CAPSELLA

COLLECTIVE AWARENESS PLATFORMS FOR ENVIRONMENTALLY-SOUND LAND MANAGEMENT BASED ON DATA TECHNOLOGIES AND AGROBIODIVERSITY

Deliverable 1.2

Data Management Plan and Support Pack

Date:	December 2016
Authors:	Maria Boile, Panagiota Koltsida, Eleni Toli (ATHENA RC)
Contributors	Panagiotis Zervas (Agroknow), Peter Paree (ZLTO), Haris Papageorgiou (ATHENA RC)
Dissemination level:	PU
Work package	WP1
Version:	1.1
Keywords:	Data, datasets, use, sharing, access, preservation, dissemination
Description:	This deliverable describes the Data Management principles for the CAPSELLA project.

Horizon 2020
European Union Funding
for Research & Innovation

ICT-10-2015 Collective Awareness Platforms for Sustainability and Social Innovation

CAPSELLA (Collective Awareness PlatformS for Environmentally-sound Land management based on data technoLogies and Agrobiodiversity)

Project No. 688813

Project Runtime: January 2016 – June 2018

Document Metadata

Quality Assurors and Contributors

Quality assuror(s):	Panagiotis Zervas (Agroknow)

Version History

Version	Date	Description
1.0	27 June	Final version submitted
1.1	28 December	Updated version: Executive Summary, Explanatory Remarks, addition of FAIR principles (Section 2), new datasets in Section 3, Conclusions

Glossary

ABBREVIATION	DEFINITION
API	Application Programming Interface
CC	Creative Commons
CC-BY	Creative Commons Attribution Licence
CC-BY-NC	Creative Commons Attribution-Non Commercial Licence
CSV	Comma Separated Values
DM	Data Management
DMP	Data Management Plan
DMS	Data Management System
DoA	Description of Action
re3data	Registry of Research Data Repositories
SW	Software
TSV	Tab Separated Values
UI	User Interface

Disclaimer

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of CAPSELLA consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (http://europa.eu.int/)

CAPSELLA is a project partially funded by the European Union

Executive Summary

People create so much data as never before in the human history. In 2013 every day we created 2.5 quintillion bytes of data[1]. 90% of the data existing in the world at that point had been created in the previous two years alone. It has been said, that data is the gold of today. Access to and use of data are key elements for scientific progress and social innovation.

Ascribing to data this value, also means that issues such as collection, analysis, access, maintenance, and preservation are increasingly important. Acknowledging this fact, the European Commission endorses open access to scientific information and encourages H2020 projects to participate to a pilot action on open access to research data. Part of this action is the obligation to draft and use of a Data Management Plan (DMP) for all datasets to be collected, processed or generated by a research project.

The current deliverable specifies how research publications and data will be collected, processed, monitored, catalogued, and disseminated during the CAPSELLA lifetime. It will be a "living document" and the manual for the CAPSELLA partners for handling of data. The first version of it was available in June 2016, containing the first group of datasets collected. It was, however, updated as the work in project and particularly in the pilots evolved. The current, second, version contains datasets collected up to December 2016.

[1] SINTEF. "Big Data, for better or worse: 90% of world's data generated over last two years." ScienceDaily. ScienceDaily, 22 May 2013. www.sciencedaily.com/releases/2013/05/130522085217.htm

Table of Contents

List of Figures

1. Explanatory Remarks and Guidelines

1.1 CAPSELLA Data management approach and practices

CAPSELLA participates in the Pilot on Open Research Data in Horizon 2020, in line with the Commission's Open Access to research data policy[2] for facilitating access, reuse and preservation of its scientific publications and research data, as they are also outlined in Articles 29.2 and 29.3 of the CAPSELLA Grant Agreement.

This first version of the Data Management Plan (DMP) followed the Guidelines on Data Management in Horizon 2020[3] (v.2.1, 15 February 2016) for the identification of Research Datasets. This first version contained an initial description of datasets collected. The current version includes the additional datasets identified in the months after the initial submission of the deliverable. Furthermore, it also answers on how CAPSELLA addresses the FAIR principles, as they have been captured in the Template H2020 Data Management Plan (v1.0 – 13.10.2016)[4].

For creating a common ground of understanding, in the current deliverable we use the following definition of dataset:

A dataset is any set of data (no matter how many files it materialises) that is worth to be considered as a unit for data management activities[5]

Any of the following can be considered as a possible dataset in the context of a project:

- Any dataset produced by aggregating data from data providers for the sake of analysing it;
- Any dataset produced by aggregating data from data providers for the sake of building an integrated dataset out of the aggregated data (e.g. this is the case of Knowledge Bases);
- The material of a training course;
- A dataset documenting and providing evidence for either a report or a publication produced in the context of project activities.

In the case of CAPSELLA we have identified four main categories of datasets:

- *Core Datasets,* i.e., datasets related to the main project activities and worth to be used by the project. The majority of these datasets pre-exist CAPSELLA and they are publicly available;

[2] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 2.1, 15 February 2016

[3] Guidelines on Data Management in Horizon 2020: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[4] http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx

[5] Renear, A.H., Sacchi, S. and Wickett, K.M. (2010) Definitions of dataset in the scientific and technical literature. Proceedings of the American Society for Information Science and Technology, 47(1): 1–4. DOI: 10.1002/meet.14504701240

- *Produced Datasets,* datasets resulting from the operation and evaluation of CAPSELLA's pilot applications. These may include but are not limited to sensor data, field data and user related datasets.
- *Project Related Data*, i.e., datasets resulting from the operation of the CAPSELLA project and produced by the CAPSELLA consortium. These datasets are collections of standard material produced by a research project, e.g. deliverables, dissemination material, training material, scientific publications;
- *Software*, i.e., datasets resulting from the software developed in the frame of CAPSELLA. These datasets are mainly either software artefacts and source code and can be used for various purposes including research tasks or the development of new software components.

After a first evaluation of the collected datasets and for the time being, it becomes clear that the majority of them belongs to the first category *Core Datasets*. This is, however, normal as the main project activities that generate data will deliver results in the upcoming months of the project.

In the next paragraphs we describe the methodology for the project coordinator and the partners, explaining how they should practically apply the guidelines during their research activities, which software tools and services they should use, and how they can align the project requirements with their institutions' standard practices and systems.

1.2 Methodology for the DMP

Apart from the list of datasets described, there are two basic elements characterising each Data Management Plan, i.e., (a) the set of information to be captured per dataset and (b) the set of dataset management practices and tools that can be used.

For capturing both, we have carried out the following workflow:

- The main information categories have been collected.
- Details have been added to each main category.
- A DMP questionnaire has been drafted, and sent out to the partners for validation.
- The final version of the questionnaire has been transferred to Google forms so as to allow the direct input from all partners.

Sections 1.3 and 1.4 describe how CAPSELLA implements these two basic elements.

1.3 Key features of each dataset

For each dataset, the following classes of information are collected: (a) dataset description, origin, use and availability, (b) data and metadata standards, (c) dataset use and sharing policies and (d) data handling and synchronisation and (e) preservation policies.

Dataset description

The provided description should characterise the dataset, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse. Requested information includes:

- ***Description:*** A summary of the dataset content;
- ***Generated/Collected:*** Indicate whether the dataset will be genuinely generated within the project or will be produced by aggregating content out of existing datasets / data sources. In case the dataset is collected indicate the origin(s);
- ***Nature***: Indicate the typology of content made available through the dataset. Datasets can either comprise items of one single typology only (e.g. images) or they can comprise items of heterogeneous typologies;
- ***Scale***: Indicate the size of the dataset (and the growth). Size can be expressed by reporting the number of entries, the disk space, and any other metric worth to use for the specific case;
- ***Potential use***: Indicate any potential use of the dataset, any community or use case that might benefit from the dataset;
- ***Scientific publication***: Indicate whether the dataset will be actually underpinning any scientific publication. If you already have, give a reference to the publication;

The figures below summarise the information on the first CAPSELLA datasets with regards the method of acquisition and the typology of the content (bullets 2 and 3 in the above list).

