Project Acronym:	PhasmaFOOD
Grant Agreement number:	732541 (H2020-ICT-2016-1 - RIA)
Project Full Title:	Portable photonic miniaturised smart system for on-the-spot food quality sensing

DELIVERABLE

Deliverable Number	**D7.2**
Deliverable Name	**Data Management Plan**
Dissemination level	PU
Type of Document	RE
Contractual date of delivery	30/4/2017
Deliverable Leader	INTRASOFT International
Status & version	1.1 Final
WP / Task responsible	INTRASOFT International (INTRA)
Keywords:	Data Management Plan, data accessibility, data repositories, data interoperability, security, Metadata Standards.
Abstract (few lines):	The deliverable describes the data management life cycle for the data to be collected, processed and/or generated by PhasmaFOOD.

Deliverable Leader:	George Koutalieris (INTRA)
Contributors:	George Koutalieris (INTRASOFT), Yannick Weesepoel (RIKILT), Gerhard Wunder (FUB), Annamaria Gerardino (CNR), Prof. George Nychas (AUA), Susanne Hintschich (IPMS), Paraskevas Bourgos (WINGS)
Reviewers:	Gerhard Wunder (FUB), George Nychas (AUA), Paraskevas Bourgos (WINGS)
Approved by:	George Koutalieris (INTRA)

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732541. It is the property of the PhasmaFOOD consortium and shall not be distributed or reproduced without the formal approval of the PhasmaFOOD Management Committee.

Executive Summary

This deliverable reports the data collecting and sharing plan in the scope of the PhasmaFOOD project. PhasmaFOOD will design and implement a parameterized, knowledge-based, multi-target food sensitive mini-portable system for on-the-spot food quality sensing and shelf-life prediction. The miniaturized smart integrated system Smart will collect spectrum images via heterogeneous micro-scale photonics. The data will be forwarded to the cloud where the spectroscopy analysis will take place using a reference database. The PhasmaFOOD reference database will be built upon laboratory measurements of the specific foods and qualities supported by the PhasmaFOOD Use Cases.

In Section 3 we provide descriptions of the Datasets that will be produced by the different laboratory experiment set-ups that will be used to collect reference measurements. These dataset descriptions provide useful insight on meat-related laboratory experiments and may be useful in other food security settings outside the project.

Furthermore, considering the IT system aspects of PhasmaFOOD, Section 4 provides a brief overview of related metadata standards that could be used during the system design and implementation in order to improve the interoperability of PhasmaFOOD with existing systems and services of the food safety domain.

The remaining sections provide the detailed viewpoints of the Data Management Plan addressing: Naming methodology (Section 2); Data Access and Sharing (Section 5); and, Archiving and Preservation (Section 6).

Deliverable updates are planned for Months 18 and 36 to reflect the updates in the data management framework of PhasmaFOOD. The next version of the Data Management Plan will include additional food and beverage measurements related to all use cases. Data regarding hardware design and specification as well as software design and implementation (cloud, mobile and embedded level) will be also addressed.

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.1	28/02/2017	INTRA	ToC
0.2	21/03/2017	INTRA	1st preliminary version, including content for each chapter
0.3	15/04/2017	INTRA, ALL	2nd preliminary version including instantiation of data set and comments from internal review
0.8	24/04/2017	INTRA, AUA, WINGS, FUB, IPMS	Final Approval, QA
1.0	28/04/2017	INTRA	Final document for submission to the EC
1.1	29/04/2017	INTRA	Final revisions on comments provided by AUA, WINGS, FUB, IPMS

Table of Contents

Definitions, Acronyms and Abbreviations

Acronym	Title
API	Application Programming Interface
CO	Confidential, only for members of the consortium (including Commission Services)
CR	Change Request
D	Demonstrator
DL	Deliverable Leader
DM	Dissemination Manager
DMS	Document Management System
DoA	Description of Action
Dx	Deliverable (where x defines the deliverable identification number e.g. D1.1.1)
EIM	Exploitation Innovation Manager
EU	European Union
FM	Financial Manager
MSx	project Milestone (where x defines a project milestone e.g. MS3)
Mx	Month (where x defines a project month e.g. M10)
O	Other
P	Prototype
PC	Project Coordinator
PM	partner Project Manager
PO	Project Officer
PP	Restricted to other programme participants (including the Commission Services)
PU	Public
QA	Quality Assurance
QAP	Quality Assurance Plan
QFD	Quality Function Deployment
QM	Quality Manager
R	Report
RE	Restricted to a group specified by the consortium (including Commission Services)
STM	Scientific and Technical Manager
TL	Task Leader
WP	Work Package
WPL	Work Package Leader
WPS	Work Package Structure

1 Data Summary

The main objective of PhasmaFOOD is to design and implement an autonomous, multifunctional and programmable optical sensing device, integrated with spectroscopy technologies for food hazard, microbial activity detection and shelf-life estimation.

