

Work Programme 8 – Management

D8.2 – Data management plan version 1

Lead Contractor: AAU

Author(s): Agneta Ghose

This document is the ALIGNED project (grant no. 101059430) Data Management Plan first version. It details what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use and how the data will be curated and preserved. The design of the core part of this document is different from ALIGNED other deliverables as it was prepared in [ARGOS](#), an online tool.

| PROJECTS DETAILS | | | |
|------------------------|---------|--|------------------------|
| Project title | | Aligning Life Cycle Assessment methods and bio-based sectors for improved environmental performance. | |
| Project acronym | ALIGNED | Start / Duration | 01/10/2022 – 36 months |
| Type of Action | RIA | Website | www.alignedproject.eu |

| DELIVERABLE DETAILS | | | |
|----------------------------|----|------------------------|------------|
| Dissemination level | PU | Nature | DMP |
| Due date (M) | M6 | Submission date | 28/03/2023 |

| DELIVERABLE CONTRIBUTORS | | | |
|--|------------------|--------------|--|
| | Name | Organisation | Job title |
| Deliverable leader | Agneta Ghose | AAU | Assistant Professor and in charge of data management for ALIGNED |
| Contributing Author(s) | Massimo Pizzol | AAU | Project coordinator |
| Reviewer(s) | Flora Champetier | AAU | Project Manager |
| Final review and quality approval | Flora Champetier | AAU | Project Manager |

| DOCUMENT HISTORY | | | |
|------------------|---------|--------------|--------------------------------------|
| Date | Version | Name | Changes |
| 21/02/2023 | 0.1 | Agneta Ghose | First version created in Argos |
| 21/03/2023 | 0.2 | Agneta Ghose | Changes implemented following review |
| 28/03/2023 | 1.0 | Agneta Ghose | Submission to EC |

TABLE OF CONTENTS

| | |
|--|----|
| ALIGNED | 4 |
| Funder | 4 |
| Grant | 4 |
| Researchers | 4 |
| Organizations | 5 |
| Datasets | 6 |
| Dataset 1 - Title: ALIGNED_he_Case study data collection | 6 |
| Dataset 2 - Title: ALIGNED_he_Sector data collection | 16 |
| Dataset 3 - Title: ALIGNED_he_WP1 Models and data | 26 |

ALIGNED

Version 1.0

Funder

European Commission | | EC

Grant

Aligning Life Cycle Assessment methods and bio-based sectors for improved environmental performance/ No 101059430

Researchers

Agneta Ghose (orcid:0000-0003-1972-1433), Søren Løkke (orcid:0000-0002-4544-1788), Massimo Pizzol (orcid:0000-0002-7462-2668)

Organizations

BLOOM BIORENEWABLES SA, Sustainable Innovations, LLC, Aalborg University, Norwegian Uni of Science and Technology, UTEXBEL, OLEON, INSA de Toulouse, BTG - Biomass Technology Group, KINGSPAN, CENTRE SCIENTIFIQUE & TECHNIQUE DE L'INDUSTRIE TEXTILE BELGE, University of Antwerp, Oleon NV, Foreco (Netherlands)

Datasets

Dataset 1 - Title: ALIGNED_he_Case study data collection

Template: [Horizon Europe](#)

The ALIGNED project has collaborated with six industry partners that provide primary data to develop case studies in work packages (WP 2-6)

The following sections will outline the type of data that will be collected and generated from specific case studies and how this data will be made accessible for re-use during the project and upon the closure of the project.

The data currently collected has not been used in any publication or research output. This will change during the course of the project. The DMP is a living document, where research outputs related to the data will also be updated.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This dataset will include 5 datasets collected from each of the five case studies. Although this data might be considered a primary data source here, it may have been used by the data providers in different research contexts. Hence, the data collected in this dataset is assumed to be re-used.

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.5 What is its format?