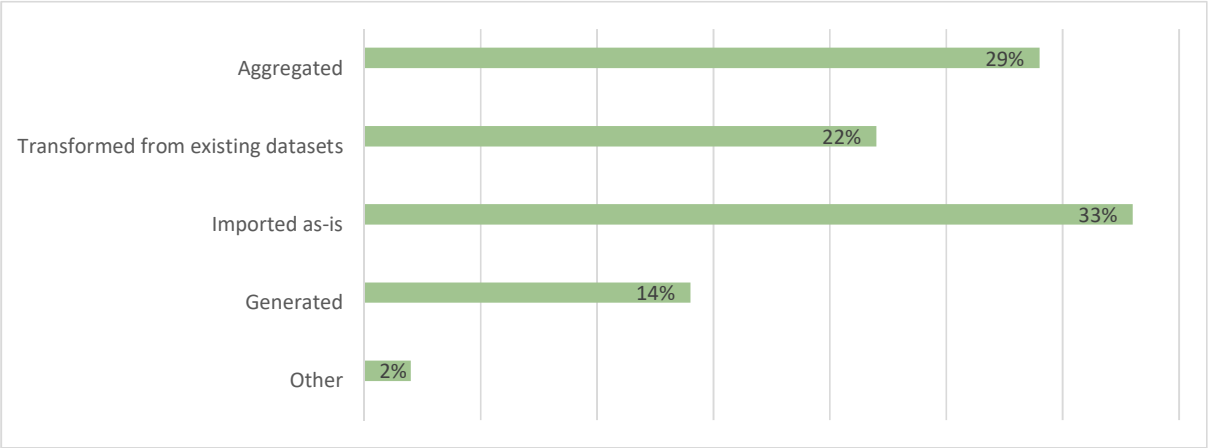

Figure 1-1: Current statistics on the method of dataset acquisition

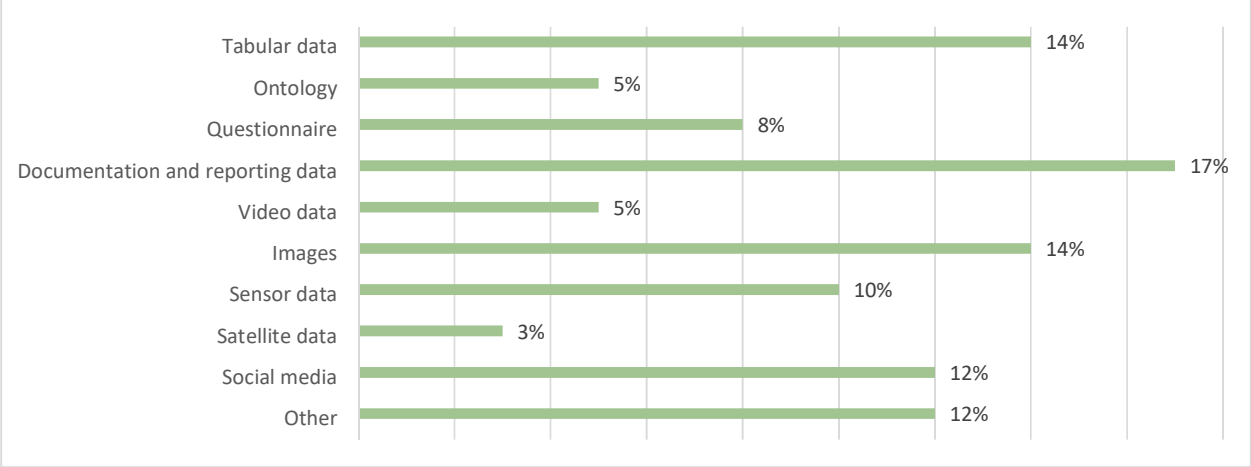

Figure 1-2: Current Statistics of datasets nature

Data and Metadata Standards

Reference to existing suitable data and metadata representation standards of the discipline. If these do not exist, an outline on how and what data and metadata will be created. The most suitable standards for representing the dataset content and the associated metadata should be indicated. Such standards are expected to promote the (re-) use of the dataset.

Dataset sharing

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related). Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). Requested information include:

- ***Access Procedure***: Indicate how a user will have access to the dataset (e.g. API, web site);
- ***Embargo Periods***: Indicate whether there is an embargo period or not;
- ***Dissemination Mechanisms***: Indicate how the availability of the dataset will be announced, e.g. publishing the dataset in the CAPSELLA Data Catalogue;
- ***Software and tools for re-use***: Indicate if there is any specific software to be used for consuming the dataset;
- ***Access rights***: Indicate what are the policies governing access to the dataset, e.g. the dataset is "open", the dataset is made available to authorised users only;
- ***Licence:*** Indicate what can be done once having accessed the dataset, e.g. CC Licences;

- ***Repository***: Indicate whether the dataset has been published in any "repository" including CAPSELLA ones;

Dataset archive and preservation

Description of the procedures and tools that will be put in place for long-term preservation of the data. Requested information include:

- ***Preservation strategy:*** Indicate the strategy to be implemented including the duration;
- ***Preservation tool:*** Indicate what are the instruments put in place to implement the strategy.

1.4 Summary of Data Management related tasks

In the following table we summarise the main Data Management (DM) tasks for CAPSELLA and provide related guidelines of how these issues should and will be handled by the project partners.

DM related tasks	CAPSELLA activity
Deposit, storage and preservation of data	Data and metadata collected, aggregated, harvested, created in the CAPSELLA project will be deposited in the CAPSELLA platform, which will be available after its first release as indicated in the DoA. Data compatible with the agINFRA thematic aggregator will also be deposited there.
Monitoring, tracking, dissemination of research publications	CAPSELLA research publications should be stored in the project workspace (Redmine). The project management tool will allow the efficient monitoring of them. They will be also stored on and promoted through the CAPSELLA platform, the Zenodo repository, agINFRA, the Registry of Research Data Repositories (re3data) and any other relevant channel and repository.
Allowing access, mining, exploitation, reproduction and dissemination to third parties	CAPSELLA will offer standard APIs, such as OAI-PMH [6]and User Interface (UI) to allow to third parties to discover and access data and related metadata.
Tools and instruments to validate the results	The datasets will be exploited by the CAPSELLA's pilots and will thus be evaluated on the produced results from the communities participating in the project.

[6] OAI-PMH protocol: https://www.openarchives.org/pmh/

produced by using the specific data sets	
Guidelines, support material, proposed workflows and tools	The CAPSELLA wiki on the project management platform contains the DMP questionnaire on the main attributes each dataset has to include in order to be added to and exploited by the CAPSELLA platform, how to collect it, and where to store it

2. FAIR Principles in CAPSELLA

In the following we present how CAPSELLA addresses the FAIR data principles (findable, accessible, interoperable, reusable), as they are defined in the October version of the Template H2020 Data Management Plan. The information we provide for each dataset (in the Section 3 below) includes in more detail, how these principles are applied for each dataset.

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)?

Each dataset stored into the system will have a unique identifier.

What naming conventions do you follow?

For discoverability reasons we keep the names the generators of datasets have given to them. Data creators have the possibility to choose the name for their dataset. System keeps information about provenance: who, when, how entered the dataset.

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

Each dataset is accompanied with a set of descriptive metadata and a number of tags. Search is available either by free-text or using facets or by tags. This user interface is available through the data catalogue portal.

Do you provide clear version numbers?

There is no need for versioning, the collection and enrichment of the datasets will be a continuous process, there is no need to work with previous versions of the same dataset.

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.

We will use a generic schema, such as Dublin Core[7] to be able to provide a common set of metadata for every dataset, independent of its nature and semantics.

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.

CAPSELLA will make use of open datasets. Project generated data will be also openly accessible by default, using the appropriate licensing schemes. Special provisions apply to RSR data.

[7] Dublic Core (DC) schema: http://dublincore.org/

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

There are no related provisions in the CA

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

All datasets are available through the CAPSELLA metadata catalogue portal which contains for each dataset at least one endpoint for accessing it. Endpoints may be either internal coming from the Capsella Data Management System (DMS) or pointing to external repositories.

All datasets will be automatically harvested by the CIARD-RING, the global catalogue of research data and software for agriculture, food and the environment.

We will be OpenAIRE compliant by offering an OAI-PMH service, which follows the OpenAIRE guidelines.

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

It depends on the nature and format of the data: Web browser for standard http urls, geoserver for shape files, REST api calls to interact with the available services, etc.

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

A dedicated page in the CAPSELLA data catalogue will provide information about the software that should be used to consume the various datatypes.

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

Not relevant

Have you explored appropriate arrangements with the identified repository?

CAPSELLA is in the process of making all necessary arrangements with the open access repositories in question.

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

Authentication and authorisation mechanisms will be put in place for restricted data.

Is there a need for a data access committee?

No

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

Machine readable license is currently not supported

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

Authentication and authorisation mechanisms will be put in place

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)?

Yes, all the above.

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

We will use OAI_DC, which is the mandatory schema to follow when you offer your data through the OAI-PMH publisher.

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

Where applicable we will use standard controlled vocabularies, such as the Internet Media Types (MIME), describing the data type. For the agricultural related datasets, we will use a reference to the AGROVOC Ontology. We have created a Glossary of CAPSELLA agroecology terms. We are assigning for each term of the glossary the AGROVOC reference URI.

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?

If a term is not present in the AGROVOC ontology, we are adding the references to the closest AGROVOC terms with the type if relationship (e.g. narrow concept; influence; is_component_of etc...). In the CAPSELLA pilot we will ensure that all the basic definition (e.g. crops, practices, inputs) will be described using an AGROVOC Term.