The project will configure a dedicated cloud platform for collecting data from the PhasmaFOOD device and to perform in-depth analysis of the data obtained from the project's device. The cloud platform will enable correlation of the device measurements with spectral analysis results collected from all connected PhasmaFOOD smart sensing devices. This will provide the opportunity for detection of trends, patterns and distribution of food contamination which can help prevent outbreaks and provide recommendation for improving food safety at different stages of farm-to-fork production chain.

PhasmaFOOD cloud platform will also host sensory and contextual database which will be used for training data analysis and machine learning models deployed on smart sensory device and as part of PhasmaFOOD mobile application (data analytics calibration). The data collected from the miniaturized device will be forwarded to the cloud where the spectroscopy analysis will take place using a reference database. The PhasmaFOOD reference database will be built upon laboratory measurements of the specific foods and qualities supported by the PhasmaFOOD Use Cases:

- **Use case 1: Detection of mycotoxins in various grains and nuts:** Aflatoxins, a special type of mycotoxins, will be detected. A simple, convenient ultraviolet test makes it possible to detect the possible presence of aflatoxin.
- **Use case 2: Detection of early sign of spoilage and spoilage in fruits, vegetables, meat, fish**: Combined with estimation on product expiration date.
- **Use case 3: Detection of food fraud:** Adulteration of alcoholic beverages, oil, milk and meat.

In Section 1 and Section 3 we provide summary and descriptions, respectively, of the Datasets that will be produced by the different laboratory experiment set-ups that will be used to collect reference measurements. These dataset descriptions provide useful insights into meat-related laboratory experiments and may be useful in other food security settings outside the project.

Deliverable updates are planned for Months 18 and 36 to reflect the updates in the data management framework of PhasmaFOOD.

2 Datasets: Reference and Name

The PhasnaFOOD project has already identified, and is still further identifying, a set of data sources that will stem from new and existing laboratory experiments that will produce the required reference measurements required by the use case scenarios.

Following an iterative process, data sources are established, encapsulating the project's use case requirements. A refinement process is foreseen to continuously take place throughout the lifetime of the project, as new laboratory experiments will be deployed to collect additional reference measurements.

2.1 Naming Methodology

2.1.1 Folder nomenclature

For each data source there is a folder containing the data source definition and the data sample. Each folder has a specific name that is composed by different parts/elements, containing information about:

- Use Case (UC1, UC2, UC3),
- Food & Beverage Type (FBT-<Name>),
- Contamination & Fraud Type (CFT-<Name>),
- Measurement Sensor (MS-<Name>),
- Laboratory (LO-<Organisation short name>).

Based on the above, a data source folder instance example would be:

UC1_FBT-Pork-Meat_CFT-SPOILAGE_MS-FTIR_LO-AU

2.1.2 Datasets nomenclature

In each folder the files containing the measurement datasets will follow a similar nomenclature, adding field information required for the further identification of the datasets. For each dataset, its specific name is composed by different parts, containing the following elements:

- Use Case (UC1, UC2, UC3),
- Food & Beverage Type (FBT-<Name>),
- Contamination & Fraud Type (CFT-<Name>),
- Measurement Sensor (MS-<Name>),
- Laboratory (LO-<Organisation short name>),
- Dataset ID (DID-<ID provided by LO>),
- Date Provided (DP-<YYYYMMDD>),
- Dataset File Extension (DFE-<filename extension>).

Based on the above, a dataset instance example would be:

UC1_FBT-Pork-Meat_CFT-SPOILAGE_MS-FTIR_LO-AU_ DID-111_ DP-20170122_ DFE-XLSX

3 Dataset Descriptions

3.1 Generic Use Case Data Sources

For the implementation of the PhasmaFOOD use cases' samples spectral databases will be built up to hold chemical reference data and used later to benchmark and validate new samples. The general schema for Data Sources descriptions in PhasmaFood is provided below.

Data Source name: Use-case 1, 2 and 3	
Data source description	
Data, used for use-cases 1, 2 and 3 (see D1.1), populating spectroscopic databases: 1. UV-Vis data – spectroscopic data 2. NIR data – spectroscopic data 3. Image data – 3D spectroscopic data Chemical reference data (e.g. nitrogen/moisture determination, GC, LC-MS) to benchmark the samples in use-cases 1, 2, and 3.	
Dataset entities	Samples used in use-cases 1, 2 and 3 for spectral data base building (WP3) and validation (WP6)
Dataset attributes	Unknown, pilot equipment is not available so far
Data type	Unknown so far, probably .csv or excel formats
Standard	NA
Direct data URI	Unknown so far
Data Size	Unknown so far
Sample size	Unknown so far
Data lifetime	WP 3 – till M 27, WP 6 till M36
Availability	Only available to consortium
Data collection frequency	On demand, when pilot equipment is available
Data quality	Unknown so far
Raw data sample	
NA	
Print screen (if possible)	
(Print screen for a data sample, if not in text format; e.g. Excel Sheet) NA	

Field Descriptions (Description of the fields within the data, whenever possible; e.g. JSON keys descriptions, Excel Sheet's column descriptions, etc.)		
Field Name	**Field Description**	**Type of Data**
Sample name	Sample name	text
Reference value	Class1/Class2	category
Wavenumber	Wavenumber	number

In the remaining sections, detailed instances of PhasmaFOOD laboratory measurements are presented. These will form the basis of the PhasmaFOOD cloud reference database and will be specifically considered during the specification (WP1) and design (WP2) project activities.

