

Next FOOD

EDUCATING THE NEXT GENERATION
OF PROFESSIONALS IN THE AGRIFOOD SYSTEM

D 6.2: Data Management Plan

WP6 – Communication, dissemination and exploitation



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 771738

The present Deliverable reflects only the author's view and the Research Executive Agency is not responsible for any use that may be made of the information it contains

Document Information

Grant Agreement	771738	Acronym	NextFOOD
Full Project Title	Educating the next generation of professionals in the agrifood system		
Start Date	15/03/2018	Duration	48
Project URL	http://www.nextfood-project.eu/		
Deliverable	D 6.2: Data Management Plan		
Working Package	WP6 – Working Package Title Here		
Date of Delivery	Contractual	31/10/2018	Actual 31/10/2018
Nature	R – Report etc.	Dissemination Level	P - Public
WP Leader	American Farm School		
Authors	Philippos Papadopoulos Vicky Krystallidou Vassilios Tsekeridis Giorgos Gkiouzepas		
Contributors	Jiří Lehejček		

Document History

Version	Issue Date	Stage	Changes	Contributor
0.1	20/10/2018	Draft	Comments on Datasets, Data Naming Processes	Jiří Lehejček
1.0		Final	Review	

Table of Contents

Executive summary.....	5
Project abstract:.....	6
1 Data summary.....	7
2 Findable, Accessible, Interoperable, Reusable Data for the NextFOOD Project 11	
2.1 Making data findable, including provisions for metadata	11
2.2 Making data openly accessible	13
2.3 Making data interoperable	16
2.4 Data re-use and licenses	16
3 Allocation of resources	18
4 Data security	20
5 Ethical aspects.....	21



List of Tables

Table 1: Recommended file formats per type of data to be used in the NextFOOD project.....	8
Table 2: Versioning rules to be used for the NextFOOD project	13

Executive summary

The following document is describing the data management cycle for the data sets that will be developed, collected, generated, and processed by the NextFOOD project. The document provided an overview of the data that will be generated and processed during the project and/or after its completion. This current document is the first version of a series and therefore is a starting point and reference document for the researchers that are going to take part in the NextFOOD project. As the project will progress more detailed and elaborate versions of this document will be developed to incorporate the developments of the projects as far data generation, development, processing, and storage are concerned.

Project abstract

NEXTFOOD will drive the crucial transition to more sustainable and competitive agrifood and forestry systems development by designing and implementing education and training systems to prepare budding or already practicing professionals with competencies to push the green shift in our rapidly changing society. NEXTFOOD will challenge the linear view of knowledge transfer as a top-down process from research to advice and practice, and support the transition to more learner-centric, participatory, action-based and action-oriented education and learning in agrifood and forestry systems. In several pioneering case studies, covering agrifood and forestry systems in Europe, Asia and Africa, farmers solve real challenges related to sustainability together with researchers, students and other relevant stakeholders while developing both green technical skills and soft collaborative competencies. NEXTFOOD will assure quality in research and education by creating a peer-review system for evaluation of practice-oriented research outputs focusing on sustainability and practical usefulness. In addition, we will develop an international accreditation framework for education and training in fields relevant to sustainable agrifood and forestry systems. An innovative action research process will guide the NEXTFOOD project's development in a cyclical manner, ensuring that the research process and actual case studies are ever-improving. This will exemplify how practice-oriented research can be instrumental to achieve: better collaboration between university and society, more innovation in the agrifood and forestry systems sector, and a progressive agrifood community ready to tackle complex sustainability challenges of the 21st century.

1 Data summary

Purpose of the data collection/generation

The NextFOOD Project's aim is to generate an innovative European science and education roadmap for sustainable agriculture along the value chain from research via fabrication into the application.

In order to achieve this goal, the project team will engage in several actions to ensure a meaningful result. These actions are organized in 7 work packages as follows

- WP1 Inventory of the skills needed for a transition to more sustainable agriculture, forestry and associated bio-value chains.
- WP2 Action research facilitation
- WP3 Future curriculum, education and training system
- WP4 Policy assessment and recommendations
- WP5 Quality assured knowledge transfer
- WP6 Communication, dissemination and exploitation
- WP7 Management
- WP8 Ethics requirements

For all the above Work packages and especially for the Work packages 1 to 5 data will be collected and produced during the NextFOOD project for

- research purposes
- development of new educational modules
- development of new educational material
- communication purposes within project partners and other interested bodies

Relation to the objectives of the project

The collection of data is closely related to the objectives of the project which are summarized as follows.