2.4 Increase data re-use (through clarifying licences)

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

We will have 3 different levels: a. Attribution-NonCommercial “CC BY-NC”, b. Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”, c. Attribution-NonCommercial-NoDerivs “CC BY-NC-ND”

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No embargo period for open data

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Datasets will be reused and exploited in particularly by the organisations participating in the CAPSELLA pilots (e.g. City of Cork, Ordr, farmers belonging to the ZLTO network). All datasets collected during the project will maintain the same qualities after the end of it.

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

For at least 2 to 5 years after the closing of the project.

Are data quality assurance processes described?

Datasets aggregated, are from trusted sources, thus quality assurance has taken place at the source. For the CAPSELLA data, the communities collecting data have described quality and validation criteria already in the collection phase of the data, so that data arrive “clean”.

3. Datasets

This section provides information about each dataset, including its origin (in case it is collected), nature and scale, possible uses, whether it underpins a scientific publication, data and metadata formats, etc.

Incomplete information for the provided datasets will be filled in the next revisions of the DMP document.

3.1 Plant Ontology

A. Description, Origin, Use, Availability	
Description	The Plant Ontology is a structured vocabulary and database resource that links plant anatomy, morphology and growth and development to plant genomics data. The PO is under active development to expand to encompass terms and annotations from all plants.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://agroportal.lirmm.fr/ontologies/PO http://plantontology.org/
Origin: licenses of the original data	Creative Commons Attribution 4.0 International License http://plantontology.org/node/279
Nature of the dataset content	Ontology
Estimated scale (size) of the dataset	1729 classes
Foreseen use: specific applications or research purposes	Exploited by the pilots in aggregation with project's datasets
References: related publications	
Data formats used (e.g. XML, CSV)	RDF/XML, OWL

Metadata standards used	RDF/XML
When will be the dataset available to the project (month/year)	Will be imported to the CAPSELLA platform after its first release as indicated in the DoA
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://agroportal.lirmm.fr/ontologies/PO
Is there a specific software required for consuming the dataset?	S/W capable of consuming ontologies
Dissemination mechanisms through which the availability of the dataset will be announced	Data catalogue portal Repositories in which CAPSELLA will deposit the data
License of the dataset	Creative Commons Attribution 4.0 International License
Dataset access policy	Open
Dataset consuming procedure for a user	Website download, URL, standard APIs
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	NA
In case of creating or collecting data in the field how will you ensure its safe transfer	NA

into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin.
Foreseen preservation period duration	
Tools to implement the preservation strategy	

3.2 IBP Crop Research Ontology

A. Description, Origin, Use, Availability	
Description	Describes experimental design, environmental conditions and methods associated with the crop study/experiment/trail and their evaluation.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://agroportal.lirmm.fr/ontologies/CO_715 http://www.cropontology.org/
Origin: licenses of the original data	Creative Commons Attribution 4.0 International License
Nature of the dataset content	Ontology
Estimated scale (size) of the dataset	256 classes
Foreseen use: specific applications or research purposes	Exploited by the pilots in aggregation with project's datasets
References: related publications	

Data formats used (e.g. XML, CSV)	OBO, RDF/XML, CSV
Metadata standards used	
When will be the dataset available to the project (month/year)	Will be imported to the CAPSELLA platform after its first release as indicated in the DoA
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://agroportal.lirmm.fr/ontologies/CO_715
Is there a specific software required for consuming the dataset?	S/W capable of consuming ontologies
Dissemination mechanisms through which the availability of the dataset will be announced	Data catalogue portal Repositories in which CAPSELLA will deposit the data
License of the dataset	Creative Commons Attribution 4.0 International License
Dataset access policy	Open
Dataset consuming procedure for a user	website download, URL, standard APIs
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	NA

In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	NA
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin.
Foreseen preservation period duration	
Tools to implement the preservation strategy	

3.3 AGRIS

A. Description, Origin, Use, Availability	
Description	AGRIS contains approximately 8.5 million agricultural datasets, provided by many different providers around the world
Acquisition	Aggregated
Origin if derived from existing datasets or imported	http://agris.fao.org/agris-search/index.do
Origin: licenses of the original data	http://agris.fao.org/content/acceptable-use-policy
Nature of the dataset content	Documentation and reporting data
Estimated scale (size) of the dataset	8,500,000 dataset records

Foreseen use: specific applications or research purposes	Find publications for specific topics of agricultural and food science
References: related publications	
Data formats used (e.g. XML, CSV)	XML, OAI-PMH targets + any machine readable formats
Metadata standards used	Mendeley
When will be the dataset available to the project (month/year)	Specific datasets will be harvested when required
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	FAO AGRIS, Triple Store, http://capsella.madgik.di.uoa.gr/dataset/agris
Is there a specific software required for consuming the dataset?	Any software capable of processing xml/rdf metadata formats
Dissemination mechanisms through which the availability of the dataset will be announced	
License of the dataset	License for publications are declared by each data provider and included in the metadata. The metadata for the publications are currently available as open dataset under CC4 BY SA
Dataset access policy	Open access. Everyone can access and view the data
Dataset consuming procedure for a user	SPARQL end point: http://202.45.139.84:10035/catalogs/fao/repositories/agris#query

In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	For the metadata, no. For the actual publications, it depends on the data provider
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The overall strategy is to continue collecting metadata from providers around the world each month and publish them in AGRIS
Foreseen preservation period duration	
Tools to implement the preservation strategy	In order to collect the metadata from providers, we have set up a Service, the AKstem AGRIS Service, where providers can register and upload their metadata. We then collect the metadata and process them in order to publish them in AGRIS

3.4 GLN

A. Description, Origin, Use, Availability	
Description	Global Learning Network, contains open educational resources for school, higher and professional educational
Acquisition	Aggregated
Origin if derived from existing datasets or imported	
Origin: licenses of the original data	License for educational resources are declared by each data provider and included in the metadata

Nature of the dataset content	Documentation and reporting data
Estimated scale (size) of the dataset	20.000 records approximately
Foreseen use: specific applications or research purposes	Find educational resources for agricultural and food sciences
References: related publications	
Data formats used (e.g. XML, CSV)	XML
Metadata standards used	Json
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	ILSE, GFSP, Organic, Edunet
Is there a specific software required for consuming the dataset?	Any code written application which can call the API and use the machine readable format returned
Dissemination mechanisms through which the availability of the dataset will be announced	
License of the dataset	License for educational resources are declared by each data provider and included in the metadata
Dataset access policy	Open access

Dataset consuming procedure for a user	Standard API
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No for the metadata. For the actual publications it depends on the data provider
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We don't since this is open data
C. Preservation	
Preservation strategy	Metadata will be preserved under a green learning network URI, the educational objects are preserved by the data providers
Foreseen preservation period duration	
Tools to implement the preservation strategy	API

3.5 Electro Magnetic Conductivity (EM) of soil

A. Description, Origin, Use, Availability	
Description	Electro Magnetic Conductivity (EM) of soil on different depths (till 3m) EM values show how much organic matter is in the soil, so when it's low, more compost is needed. But we should correct on other aspects, suc as elevation and historical harvest data
Acquisition	Imported as-is

Origin if derived from existing datasets or imported	
Origin: licenses of the original data	
Nature of the dataset content	Sensor data
Estimated scale (size) of the dataset	400.000 measurements
Foreseen use: specific applications or research purposes	Exploited by the precision farming pilot
References: related publications	
Data formats used (e.g. XML, CSV)	CSV, Shape files, DBF
Metadata standards used	
When will be the dataset available to the project (month/year)	Will be imported to the CAPSELLA platform after its first release as indicated in the DoA
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Geospatial tools for consuming shape files or any other tool or software capable of consuming CSV files
Dissemination mechanisms through which the availability of the dataset will be announced	

License of the dataset	
Dataset access policy	
Dataset consuming procedure for a user	Standard APIs
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin.
Foreseen preservation period duration	
Tools to implement the preservation strategy	

3.6 WheatPhenotype

A. Description, Origin, Use, Availability	
Description	WheatPhenotype is an ontology of wheat traits and environmental factors that affect these traits. They include resistance to disease, development nutrition, bread quality, etc. Environmental factors include biotic and abiotic factors
Acquisition	Imported as-is

Origin if derived from existing datasets or imported	http://agroportal.lirmm.fr/ontologies/WHEATPHENOTYPE http://genome.jouy.inra.fr/bibliome/WheatPhenotypeOntology
Origin: licenses of the original data	Open
Nature of the dataset content	Ontology
Estimated scale (size) of the dataset	300-499 Phenotype and Trait
Foreseen use: specific applications or research purposes	
References: related publications	http://maiage.jouy.inra.fr/?q=fr/ontologies-corpus
Data formats used (e.g. XML, CSV)	OBO, RDF/XML, CSV
Metadata standards used	
When will be the dataset available to the project (month/year)	Will be imported to the CAPSELLA platform after its first release as indicated in the DoA
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	S/W capable of consuming ontologies
Dissemination mechanisms through which the availability of the dataset will be announced	