NOTE: The measuring methods of the PhasmaFOOD demonstrator will include spectral data (from visible and near-infrared spectroscopy) and image data (visible camera). The following data sets are examples from FTIR spectroscopy and multispectral imaging (MSI). Although these exact measuring methods will not be implemented into the PhasmaFOOD demonstrator, the data format of FTIR spectra is representative of our visible and NIR spectra, and MSI is used to demonstrate the data format of our visible images.

3.2 Beef meat spoilage experiment with FTIR

Data Source name: Beef meat spoilage experiment with FTIR	
Data source description	
Beef meat spoilage using FTIR spectroscopy. Aerobic storage at chill (0, 5 °C) and abuse (10, 15, and 20 °C) temperatures. Prediction of the microbial load on the surface of meat samples directly from FTIR spectral data.	
Dataset entities	Beef meat spoilage
Dataset attributes	FTIR spectral data
Data type	Comma Separated values text files: *.csv
Standard	Csv
Direct data URI	N/A
Data Size	5 (storage temperatures) x [1,7] MB
Sample size	[1,7] MB for each storage temperature. Each csv file is about [70,100] KB.
Data lifetime	Overall period of 350 h
Availability	Upon Request
Data collection frequency	Meat samples stored at 0 and 5 °C were analyzed every 24 h, whereas samples stored at 10, 15, and 20 °C were analyzed every 8, 6, and 4 h, respectively.

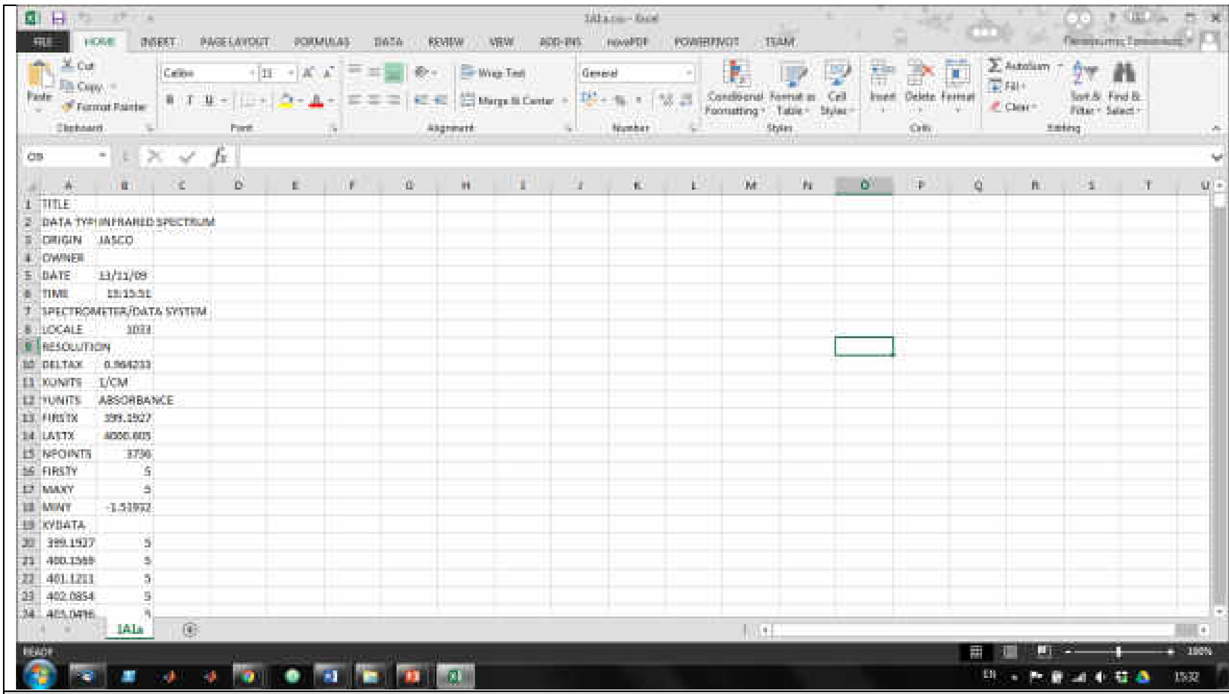
Data quality	Complete and published: Argyri et al. (2010)[1] Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks Sensors and Actuators B: Chemical 145, 146-154.