- Create an inventory of the skills and competencies needed for a transition to more sustainable agriculture, forestry and associated bio-value chains,
- Facilitate case studies to identify gaps and needs
- Test new relevant curricula and training methods
- Identify policy instruments that support the transition towards action-oriented, and practice-oriented learning methods
- Peer-review tools for evaluating the quality of the practice-oriented research
- Create a platform for knowledge sharing

An Annex will be provided by case studies' leader that will summarizing the Dataset of the project explaining in more detail the rationale of compiling the dataset and its purpose within the scope of the project

Types and formats of data generated/collected

The NextFOOD project will employ a number of Quantitative and Qualitative methods in order to realize its objectives.

For Qualitative methods, data created in audio and/or video will be transcribed and anonymized. Any hardcopy data products, such as participatory design outputs will be documented in a proper electronic medium (e.g. photographed).

Notetaking, journals etc. should be produced in an electronic textual data form. When this is not possible, originals should be documented in a proper electronic medium (e.g. photographed) and an electronic textual data form should also be produced.

To ensure the greater possible data sharing, reuse and preservation will adhere to the UK Data Service guidance on recommended formats summarized in the following table.

Table 1: Recommended file formats per type of data to be used in the NextFOOD project

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) Delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tiff) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages
Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti

Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)
Video data	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Documentation and scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHTML 1.0

Existing data to be re-used

No re-use of existing data is envisaged for the project so far.

Origin of the data

The data for the NextFOOD project will be generated by the project team through qualitative and quantitative methods. In particular, the core of the project is utilizing a participatory action research protocol, especially in WP2 case studies, in order to guide the NextFOOD project's development in a cyclical manner, ensuring that the research process and actual case studies are ever-improving.

In particular, during the participatory action research the following types of data collection are expected to be utilized:

participation, observation, recordkeeping, notetaking, surveying and profiling, semi-structured and informal interviewing, analysis of key reports, running focus groups, photographing and videoing, and journaling.

In addition to the above data generated by the NextFOOD project will include quantitative surveys through questionnaires.

Expected size of the data

The expected size of the data is not known.

Data utility

Data generated from the NextFOOD project is expected to be useful to the following entities:

- Educational Institutions
- Agricultural Advisory Services
- Policy and decision makers
- Agrifood industry
- Farmers, Farmers organisations
- Forestry Associations
- Research community
- Evaluators
- Project officers and project administration offices

2 Findable, Accessible, Interoperable, Reusable Data for the NextFOOD Project

2.1 Making data findable, including provisions for metadata

Discoverability of data and metadata provision

There are various disciplinary metadata standards, the NextFOODProject will follow OpenAIRE Guidelines concerning the availability of metadata.

All datasets will include metadata defining the what, where, when, why, and how data of the data. Where appropriate, the data creators will assign Digital Object Identifiers (DOIs).

For all data uploaded in a database/repository, the metadata will be in JSON-format according to a defined JSON schema. The repository will allow for the metadata exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the OpenAIRE Guidelines).

Metadata will be provided in a Project and Dataset level and will include sufficient information in order to link it to the research publications/outputs, identify the funder and discipline of the research, and will include appropriate keywords to help external and internal users to locate the data.

Final responsibility to comply with EU regulations will be the responsibility of each partner producing and providing data.

To the extent, possible surveys and coding will be shared between case studies.

Identifiability of data and standard identification mechanism

NextFOOD will use an Internal project Identifier for the data set to be produced. This will follow the format:

WPNumber_TaskNumber__PartnerName_DataSubset_DatasetName_Version__DateOfStorage,

where the project name is NextFOOD, the PartnerName represents the name of the data custodian (WP Lead/ Task Leader).

An example of this naming format would be:

WP2_T2.1_NMBU_Subset1_UserRequirements_V1.0_20.09.18

Data uploaded in repositories (e.g. Zenodo) will have their own Digital Object Identifier issued by the Repository that will be utilized. Similarly, all Journal Articles and the Monograph will have their own digital identifier issued by the respective publisher.

Naming conventions used

The following naming conventions should be used across all different case studies and project partners and irrespective to whether documents, data sets, etc. are finalized or not.

Folders

Aside from the predefined folder names, new folders only shall be created after approval by the project coordinator.

Preferably folder names shall be built of a number and a short descriptive name in order to have a consistent sorting and that the folder can be found in the same place at any time.

Files

File names should be developed in a way such that on the one hand they provide information about the content of the file, on the other hand, they should enable a chronological order and the creation of variants.

Repetitions in files and/or folder names shall be avoided if possible.

For maximum compatibility between Windows/Linux/Mac etc. the following characters **must be avoided** in both, folder and file names: < > : " \ / | * ? .

Following characters **should be avoided** in both, folder and file names: [] = % \$ + , ; “-“ at the beginning of a name is not allowed.

The length of the entire path shall not exceed 256 characters including empty spaces.

Naming conventions for Survey Data and Coding schemes

Any surveys that will be conducted during the Project will include a detailed explanation of the naming convention and a documentation file when needed. Survey naming conventions should be common across cases to the extent that is possible.