License of the dataset	No specific licence
Dataset access policy	Open
Dataset consuming procedure for a user	Website download, URL, standard APIs
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.7 Eurostat statistics

A. Description, Origin, Use, Availability	
Description	Eurostat database contains more than 4 600 datasets with more than 1.2 billion statistical data values
Acquisition	Imported as-is

Origin if derived from existing datasets or imported	http://ec.europa.eu/eurostat/data/database
Origin: licenses of the original data	Eurostat License: http://ec.europa.eu/eurostat/about/our-partners/copyright
Nature of the dataset content	Tabular data, with various data types
Estimated scale (size) of the dataset	Depends on the specific dataset
Foreseen use: specific applications or research purposes	Eurostat statistics at European level could enable stakeholders to perform comparisons between countries and regions
References: related publications	
Data formats used (e.g. XML, CSV)	CSV / TSV
Metadata standards used	
When will be the dataset available to the project (month/year)	Specific datasets will be imported to the CAPSELLA platform whenever required
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/eurostat
Is there a specific software required for consuming the dataset?	Any tool or software capable of consuming CSV / TSV files
Dissemination mechanisms through which the availability of the dataset will be announced	Eurostat's website and CAPSELLA data catalogue

License of the dataset	Origin's dataset remains: http://ec.europa.eu/eurostat/about/our-partners/copyright
Dataset access policy	Open
Dataset consuming procedure for a user	Website download, URL, standard API
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.8 The Yelp Challenge Dataset

A. Description, Origin, Use, Availability	
Description	The Yelp Challenge Dataset includes information about local businesses in 10 cities across 4 countries. It consists of 2.2M reviews and 591K tips by 552K Yelp users. 566K business attributes and 200,000 pictures are available

Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://www.yelp.com/dataset_challenge
Origin: licenses of the original data	The dataset is available for academic purposes under the YELP terms stated in the "Dataset Terms of Use" document accessed through this URL: https://www.yelp.com/html/pdf/Dataset_Challenge_Academic_Dataset_Agreement.pdf
Nature of the dataset content	Social media data
Estimated scale (size) of the dataset	The dataset size is in the millions.
Foreseen use: specific applications or research purposes	The foreseen use of the dataset will be specified in the context of the Pilots and the Food research prototypes
References: related publications	Hundreds of academic papers written
Data formats used (e.g. XML, CSV)	Json
Metadata standards used	N/A
When will be the dataset available to the project (month/year)	Available through https://www.yelp.com/dataset_challenge/dataset
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	N/A
Is there a specific software required for consuming the dataset?	Json readers are needed in order to read and process (parts of) the data.

Dissemination mechanisms through which the availability of the dataset will be announced	N/A
License of the dataset	The dataset is available for academic purposes under the YELP terms stated in the "Dataset Terms of Use" document accessed through this URL: https://www.yelp.com/html/pdf/Dataset_Challenge_Academic_Dataset_Agreement.pdf
Dataset access policy	See Terms of Use above
Dataset consuming procedure for a user	See Terms of Use above
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	See Terms of Use above
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	See Terms of Use above
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	The foreseen preservation period duration is determined by the policy processes set by the origin
Tools to implement the preservation strategy	The tools needed to implement the preservation strategy are set and supported by the origin

3.9 The Capsella Food Dataset

A. Description, Origin, Use, Availability	
Description	The Capsella Food Dataset provides the data that is needed in order to investigate food consumption patterns and the relationship between food consumption and social media in general. Data will be collected from a rich set of OSNs (indicatively: Foursquare, Instagram, Facebook, Reddit etc.)
Acquisition	Aggregated Filtered based on the CAPSELLA research requirements
Origin if derived from existing datasets or imported	N/A
Origin: licenses of the original data	
Nature of the dataset content	Social media data
Estimated scale (size) of the dataset	The estimated size is in the millions
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	Heterogeneous, depending on the OSN; mostly in json. XML / CSV/ TXT will be supported through specific converters
Metadata standards used	N/A
When will be the dataset available to the project (month/year)	
B. Dataset use and sharing	

Repositories that the dataset or/and its metadata have been published in	N/A
Is there a specific software required for consuming the dataset?	The Dataset will be consumed through Apache Kafka consumers encapsulated in asynchronous processing workflows executing a series of tasks on the data in a distributed manner.
Dissemination mechanisms through which the availability of the dataset will be announced	Not specified yet
License of the dataset	Under the specific Terms of Use of each OSN
Dataset access policy	
Dataset consuming procedure for a user	Not specified yet
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	
Tools to implement the preservation strategy	

3.10 The Capsella Twitter Dataset

A. Description, Origin, Use, Availability	
Description	The Capsella Twitter Dataset is a dynamic stream collection related to the major domains of the project: namely, agrobiodiversity and the food supply chain and its emerging topics of interest. Its primary focus is the support of conducting research in order to investigate correlations between food consumption patterns and perceptions and sentiments expressed
Acquisition	Collected through the Twitter APIs and filtered on the Capsella topics of interest
Origin if derived from existing datasets or imported	Twitter (through its APIs) is the source of this dataset
Origin: licenses of the original data	The Capsella Twitter dataset consists of public tweets; there are restrictions on the distribution of the dataset following the Twitter policy
Nature of the dataset content	Social media data
Estimated scale (size) of the dataset	The estimated size is in the millions
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	N/A
Data formats used (e.g. XML, CSV)	The dataset is encoded in json messages. XML/CSV/TXT types will be supported based on specific converters and in compliance with the project requirements
Metadata standards used	Metadata related to the tweeps, the authors of the tweets are also stored in json format

When will be the dataset available to the project (month/year)	Will be imported to the CAPSELLA platform after its first release as indicated in the DoA
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	N/A
Is there a specific software required for consuming the dataset?	Json readers are needed in order to read and process (parts of) the data.
Dissemination mechanisms through which the availability of the dataset will be announced	Not specified yet
License of the dataset	Under the Terms of Use stated on Twitter
Dataset access policy	https://dev.twitter.com/overview/terms/agreement-and-policy
Dataset consuming procedure for a user	Through analytics REST API services
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	See Terms of Use above
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	See Terms of Use above
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin

Foreseen preservation period duration	CAPSELLA safeguards the long-term preservation of the curated datasets
Tools to implement the preservation strategy	Preservation strategy is facilitated through the CAPSELLA platform

3.11 Compost

A. Description, Origin, Use, Availability	
Description	Information about soil conditions, measures, results (quality and kgs) and consumer/citizen appreciation of the sustainable production method
Acquisition	Transformed from existing datasets and aggregated
Origin if derived from existing datasets or imported	Parcel&Growth: Link to the database in Crop-r.nl via the new farmcloud app (under construction), selecting the parcels that are in the experiment. Aggregate context satellite data, photo & video on the parcels and working methods/machines. Aggregate consumers preference data from platforms with this theme and from social media research. Combine the 2 in a relevant (for consumers stories) set of data from farming and context. visualise these data with GIS, virtual reality and stories. Option: check the consumers preferences via questionnaire
Origin: licenses of the original data	Under construction, within Crop-r, account can be generated
Nature of the dataset content	Social media data, satellite data, sensor data, images, video data, documentation and reporting data, questionnaire results data
Estimated scale (size) of the dataset	Terrabite
Foreseen use: specific applications or research purposes	In the testing period: 25 users
References: related publications	www.vandeborneaardappelen.com / digitale boerderij

Data formats used (e.g. XML, CSV)	Soil sensor data: CSF, with field names in the top line
Metadata standards used	Editeelt (uniformity managed by Agriconnect.nl)
When will be the dataset available to the project (month/year)	December 2016
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Any GIS, and visualisation software
Dissemination mechanisms through which the availability of the dataset will be announced	Sites of participants
License of the dataset	
Dataset access policy	Consumer info open, other data for registered users
Dataset consuming procedure for a user	Not specified yet
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main	Upload in the existing systems of vd borne and transfer them to the CAPSELLA platform

data management system?	
C. Preservation	
Preservation strategy	Include in Crop-r and vandenborneaardappelen.com, evt ZLTO.nl
Foreseen preservation period duration	5 years, maybe earlier incorporated in bigger concept
Tools to implement the preservation strategy	Direct maintenance by vd Borne, Crop-r, ZLTO

3.12 Farmersfood Experience ZLTO

A. Description, Origin, Use, Availability	
Description	Data for optimal storytelling of products and production methods
Acquisition	Generated – Transformed from existing datasets - Aggregated
Origin if derived from existing datasets or imported	1a. social groups in 30 km circle around test location: Barendonck; 1b consumer preferences of these groups on: 2a product data (nutritional value, etc) 2b production data (relate to list in description HACCP, etc), 2c event/experience data (list in description) 2d existing stories. This results in: 3a links with ratings per link (between 1 and 2) and 3b stories (speach or video) edited on the high rating links
Origin: licenses of the original data	1a: geonovum.nl has open datasets on geo basis (and an overview of data portals http://www.geonovum.nl/wegwijzer/dataportalen-0); 2a:https://www.levensmiddelendatabank.nl/Security/Login.aspx?ReturnUrl=/ 2b (1st version to be provided by farmer); 2c see description; 3a to be generated; 3b to be edited
Nature of the dataset content	Social media data, images, video data, documentation and reporting data, questionnaire results data

