuse in the Project

The purpose of data generation and re-use in the ArcSolution project is to enhance understanding of the environmental health of the Arctic and its interconnectedness with human and wildlife health, following a holistic One Health approach. This comprehensive knowledge supports the project's objectives by integrating local and Indigenous knowledge with environmental, health, technical, and social science research across four key locations: Northern Finland, Svalbard, Northern Norway, and Greenland.

Data will be collected from these sites, including both new data and information drawn from previously published studies and available datasets, to provide a thorough analysis of priority pollutants. By identifying knowledge gaps and synthesizing relevant information, the project will inform citizen science initiatives focused on locally prioritized issues, pollution sources, and human exposure.

The findings will contribute to developing sustainable solutions in circular economy, technology, and chemicals management while being communicated to policymakers, industry, and educational institutions. The Data Management Plan (DMP) outlines the framework for data collection, processing, sharing, and storage, ensuring compliance with ethical standards and facilitating collaboration among project partners, thus further reinforcing the project's objectives of fostering sustainable actions in the Arctic.

2.2 Data Utility and Stakeholder Engagement

The data generated from the ArcSolution project will be invaluable to a diverse array of stakeholders, including:

- Scientific communities
- Governmental agencies
- Municipalities
- Policymakers
- Schools and educational institutions
- Local populations in the study areas, including Indigenous groups and immigrants
- Media outlets

This broad audience will benefit from the project's insights into the Arctic's environmental, human, and wildlife health, fostering informed decision-making, enhancing educational opportunities, and raising public awareness about critical issues affecting the region.

3 FAIR data

General on Open Research Data in ArcSolution:

Personal data within ArcSolution will only be shared as anonymized if there is no possibility that combination of variables can lead to the identification of individuals or otherwise as pseudonymized form. No personal data will be exchanged between partners unless a joint data controller agreement is in place. Personal data will remain within the country they are collected. Data involving Indigenous populations are owned by the respective Indigenous communities.

All publicly accessible data will be made available in Open Access, either through Zenodo under our ArcSolution community page or through Nord University's institutional repository (DataverseNO).

3.1 Making data findable, including provisions for metadata:

1. **How will data be identified by a persistent identifier?**

To increase findability of research data, ArcSolution will use data repositories that assign globally unique and persistent identifiers to data (e.g., DOI).

The data files and associate metadata will be assigned persistent IDs (e.g., DOI) while archiving at trustworthy, indexed data repositories.

Open (green) data and anonymized/pseudo anonymized datasets (i.e., after removal of personal information) can be archived https://dataverse.no/, which is also Nord University's open repository for research data. Alternatively, a discipline-specific open archive can be used to make green data available.

With valid written consent for reuse of data for research purposes after the end of the project, personal data (not belonging to special categories) can be archived at https://sikt.no/en/archiving-research-data, after an agreement is entered into by Nord University and Sikt. Alternatively, personal data must be deleted at the end of the project while anonymized data can be archived. Sikt also assigns DOIs to datasets and metadata.

Data containing special categories of personal information, such as health data, will be managed via Service for Sensitive Data (TSD) at UiO.

2. **Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.**

To allow discoverability of data, internationally recognized metadata standards such as DDI (https://ddialliance.org/) and Dublin core (https://www.dublincore.org/) will be followed. Dataset- and file-level metadata will be created to facilitate reuse of the data by clarifying contexts and data provenances.

3. **Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?**

Keywords will be added to the metadata to increase the findability of the data and potential for reuse.

4. **Will metadata be offered in such a way that it can be harvested and indexed?**

The metadata will be provided in machine-readable formats via indexed repositories such as SIKT and DataverseNO. At Sikt, anonymized metadata will remain openly accessible via Surveydatabanken.

3.2 Making data accessible - Repository:

1. **Will the data be deposited in a trusted repository?**

The data will be deposited in trusted repositories. Restricted data will be archived at SIKT whereas green data can be made openly available through the Core Trust Seal certified database DataverseNO. Health data and special categories of personal data will be handed via TSD.

Whenever possible, data will be archived for long-term preservation in a certified repository, under open licenses (CC 0 or CC BY or a license with equivalent rights). Further, any information necessary to understand and re-use the data will be provided (e.g., information on methodology, data cleaning processes, variable definitions, and data quality assurance processes), as well as any tools or software necessary to interpret and re-use the data.

NORD's institutional repository (DataverseNO) has this feature and will be considered for open research data. DataverseNO is a curated, CoreTrustSeal-certified, and FAIR-compliant repository for open research data from all disciplines. Data and metadata published in the repository are securely archived and available for at least 10 years from the date of publication. Also, the repository's metadata are harvested/indexed by several discovery services, including B2FIND, BASE, DataCite Search, Google Dataset Search, and ExLibris Primo. NORD's institutional repository uses standard-compliant metadata, which helps increase the interoperability of research data.