- Microsoft Excel 2000-2003 Workbook (xls)
- Acrobat PDF 1.0 - Portable Document Format

- Comma Separated Values

1.1.6 What is its expected size?

The expected volume of the data is presently unknown, but it is foreseen that it will stay small enough to be stored in the standard available servers used in the project.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To make informed decisions
- To improve a product
- To combine with other data

1.1.8 What is its origin / provenance?

Case study data obtained with private correspondence from:

WP2 Kingspan (Thermal insulation alternatives)

WP3 Foreco (Timber for facade and fencing)

WP4 Utextbel and Centexbel (Textile recycling)

WP5 BLOOM Biorenewables (Lignin valorisation)

WP6 OLEON and OLEON NV (oleochemical production)

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers
- Industry

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Data identifiers
- Researchers identifiers
- Projects identifiers
- None

DOI

ORCID

Cordis

All relevant metadata will be registered on Zenodo which generate DOI. In addition, all datasets and research outputs linked to Zenodo will also be linked to the [ALIGNED community](#) to ensure easy findability.

Researchers associated with developing a dataset will need to link their ORCID ID with datasets stored on Zenodo.

The [Cordis ID](#) for the project will be referred to in the metadata of each output.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Reference

Two types of metadata will be provided:

1) The meta data mandatory for datasets deposited in [Zenodo](#) - is compliant with DataCite's Metadata Schema minimum and recommended terms with additional enrichments.

2) In addition, all LCI data will adhere to [mandatory metadata descriptors](#) defined by GLAD ([Global Data Access network](#)).

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

3.1.1.5 Please provide URL/Description of used vocabularies

<https://www.globallcadataaccess.org/sites/default/files/Table%201-proposed%20metadata%20descriptors.pdf>

The link above provides a description of the GLAD metadata descriptors.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository

Ensuring the LCI data is defined using GLAD metadata descriptors, enables the dataset to be linked to GLAD which acts as a catalogue of LCI specific data.

In addition, the data is also searchable using Zenodo's metadata description, which allows the registration of keywords. However, the Zenodo metadata schema is much broader and generic.

3.1.1.8 Are keywords provided in the metadata?

Yes

bio-based, pulp and paper, lignin, textile, construction, oleochemicals, woodworking, insulation

3.1.1.9 Are metadata harvestable?

Yes

As the datasets will be deposited in [Zenodo](#), *"the metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name."*

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

Zenodo

Once finalized datasets will be available on Zenodo. All uploads can be found on <https://zenodo.org/communities/aligned-he/>

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy
- Supports back up
- Assigns PIDs
- Follows metadata standards

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

All datasets uploaded will be linked to the Zenodo community, making all research outputs easy to search <https://zenodo.org/communities/aligned-he/>.

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

PIDs to be used as defined in section 3.1.1.1

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

HE_ALIGNED_CaseStudyData

3.2.2.2 How is the dataset / output shared?

Closed

Data collected from case study will have closed or restricted access. If necessary, the closed data can be accessed by applying the principle of least privilege, i.e., access will be strictly provided to data concerned and not the entire dataset.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Primary data collected from participating industries is sensitive and opening their data goes against their legitimate interests. Provision to keep this data closed or restricted was mentioned in the Grant agreement.

However, data generated/calculated using the case study data will be made available using an open license ([CC-BY-4.0](#)).

Based on the agreement with project partners primary data in certain instances could be anonymized or shared under a restrictive license such as shared only by attribution for non-commercial and non-derivative purposes ([CC-BY-NC-ND](#))

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

Requests can be made to obtain access to the data from case studies by contacting project participants responsible for data management. These requests will be evaluated based on internal discussions directly with project partners in accordance with GA and CA provisions and granted only if the partners who provided the data approve the request.

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

During the project, information on access to the dataset from case studies can be obtained from:

1) Agneta Ghose (agneta@plan.aau.dk)

OR

2) Flora Champetier (fch@adm.aau.dk)

Data access for primary data from case studies after the project ends has not yet been determined and will be discussed.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes, all meta data will be available on Zenodo and GLAD to ensure findability and possibility to access data if deemed necessary (e.g., a review of models generated using this data)

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

Access to restricted or closed files is limited, and these files will not be made available publicly. However, access to datasets if necessary could be made possible by the approval of the dataset curator/depositor. The name and contact details of data depositor and curator are provided in the metadata.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Meta data is mandatory even for closed and restricted access dataset (s) uploaded on Zenodo. All datasets (and metadata) published on [Zenodo](#) will be maintained for the lifetime of the repository (approximately 20 years).

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

DataCite Metadata Schema

Besides using the DataCite Meta data schema, the [GLAD metadata descriptors](#) will also be used.

3.3.5 What is the methodology followed?

1. First the dataset will be uploaded on Zenodo. All uploads will provide the persistent identifiers and/or URL to the dataset. The URLs are a mandatory metadata defined by GLAD.

