DATA MANAGEMENT PLAN

Deliverable 9.3 – D43 – WP9

DATE OF PUBLICATION: 1.12.2019

RESPONSIBLE PARTNER: LUKE

AUTHORS: HANNA LINDQVIST AND KARI YLIVAINIO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 818309 (LEX4BIO). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein

OPTIMISING BIO-BASED FERTILISERS IN AGRICULTURE – PROVIDING A KNOWLEDGE BASIS FOR NEW POLICIES

Project funded by the European Commission within the Horizon 2020 programme (2014-2020)

Deliverable 9.3 – D43 – Version 1

Work-package n°9

Nature of the deliverable		
R	Report	X
Dec	Websites, patents, filling etc.	
Dem	Demonstrator	
O	Other	

Dissemination Level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

Acknowledgement

This report forms part of the deliverables from the LEX4BIO project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 818309. The Community is not responsible for any use that might be made of the content of this publication.

LEX4BIO aims to reduce the dependence upon mineral/fossil fertilisers, benefiting the environment and the EU's economy. The project will focus on collecting and processing regional nutrient stock, flow, surplus and deficiency data, and reviewing and assessing the required technological solutions. Furthermore, socioeconomic benefits and limitations to increase substitution of mineral fertiliser for BBFs will be analysed. A key result of LEX4BIO will be a universal, science-based toolkit for optimising the use of BBFs in agriculture and to assess their environmental impact in terms of non-renewable energy use, greenhouse gas emissions and other LCA impact categories. LEX4BIO provides for the first-time connection between production technologies of BBFs and regional requirements for the safe use of BBFs.

The project runs from June 2019 to May 2023. It involves 21 partners and is coordinated by Luke (Luonnonvarakeskus - Natural Resources Institute Finland).

More information on the project can be found at: http://www.lex4bio.eu

TABLE OF CONTENTS

Version Log		
Version	**Publication date**	**Change / Author**
1.0	11/2019	▪ First draft/Hanna Lindqvist
1.1	11/2019	▪ Review according to quality check by Manon Ballester/EP

D9.3: DATA MANAGEMENT PLAN

I. DATA SUMMARY

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The purpose of the data collection is to evaluate the suitability of different types of bio-based fertilisers (BBF) for replacing mineral phosphorus and nitrogen fertilisers in different climatic conditions across the EU.
Mapping of the regional distribution of nutrient-rich side-streams (NRSS) and technologies for producing BBFs out of NRSS requires collection of the available data. Laboratory and growth experiments for evaluating BBFs effect on soil quality and crop growth produces data for providing information of their safe and efficient use, following by life cycle analyses for their environmental impact when considering the whole chain from manufacturing BBFs up to the crop production, followed by the people's perception towards BBFs. Taken these together, policy recommendations are provided for enhancing the use of BBFs and thus closing the nutrient cycles.
Data is also collected to boost the exploitation of the project results; these include also personal data, which are handled in compliance with the GDPR regulation. These issues are covered in the ethics review as well.

What types and formats of data will the project generate/collect?

The size of the data can only be estimated, based on the current technology and knowledge. Here a forecast of the expected size of the data by WP:

WP1: Dataset and related report on nutrient rich side streams (NRSS) available for producing BBFs in EU countries (national and, where available, regional level) – Excel file and Word document, sizes unknown, they will be defined later in the project.
Dataset and report on mineral NP fertilizer use in EU countries, based on official fertilizer statistics - Excel file and Word document (sizes unknown). Distribution maps of NRSS and mineral fertilisers in the EU – file type and size unknown
Dataset and report on legal regulations concerning BBFs in the EU (national and EU regulations) – Excel file and Word document, sizes unknown
Dataset and related report on case studies on existing interregional and transboundary exchange of NRSS/BBFs for selected regions – Excel file and Word document, sizes unknown

WP2: Expected size probably few GB; replies to questionnaire, database on BBFs, database on soil quality, results and pictures from container and phenotyping studies -> mainly .xls and .jpg files; presentations on conferences etc.; publications

WP3: Expected size about 5 GB. Experimental results in Excel, reports in Word, presentations in Powerpoint. Chloropleth maps.