Raw data sample

TITLE	
DATA TYPE	INFRARED SPECTRUM
ORIGIN	JASCO
OWNER	
DATE	13/11/09
TIME	15:15:51
SPECTROMETER/DATA SYSTEM	
LOCALE	1033
RESOLUTION	
DELTAX	0.964233
XUNITS	1/CM
YUNITS	ABSORBANCE
FIRSTX	399.1927
LASTX	4000.6047
NPOINTS	3736
FIRSTY	5
MAXY	5
MINY	-1.51932
XYDATA	
399.1927	5
400.1569	5
401.1211	5
402.0854	5
403.0496	5
404.0138	5
404.9781	5
405.9423	1.17363
406.9065	0.86412
407.8708	0.715091

Print screen (if possible)

[1] http://www.sciencedirect.com/science/article/pii/S0925400509009174



Field Descriptions (Description of the fields within the data, whenever possible; e.g. JSON keys descriptions, Excel Sheet's column descriptions, etc.)

Field Name	Field Description	Type of Data
XYDATA	Spectral data	Wavelengths (cm^{-1})
Second column (next to XYDATA)	Spectral data	Absorbance at each specific wavelength

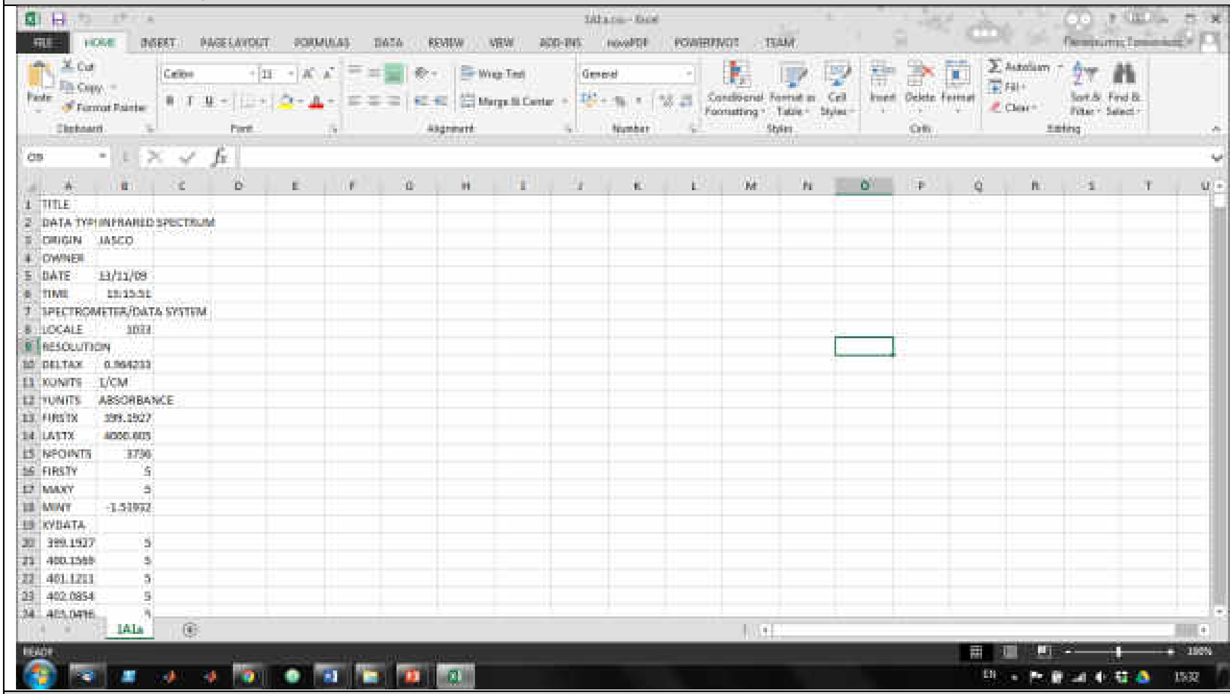
3.3 Pork meat spoilage experiment with FTIR

Data Source name: Pork meat spoilage experiment with FTIR	
Data source description	
Spectral data from FTIR with minced pork meat spoilage during aerobic storage of meat samples at different storage temperatures (0, 5, 10, and 15 °C). Prediction of the microbial load on the surface of meat samples directly from FTIR spectral data.	
Dataset entities	Pork minced meat spoilage
Dataset attributes	FTIR spectral data
Data type	Comma Separated values text files: *.csv
Standard	csv

Direct data URI	N/A
Data Size	4 (storage temperatures) x [1,7] MB.
Sample size	[1,7] MB for each storage temperature. Each csv file is about [70,100] KB.
Data lifetime	Overall period of 350 h
Availability	Upon Request
Data collection frequency	Samples stored at 0 and 5 °C were analyzed approximately every 24 and 12 h, respectively, whereas samples stored at 10 and 15 °C were analyzed every 6–7 h.
Data quality	Complete and published: Papadopoulou et al. (2011)[2] Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage. Food Research International 44, 3264.
Raw data sample	

