# Data Management Plan

Deliverable D.1.2
WP1. Management and coordination

**Towards a sustainable agro-food INDUstry: Capacity building programmes in Energy efficiency**
Grant agreement: 785047

INDUCE

# Project factsheet

Acronym:        **INDUCE**

Title:            **Towards a Sustainable Agro-Food Industry. Capacity Building Programmes in Energy Efficiency.**

Coordinator:    CIRCE -  Research Centre for Energy Resources and Consumption

Reference:      785047

Type:            Coordination and Support Action

Program:        HORIZON 2020

Call:            H2020-EE-2017-CSA-PPI

Start:            1st February 2018

Duration:        30 months

Website:        www.induce2020.eu

Consortium:    Research Energy Centre of the Netherlands, **ECN/TNO**[1]

Fraunhofer Institute Systems and Innovation Research, **FhISI**

Spanish Food and Drink Industry Federation, **FIAB**

Federatie Nederlandse Levensmiddelen Industrie, **FNLI**

Services Coop de France, **SCDF**[2]

Food-Processing Initiative, **FPI**

KWA Bedrijfsadviseurs, **KWA**

ÖKOTEC Energiemanagement, **ÖKOTEC**

SYNYO GmbH**,** Austria **SYNYO**

Association de Coordination Technique pour l'Industrie Agro-alimentaire, France **ACTIA**

---

[1] ECN informed the coordinator about its partial takeover by TNO on May 1st 2018. Until the Grant Agreement is officially amended this partner will be referred as ECN/TNO.

[2] SCDF informed the coordinator about its resignation from the project on March 19th 2018. After approval by the General Assembly, another French entity (ACTIA – Le réseau français des instituts techniques de l'agro-alimentaire) started in the project on May 1st 2018.  However, the document keeps the reference to SCDF as the amendment to the Grant Agreement concerning such change is not official yet. ACTIA is participating in the project according to this Project Handbook.

# Deliverable factsheet

Number:                         **D1.2**

Title:                          **Data Management Plan**

Responsible partner:            CIRCE

Work package:                   WP1. Management and coordination

Task:                           T1.3 Quality Assurance and Data Management Plan

Dissemination level[3]:         PU

Contributors:                   All partners


Document history:

| Revision | Date | Main modification | Author |
|----------|------|-------------------|--------|
| 1 | 08/08/2018 | First Version | CIRCE |
| 2 | 17/09/2018 | Review | CIRCE |

# Disclaimer of warranties

*"This project has received funding from the European Union's Horizon 2020, research and innovation programme, under Grant Agreement No 785047"*

This document has been prepared by INDUCE project partners as an account of work carried out within the framework of the EC-GA contract no 785047.

Neither Project Coordinator, nor any signatory party of INDUCE Project Consortium Agreement, nor any person acting on behalf of any of them:

- (a) makes any warranty or representation whatsoever, express or implied,
    - (i). with respect to the use of any information, apparatus, method, process, or similar item disclosed in this document, including merchantability and fitness for a particular purpose, or
    - (ii). that such use does not infringe on or interfere with privately owned rights, including any party's intellectual property, or
    - (iii). that this document is suitable to any particular user's circumstance; or

- (b) assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if Project Coordinator or any representative of a signatory party of the INDUCE Project Consortium Agreement, has been advised of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

# Abbreviations

CA: Consortium Agreement

DoA: Description of Action

EC: European Commission

GA: General Assembly

IPR: Intellectual Property Right

PC: Project Coordinator

PH: Project Handbook

DMP: Data Management Plan

H2020: Horizon 2020

WP: Work Package

# Executive Summary

This deliverable provides the INDUCE data management plan version 1. The deliverable outlines how the research data collected or generated will be handled during the project. It describes which standards and methodology for data collection and generation will be followed, and whether and how the data will be shared.

This document follows the template provide by the European Commission in the Participant Portal.

The INDUICE project will develop an open access platform (INDUCE toolkit) where training material, online lessons, guidelines and tools will be available for companies aiming to increase their energy efficiency. Besides, the INDUCE methodology will be integrated in the platform with the objective to design a specific capacity building programme for every company accessing the INDUCE toolkit, which will result in ad-hoc training adapted to the company's needs and opportunities. The INDUCE methodology and the toolkit will be both tested and validated in 15 pilot companies from the food and beverage sector in four countries that represent over 45% of the EU companies in this sector: Spain, France, Netherlands and Germany. In a second replication step, an INDUCE community of trainers will be established by certifying 60 INDUCE trainers. They will be in charge of implementing the INDUCE methodology in another 300 companies, hence increasing INDUCE's impact up to 106 GWh/year of energy savings.