Estimated scale (size) of the dataset	N/A
Foreseen use: specific applications or research purposes	Farmers: make stories that fit to audience, Consumers: find information about (produce of) their food, ZLTO/research: build up story database
References: related publications	see a5
Data formats used (e.g. XML, CSV)	N/A
Metadata standards used	N/A
When will be the dataset available to the project (month/year)	To be built in the pilot
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	See above
Is there a specific software required for consuming the dataset?	GIS for geonovum, rest is web
Dissemination mechanisms through which the availability of the dataset will be announced	Capsella, ZLTO.nl, news items
License of the dataset	N/A
Dataset access policy	For applicants, 1st consumer entry open
Dataset consuming procedure for a user	See possibilities, subscribe and download app, consume stories with beacons, do rating
In case of a generated dataset, will there be an embargo period for providing	Subscription is technically necessary

(open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	Good coverage of gsm
C. Preservation	
Preservation strategy	It should be a self-learning platform, expanding from consumers and producers' needs
Foreseen preservation period duration	5 years
Tools to implement the preservation strategy	ZLTO will connect farmers, new costs will be covered in a subscription fee / Automatic backups of the CAPSELLA platform repository

3.13 The Capsella Recipes Collection

A. Description, Origin, Use, Availability	
Description	The Capsella Recipes collection aggregates recipes from all over the world. Recipes contain the list of ingredients, the preparation steps to follow, cook time, nutrition info, the number of servings produced and photos of the prepared dish. In addition, user ratings and reviews with comments on the quality of the recipe and suggested improvements
Acquisition	Transformed from existing datasets and aggregated
Origin if derived from existing datasets or imported	The collection comprises datasets from recipe-sharing websites like allrecipes.com and its international sites, cookpad.com, yummly.com etc.
Origin: licenses of the original data	Under the Terms of service of the specific recipe sharing websites

Nature of the dataset content	Social media data, images
Estimated scale (size) of the dataset	The estimated size is in the millions
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	Freeform text and / or transformed in CSV format
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	None
Is there a specific software required for consuming the dataset?	Tools or Software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	
License of the dataset	
Dataset access policy	Per recipe's initial license
Dataset consuming procedure for a user	website download, URL

In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.14 RSR' Variety data

A. Description, Origin, Use, Availability	
Description	Data on varieties includes the varieties' names, their origin or area of traditional cultivation; data on farms and farmers describes who currently holds the variety, either having contributed seeds or information to RSR or having received seeds and information from RSR or other farmers in the network; data on the farms describes their locations, mode of management (organic, biodynamic..., with or without a processing infrastructure, etc.) and information (where farmers have contributed it) about agronomic practices used (rotations, fertilizations, machinery used).
Acquisition	Transformed from existing datasets and aggregated
Origin if derived from existing datasets or imported	This dataset corresponds to what is presently stored in RSR's MySQL database of varieties, farms and use types.
Origin: licenses of the original data	The data provided by farmers (is private) or collected from publicly available data (e.g. scientific or grey literature, personal communications from researchers)
Nature of the dataset content	Scientific data and information, nomenclature, geographical data, images
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	The info types and structure of the data will guide the development of applications and services for the project's Seed scenario
References: related publications	Bioversity International crop descriptors
Data formats used (e.g. XML, CSV)	Relational
Metadata standards used	Not specified

When will be the dataset available to the project (month/year)	In progress
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	The data is not published anywhere, since it is mostly of restricted use; however, it is stored in RSR' server
Is there a specific software required for consuming the dataset?	Software capable of consuming relational data
Dissemination mechanisms through which the availability of the dataset will be announced	
License of the dataset	To be defined
Dataset access policy	Under definition
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	The data currently in RSR's database is partly "generated" in the sense that it is collected among farmers. This data is considered traditional knowledge and is protected by international conventions, most of all the Convention on Biological Diversity (CBD). Therefore an embargo period is not foreseen for releasing this information, which is under the control of the community which provides it and can decide under what conditions to make it available to others.
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	This will be supported by a dedicated CAPSELLA pilot, which will allow the onsite and online collection of data, and their safe storage on the CAPSELLA data platform. Communication between services and the data platform will be secured by exploiting standard communication mechanisms, like authentication tokens. Data collected through RSR staff in the field are currently collected on paper and needs to be transferred in an electronic, online language.

C. Preservation	
Preservation strategy	The master version of the dataset is maintained by RSR on their server
Foreseen preservation period duration	Until RSR is in operation
Tools to implement the preservation strategy	Server maintenance and regular updates or management operations performed by RSR staff or hired ICT administrators

3.15 RSR' experimental field data

A. Description, Origin, Use, Availability	
Description	RSR's experimental fields have been running on-farm since 2010. They have a specific statistical design behind them and are hosted on farms in different environments across the Italian territory (north, centre and south). Different varieties, mixtures and populations are tested and compared in these fields. Farmers in some cases, or technical staff from RSR in others, monitor the fields throughout the cropping season, taking repeated measures from the emergence stage to complete maturation. After the harvest, additional measurements are taken on the grain. Starting in 2016, RSR is embarking on a new model of participatory research, which consists in splitting experimental fields (which at present are quite large and labour intensive) into portions, which are handed out to different farmers, requiring less maintenance and monitoring effort from then and therefore making the experimentation process more accessible even to smaller farmers. Ideally, all farmers taking part to this new stage of the participatory research conducted by RSR, will be able to record the data that until now has been collected on only 6 experimental fields, and that this will contribute to the growth of this dataset.
Acquisition	Generated
Origin if derived from existing datasets or imported	This dataset corresponds to what is presently stored in RSR's paper sheets (excel sheets in some cases) used in the field days.
Origin: licenses of the original data	
Nature of the dataset content	Qualitative and quantitative data on morphology and agronomic performance of varieties, mixtures and populations in the field
Estimated scale (size) of the dataset	

Foreseen use: specific applications or research purposes	The info types and structure of the data will guide the development of applications and services for the project's Seed scenario
References: related publications	Bioversity International crop descriptors
Data formats used (e.g. XML, CSV)	XML, paper
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	In progress
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	The data is not published anywhere, since it is mostly of restricted use; however, it is stored in RSR' records and server (when in electronic form)
Is there a specific software required for consuming the dataset?	Tools or software capable of consuming xml data
Dissemination mechanisms through which the availability of the dataset will be announced	
License of the dataset	To be defined
Dataset access policy	Under definition
Dataset consuming procedure for a user	Consultation of paper or XML records
In case of a generated dataset, will there be an embargo period for providing	The data may be made available after publication of the scientific results of the experimentation but would be released only under specific conditions agreed within the community, since it is data collected on traditional genetic resources maintained on-farm and

(open) access to the dataset? If yes, how long (in months)?	hence falls within the obligations of international conventions in these thematic areas.
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	This will be supported by a dedicated CAPSELLA pilot, which will allow the onsite and online collection of data, and their safe storage on the CAPSELLA data platform. Communication between services and the data platform will be secured by exploiting standard communication mechanisms, like authentication tokens. Data collected through RSR staff in the field are currently collected on paper and needs to be transferred in an electronic, online language.
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by RSR in paper records or, when in excel, on their server
Foreseen preservation period duration	Until RSR is in operation
Tools to implement the preservation strategy	Server maintenance and regular updates or management operations performed by RSR staff or hired ICT administrators

3.16 FDA: Daily Values for Infants, Children Less Than 4 Years of Age, and Pregnant and Lactating Women

A. Description, Origin, Use, Availability	
Description	The datasets contain sets of reference values for nutrients. It helps consumers determine the level of various nutrients in a standard serving of food in relation to their approximate requirement for it. The label actually provides the %DV so that anyone can see how much (what percentage) a serving of the product contributes to reaching the DV.
Acquisition	Transformed from existing datasets, Aggregated
Origin if derived from existing datasets or imported	http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/LabelingNutrition/ucm064930.htm

Origin: licenses of the original data	Most of the information available from the Office of Dietary Supplements (ODS) Web site is within the public domain and unless stated otherwise, may be freely downloaded and reproduced, provided the content has not been changed or modified
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/fda-daily-values-dv-for-adults-and-children-4-or-more-years-of-age
Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access

Dataset access policy	https://ods.od.nih.gov/Health_Information/ODS_Frequently_Asked_Questions.aspx
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.17 FDA: Daily Values for adults and children 4 or more years of age

A. Description, Origin, Use, Availability	
Description	The datasets contain sets of reference values for nutrients. It helps consumers determine the level of various nutrients in a standard serving of food in relation to their approximate requirement for it. The label actually provides the %DV so that anyone can see how much (what percentage) a serving of the product contributes to reaching the DV.
Acquisition	Transformed from existing datasets, Aggregated

Origin if derived from existing datasets or imported	http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/LabelingNutrition/ucm064928.htm
Origin: licenses of the original data	Most of the information available from the Office of Dietary Supplements (ODS) Web site is within the public domain and unless stated otherwise, may be freely downloaded and reproduced, provided the content has not been changed or modified
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/fda-daily-values-for-infants-children-less-than-4-years-of-age-and-pregnant-and-lactating-women
Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the	CAPSELLA website, social media, project events (e.g. hackathons), and other material

availability of the dataset will be announced	
License of the dataset	Open access
Dataset access policy	https://ods.od.nih.gov/Health_Information/ODS_Frequently_Asked_Questions.aspx
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.18 WHO Child Growth Standards

A. Description, Origin, Use, Availability	
Description	The dataset includes the child growth standards that were developed using data collected in the WHO Multicentre Growth Reference Study.