TITLE	
DATA TYPE	INFRARED SPECTRUM
ORIGIN	JASCO
OWNER	
DATE	13/11/09
TIME	15:15:51
SPECTROMETER/DATA SYSTEM	
LOCALE	1033
RESOLUTION	
DELTAX	0.964233
XUNITS	1/CM
YUNITS	ABSORBANCE
FIRSTX	399.1927
LASTX	4000.6047
NPOINTS	3736
FIRSTY	5
MAXY	5
MINY	-1.51932
XYDATA	
399.1927	5
400.1569	5
401.1211	5
402.0854	5
403.0496	5
404.0138	5
404.9781	5
405.9423	1.17363
406.9065	0.86412
407.8708	0.715091

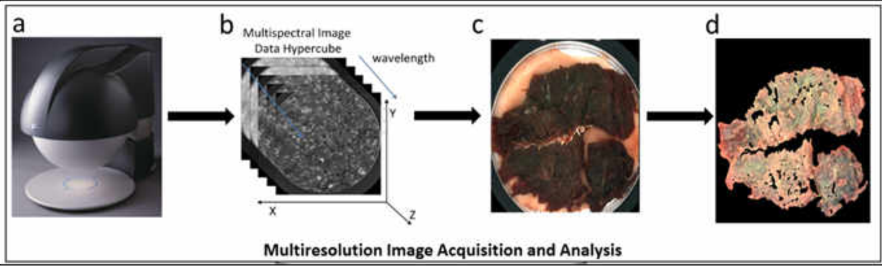
[2] http://www.sciencedirect.com/science/article/pii/S0963996911005424

Print screen (if possible)		
		
Field Descriptions (Description of the fields within the data, whenever possible; e.g. JSON keys descriptions, Excel Sheet's column descriptions, etc.)		
Field Name	**Field Description**	**Type of Data**
XYDATA	Spectral data	Wavelengths (cm^{-1})
Second column (next to XYDATA)	Spectral data	Absorbance at each specific wavelength

3.4 Beef meat spoilage experiment with MultiSpectral Imaging (MSI)

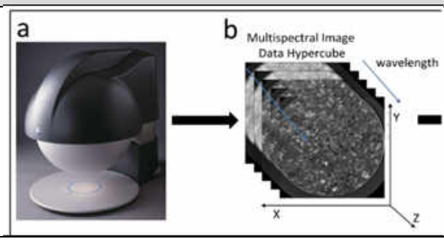
Data Source name: Beef meat spoilage experiment with MultiSpectral Imaging (MSI)	
Data source description	
Beef meat spoilage using Multispectral imaging. Aerobic storage at 2, 8 and 15 °C, sterile and naturally contaminated samples. Prediction and mapping of the microbial load on the surface of meat samples directly from Multispectral Imaging data.	
Dataset entities	Beef meat spoilage.
Dataset attributes	Multispectral images. Each sample consists of 18 images captured at 18 different wavelengths.

Data type	*.hips files. Graphic classification file.
Standard	Hips
Direct data URI	N/A
Data Size	(105 for sterile + 114 for naturally contaminated) x 100MB ~= 22 GB
Sample size	~100MB for each sample
Data lifetime	Overall period of 350 h
Availability	Upon Request due to the large size
Data collection frequency	Meat samples stored at 2 and 8 °C were analyzed every 24 h, whereas samples stored at 15 °C were analyzed every 6 h.
Data quality	Complete and published: Tsakanikas et al (2016) Exploiting multispectral imaging for non-invasive contamination assessment and mapping of meat samples. Talanta 161, 604. http://www.sciencedirect.com/science/article/pii/S0039914016306889
Print screen (if possible)	



3.5 Pork minced meat spoilage experiment with MultiSpectral Imaging (MSI)

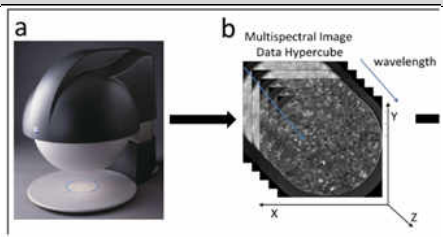
Data Source name: Pork minced meat spoilage experiment with MultiSpectral Imaging (MSI)	
Data source description	
Pork minced meat spoilage using Multispectral imaging. Aerobic and MAP (modified atmospheres packaging) storage at 0, 5, 10, 15 and 20 °C. Prediction and mapping of the microbial load on the surface of meat samples directly from Multispectral images data.	
Dataset entities	Pork minced meat spoilage.
Dataset attributes	Multispectral images. Each sample consists of 18 images captured at 18 different wavelengths.
Data type	*.hips files. Graphic classification file.
Standard	Hips
Direct data URI	N/A