# Table of Contents

# 1. INTRODUCTION

## 1.1. PURPOSE OF THE INDUCE DATA MANAGEMENT PLAN (DMP)

INDUCE is a Horizon 2020 project participating in the Open Research Data Pilot. This pilot is part of the Open Access to Scientific Publications and Research Data programme in H2020. The goal of the program is to foster access to data generated in H2020 projects.

Open Access refers to a practice of giving access to all scholarly disciplines information that is free of charge to the end-user. In this way data becomes re-usable, and the benefit of public investment in the research will be improved.

The EC provided a document with guidelines for projects participants in the pilot. The guidelines address aspects like research data quality, sharing and security. According to the guidelines, project participants will need to develop a DMP.

The DMP describes the types of data that will be generated or gathered during the project, the standards that will be used, the ways how the data will be exploited and shared for verification or reuse, and how the data will be preserved.

This document has been produced following these guidelines and aims to provide a consolidated plan for INDUCE partners in the data management plan policy that the project will follow. This document is the first version delivered in M4 of the project. The DMP will be updated during the lifecycle of the project.

## 1.2. BACKGROUND OF THE INDUCE DMP

The INDUCE DMP will be written in reference to the Article 29.3 in the Model Grant Agreement called "Open access to research data". Projects participants must deposit their data in a research data repository and take measures to make the data available to third parties. The third party should be able to access, mine, exploit, reproduce and disseminate the data. This should also help to validate the results presented in scientific publications. In addition, Article 29.3 suggest that participants will have to provide information, via the repository, about tools and instruments needed for the validation of project outcomes.

The DMP will be important for tracking all data produced during the INDUCE project. Article 29 states that project beneficiaries do not have to ensure access to parts of research data if such access would be lead to a risk for the project´s goals. In such cases, the DMP must contain the reason for not providing access.

Regarding the nature of the data, in order to fulfil the required security and privacy requirements in this project, which are set by the Data Protection Directive (Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data), the project assumes the differentiation set in this Directive between Personal and No Personal data. Data are considered as personal data "when someone is able to link the information to a person, even if the person holding the data cannot make this link". Any data susceptible of being considered as Personal Data will be managed according to this Directive.

# 2. INDUCE DATA SUMMARY

Being in line with the EU's guidelines regarding the DMP, this document should address for each data set collected, processed and/or generated in the project the following characteristics: Dataset description, reference and name, Standards and metadata, Data sharing, archiving and preservation. At this point in time, an estimation of the size of the data cannot be given. To this end, the consortium develops a number of strategies that will be followed in order to address the above elements.

This section, shall be provided a detailed description of these elements in order to ensure their understanding by the partners of the consortium. For each element, we also describe the strategy that will be used to address it.

## 2.1. DATA SET DESCRIPTION, REFERENCE AND NAME

In order to be able to distinguish and easily identify data sets, each data set will be assigned with a unique name. This name can also be used as the identifier of the data sets.

All data files produced, including emails, include the term "INDUCE", followed by file name which briefly describes its content, followed by a version number (or the term "FINAL"), followed by the short name of the organisation which prepared the document (if relevant).

Each data set that will be collected, processed or generated within the project will be accompanied by a brief description.

## 2.2. STANDARDS & METADATA

This version of the INDUCE DMP does not include a compilation of all the metadata about the data being produced in INDUCE project, but there are already several domains considered in the project which follows different rules and recommendations. This is a very early stage identification of standards:

- Microsoft Word 2010 for text based documents (or any other compatible version). doc, .docx, .xls, .xlsx, .ppt, .pptx. Also, especially where larger datasets need to be dealt with, .csv and .txt file formats will be used. All finished and approved documents will also be made available as .pdf documents.
- Illustrations and graphic design will make use of Microsoft Visio (Format: .vsd), Photoshop (Format: different types possible, mostly .png), and will be made available as .jpg, .psd, .tiff and .ai files.
- MP3 or WAV for audio files.
- Quicktime Movie or Windows Media Video for video files.

These file formats have been chosen because they are accepted standards and in widespread use. Files will be converted to open file formats where possible for long-term storage.

Metadata will be comprised of two formats – contextual information about the data in a text based document and ISO 19115 standard metadata in an xml file. These two formats for metadadata are chosen to provide a full explanation of the data (text format) and to ensure compatibility with international standards (xml format).