Acquisition	Transformed from existing datasets, Aggregated
Origin if derived from existing datasets or imported	http://www.who.int/childgrowth/standards/en/
Origin: licenses of the original data	Copyright of Data. The Data originates from the WHO Health Equity Monitor database, and is © WHO. For the avoidance of any doubt, WHO hereby asserts its copyright in the Data, and reserves all rights in the Data.
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	http://www.who.int/childgrowth/publications/en/
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Will be imported in February 2017
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Tools and Software capable of consuming CSV files

Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access. Everyone can access and view the data
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.19 Body Mass Index adults normal

A. Description, Origin, Use, Availability

Description	The dataset includes the values that are commonly used to classify normal weight in adults. BMI is defined as the weight in kilograms divided by the square of the height in metres (kg/m2).
Acquisition	Transformed from existing dataset, Aggregated
Origin if derived from existing datasets or imported	http://apps.who.int/bmi/index.jsp
Origin: licenses of the original data	Copyright of Data. The Data originates from the WHO Health Equity Monitor database, and is © WHO. For the avoidance of any doubt, WHO hereby asserts its copyright in the Data, and reserves all rights in the Data.
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/global-database-on-body-mass-index

Is there a specific software required for consuming the dataset?	Tools and Software capable of consuming CSV files
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access. Everyone can access and view the data
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.20 Body Mass Index adults normal

A. Description, Origin, Use, Availability	
Description	The dataset includes the values that are commonly used to classify normal weight in adults. BMI is defined as the weight in kilograms divided by the square of the height in metres (kg/m2).
Acquisition	Transformed from existing dataset, Aggregated
Origin if derived from existing datasets or imported	http://apps.who.int/bmi/index.jsp
Origin: licenses of the original data	Copyright of Data. The Data originates from the WHO Health Equity Monitor database, and is © WHO. For the avoidance of any doubt, WHO hereby asserts its copyright in the Data, and reserves all rights in the Data.
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	For the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	

Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access. Everyone can access and view the data
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.21 FDA Food Safety Recalls

A. Description, Origin, Use, Availability	
Description	The datasets provides information gathered from press releases and other public notices about certain recalls of FDA-regulated products.
Acquisition	Transformed from existing datasets, Aggregated
Origin if derived from existing datasets or imported	http://www.fda.gov/Safety/Recalls/ https://api.fda.gov/food/enforcement.json http://www.fda.gov/AboutFDA/ContactFDA/StayInformed/RSSFeeds/FoodSafety/rss.xml
Origin: licenses of the original data	http://www.fda.gov/AboutFDA/AboutThisWebsite/WebsitePolicies/default.htm#third
Nature of the dataset content	RSS and/or JSON
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	For the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	XML, JSON
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/fda-food-safety-recalls

Is there a specific software required for consuming the dataset?	Tools and Software capable of consuming XML files
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.22 USDA National Nutrient Database for Standard Reference

A. Description, Origin, Use, Availability	
Description	This is a database of foods, their nutrient values, and weights for typical food portions. It is used to code food intake data and to calculate nutrient intakes based on the foods and amounts reported.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	https://ndb.nal.usda.gov/ndb/
Origin: licenses of the original data	Open access
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	For the demonstrators to be developed in the context of the project
References: related publications	https://www.ars.usda.gov/northeast-area/beltsville-md/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/articles-and-presentations-by-ndl-staff/
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	

Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/usda-food-composition-databases
Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.23 USDA Branded Food Products Database

A. Description, Origin, Use, Availability	
Description	The database is the result of a Public-Private Partnership, whose goal is to enhance public health and the sharing of open data by complementing the USDA National Nutrient Database for Standard Reference (SR) with nutrient composition of branded foods and private label data provided by the food industry. The USDA Branded Food Products Database includes the list of ingredients for each product.
Acquisition	Aggregated
Origin if derived from existing datasets or imported	https://ndb.nal.usda.gov/ndb/
Origin: licenses of the original data	Open access
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	For the demonstrators to be developed in the context of the project
References: related publications	https://www.ars.usda.gov/northeast-area/beltsville-md/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/articles-and-presentations-by-ndl-staff/
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	

Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/usda-food-composition-databases
Is there a specific software required for consuming the dataset?	N/A
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.24 AllergenOnline

A. Description, Origin, Use, Availability	
Description	AllergenOnline provides access to a peer reviewed allergen list and sequence searchable database intended for the identification of proteins that may present a potential risk of allergenic cross-reactivity
Acquisition	Transformed from existing datasets, Aggregated
Origin if derived from existing datasets or imported	http://www.allergenonline.com/
Origin: licenses of the original data	Open access
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the demonstrators to be developed in the context of the project
References: related publications	http://www.allergenonline.com/about.shtml
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/allergenonline

Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.25 Soil health pilot data

A. Description, Origin, Use, Availability	
Description	The dataset includes the data collected in the frame of the soil health pilot
Acquisition	Generated
Origin if derived from existing datasets or imported	
Origin: licenses of the original data	
Nature of the dataset content	Questionnaire results data
Estimated scale (size) of the dataset	1-10 MB
Foreseen use: specific applications or research purposes	To support the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	CAPSELLA platform

Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	At least 2 years after the closing of the project
Foreseen preservation period duration	
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.26 European Soil Database

A. Description, Origin, Use, Availability	
Description	The dataset contains the answers collected from farmers about the soil health issue.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	https://drive.google.com/open?id=1rYeptAAnnR8fXjtW23_bs8clf5ATLKFKdacgH69iSIk
Origin: licenses of the original data	Open access
Nature of the dataset content	Questionnaire results data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	

Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.27 Aegilops questionnaires

A. Description, Origin, Use, Availability	
Description	The dataset contains the farmers' results data of the CAPSELLA survey about the soil health issue
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	https://drive.google.com/open?id=1rYeptAAnnR8fXjtW23_bs8clf5ATLKFKdacgH69iSIk
Origin: licenses of the original data	Open access
Nature of the dataset content	Questionnaire results data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/european-soil-data-centre-esdac

Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.28 European Soil Datasets

A. Description, Origin, Use, Availability	
Description	The European Soil Data Centre (ESDAC) contains currently many soil data and information; most of the offered data are at European scale, while, when possible, links to national or global datasets are provided. Datasets are organized in some broad categories: - A first category contains the European Soil Database (ESDB), datasets that have been derived with the help of the ESDB and general European datasets that contain soil properties. - A second category offers data that are related to soil threats (erosion, soil organic carbon, landslides, compaction, salinization, soil biodiversity, contaminated sites, soil sealing, etc.) - A third category offers soil point data (LUCAS, SPADE, etc) - A fourth category contains data that stem from projects.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://esdac.jrc.ec.europa.eu/
Origin: licenses of the original data	Some datasets can be freely downloaded; others will be made accessible after prior registration, through a fill-in form
Nature of the dataset content	Raster data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	Access a map of soil health threat
References: related publications	- Panagos P., Van Liedekerke M., Jones A., Montanarella L. European Soil Data Centre: Response to European policy support and public data requirements. (2012) Land Use Policy, 29 (2), pp. 329-338, doi:10.1016/j.landusepol.2011.07.003

	- The European Soil Data Centre: a one-stop-shop for soil science "Science for Environment Policy": European Commission DG Environment News Alert Service (9 February 2012)
Data formats used (e.g. XML, CSV)	
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field	We do not consider "safe transfers" as those are Open Access data

how will you ensure its safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.29 Worldclim

A. Description, Origin, Use, Availability	
Description	WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km2. These data can be used for mapping and spatial modelling.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://www.worldclim.org/
Origin: licenses of the original data	Open access
Nature of the dataset content	Raster data
Estimated scale (size) of the dataset	(depend on the scale)
Foreseen use: specific applications or research purposes	We could use the climatic average to estimate the mineralisation of organic matter