Qualitative data coding schemes should also be documented and follow clear naming conventions documented to a separate file when needed. As in the case of surveys coding schemes naming conventions should be common across cases, to the extent that this is possible.

Approach to search keywords

The NextFOOD project will assign keywords based on the UK Data Archive [HASSET Thesaurus](#) and the [ELSST multilingual thesaurus](#).

Outline the approach for clear versioning

Versioning

No more than 10 versions of a file should be kept. Versions of previous Substantive changes should be kept. When a new version due to substantive changes (milestones versions) is produced minor versions of the previous edition can be deleted.

Different versions should be identified by numbering v1.0, v.1.1 etc. and any changes made to a file when a new version is created should be recorded by a different colour. Moreover, relationships between items (e.g. code and data files) should be recorded.

Lastly, files saved in different locations should be synchronized, and master versions should be uploaded in the Project Sharing drive.

Version numbering in file names can be through discrete or continuous numbering depending on minor or major revisions.

Example:

Table 2: Versioning rules to be used for the NextFOOD project

File name	Changes to file
Interviewschedule_1.0	Original document
Interviewschedule_1.1	Minor revisions made
Interviewschedule_1.2	Further minor revisions
Interviewschedule_2.0	Substantive changes

Standards for metadata creation

The NextFOOD project strongly suggests to the partners to utilize the DDI standard in the development of their surveys. More Information about the DDI can be found [here](#).

2.2 Making data openly accessible

Data that will be made openly available

The NextFOOD project will facilitate the sharing of results and deliverables, both within and beyond the consortium. Results will be widely shared with the interested communities, including but not limited to the scientific community, policy and decision makers through publications in scientific journals and presentations at conferences, as well as through open access data repositories. Overall an open access policy will be applied, following the rules outlined in the Grant and Consortium Agreements.

All data will be considered by default openly available, with the exception of datasets that include personalized data. In the latter case data should be anonymized before being considered to be openly available.

All data will be made available for verification and re-use unless the task leader can justify why data cannot be made openly accessible.

The Steering Committee will assess the reasoning of the justification and make the final decision based on examination of the following elements regarding confidentiality of datasets:

- (i) Commercial sensitivity of datasets
- (ii) Data confidentiality for security reasons
- (iii) Conflicts between open-access rules and national and European legislation (e.g. data protection regulations).
- (iv) Sharing data would jeopardise the aims of the project
- (v) Other legitimate reasons, to be validated by the IPR Committee

Availability guarantees

All NextFOOD project datasets will be made available in an online open repository (e.g. Zenodo) including the relevant metadata to identify the project, funder, scope etc. of the project unless the Steering committee accepts that there is sufficient reason and justification not to.

Methods or software tools needed to access the data and documentation about the software

NextFOOD will ensure that it will follow the UK archive guidelines on recommended file formats to ensure accessibility. The following documentation on a study and Data level will be also provided for this reason.

Study-Level

On a study-level the following documentation should be provided:

- research design and context of data collection: project history, aims, objectives, hypotheses, investigators and funders
- data collection methods, data collection protocols, sampling design, sample structure and representation, workflows, instruments used, hardware and software used, data scale and resolution, temporal coverage and geographic coverage, and digitisation or transcription methods used
- structure of data files, with a number of cases, records, files and variables, as well as any relationships among such items
- secondary data sources used and provenance, for example, for transcribed or derived data

- data validation, checking, proofing, cleaning and other quality assurance procedures carried out, such as checking for equipment and transcription errors, calibration procedures, data capture resolution and repetitions, or editing, proofing or quality control of materials
- modifications made to data over time since their original creation and identification of different versions of datasets
- for time series or longitudinal surveys, changes made to methodology, variable content, question text, variable labelling, measurements or sampling, and how panels were managed over time and between waves
- information on data confidentiality, access and any applicable conditions of use
- publications, presentations and other research outputs that explain or draw on the data

Important data documentation should include original questionnaires, interviewer instructions, interview topic guides or experimental protocols.

Data-Level

On a Data-level for Survey and Transcription data partners should follow the guidelines of the UK Data Survey available [here](#) and [here](#) respectively.

Based on the acceptable file formats SPSS, Stata, NVivo, etc software might be needed to access the original data. Nevertheless, partners are encouraged to provide DDI XML data and metadata.

The above should be also documented in specific "readme" files in the respective folder.

Specify where the data and associated metadata, documentation and code are deposited

The NextFOOD project will utilize the Zenodo or a similar open Repository for the data that will make available publicly.

Original non-anonymized data or data that is decided not to be published will be stored under the responsibility of the task leader in a closed repository that will adhere to all the necessary legal and ethical requirements.