## 2.3. DATA SHARING, ACCESS AND PRESERVATION

The digital data created by the project will be diversely curated depending on the sharing policies attached to it. For both open and non-open data, the aim is to preserve the data and make it readily available to the interested parties for the whole duration of the project and beyond. A public API will be provided to registered users allowing them the access to the platform. The database compliance aims to ensure the correct implementation of the security policy on the databases verifying vulnerability and incorrect data. The target is to identify excessive rights granted to users, too simple passwords (or even the lack of password) and finally to perform an analysis of the entire database. At this point, we can assure that at least the following measures will be considered for assuring a proper management of data:

- Dataset minimisation. The minimum amount of data needed will be stored so as to prevent potential risks.
- Access control list for user and data authentication. Depending on the dissemination level of the information an Access Control List will be implemented reflecting there for each user the data sets that can be accessed.
- Monitoring and Log of activity. The activity of each user in the project platform, including the data sets accessed, is registered in order to track and detect harmful behaviour of users with access to the platform.
- Implementation of an alert system that informs in real time of the violation of procedures or about hacking attempts.
- Liability. Identification of a person who is responsible for keeping safe the information stored,
- When possible, the information will be also made available in the initiative that the EC has launched for open data sharing from research, which is ZENODO.ORG.

The mechanisms explained in this document aim at reducing to the maximum the risks related to data storage. However, due to the activities that are going to be carried out in the project, it is still not defined the amount of time that data will be stored in the platform since Big Data analysis and services run data analytics procedures and depending on the accuracy of results based on the size of the sets considered.

### 2.3.1. Non-Open research data

The non-open research data will be archived and stored long-term in the EMDESK portal administered by CIRCE. The CIRCE platform is currently being employed to coordinate the project's activities and to store all the digital material connected to INDUCE. If certain datasets cannot be shared (or need restrictions), legal and contractual reasons will be explained.

### 2.3.2. Open research data

The open research data will be archived on the Zenodo platform (http://www.zenodo. org). Zenodo is a EU-backed portal based on the well-established GIT version control system (https://git-scm.com) and the Digital Object Identifier (DOI) system (http://www.doi.org). The portal's aims are inspired by the same principles that the EU sets for the pilot; Zenodo represents thus a very suitable and natural choice in this context. The repository services offered by Zenodo are free of charge and enable peers to share and preserve research data and other research outputs in any size and format: datasets, images, presentations, publications and software. The digital data and the associated meta-data is preserved through well-established practices such as mirroring and periodic backups. Each uploaded data-set is assigned a unique DOI rendering each submission uniquely identifiable and thus traceable and referenceable.

# 3. ALLOCATION OF RESOURCES

Data management in INDUCE will be done as part of the WP1 and CIRCE, as project coordinator, will be responsible for data management in INDUCE project. CIRCE has allocated a part of the overall WP1 budget and person months to these activities. For the time being, the project coordinator is responsible for FAIR data management. Costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). Resources for long term preservation, associated costs and potential value, as well as how data will be kept beyond the project and how long, will be discussed by the whole consortium during General Assembly meetings.

# 4. DATA SECURITY

For the duration of the project, datasets will be stored on the responsible partner's storage system. Every partner is responsible to ensure that the data are stored safely and securely and in full compliance with European Union data protection laws. After the completion of the project, all the responsibilities concerning data recovery and secure storage will go to the repository storing the dataset.

All data files will be transferred via secure connections and in encrypted and password-protected form (for example with the open source 7-zip tool providing full AES-256 encryption: http://www.7-zip.org/ or the encryption options implemented in MS Windows or MS Excel). Passwords will not be exchanged via e-mail but in personal communication between the partners.

# 5. ETHICAL AND LEGAL ASPECTS

This section deals with ethical and legal compliance issues, like the consent for data preservation and sharing, protection of the identity of individuals and companies and how sensitive data will be handled to ensure it is stored and transferred securely. Data protection and good research ethics are major topics for the consortium of this project. Good research ethics meet all actions to take great care and prevent any situation where sensitive information could get misused. This is what the consortium wants to guarantee for this project. Research data which contains personal data will just be disseminated for the purpose for which it was specified by the consortium. Furthermore, all processes of data generation and data sharing have to be documented and approved by the consortium to guarantee the highest standards of data protection.

INDUCE partners have to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with:

- Ethical principles (including the highest standards of research integrity as set out, for instance, in the European Code of Conduct for Research Integrity and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct).
- Applicable international, EU and national law (in particular, EU Directive 95/46/EC).

Data collection, storage, protection, retention and destruction will be carried out through the intranet system of the project: EMDESK. Interviewees/beneficiaries/recipients will be informed about data security, anonymity and use of data as well as asked for accordance. Participation happens on a voluntary basis

# 6. LIST OF THE DATA-SETS

This section will list the data-sets produced within the INDUCE project.

# 7. REFERENCES

Zenodo platform. https://zenodo.org

GIT version control system. https://git-scm.com

Digital Object Identifier (DOI) system. http://www.doi.org

Metadata Standards Directory. https://www.rd-alliance.org/metadata-standards-directory

EUDAT B2SHARE tool. https://b2share.eudat.eu/