Data Size	(160 for aerobic storage + 150 for map storage) x 100MB ~= 300 GB
Sample size	~100MB for each sample.
Data lifetime	Overall period of 350 h.
Availability	Upon Request due to the large size.
Data collection frequency	Meat samples stored at 0 and 5 °C were analyzed every 24 h, samples stored at 8 °C were analyzed every 8 h, whereas samples stored at 15 and 20 °C were analyzed every 5 h.
Data quality	Complete and published: Dissing et al. (2013) Using Multispectral Imaging for Spoilage Detection of Pork Meat. Food Bioprocess Technol (2013) 6:2268- 2279[3]
Print screen	
	

3.6 Adulteration experiment Beef & Pork meat with MultiSpectral Imaging (MSI)

Data Source name: Adulteration experiment Beef & Pork meat with MultiSpectral Imaging (MSI)	
Data source description	
Beef and pork-minced meat was mixed in order to achieve nine different proportions of adulteration and two categories of pure pork and beef. Detection of minced beef fraudulently substituted with pork and vice versa using Multispectral imaging.	
Dataset entities	Detection of minced beef fraudulently substituted with pork and vice versa
Dataset attributes	Multispectral images. Each sample consists of 18 images captured at 18 different wavelengths.
Data type	*.hips files. Graphic classification file.
Standard	Hips
Direct data URI	N/A

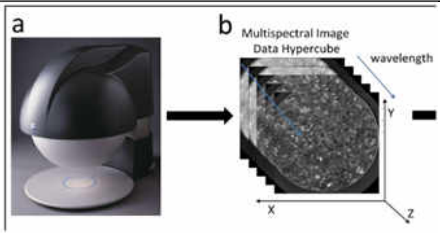
[3] http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6769/pdf/imm6769.pdf

Data Size	220 meat samples in total from four independent experiments (55 samples per experiment) x 100MB ~= 22 GB
Sample size	~100MB for each sample
Data lifetime	N/A
Availability	Upon Request due to the large size
Data collection frequency	On demand. Storage is not applicable here.
Data quality	Complete and published: Ropodi et al. (2015) Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. Food Research International 67, 12-18[4].
Print screen (if possible)	
	

3.7 Adulteration experiment Beef & Horse meat with MultiSpectral Imaging (MSI)

Data Source name: Adulteration experiment Beef & Horse meat with MultiSpectral Imaging (MSI)	
Data source description	
Detection of minced beef adulteration with horsemeat, as well as during storage in refrigerated conditions. For this reason, multispectral images of 110 samples from three different batches of minced beef and horsemeat in 18 wavelengths were acquired. Images were taken again after samples were stored at 4 °C for 6, 24 and 48 h.	
Dataset entities	Detection of minced beef adulteration with horsemeat.
Dataset attributes	Multispectral images. Each sample consists of 18 images captured at 18 different wavelengths.
Data type	*.hips files. Graphic classification file.

[4] http://www.sciencedirect.com/science/article/pii/S0963996914006887

Standard	Hips
Direct data URI	N/A
Data Size **Sample size**	(110 samples from three different batches of minced beef and horsemeat) x 100MB ~= 11 GB ~100MB for each sample
Data lifetime	N/A
Availability	Upon Request due to the large size
Data collection frequency	Sampling at 6, 24 and 48 h of storage at 4°C.
Data quality	Complete and published: Ropodi et al. (2017) Multispectral imaging (MSI): A promising method for the detection of minced beef adulteration with horsemeat. Food Control 73, 57-63[5].
Print screen (if possible)	
	

[5] http://www.sciencedirect.com/science/article/pii/S0956713516302912

4 Standards and Metadata

The food sector is highly characterized by non-centralized sources of dynamic and heterogeneous data which increase the information about, for example, meat quality in a particular region, but typically, decrease the effectiveness of sharing that information across stakeholders. The harmonization and standardization of data structures and data exchange services are fundamental challenges for both the information society as a whole, as well as for Food Security applications.

The main focus of PhasmaFOOD is to provide the required data interoperability and adaptability in a variety of food safety settings. PhasmaFOOD will seek to establish liaisons within the food security value chain, with a particular interest in standardizing the use of food related Open Data in the food security sector and standardizing the types of food related analytics that can be sought from big data platforms.

4.1 Rapid alert system for food and feed (RASFF)

Commission Regulation (EU) No. 16/2011 specifies the framework for implementing measures for the rapid alert system for food and feed (RASFF). The RASFF notifications are generated from templates that provide guidelines on how different fields in the notifications are used. In the table below we provide a sample example of the RASFF fields that are closely related to the PhasmaFOOD application scenarios. Detailed analysis is provided in the project specification document D1.2 *"Functional and System Specifications"*.