2. All mandatory GLAD metadata descriptors will be defined using the [template provided on the GLAD website](#).

3. Among the various compulsory descriptors for GLAD, it is necessary to provide a unique identification number (besides the DOI to the dataset.). This will be generated from <https://www.uuidgenerator.net/>

4. The completed template will be uploaded on the GLAD website.

3.3.6 What community-endorsed interoperability best practices are followed?

Currently GLAD is the largest LCA network supporting accessibility and interoperability. By linking the project outputs to GLAD is the current best practice.

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

All research outputs linked to the case study datasets will be cross referenced. There is a provision to specify identifiers of related publications and datasets on the DataCite metadata schema provided by Zenodo.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution-NonCommercial 4.0

Given that most of the case study datasets will be closed or restricted access, the data will not be available publicly. If the access to data is provided, it will be by Attribution and No commercial, i.e., the data from case studies cannot be used for other commercial purposes.

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Data cleaning
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

As mentioned all the data outputs will be stored on the Zenodo.

This repository was chosen as its infrastructure, policies and principles support the possibility to adhere to FAIR data practices.

3.4.5 Is provenance well documented?

Yes

3.4.6 What documented procedures for quality assurance do you have in place?

- Set up of scientific and technical committee
- Consistency verified with data models and standards

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

10000 Euro

- Archiving
- Other

person hours for metadata description

Direct cost

The costs mentioned here is a rough estimation of time required by researchers (level 2) for data collection, documentation (including data description and metadata)

There is no indication of a need to budget for repository storage outside of the scope of what is accepted by repositories (in the case of this project; internal AAU/project partner server) within their free limit quotas.

4.1.2 How will this cost be covered?

Infrastructure Grant

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Agneta Ghose (orcid:0000-0003-1972-1433)

5.1 Data Security

5.1.1 What security measures are followed?

- **Dataset labelling** - All confidential information must be labelled so that it is protected against unintentional disclosure. Documents must be labelled on the cover sheet, the file or folder name.

- **Dataset Access security** - Electronic access to the dataset is available only to project partners on password protected web-based platform (Sharepoint).

- **Dataset sharing security** - It is recommended that internal data be encrypted before transmission. In addition, senders must ensure that the recipient is aware of the rules governing the processing of the information received.

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data storage (long term) security - long term data storage to be made on Aalborg University (AAU) server or by partners with whom an AAU-approved data processing agreement for the storage of confidential data has been made.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

No

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary
- Data accompanied by informed consent statements
- Privacy policies

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Dataset 2 - Title: ALIGNED_he_Sector data collection

Template: Horizon Europe

The ALIGNED project focusses in work packages (WP 2-6) on five industrial sectors which use bio-based materials. These sectors are: construction, wood working, textile, pulp and paper, and biochemicals.

The following sections will outline the type of data that will be collected and generated from each sector and how this data will be made accessible for re-use during the project and upon the closure of the project.

The data currently collected has not been used in any publication or research output. This will change during the course of the project. The DMP is a living document, where research outputs related to the data will also be updated.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This dataset will include sector wide data on the environmental impact each of the five sectors to be assessed in this study. This would include information on both conventional and bio-based products/alternatives. Focus will be on defining the current and expected market trends including the involved product portfolio, resources used, co-products generated and overall outputs and impacts of the sector.

1.1.4 What is the type of the described dataset?

Derived or compiled

Data will be derived from market foresights, inputs from venture funds specific to the sector, scientific and other external reports (such as CEPI, JRC, EU BBI studies, Ellen MacArthur Foundation) and open access databases such as Eurostat and EC Cordis.

1.1.5 What is its format?

- Microsoft Excel 2000-2003 Workbook (xls)
- Acrobat PDF 1.0 - Portable Document Format
- Comma Separated Values
- Microsoft Word for Windows Document

Numerical data or tabular data with minimal metadata will be saved in CSV formats. While Tabular data with extensive metadata will be saved in XLS format. Other options such as JSON and rdf format may be considered in the later stages of the project.