Zenodo, a general-purpose data repository, supports all forms of research output, from data files to presentation materials. Built on open-source software, it is part of the OpenAIRE project, which was initiated by the European Commission to support Open Data policies by offering a repository for EC-funded research. Zenodo, launched in May 2013 by CERN, an OpenAIRE partner and open-source pioneer, is funded by CERN's core resources and EU research grants.

Both repositories are compatible with, for instance, DDI Lite, DDI 2.5 Codebook, and Dublin Core standards. Further, controlled vocabularies will be used when available. For social science data, for instance, the European Language Social Science Thesaurus (ELSST) will be considered.

2. **Have you explored appropriate arrangements with the identified repository where your data will be deposited?**

DataverseNO can publish open (green) data, including anonymized data and Nord University has agreements in place for archiving data at DataverseNO. Regarding the archiving of data containing personal information, we have obtained an overview of the processes SIKT has in place. Suitable permissions will be sought for processing of personal information in accordance with the correspondence of the present project/Nord University with SIKT. Sikt will, on behalf of Nord university, evaluate the strategy for handling personal data before data collection can begin. During the course of evaluation, Sikt can suggest measures for improving the handling of personal data.

3. **Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?**

The selected archives assign digital object identifiers to the data.

3.3 Making data accessible - Data:

1. **Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions.**

Research data from ArcSolution will be made openly accessible for further use for all relevant users, except when there are legal, ethical, security-related or commercial reasons for not doing so. Reasons for restricting the availability of research data include: security concerns (when making the data available may threaten the security of individuals or national security); confidentiality of personal data (when making the data available is in violation of applicable

regulations concerning personal data); and IPR considerations (for data with commercial value). In other words, the project will follow the principle of "as open as possible and as closed as necessary".

2. **If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.**

An embargo may be applied to ensure the protection of intellectual property rights (IPR) and to provide adequate time for publishing results or filing for patents if applicable. The embargo duration will be flexible and depend on the data type and the nature of the research. Sensitive data must remain protected to safeguard unpublished findings and any associated cultural or ethical considerations. The embargo will last only as long as necessary to secure IPR or complete the publication process, with the aim of making the data publicly available as soon as these conditions are met.

3. **Will the data be accessible through a free and standardized access protocol?**

Both SIKT and dataverseNO provide free and standardized protocols for accessing data.

4. **If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?**

Access to the data will be granted to project partners in accordance with the "Joint Controller Agreements." Once the data are archived, access for reuse by others will be managed through SIKT, subject to the necessary consents and permissions for reusing personal data. All personal data stored in TSD will be deleted at the conclusion of the project.

As a general rule, only anonymized data will be shared among partners, unless there is a specific reason to share personal data.

5. **How will the identity of the person accessing the data be ascertained?**

Login credentials will be used for identifying the persons reusing the data.

6. **Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?**

Access to restricted data will be administered in consultation with SIKT. In general, all access requests will be followed-up by the project coordinators office.

3.4 Making data accessible - Metadata:

1. **Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?**

Yes, wherever possible.

2. **How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?**

Archived data and metadata will remain available for a long period (at least 10 years). The metadata will remain available even after the data is no longer available.

3. **Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?**

Not applicable. The data and metadata will be available in machine-readable non-proprietary formats.

3.5 Making data interoperable:

1. **What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?**

Adressed under 3.1 – making metdata findable.

2. **In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?**

Yes

3. **Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?**

[1]A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between(meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)

Yes

3.6 Increase data re-use:

1. **How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?**

We will provide readme files with information on background, context and methodology, codebooks syntax files and variable definitions.

2. **Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?**

Green data and anonymized data (after removal of personal, health, Indigenous data and sensitive information) will be made openly available for re-use under CC0 License. A suitable license will be selected later for the re-use of restricted data.

3. **Will the data produced in the project be useable by third parties, in particular after the end of the project?**

Yes, we will archive the data in trustworthy archives – DataverseNO and SIKT.

4. **Will the provenance of the data be thoroughly documented using the appropriate standards?**

Yes

5. **Describe all relevant data quality assurance processes.**

The data quality assurance processes will include standardized protocols for the collection of samples intended for chemical analyses, ensuring compliance with QA/QC standards.

Additionally, standardized protocols will be applied for the collection and handling of social science data to maintain consistency and reliability.

6. **Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.**

Wherever possible, academic work will be published open access.

4 Other research outputs

1. **In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).**

We are developing various protocols and procedures, such as those for sampling and methods, as part of the research outputs. These will be made openly available to ensure accessibility and usability by others, contributing to transparency and knowledge sharing in the scientific community.

2. **Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.**

Beyond datasets, we will apply these principles to other research outputs, such as reports, publications, and any tools or models developed. These will be made openly available when appropriate, ensuring proper licensing and permissions for reuse.