RASFF Fields

RASFF Fields	Explanation
notification classification	Classification of the notification according to the definitions given in Regulation 16/2011 and to the guidance given in SOP 5.
information source	Specific source of the information contained in the notification if this is relevant to the understanding of the content of the notification, e.g. a food control body in a third country or a consumer association.
risk decision	Gives information about the evaluation of the risk: - whether the risk is considered to be serious, not serious or undecided; - motivate: why was the risk evaluated as serious (only to be added when the evaluation as serious risk is not straight forward).

product category	Choose the product category from one of the two lists (alphabetical order) or enter it into the other field if the category is not among the entries of the lists or if there are more than one (for more than one product belonging to different categories).
product name(s) (on label)	Precise product name(s), characterising the product(s), without using any commercial name; often the product name on the label that can be found on the packaging.
product CN code	Enter the Common Nomenclature code for the product concerned.
product aspect	Here you should enter important characteristics of the product such as the temperature at which it is kept but also e.g. the kind of packaging, etc.
sampling dates	6 separate fields are provided for a maximum of 6 separate values to be entered
sampling info	Make a reference to a compulsory sampling methodology or inform about the circumstances in which the sample was taken (esp. if the sample was taken from an opened packaging of the product etc.).
sampling place	Place where the samples were taken: use the list box provided or the field other if the place is not among the list entries or to specify the name of the operator.
Analytical method(s)	If a specific analytical method was applied, e.g. one described in legislation or in an EN or international standard, enter it here.
hazards identified	Enter the hazards that were evaluated as non-compliant (according to legislation or risk evaluation) as a result of the analysis or analyses.

RASFF List of values

In RASFF when there is mention of an "open list", it is a list of entities to which new entities could be added. The table below outlines the main lists of values used in RASFF.

RASFF Lists of Values	Explanation
notification type	Food, food contact material, feed
notification classification	Alert notification, border rejection notification, information notification for Attention, information notification for follow-up, news
product relation	Additional lots, different variety, ingredient, processed product, raw material
risk decision	Serious, not serious, undecided
impact on	Human health, animal health, environment
unit weight/volume	Closed list of units for weight/volume: g, kg, l, ml
temperature	Ambient, chilled, frozen
hazard	Closed list, see annex with an extracted hazards list from RASFF Access database (where the master data for hazards are kept)

durability date	Best before, sell-by, use-by

4.2 World Health Organization (WHO) FOSCOLLAB

The World Health Organization (WHO), through its Department of Food Safety and Zoonoses (FOS), initiated a project named FOSCOLLAB to improve ways of sharing food safety data and information to support risk assessment and decision-making in food safety. FOSCOLLAB is a platform accessible from internet and displaying together within dashboards various data (quantitative and qualitative) and information (e.g. expert advice) useful for food safety professionals. FOSCOLLAB allows linkages between databases using four criteria: food name, hazard name, country of origin and year for data generation.

FOSCOLLAB Element	Explanation
Sample collection, prep and analysis	Important context can be added to sampling information by also Reporting the sample size, including units, and the sample's representativeness. Where the user is interested in knowing the prevalence of an analyte, or knowing that an analyte is not present with an estimated level of confidence, information about the representativeness of the sample will be very important. In some cases, this is not necessary, for example, where the user of FOSCOLLAB is only seeking an indication of the presence of an analyte
Country of origin of the sample	Country of origin is necessary in identifying the country where contamination occurred
Why sample was collected	Outbreak investigation, recall verification, compliance, random sampling/surveillance, monitoring, baseline studies…
Action Description	Action taken based on laboratory result; e.g. International Health Regulation (IHR) risk assessment/notification
Instrument name	Analytical instrument used to identify analyte ex: Whole-Genome Sequencing (WGS), platforms, test kits, etc.

4.3 Interoperability

PhasmaFOOD will investigate data interoperability with the aforementioned standard food safety data models and will also consider additional available food data models.

Data Interoperability will have to be considered at the layer of the PhasmaFOOD Cloud Platform and the specification of the platform APIs. Adapting the existing data models will enable the PhasmaFOOD application to operate in different environments (e.g. food security checks) and exchange information with existing systems, thus growing the application potential.

5 Data Access and Sharing

Due to the nature of the data involved, some of the results that will be generated by each project phase will be restricted to authorized users, while other results will be publicly available. As is our commitment, data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules. In the case end-user testing will be performed, PhasmaFOOD users would be required to pre-register and consent using the system. Then they will need to authenticate themselves against a user database. If successful, the users will have roles associated with them. These roles will determine the level of access that a user will be given and what they will be permitted to do.

As the raw data included in the data sources will be gathered from the closed and controlled laboratory experiments, collected measurements will be seen as highly commercially-sensitive. Therefore, access to raw data can only take place through the partners involved in the performance of the laboratory measurements. For the models to function correctly, the data will have to be included into the PhasmaFOOD cloud database. The results of the food analytics will be secured and all privacy concerns will be catered during the design phase. In the cases of trend analytics, anonymization methods will be applied as part of the built-in cloud platform features.