1.1.6 What is its expected size?

The expected volume of the data is presently unknown, but it is foreseen that it will stay small enough to be stored in the standard available servers used in the project. There is no indication of a need to budget for repository storage outside of the scope of what is accepted by repositories (in the case of this project; Zenodo) within their free limit quotas.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To make informed decisions
- To combine with other data

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers
- Industry

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Data identifiers
- Researchers identifiers
- Projects identifiers

DOI

ORCID

Cordis

All datasets will be registered on Zenodo which generate DOI. In addition, all datasets and research outputs linked to Zenodo will also be linked to the [ALIGNED community](#) to ensure easy findability.

Researchers associated with developing a dataset will need to link their ORCID ID with datasets stored on Zenodo.

The [Cordis ID](#) (Grant number) for the project will be referred to in the metadata of each output.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Reference

Two types of metadata will be provided:

1) The meta data mandatory for datasets deposited in [Zenodo](#) - is compliant with DataCite's Metadata Schema minimum and recommended terms with additional enrichments.

2) In addition, all LCI data will adhere to [mandatory metadata descriptors](#) defined by GLAD ([Global Data Access network](#))

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

3.1.1.5 Please provide URL/Description of used vocabularies

<https://www.globalcadataaccess.org/sites/default/files/Table%201-proposed%20metadata%20descriptors.pdf>

The link above provides a description of the GLAD metadata descriptors.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository

Ensuring the LCI data is defined using GLAD metadata descriptors, enables the dataset to be linked to GLAD which acts as a catalogue of LCI specific data.

In addition, the data is also searchable using Zenodo's metadata description, which allows the registration of keywords. However, the Zenodo metadata schema is much broader and generic.

3.1.1.8 Are keywords provided in the metadata?

Yes

bio-based, pulp and paper, lignin, textile, construction, oleochemicals, woodworking, insulation

Manufacturing and processing, Materials (wood), Eco-design, Life Cycle Analysis, Environment, resources and sustainability, Bio-based products (products), bio-based materials, bio-based plastics, biofuels, bio-based and bio-derived bulk and fine chemicals, bio-based and bio-derived novel materials

3.1.1.9 Are metadata harvestable?

Yes

As the datasets will be deposited in [Zenodo](#), *"the metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name."*

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

Zenodo

Once finalised datasets will be available on Zenodo. All uploads can be found on <https://zenodo.org/communities/aligned-he/>

In addition to depositing the data on Zenodo, the datasets will be searchable on the GLAD website which is a directory of LCA data. This will be possible as the datasets will be defined using [GLAD](#) recommended metadata descriptors.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy
- Supports back up
- Assigns PIDs
- Follows metadata standards

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

All datasets uploaded will be linked to the Zenodo community, making all research outputs easy to search <https://zenodo.org/communities/aligned-he/>

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

PIDs to be used as defined in section 3.1.1.1

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

HE_ALIGNED_SectorData_<WP title and sector name>

3.2.2.2 How is the dataset / output shared?

Open

Data collected from sectors will be shared as open access.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

While the data is being collected by project partners during the project, it will be shared and managed internally using Sharepoint.

During the project, information on specific access to the dataset (e.g., specific format for use in software) can be obtained from:

3) Agneta Ghose (agneta@plan.aau.dk)

OR

2) Flora Champetier (fch@adm.aau.dk)

Data access from sectors after the project ends will be maintained on the Zenodo repository.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

The longevity of the Zenodo repository is approximately 20 years.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

Yes, all meta data will be available on Zenodo and GLAD to ensure findability and possibility to access data if deemed necessary (e.g., a review of models generated using this data)

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

The name and contact details of data depositor and curator are provided in the metadata. In addition, the metadata include LCI specific metadata that will enable reusability specifically for future LCA research related to these sectors.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Meta data is mandatory for all dataset(s) uploaded on Zenodo. All datasets (and metadata) published on [Zenodo](https://zenodo.org/) will be maintained for the lifetime of the repository (approximately 20 years).

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

DataCite Metadata Schema

Besides using the DataCite Meta data schema, the [GLAD metadata descriptors](#) will also be used.

3.3.5 What is the methodology followed?

1. First the dataset will be uploaded on Zenodo. All uploads will provide the persistent identifiers and/or URL to the dataset. The URLs are a mandatory metadata defined by GLAD.
2. All mandatory GLAD metadata descriptors will be defined using the [template provided on the GLAD website](#).
3. Among the various compulsory descriptors for GLAD, it is necessary to provide a unique identification number (besides the DOI to the dataset.). This will be generated from <https://www.uuidgenerator.net/>
4. The completed template will be uploaded on the GLAD website.