WP4: ca. 1,75 TB, mainly satellite imaging data.
WP5: Data file on organic contaminations in the collected bio-based Fertilizers (BBFs), data file on heavy metal levels in crops after use of BBFs. The size of the files will be defined later in the project.

WP8: < 3GB, mainly database of contacts, event participants, list of communication activities and stakeholders involved, database of pictures

WP6, WP7, WP9: Will be defined later

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

The data are useful for the scientific community as well as to stakeholders benefiting from them.

Will you re-use any existing data and how?

LEX4BIO will reuse the following existing data belonging to the following organisations, as described in Attachment 1 (Background included) in the consortium agreement:

LUKE: Nutrient, heavy metal and carbon concentrations from long-term soil monitoring study (National monitoring of arable soil chemical quality, Valse) dataset, sampled in years 1974, 1987, 1998, 2009 and 2018.

JKI: National N and P balances compiled for the German ministry of agriculture (BMEL) and for EUROSTAT.

UCPH: Soil nutrient, trace element and carbon concentrations as well as yield data from the Long-term CRUCIAL (https://plen.ku.dk/english/research/plant_soil/sf/crucial/) field trial with agricultural and urban biobased fertilisers, dataset from the period 2002-2018. Soil N&P availability as well as yield data from the Long-term Nutrient Depletion (https://plen.ku.dk/english/research/plant_soil/sf/long-term-trial/) field trial with NPK and animal manure application, dataset from the period 1996-2018.

UHOH: Yield data, soil data and soil samples from previous field (2010-2020) and greenhouse (2013-2016) experiments, mainly with regard to P concentrations.

USE: Soil samples for analysis.

IGSMIE PAN: Data coming from statistic and fertilizers-related regional/ national/ international offices, reviewed articles.

RUOKAVIRASTO: Dietary exposure data from published (Suomi et al. 2015) and ongoing (Suomi et al., manuscript) risk assessments as well as from EFSA reports. Heavy metal occurrence data collected in national monitoring programs. Unknown at this point how much of the raw data can be made public for confidentiality reasons; the reports mentioned are already publicly available / will be available after completion.

What is the origin of the data?

LUKE: National monitoring of arable soil chemical quality, Valse
JKI: Data from national and international statistics (DESTATIS, EUROSTAT) and reviewed papers, from agricultural services and personally collected Data produced and compiled by JKI from several official national sources
UCPH: Previous field experiments
UHOH: Previous field (2010-2020) and greenhouse (2013-2016) experiments
USE: Previous field and laboratory experiments
IGSMIE PAN: Previous socioeconomic research on stakeholders related to fertilizers, data from national/ international statistics and reviewed articles
RUOKAVIRASTO: Published and ongoing risk assessments (reports and raw data thereof), occurrence data from national monitoring programs, literature

II. FAIR DATA

2.1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

The data produced and used in the project will either be confidential or public. Both types of data will be identifiable and locatable by the standard identification mechanisms, however, confidential data will not be available to any third parties.

What naming conventions do you follow?

Naming conventions for the data produced will be defined by the WP leaders.

WP1: not decided yet – will follow central decision made by executive board and project coordinator.

WP2: Folder with all data generated in WP2, subfolders for subtasks. All experiments are assigned a number. File names describe content as detailed as possible and contain dates of collection or creation.

WP4: The folder with all data generated in WP4 will be structured as consistent as possible through the whole duration of the project. The folder will contain sub-folders with the project name, experiment order, the name of the experiment and the date (YYYYMMDD). The conventions followed for files will be the same as for folders.

WP7: The folder with all data collected in WP7 will be updated according to the progress of the project realisation. The folder will contain sub-folders with the sub-tasks name as WP7.1,…, WP7.5. Sub-folders will include all files needed during the project realisation, with the specific dates of up-dating the files (YYYYMMDD).

WP8: The folder with all data collected in WP8 will be updated according to the progress of the project realisation. All information related to contacts and participants will be stored in a specific folder for task T8.3.2 identifying the purpose of their collection. In addition, a specific folder "External Communication" will include several sub-folders with pictures and results from participation to events and communication activities where will be specified the date, name of the event, participants, country of venue and if needed, credentials for the picture reuse.