References: related publications	
Data formats used (e.g. XML, CSV)	Shape files / WMS
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Software capable of processing shape files
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its	We do not consider "safe transfers" as those are Open Access data

safe transfer into the main data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.30 ECMWF reanalysis

A. Description, Origin, Use, Availability	
Description	The dataset includes climatic data generated by a reanalysis process
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/
Origin: licenses of the original data	All Intellectual Property Rights of the Archive Products owned by ECMWF or its licensors shall remain the property of ECMWF or its licensors and the Licensee acknowledges the full title and ownership by ECMWF or its licensors of all the Archive Products supplied.
Nature of the dataset content	Sensor data
Estimated scale (size) of the dataset	Depends on spat & temp extension
Foreseen use: specific applications or research purposes	As a source of weather data for modelling purpose
References: related publications	

Data formats used (e.g. XML, CSV)	WMS
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Software capable of consuming shape files
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL through the CAPSELLA platform
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main	We do not consider "safe transfers" as those are Open Access data

data management system?	
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.31 StatLine dataset

A. Description, Origin, Use, Availability	
Description	StatLine is the electronic databank of Statistics Netherlands. It enables users to compile their own tables and graphs. The information can be accessed, printed and downloaded free of charge.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://opendata.cbs.nl/
Origin: licenses of the original data	Open data. All tables in StatLine are available as open data in the form of datasets. By using web services, the most recent data about a specific topic can be retrieved, filtered and combined. The data is identical in content to the tables which can be retrieved and downloaded from StatLine. The only difference is in the presentation. In this way, Statistics Netherlands aims to promote the widespread use of its statistical data.
Nature of the dataset content	Documentation and reporting data
Estimated scale (size) of the dataset	Depending on the Dataset

Foreseen use: specific applications or research purposes	The data could be used for the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	Tools or software capable of consuming CSV data
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing	No

(open) access to the dataset? If yes, how long (in months)?	
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.32 Spotzi Geo data

A. Description, Origin, Use, Availability	
Description	Spotzi has a great variety of datasets with information about weather and climate conditions, finance and economical development, environment, culture, travel etc. The data is gained from different trustable sources, like the Data Worldbank, national environmental institutions etc. In addition, it has data with information about world boundaries. This region data consists of information like safety- and administrative regions, but also zip code data that we converted in unique shapefile datasets (CSV, XLS(X), GeoJSON, ZIP, KML, GPX).
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://spotzi.com/en/

Origin: licenses of the original data	The Spotzi logo in the lower left corner of a map has to be shown in all applications to be built and all maps to be shared. With all other FREE SERVICES Spotzi must be credited. Credits should also be given to the original owner of the content when Spotzi has mentioned the original owner next to the content.
Nature of the dataset content	Social media data
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	The data could be used for the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV, XLS(X), GeoJSON, ZIP, KML, GPX
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.33 Storytelling dataset

A. Description, Origin, Use, Availability	
Description	The datasets contains farm educational material as video's, coloring, assignments, etc.
Acquisition	Transformed from existing datasets

Origin if derived from existing datasets or imported	http://www.hetkleineloo.nl/ http://www.schoolbordportaal.nl/lespakket-zuivel-melk-kaas-en-boter-van-boerderij-tot-voeding.html
Origin: licenses of the original data	
Nature of the dataset content	Images and video data
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	The data could be used for the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	…
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	…
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	Open access
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.34 Precision Agriculture Compost Overview

A. Description, Origin, Use, Availability	
Description	It contains measurement results of an experiment took place in 2016 in precision application of compost in potato
Acquisition	Generated

Origin if derived from existing datasets or imported	https://www.dropbox.com/sh/ssb37598q9b2kos/AAAYe5qa7x-9MFqLCmWCdgqZa?dl=0
Origin: licenses of the original data	Attribution-NonCommercial-ShareAlike "CC BY-NC-SA"
Nature of the dataset content	Images, documentation and reporting data
Estimated scale (size) of the dataset	2MB
Foreseen use: specific applications or research purposes	Getting more insight in effectiveness of precision application of compost, derive decision support from combi of growth, harvest and sensor data
References: related publications	N/A
Data formats used (e.g. XML, CSV)	pdf (descriptions, images), xlsx (measurements)
Metadata standards used	where needed: agriconnect.nl standards for arable farming
When will be the dataset available to the project (month/year)	December 2016
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	N/A
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	To be discussed
Dataset access policy	Open, in the frame of the licensing scheme above
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.35 Precision Agriculture Compost, pictures

A. Description, Origin, Use, Availability	
Description	Pictures of soil, crop and testing harvest on measuring data in PA compost experiment
Acquisition	Generated

Origin if derived from existing datasets or imported	
Origin: licenses of the original data	Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”
Nature of the dataset content	Images
Estimated scale (size) of the dataset	100 MB
Foreseen use: specific applications or research purposes	For the pilots deployment
References: related publications	N/A
Data formats used (e.g. XML, CSV)	jpg
Metadata standards used	
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	N/A
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	To be discussed
Dataset access policy	Open, in the frame of the licensing scheme above
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	5 years
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.36 Precision Agriculture Compost EC Measurements

A. Description, Origin, Use, Availability	
Description	Conductivity (= water content = organic matter content) in sandy soil of Precision Agriculture Compost experiment.
Acquisition	Generated

Origin if derived from existing datasets or imported	
Origin: licenses of the original data	Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”
Nature of the dataset content	Data from scanning with DualM21 soil scanner including visualization, sensor data
Estimated scale (size) of the dataset	20 MB
Foreseen use: specific applications or research purposes	For the compost demonstrator and analysis of growth and harvest. Scientific use, hackathons.
References: related publications	Available at Univ of Ghent, Belgium
Data formats used (e.g. XML, CSV)	CSV, dbf, shp, pdf
Metadata standards used	
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	N/A
Is there a specific software required for consuming the dataset?	GIS
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”
Dataset access policy	Open, in the frame of the licensing scheme above
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.37 Precision Agriculture Compost, potato growth

A. Description, Origin, Use, Availability	
Description	PA Compost: Potato growth measured in subfields on 3 dates with Fritsmeier sensor
Acquisition	Generated

Origin if derived from existing datasets or imported	Imported from Fritsmeier sensor set on tractor (pictures see vandenborneaardappelen.com)
Origin: licenses of the original data	Attribution-NonCommercial-ShareAlike "CC BY-NC-SA"
Nature of the dataset content	Sensor data
Estimated scale (size) of the dataset	10 MB
Foreseen use: specific applications or research purposes	For the compost demonstrator and analysis of growth and harvest. Scientific use, hackathons.
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	All sensor data are also stored in Cloudfarm database, supported by Crop-R.nl
Is there a specific software required for consuming the dataset?	GIS
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”
Dataset access policy	Open, in the frame of the licensing scheme above
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.38 Precision Agriculture Compost, satellite

A. Description, Origin, Use, Availability	
Description	Sentinel-2 images of PA Compost experiment: processed field images 26/8/2016
Acquisition	Transformed of existing datasets

Origin if derived from existing datasets or imported	Preprocessed data are derived from: http://www.satellietbeeld.nl/ or from sentinel hub: https://scihub.copernicus.eu/dhus/#/home
Origin: licenses of the original data	satellietbeeld.nl: open after login; sentinel: open after login
Nature of the dataset content	Satellite data
Estimated scale (size) of the dataset	1 MB
Foreseen use: specific applications or research purposes	Basis for bigger areas where influence of compost on potato growth is estimated
References: related publications	background information: http://g4aw.spaceoffice.nl/en/
Data formats used (e.g. XML, CSV)	Screenshots in word office
Metadata standards used	
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	All sensor data are also stored in Cloudfarm database, supported by Crop-R.nl
Is there a specific software required for consuming the dataset?	GIS
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”
Dataset access policy	Open, in the frame of the licensing scheme above
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	
Tools to implement the preservation strategy	

3.39 Precision Agriculture Compost, plane images

A. Description, Origin, Use, Availability	
Description	HD plane images showing growth conditions (draught, diseases) in potato
Acquisition	Generated

Origin if derived from existing datasets or imported	Owner is vandenborneaardappelen.com, advisor is Vigilance / j.souren@delphy.nl
Origin: licenses of the original data	
Nature of the dataset content	Sensor data and images
Estimated scale (size) of the dataset	1500 MB
Foreseen use: specific applications or research purposes	Analyse growth conditions at different stages and relate that to harvest and compost application.
References: related publications	
Data formats used (e.g. XML, CSV)	csv, kmz, pdf (illustrate and describe)
Metadata standards used	
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	All sensor data are also stored in Cloudfarm database, supported by Crop-R.nl
Is there a specific software required for consuming the dataset?	GIS
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material