3.3.6 What community-endorsed interoperability best practices are followed?

Currently GLAD is the largest LCA network supporting accessibility and interoperability. By linking the project outputs to GLAD is the current best practice.

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

All research outputs linked to the case study datasets will be cross referenced. There is a provision to specify identifiers of related publications and datasets on the DataCite metadata schema provided by Zenodo.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Data cleaning
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

As mentioned all the data outputs will be stored on the Zenodo.

This repository was chosen as its infrastructure, policies and principles support the possibility to adhere to FAIR data practices.

3.4.5 Is provenance well documented?

Yes

3.4.6 What documented procedures for quality assurance do you have in place?

- Set up of scientific and technical committee
- Consistency verified with data models and standards

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

10000

Euro

- Archiving
- Other

person hours for metadata description

Direct cost

The costs mentioned here is a rough estimation of time required by researchers (level 2) for data collection, documentation (including data description and metadata).

There is no indication of a need to budget for repository storage outside of the scope of what is accepted by repositories (in the case of this project; Zenodo) within their free limit quotas.

4.1.2 How will this cost be covered?

Infrastructure Grant

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Agneta Ghose (orcid:0000-0003-1972-1433)

5.1 Data Security

5.1.1 What security measures are followed?

Other

Data security related to access and storage not required as the data will be available in the public domain and disclosure is not harmful to involved project partners.

5.1.2 What conditions do the security measures meet?

- Data access
- Data recovery
- Data sharing

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

No

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary
- Data accompanied by informed consent statements
- Privacy policies

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Dataset 3 - Title: ALIGNED_he_WP1 Models and data

Template: Horizon Europe

The ALIGNED project aims to improve, harmonise and align LCA methodology for bio-based industries. The project's first work package (WP1) will develop a framework and harmonised models such as integrated assessment models to develop scenarios, impact assessment including dynamic and time dependent carbon accounting, guide to uncertainty analysis, socio-economic assessment etc.

The following sections will outline how the models developed in this study will be shared and how they will be made accessible for re-use during the project and upon the closure of the project.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Models

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

Existing models relevant in the bio-based sector such as those used for developing scenarios (e.g., integrated assessment models), time dependent carbon accounting, land use, biodiversity indicators will be reviewed and aligned to develop a new framework that could be easily implemented with respect to LCA.

1.1.4 What is the type of the described dataset?

Derived or compiled

WP1 will build a common approach in terms of modelling choices and tools for the foreground and background modelling of bio-based processes common to all five sectors.

Building on a consequential LCA framework, the task will make operational analytical methods and tools for 1) identification of constraints to supply of biomass, considering four factors: geography, production capacity, co-production, and policy, by automating via computer simulation network and regression analysis on trade and production data respectively, and

2) consistent, time-dependent, and mass-balanced carbon accounting. The methods and tools will be made operational for comfortable use in the project case studies and easy integration with LCA software.

Moreover, WP 1 will build an advanced LCA model interfacing with outputs from Integrated Assessment Models (IAMs). This will be achieved through a set of programming routines updating life cycle inventories of production processes (energy, electricity, transport, etc.) with open source IAM-derived projections of changes in technologies and energy systems.

A common comprehensive set of impact categories will be used in the assessment from existing LCIA methods (e.g., Recipe, ILCD), time and space dependent characterization factors for Global Warming Potential (GWP) and Global Temperature Potential (GTP) impacts will be generated to match the time-dependent inventories. This is possible with existing models that, however, need updating and especially need to be made operational for comfortable use in LCA software and in case studies.

1.1.5 What is its format?

- Microsoft Excel 2000-2003 Workbook (xls)
- Acrobat PDF 1.0 - Portable Document Format
- Comma Separated Values
- Microsoft Word for Windows Document

Numerical data or tabular data with minimal metadata will be saved in CSV formats. While Tabular data with extensive metadata will be saved in XLS format. Other options such as JSON and rdf format may be considered in the later stages of the project.

1.1.6 What is its expected size?

The expected volume of the data is presently unknown, but it is foreseen that it will stay small enough to be stored in the standard available servers used in the project. There is no indication of a need to budget for repository storage outside of the scope of what is accepted by repositories (in the case of this project; Zenodo) within their free limit quotas.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To make informed decisions
- To combine with other data

1.1.8 What is its origin / provenance?

All sources used to develop the model will be reported in the metadata.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers
- Industry

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Data identifiers
- Researchers identifiers
- Projects identifiers

DOI

ORCID

Cordis

All datasets (including models) will be registered on Zenodo which generate DOI. In addition, all datasets and research outputs linked to Zenodo will also be linked to the [ALIGNED community](#) to ensure easy findability.