WP3, WP5, WP6, WP9: Will be defined later

Will search keywords be provided that optimize possibilities for re-use?

WP7: The newest literature (update from proposal submission to project realization) will be searched and associated with the use of a few keywords: 'waste management', 'waste', 'wastewater', 'sewage', 'circular economy', 'CE', 'reuse', 'recycling', 'removal', 'reclamation', 'recovery', 'fertilizer', 'bio based fertilizer'.

Other WPs: Will be defined later

Do you provide clear version numbers?

Version numbers will be provided when organisational or standards provide methodology for the version handling.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

WP1: Dataset on availability and nutrient contents of NRSS in the EU; Dataset on legal regulations concerning the use of BBFs in the EU

WP3: Metadata will be based on the needs of users to find, select and use the data as well as to facilitate the management of the data after the project is finished. In each experiment folder, a readme file will be included with specifications of the raw and treated data.

WP4: Metadata will be based on the needs of users to find, select and use the data as well as to facilitate the management of the data after the project is finished. In each experiment folder, a readme file will be included with specifications of the raw and treated data.

WP5: Data file on organic contaminations of BBFs; Toxicological data of BBFs

WP8: Metadata will mainly apply to the pictures collected by the partners. It will include technical metadata generated automatically by the camera (date and time, author, camera …) and administrative/descriptive metadata as picture name, comments, license & usage information, that will be collected through a one-page documentation of the communication activity related to the picture.

WP9: The metadata of the collected data at Luke will be documented at the Luke metadata service RADAR that is openly available at the Luke website (radar.luke.fi). RADAR allows flexible searching for

the metadata descriptions, which include the core information about the dataset to facilitate the future use of the data. RADAR uses ISO19115 and Inspire standards for the metadata descriptions. The data includes the detailed technical documentation, according to which the data can be reused if needed.

WP2, WP6, WP7: Will be defined later

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

According to the publication and data policies of the consortium, as agreed in the Consortium Agreement, data produced during this project will mainly be open to the scientific community and society as a whole.

Reasons for data not being shared may include commercial exploitation, protection of IPR, privacy policy or other legislation.

Scientific publications will be published in Open Access publications, according to general H2020 guidelines.

Individual partners will be responsible for the collection, transfer and storage of data from their own research activities, within the framework of the DMP. The DMP will outline access rights for consortium participants to data held by other consortium members as agreed in the Consortium Agreement.

How will the data be made accessible (e.g. by deposition in a repository)?

This is to be updated as the project goes along. Regarding public deliverables and results, the project website acts a repository and as a way to share project publications and key outcomes.

What methods or software tools are needed to access the data?

This is to be updated as the project goes along.

Is documentation about the software needed to access the data included?

For most cases at the moment, there is no documentation about the software needed to access data. This is to be updated as the project goes along.

Is it possible to include the relevant software (e.g. in open source code)?

For most WP, this is not currently relevant. This is to change as the project goes along.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The data collected and intended as open data produced at **Luke** will be made accessible via downloading from a special service called Luke Open Data (opendata.luke.fi). Datasets will receive permanent identifiers in the downloading service. The documented research data collected in Luke will be preserved permanently in cooperation with the National Archives of Finland. Other repositories are used for long-term preservation and sharing when possible.

Data produced at **UCPH** will be deposited in the University's Electronic Research Data Archive (ERDA), providing possibilities to store and share research data. If a dataset is made public, it is possible to get a DOI for the dataset.

Have you explored appropriate arrangements with the identified repository?

Where needed, it will be explored.

If there are restrictions on use, how will access be provided?

Currently, there is no limited restriction noted for most WPs.

Is there a need for a data access committee?

Currently, there is no need for implementing a data access committee.

Are there well described conditions for access (i.e. a machine readable license)?

N/A

How will the identity of the person accessing the data be ascertained?

Most open data will be accessible unrestricted and there is no need for registration/identification. When restricted access is needed, the appropriate protocols for data protection and safety will be followed.