License of the dataset	Attribution-NonCommercial-ShareAlike “CC BY-NC-SA”
Dataset access policy	Open, in the frame of the licensing scheme above
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	
Foreseen preservation period duration	
Tools to implement the preservation strategy	Automatic backups of the CAPSELLA platform repository

3.40 Database on kCal consumption

A. Description, Origin, Use, Availability	
Description	It includes values regarding the kCal consumption per activity
Acquisition	Transformed from existing dataset

Origin if derived from existing datasets or imported	http://apps.who.int/bmi/index.jsp
Origin: licenses of the original data	Copyright of Data. The Data originates from the WHO Health Equity Monitor database, and is © WHO. For the avoidance of any doubt, WHO hereby asserts its copyright in the Data, and reserves all rights in the Data.
Nature of the dataset content	Tabular data
Estimated scale (size) of the dataset	<1MB
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the	CAPSELLA website, social media, project events (e.g. hackathons), and other material

availability of the dataset will be announced	
License of the dataset	Open access. Everyone can access and view the data
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.41 RASFF – the Rapid Alert System for Food and Feed

A. Description, Origin, Use, Availability	
Description	The EU has one of the highest food safety standards in the world – largely thanks to the solid set of EU legislation in place, which ensures that food is safe for consumers. A key tool to ensure the cross-

	border follow of information to swiftly react when risks to public health are detected in the food chain is RASFF
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://ec.europa.eu/food/safety/rasff_en
Origin: licenses of the original data	https://webgate.ec.europa.eu/rasff-window/help/disclaimer.pdf
Nature of the dataset content	
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	To support the Food Apps and the Pilots to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	HTML
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Already available
B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/rasff
Is there a specific software required for consuming the dataset?	

Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	Open access. Everyone can access and view the data
Dataset access policy	
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A
Tools to implement the preservation strategy	N/A

3.42 CIARD RING datasets

A. Description, Origin, Use, Availability

Description	The RING is a global directory of web-based information services and datasets for agricultural research for development (ARD). It is the principal tool created through the CIARD initiative to allow information providers to register their services and datasets in various categories and so facilitate the discovery of sources of agriculture-related information across the world. The CIARD RING stores and makes publicly accessible only the metadata about the registered datasets. The actual content of the registered datasets and the related methods of access (download, special protocol, queries...) are subject to usage rights established by their owners.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://ring.ciard.net/datasets
Origin: licenses of the original data	https://creativecommons.org/licenses/by/4.0/
Nature of the dataset content	Documentation and reporting data
Estimated scale (size) of the dataset	
Foreseen use: specific applications or research purposes	
References: related publications	
Data formats used (e.g. XML, CSV)	XML, OAI-PMH targets + any machine readable formats
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Specific datasets will be harvested when required
B. Dataset use and sharing	

Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/ciard-ring
Is there a specific software required for consuming the dataset?	Any software capable of processing xml/rdf metadata formats
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	License are declared by each data provider and included in the metadata.
Dataset access policy	Open access. Everyone can access and view the data
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A

Tools to implement the preservation strategy	N/A

3.43 DATA.GOV

A. Description, Origin, Use, Availability	
Description	Data.gov federates data from hundreds of sources, including federal government, cities, counties, states, and universities. All contributing organizations are listed under "Organizations" in the top right.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	http://catalog.data.gov/dataset
Origin: licenses of the original data	https://www.data.gov/privacy-policy
Nature of the dataset content	Documentation and reporting data
Estimated scale (size) of the dataset	193,141 datasets
Foreseen use: specific applications or research purposes	For the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV, XML, OAI-PMH targets + any machine readable formats
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Specific datasets will be harvested when required

B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/data-gov
Is there a specific software required for consuming the dataset?	Any software capable of processing xml/rdf metadata formats
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	License are declared by each data provider and included in the metadata.
Dataset access policy	Open access. Everyone can access and view the data
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A

Tools to implement the preservation strategy	N/A

3.44 European Union Open Data Portal datasets

A. Description, Origin, Use, Availability	
Description	The EU Open Data Portal is a single point of access to a growing range of data produced by the institutions and other bodies of the European Union.
Acquisition	Imported as-is
Origin if derived from existing datasets or imported	https://data.europa.eu/euodp/en/data
Origin: licenses of the original data	Data are free to use, reuse, link and redistribute for commercial or non-commercial purposes, https://data.europa.eu/euodp/en/data/
Nature of the dataset content	Documentation data
Estimated scale (size) of the dataset	9255 datasets
Foreseen use: specific applications or research purposes	For the demonstrators to be developed in the context of the project
References: related publications	
Data formats used (e.g. XML, CSV)	CSV, XML, OAI-PMH targets + any machine readable formats
Metadata standards used	Not specified
When will be the dataset available to the project (month/year)	Specific datasets will be harvested when required

B. Dataset use and sharing	
Repositories that the dataset or/and its metadata have been published in	http://capsella.madgik.di.uoa.gr/dataset/europen-union-open-data-portal
Is there a specific software required for consuming the dataset?	
Dissemination mechanisms through which the availability of the dataset will be announced	CAPSELLA website, social media, project events (e.g. hackathons), and other material
License of the dataset	License are declared by each data provider and included in the metadata.
Dataset access policy	Open access. Everyone can access and view the data
Dataset consuming procedure for a user	website download, URL
In case of a generated dataset, will there be an embargo period for providing (open) access to the dataset? If yes, how long (in months)?	No
In case of creating or collecting data in the field how will you ensure its safe transfer into the main data management system?	We do not consider "safe transfers" as those are Open Access data
C. Preservation	
Preservation strategy	The master version of the dataset is maintained by its origin
Foreseen preservation period duration	N/A

Tools to implement the preservation strategy	N/A

4. Project Related Data

4.1 Deliverables

This dataset is the collection of deliverables generated within the context of CAPSELLA. The estimated scale is less than 100 objects, both PDF and Word files.

Deliverables are published via the project website for the audience to be informed on project activities and results, and are stored in the Redmine management platform, which has been created for supporting the project operation. The workspace of CAPSELLA Redmine contains tickets where deliverables are stored as soon as they are produced (actually submitted to EC). Access to this platform is granted to project members only.

Public deliverables will be also published to other repositories such as Zenodo or OpenAIRE and will be accompanied by a set of metadata required by each repository.

CAPSELLA deliverables of type “Other” are available through the public space in Redmine and accessible by everyone with the link.

4.2 Questionnaires

This dataset is the collection of questionnaires generated and collected within the context of CAPSELLA surveys. The estimated scale is less than 500 objects. Regarding the data and metadata standards, a Dublin Core record will be associated with each item.

Questionnaires are published via the project website for stakeholders to complete them. The filled questionnaires are collected and stored in the Redmine management platform. The workspace of CAPSELLA Redmine contains a folder where questionnaires are stored as soon as they are collected. Access to this platform is granted to project members only.

4.3 Scientific Papers

At this stage of the project we can identify the following high level datasets that CAPSELLA either collects or generates for research purposes:

1. Scientific publications
2. Automatically and manually generated annotations
3. Consortium publications
4. Metadata for all the above

Scientific papers are mainly available / accessible via the Publishers’ web site according to the associated access method. CAPSELLA promotes “Open Access” thus a machine-readable electronic copy of every publication is expected to be deposited in suitable Open Access repositories. In addition, they will be disseminated through the CAPSELLA Project Website as well as through scholarly communication channels, e.g., Publishers/Journals web sites, Institutional Repositories, scholarly communication networks (ResearchGate, Google Scholar).

Beneficiaries will deposit the published version or the final peer-reviewed manuscript accepted for publication in at least an "OpenAIRE compliant" repository. Authors may rely on their Institutional Repositories (if any) as well as on Zenodo. Moreover, a copy of each paper must be uploaded in the project workspace.

4.4 Dissemination Material

This type of dataset is a collection of dissemination material produced for the outreach of the project, including videos, slides, posters, fliers, news, etc. It will be generated during the project lifetime and will be made available through the CAPSELLA project website. Dataset material is actually stored in several repositories including the workspace of the CAPSELLA Redmine created to support the project, the institutional repositories of project beneficiaries and/or third party services.

5. Software

The software produced in the CAPSELLA project will be publicly available for accessing and downloading, both its artefacts and the source code and will be managed with the Apache Maven build tool[8].

All software components are licensed under the Creative Commons and Attribution (CC BY) license, unless is specified differently.

Open repositories (such as Github or BitBucket) or private with public read access will be used to store the software components.

[8] Apache Maven Build Tool: https://maven.apache.org/

6. Conclusions

This deliverable is the updated version of the CAPSELLA Data Management Plan and contains datasets collected in the first year of the project. The described datasets may be of value for the project and will be exploited by the different scenarios and pilots through the CAPSELLA platform. In the coming months the list of will be enriched, with new datasets either identified or created by the project. Dataset use, sharing, preservation and dissemination issues will be also further specified. All this updated information will be included in the future versions and revisions of the current document.