Researchers associated with developing a dataset will need to link their ORCID ID with datasets stored on Zenodo.

The [Cordis ID](#) (Grant number) for the project will be referred to in the metadata of each output.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Reference

The meta data mandatory for datasets deposited in Zenodo - is compliant with DataCite's Metadata Schema minimum and recommended terms with additional enrichments.

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository

The models will be searchable using Zenodo's metadata description, which allows the registration of keywords. In addition, all deposits on Zenodo will be linked to the ALIGNED community built on Zenodo (<https://zenodo.org/communities/aligned-he/>)

3.1.1.8 Are keywords provided in the metadata?

Yes

bio-based, pulp and paper, lignin, textile, construction, oleochemicals, woodworking, insulation

Manufacturing and processing, Materials (wood), Eco-design, Life Cycle Analysis, Environment, resources and sustainability, Bio-based products (products), bio-based materials, bio-based plastics, biofuels, bio-based and bio-derived bulk and fine chemicals, bio-based and bio-derived novel materials

3.1.1.9 Are metadata harvestable?

Yes

As the datasets will be deposited in [Zenodo](#), *"the metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name."*

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

Zenodo

Once finalized datasets will be available on Zenodo. All uploads can be found on <https://zenodo.org/communities/aligned-he/>.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy
- Supports back up
- Assigns PIDs
- Follows metadata standards

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

All datasets uploaded will be linked to the Zenodo community, making all research outputs easy to search <https://zenodo.org/communities/aligned-he/>

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

PIDs to be used as defined in section 3.1.1.1

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

ALIGNED_he_WP1<model name>

3.2.2.2 How is the dataset / output shared?

Open

Models and workflows generated will be shared as open access.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

All data related to model and workflow development are being collected by project partners during the project, it will be shared and managed internally using Sharepoint.

During the project, information on specific access to the dataset (e.g., specific format for use in software) can be obtained from:

1) Agneta Ghose (agneta@plan.aau.dk)

OR

2) Flora Champetier (fch@adm.aau.dk)

Data access from sectors after the project ends will be maintained on the Zenodo repository.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

The longevity of the Zenodo repository is approximately 20 years.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

Yes, all meta data will be available on Zenodo to ensure findability and possibility to access data if deemed necessary (e.g., a review of models generated using this data)

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

The name and contact details of data depositor and curator are provided in the metadata. In addition, the metadata will include LCI specific metadata that will enable reusability specifically for future LCA research related to these sectors.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Meta data is mandatory for all dataset(s) uploaded on Zenodo. All datasets (and metadata) published on [Zenodo](https://zenodo.org/) will be maintained for the lifetime of the repository (approximately 20 years).

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

DataCite Metadata Schema

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

All research outputs will be cross referenced. There is a provision to specify identifiers of related publications and datasets on the DataCite metadata schema provided by Zenodo.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Data cleaning
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

As mentioned all the data outputs will be stored on the Zenodo.

This repository was chosen as its infrastructure, policies and principles support the possibility to adhere to FAIR data practices.

3.4.5 Is provenance well documented?

Yes

3.4.6 What documented procedures for quality assurance do you have in place?

- Set up of scientific and technical committee
- Consistency verified with data models and standards

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

5000

Euro

- Archiving
- Other

person hours for metadata description

Direct cost

The costs mentioned here is a rough estimation of time required by researchers (level 2) for data collection, documentation (including data description and metadata).

There is no indication of a need to budget for repository storage outside of the scope of what is accepted by repositories (in the case of this project; Zenodo) within their free limit quotas.

4.1.2 How will this cost be covered?

Infrastructure Grant

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Agneta Ghose (orcid:0000-0003-1972-1433)

5.1 Data Security

5.1.1 What security measures are followed?

Other

Data security related to access and storage not required as the data will be available in the public domain and disclosure is not harmful to involved project partners.

5.1.2 What conditions do the security measures meet?

- Data access
- Data recovery
- Data sharing

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

No

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary
- Data accompanied by informed consent statements
- Privacy policies

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by