Access provision for restricted datasets

All partners will have access to data produced by the NextFOOD project (internal), with the exception of non-anonymized data.

All published data will be openly accessible to all internal and external data requests.

In the case of restricted access, the steering committee will assess external data requests or internal data requests for non-anonymized access.

Additionally, to the above task leaders should set an embargo period for the data produced by the project that cannot be longer than 24 months after the completion of the project.

2.3 Making data interoperable

Interoperability of Data

The Datasets produced by the NextFOOD project will have high interoperability, taking into account the type and discipline of the project.

The project will be using the UK Data Service Guidelines concerning data and metadata vocabularies, partners are expected to follow the recommended file formats of the UK data archive and are encouraged to follow the DDI standards to assess interoperability.

Standard vocabularies

To the extent, possible partners are expected to use standard vocabulary for all data types present in the datasets.

In the case that this won't be possible partners should provide adequate documentation to allow interoperability.

2.4 Data re-use and licenses

Data licensing and Permissions

All project data will follow a Creative Commons Attribution Non-Commercial licenses.

The task leader may opt out proposing a different license by providing the necessary reasoning and justification to the Steering Committee.

Data availability and disclosure

All data sets should be made public before the finalization of the project in an open repository.

In the case of embargoed data, the embargo time will not exceed a period of 24th months after the completion of the project.

Data usability by third parties

Datasets open and usable by third parties, since they are going to be uploaded under a Creative Commons Attribution Non-Commercial license.

In some cases, task leaders may opt out of publishing and publicly releasing data under the above-mentioned license. In this case, sufficient justification will be provided and the Data management Plan will be amended accordingly.

Non-open data will also be reusable (except non-anonymized data), provided that the data request is reasonable and justifiable.

Data quality assurance processes

Survey data

Surveys should be developed collectively with the participation of all relevant partners that will ensure through internal peer review the quality of the planning, data collection and documentation of the data.

A common survey and metadata registry in a data-level will be developed for each survey to the extent that a common survey can be applicable in all the different case studies.

All surveys common or not should provide sufficient documentation on a Data-Level (as described above).

Transcripts and participatory research data

Similarly to the common developed protocols for survey research, specific guidelines, coding schemes etc will be developed for transcripts and participatory research data. All of the above should be common across all case studies of the projects to the extent possible.

For each dataset sufficient documentation should be provided on a Data-Level (as described above).

The Data Quality assurance processes are the responsibility of each Task Leader.

Length of time for which the data will remain re-usable

Published data in the open repository will remain re-usable for as long as the repository specifies.

3 Allocation of resources

No specific costs are allocated for making the NextFOOD data FAIR. Each partner will be responsible for the data management costs of the data that they will produce.

Responsibilities for data management in your project

The project coordinator will be responsible for keeping partners accountable as far as data management is being concerned.

The steering committee will be responsible to assess whether data should be deemed public or not, as well as under which licence the data will fall.

Task leaders will be responsible to ensure that

- sufficient documentation is produced for each dataset
- the quality assurance processes are being implemented
- the datasets are uploaded to public repositories or for non-public datasets stored safely in the chosen repository by the organization.

Case study leaders will be responsible for the collection, documentation, as well as legal and ethical rules, are being followed as they are specified in the respective documents

In the case of external contractors concerning data collection, data entry, transcribing, processing or analysis, the organization making the subcontracting must assure that all the legal and ethical rules are being followed as they are specified in the respective documents.

A specific folder structure for internal data storage will be developed by the project coordinator with the support of the AFS.

Costs and potential value of long-term preservation

There will not be additional costs for the project for a long-term preservation of the data.

Data generated from the NextFOOD project is expected to be valuable to the following entities:

- Educational Institutions
- Agricultural Advisory Services
- Policy and decision makers
- Agrifood industry
- Farmers, Farmers organisations
- Forestry Associations
- Research community
- Evaluators

- Project officers and project administration offices

4 Data security

Data recovery, secure storage and transfer of sensitive data

Case study leaders and task leaders should keep a back-up of all data and documentation that they produce.

They are also responsible to upload and update relevant data in the internal repository of the project.

We recommend that any storage should involve at least two different forms of storage, for example on a hard drive and on DVD and that the data integrity should be checked periodically.

Personal information should be removed from data files and stored separately under more stringent security measures. Any digital files or folders which contain sensitive information and data should be encrypted.

Cloud data storage should be avoided for high-risk information such as files that contain personal or sensitive information, information that is covered by the law. Encryption should be used to safeguard data files to a certain degree, but partners should keep in mind that in some cases it would still not meet the requirements of data protection legislation.

5 Ethical aspects

Ethical issues concerning the NextFOOD project are tackled in the respective deliverables.

Additional information concerning ethical aspects such as compliance etc. can be found [here](#).