This approach is further addressed in our Communication, Dissemination, and Exploitation Plan, where we outline how our research outputs will be shared, communicated, and made available to the broader community. We will also ensure transparency in the management and reuse of data, enabling long-term access and fostering collaboration with other researchers and stakeholders.

5 Allocation of resources

1. **What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?**

Archiving data in SIKT and DataverseNO are free of charge. TSD costs NOK 25000 per year from 2025. The costs of open access publication of research results will be covered by institutional agreements with publishers.

2. **How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)**

All partners have allocated dedicated funds in their budgets for data management and ensuring open access to both data and results.

3. **Who will be responsible for data management in your project?**

The project coordinator will coordinate data management in accordance with any of the Joint Data Controller Agreement for partners. Each partner is responsible for generating their specific data and ensuring that the processing, sharing, and storage of this data are conducted in full compliance with the Data Management Plan (DMP).

Nord University, as the main project coordinator, holds the overall responsibility for the DMP and ensures that all partners are aware of and adhere to the established data management standards and guidelines throughout the project.

4. **How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?**

Open data and anonymized data will be preserved in trustworthy archives (SIKT/dataverseNO) over the long-term. Data containing personal information will be archived at SIKT with restricted access. These archives will ensure availability of the data for at least 10 years from the date of publication.

6 Data security and protection measures

1. **What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?**

Data will be collected using equipment and software approved by the partner institutions. During the research process, data will be stored in secure storage systems managed by the involved partner universities, primarily on laptops and cloud services administered by the respective

institutions. Sensitive/red data will be encrypted using approved software to ensure confidentiality.

Access to the data will be restricted to the researchers responsible for acquiring it and overseeing its use, unless a Joint Controller Agreement has been signed. All data will be regularly backed up by the IT departments of the involved universities to prevent loss.

Personal data will be shared between partners within the same country signing joint controller agreements. Access will be granted only to consortium members explicitly designated for this purpose. These members will have access to the specific data and will be responsible for managing the data within their respective work packages.ni

2. **Will the data be safely stored in trusted repositories for long term preservation and curation?**

Green data and anonymized data will be preserved in trustworthy curated archives (SIKT/dataverseNO) over long-term. Data containing personal information can be archived in SIKT with restricted access.

To ensure robust data security, including data recovery, secure storage, archiving, and transfer of sensitive data, all qualitative and quantitative data will be processed in accordance with applicable regulations and guidelines for data protection, such as GDPR, as well as ethical guidelines. Data will be classified based on sensitivity levels, and processing and storage will comply with the protection requirements associated with each classification.

Furthermore, potential risks related to data production and handling will be continuously assessed, and proactive measures will be implemented to mitigate these risks. This may include regular meetings among project participants to discuss data management responsibilities.

Data will be securely stored at various institutions following strict security regulations and internal guidelines. The merging of data will only be considered for anonymized data and will require approval from regional ethical bodies.

For medical data, approvals are granted only for a specified period, as they have an expiry date. However, we will ensure that the approval timeline is sufficiently long to allow for thorough analysis and publication.

7 Ethics

1. **Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).**

Where personal or sensitive data is collected in the work packages, the necessary approvals will be obtained in advance from the relevant ethics committees or authorities.

Furthermore, ArcSolution will be supported throughout the project by an internal ethical advisory board. The board's role is to ensure that the project complies with the ethical obligations outlined in the Grant Agreement, assess the ethical merits through regular discussions with consortium members, and provide valuable insights. The board will oversee all official deliverables produced within the

project, ensuring alignment with ethical guidelines and expectations. Additionally, the board will specifically guide the involvement of local and indigenous populations in the project, ensuring that their rights, knowledge, and cultural values are respected and properly integrated into the research processes. Research data related to Indigenous peoples will be owned by them, and their consent will be obtained for its use and sharing, in line with ethical standards and community agreements.

2. **Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?**

Informed Consent for Data Sharing and Long-Term Preservation

A standardized consent form for personal data collection has been developed and will be utilized in all instances involving the collection of personal data, such as interviews, workshops questionnaires, or similar methods. This approach ensures compliance with ethical standards and legal requirements. Templates for information letter to participants include *e.g.* the public interest as legal basis allowing for flexibility while reusing data/archiving (Information for participants in research projects).

Both the general consent form and a dedicated consent form for collecting medical and lifestyle data will include information about when personal data will be deleted to ensure transparency. To guarantee that all participants fully understand the forms, they will be translated into the local languages of the study sites, including a version for the Thai population residing in Longyearbyen.

This strategy promotes clarity and fosters informed consent among diverse populations, including inhabitants from other countries.

8 Other issues

1. **Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?**

The University Library at Nord University will assist in data management (research-data@nord.no)