2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

After an initial internal evaluation by the LEX4BIO consortium, findings will be sorted in open access, consisting of data that can be freely used outside the project and data/findings of potential patentable commercial interest. The later will be made available after publication or after patents have been filed.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

This is to be updated as the project goes along.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

For all WPs: Standard vocabulary will be used, meaning that the data can be understood and used in new contexts and research areas different from the one that it was created in

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?

This is to be updated as the project goes along.

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

This is to be updated as the project goes along.

When will the data be made available for re-use?

Unless otherwise stated, data will be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?

The majority of the data will be open to third parties, both during and after the execution of the project. Some data will be restricted. Sole rights for exploitation of the restricted data will be retained by the respective partners that created the data. In instances where the data is restricted it will be to ensure that the respective partners are able to properly exploit the data commercially in their own interests and for development of their own products.

How long is it intended that the data remains re-usable?

The data will remain re-usable for the foreseeable future.

Are data quality assurance processes described?

Data quality assurance processes will be clarified at various stages later in the project, and will be included in updated versions of the DMP where necessary.

III. ALLOCATION OF RESOURCES

What are the costs for making data FAIR in your project?

LUKE: The data stored to the Luke Open Data is free of charge.

UCPH: The repository of Copenhagen University is free of charge.

Regarding other repositories and related costs to make the data Findable, Accessible, Interoperable and Reusable, it will be clarified later.

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

For open access publication costs, 10000 € is reserved for all partners, included in Luke´s budget.

Who will be responsible for data management in your project?

Each partner is responsible for the data collection and management, with the coordinator being overall responsible.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

This is to be updated as the project goes along.

IV. DATA SECURITY

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? Address data recovery as well as secure storage and transfer of sensitive data

UCPH: Apart from being saved in the repository, the data will be saved in the server of the Plant and Environmental Science Department at the Faculty of Science, Copenhagen University, in order to secure storage and for possible data recovery.

The data (personal data and storage thereof) used internally within the project is stored on a separate share point Tiimeri. Tiimeri is a collaboration service environment for flexible and secure collaboration across organizations and ecosystem boundaries globally. Every user is identified and access is restricted.

Is the data safely stored in certified repositories for long term preservation and curation?

This is to be updated as the project goes along.

V. ETHICAL ASPECTS

Are there any ethical 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).

LEX4BIO research in some cases involves personal data collection and processing. The project partners ensure the respect of personal data (e.g. survey participants' names and contact details). The survey method in WP7 and the collection of personal data for communication campaign (especially newsletter contact list) complies with generally accepted ethical standards and principles, applicable international, EU and national law (in particular, EU Directive 95/46/EC) with GDPR (Regulation (EU) 2016/679) and the necessary authorisations for collecting and processing the data.

The personal data will only be collected and processed to the degree necessary for the research, and unauthorised access to the data will be prevented.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

A project-specific privacy notice, a record of processing activities and a risk assessment will be formulated to open and clarify the use of personal data. The personal data collected for the project will only be stored as long as necessary in relation to the original purposes and will be protected from security breach.
In the framework of WP8 regarding the collection of personal data for the creation of the newsletter database, the contacts will be stored in an internal database, shared only between PC and WP8 leader. These contacts will be used only to require a formal and active consent on integrating their information in largest dissemination & communication campaign. Through this consent formal request, a strict policy will be applied regarding the suppression of all email addresses and personal information from the internal database that will disagree with the use of their information in full compliance with GDPR rules. The contacts not responding to the consent campaign will be also rejected as positive, active consent will be required.

VI. OTHER ISSUES

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

The DMP will ensure that data management and protection is compliant with EU principles and standards, and with relevant national data protection laws and institutional data management policies. Produced data will be treated according to the management guidelines.

According to the publication and data policies of Luke (the data policy is available at www.luke.fi/palvelut/luken-aineistopolitiikka/), research data produced in this project will be made (fully or partially) available both for the scientific community and the society as a whole. The data producer's (members of the research team) right to the use of research data is reserved when providing open access. Right to use here refers to the data producer's right to execute the original project plan (including possible protection measures) before opening the data for further use.