Publications will be released and disseminated through the project dissemination and exploitation channels to make external research and market actors aware of the project as well as appropriate access to the data.

Within the project, our produced conference papers and journal publications will be Green Open Access and stored in an appropriate repository – such as OpenAIRE (European Comission, 2015), Registry of Research Data Repositories (German Research Foundation, 2015) or Zenodo (CERN Data Centre, 2015).

Since the Data Management is expected to mature during the course of the project, an updated release of this document will follow on M18, where the repositories for data storage will be specified and more detailed information on how these data can be accessed by the wider research community will be provided.

6 Archiving and Preservation

Short Term

All original raw data files and data source processing programs will be versioned over time and maintained in a date-stamped file structure with text files documenting the background. As the data will be stored in the PhasmaFOOD cloud repositories, the data will be automatically backed up based on a standardized schedule. These backups could be brought back online within a reasonable timeframe that will ensure that there is no detrimental effect of the data being lost or corrupted.

Long Term

It is in the project's intentions that the long term high quality final data product generated by PhasmaFOOD will be available for use by the research community as well as industry peers. We will identify appropriate archiving institutions that might serve as long term data preservation.

7 Data Management Plan Checklist

At the end of the project, we will be carrying out the following checklist to ensure that we are meeting the criteria to have successfully implemented an Open Access Data Management Plan. By adhering to the items below, we are confident that the project will provide open access to the appropriate data and software, and thereby, enable researchers to utilize the findings of this project to further expand their knowledge capacity and personal gains as well as to provide the ITS industry with the necessary tools to advance their business and processes.

1. Discoverable:
    a. Are the relevant data that are to be made available, our project publications or any Open software that has been produced or used in the project, easily discoverable and readily located?
    b. Have we identified these by means of a standard identification mechanism?
2. Accessible:
    a. Are the data and associated software in the project accessible, where appropriate, and what are the modes of access, scope for usage of this data and what are the licensing frameworks, if any, associated with this access (e.g. licensing framework for research and education, embargo periods, commercial exploitation, etc.)?
3. Useable beyond the original purpose for which it was collected:
    a. Are the data and associated software, which are made available, useable by third parties even after the collection of the data?
    b. Are the data safely stored in certified repositories for long term preservation and curation?
    c. Are the data stored along with the minimum software, metadata and documentation to make them useful?
4. Interoperable to specific quality standards:
    a. Are the data and associated software interoperable, allowing data exchange between researchers, institutions, organizations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?

8 Conclusions

This deliverable has provided an overview on how to build the data collecting - and sharing plan during the course of the PhasmaFOOD project and after the project will be finished. This deliverable is regarded as a live document which will be updated incrementally as the project progresses. This version sets the overall framework that will form the basis for two additional iterations on M18 and M36, towards the overall delivery of a comprehensive document at the end of the project.

In this version of the deliverable, we outlined the descriptions of the Use Case related Datasets, which are still being collected as part of controlled laboratory measurements. Standardization and interoperability aspects have been introduced as well as sharing and access procedures of the project data.

It should be noted that an accurate description of the datasets to be produced (but also in some cases collected and processed) during the early months of the project is challenging. It is therefore not feasible to exhaustively list the datasets that will be subsequently used and/or produced, but to highlight datasets that the consortium agrees are the most relevant at this stage of the poject. For these reasons, regular updates of the Data Management Plan are expected. More specifically, these updates will be done in PhasmaFOOD as part of the mid- term and final project reviews and at other moments as decided by the PhasmaFOOD consortium. Updates to the DMP will appear within the dissemination reports of the project.

The upcoming revisions of this deliverable will focus -among other- to a fuller presentation of different laboratory experiments data collection, description of the PhasmaFOOD database characteristics, update of data access and sharing and update of data interoperability priorities. Data regarding hardware design and specification as well as software design and implementation (cloud, mobile and embedded level) will be also addressed.

9 References

"RASFF annual report 2015", European Commission — Health and Food Safety — 2016, Retrieved from: https://ec.europa.eu/food/sites/food/files/safety/docs/rasff_annual_report_2015.pdf

"The report of the FOSCOLLAB Data Quality Working Group" Retrieved from WHO: http://www.who.int/entity/foodsafety/databases/dqwg_report.pdf?ua=1

"Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", Version 3.2, 21 March 2017. Retrieved from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

"Guidelines on FAIR Data Management in Horizon 2020", Version 3, July 2016. Retrieved from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Comission. (2015, October 26). *OpenAIRE*. Retrieved from OpenAIRE: www.openaire.eu

German Research Foundation. (2015, October 26). *re3data.org*. Retrieved from re3data.org: www.re3data.org

Google Inc. (2015). *Static Transit*. Retrieved from Google Developers: https://developers.google.com/

Government Digital Service -UK. (2015). *Guidance, National public transport access nodes*. Retrieved from Gov.UK: http://www.naptan.org.uk/