



Horizon 2020 Societal challenge 5  
Climate action, environment, resource  
Efficiency and raw materials

# D4.2: DATA MANAGEMENT PLAN - DMP – 2<sup>ND</sup> RELEASE-UPDATE *(A Living Document)*

LEAD AUTHOR: Marc Bonazountas

OTHER AUTHORS:

Elena Krikigianni (EPSILON)

Chrysovalantis Tsiakos (EPSILON)

Lluís Echeverría (EURECAT)

Gabriel Anzaldi Varas (EURECAT)

Xavier Domingo Albin (EURECAT)

Despina Kallidromitou (EPSILON)

Angelos Anagnostopoulos (EPSILON)

<b>PROJECT</b>	Sustainable Integrated Management FOR the NEXUS of water-land-food-energy-climate for a resource-efficient Europe (SIM4NEXUS)
<b>PROJECT NUMBER</b>	EC/689150 SIM4NEXUS/2016
<b>TYPE OF FUNDING</b>	RIA (Research Innovation Action)
<b>DELIVERABLE</b>	D.4.2 Data Management Plan (DMP): 2 <sup>nd</sup> Release-update
<b>WP NAME/WP NUMBER</b>	Serious Game development and testing (WP 4)
<b>TASK</b>	Task 4.8 Data Management
<b>VERSION</b>	Deliverable 4.2 - 2nd Release v6.1
<b>DISSEMINATION LEVEL</b>	Public
<b>DATE</b>	26/11/2017 (Date of this version)
<b>LEAD BENEFICIARY</b>	EPSILON Malta Ltd
<b>RESPONSIBLE AUTHOR</b>	Marc Bonazountas (EPSILON) & Elena Krikigianni (EPSILON) & Chrysovalantis Tsiakos (EPSILON)
<b>AUTHORS</b>	supported by: Angelos Anagnostopoulos (EPSILON), Despina Kallidromitou (EPSILON), Xavier Domingo (EURECAT), Floor Brouwer (WUR), Lluís Echeverría (EURECAT), Gabriel Anzaldi Varas (EURECAT), Valantis Tsiakos (EPSILON)
<b>INTERNAL REVIEWER</b>	Floor Brouwer (WUR), Elena Krikigianni (EPSILON)

#### DOCUMENT HISTORY

VERSION	INITIALS/NAME	DATE	COMMENTS-DESCRIPTION OF ACTIONS
1.0	EPSILON	27/10/2016	1 <sup>ST</sup> VERSION BY OLYMPIA DASKALOU
1.1	EPSILON	27/10/2016	CONTRIBUTION BY MARC BONAZOUNTAS
2.0	EPSILON	31/10/2016	EDIT – INTEGRATE TO S4N TEMPALTE - 2 <sup>ND</sup> RELEASE SEND TO WP4 LEADER FOR REVIEW
2.1	EURECAT	8/11/2016	FEEDBACK RECEIVED FROM EURECAT & WUR
3.0	EPSILON	10/11/2016	UPDATED VERSION – 3 <sup>RD</sup> RELEASE SEND TO THE COORDINATOR
3.1	EURECAT	23/11/2016	FEEDBACK RECEIVED FROM EURECAT
3.2	WUR	26/11/2016	FEEDBACK RECEIVED FROM WUR (COORDINATOR)
4.0	EPSILON	28/11/2016	FINAL VERSION OF DELIVERABLE
5.0	EPSILON	22/05/2017	2 <sup>ND</sup> RELEASE- 2 <sup>ND</sup> VERSION BY EPSILON
5.1	EURECAT	24/05/2017	FEEDBACK RECEIVED FROM EURECAT
5.2	EPSILON	25/05/2017	UPDATED VERSION- 3 <sup>RD</sup> RELEASE SEND TO THE COORDINATOR
6.0	EPSILON	10/11/2017	2 <sup>ND</sup> RELEASE- UPDATED VERSION- BASED ON EC COMMENTS
6.1	EURECAT	26/11/2017	GENERAL COMMENTS AND IMPROVEMENTS DATA ACCESS, PERMISSIONS, ETC. TO FILES SPECIFIED OPEN ACCESS DETAILED AS THE PUBLICATION MODE FOR SCIENTIFIC COMMUNITY RELEVANT OUTCOMES LINK WITH D4.3 CLARIFIED

# Table of Contents

Figures .....	4
Tables .....	5
Glossary / Acronyms .....	6
Executive Summary .....	7
1. Introduction.....	9
1.1 Scope .....	9
1.2 Structure of the document .....	9
2 Background on SIM4NEXUS DMP .....	10
2.1 General .....	10
2.2 Why is a Data Management Plan needed? .....	11
2.3 Who is responsible for the implementation of the DMP? .....	12
2.4 What kind of data will be affected by DMP?.....	12
3 Digital data sets generated or collected in SIM4NEXUS.....	13
3.1 Data set reference and name.....	14
3.2 Data set description .....	15
3.3 Standards and Metadata.....	15
3.4 Archiving and Preservation .....	16
3.4.1 Web catalogue service description.....	16
3.4.2 Zenodo Repository .....	19
4 Data Summary – Specifics.....	20
4.1 System Dynamics models data sets.....	20
4.2 Semantic repository .....	22
4.3 3D map and terrain data in serious game .....	22
4.4 Other considerations.....	23
5 FAIR Data – Specifics.....	24
5.1 Making data findable, including provisions for metadata .....	24
5.2 Making data openly accessible.....	25
5.3 Data interoperability .....	25
5.4 Increased data re-use.....	25
6 Ethical & Security aspects .....	27
6.1 General .....	27
6.2 Intellectual Property Rights (IPR) .....	27
7 Questions & Answers on FAIR data.....	28
References .....	30

# Figures

- Figure 1: Open Access and dissemination of scientific data (EC, 2017) ..... 10
- Figure 2: DMP Time plan ..... 14
- Figure 3: Example of metadata information in each subfolder ..... 21
- Figure 4 Structure of the SIM4NEXUS repository ..... 21

# Tables

Table 1. WPXX datasets .....	13
Table 2. Assign persons .....	14
Table 3. SIM4NEXUS data characteristic description.....	28

# Glossary / Acronyms

As the document is being written, terms and glossary will be added here as needed. Before the last version is submitted this list will be re-arranged alphabetically by the lead author.

TERM	EXPLANATION / MEANING
CA	Consortium Agreement
CSW	Catalogue Service for the Web
DAP	Data Access Protocol
DESCA	Consortium Agreement Model
DMP	Data Management Plan
DoA	Document of Action
ERC	European Research Council
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reused
GA	Grant Agreement
HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure as a Service
IPR	Intellectual Property Rights
ISO	International Standards Organization
JSON	JavaScript Object Notation
KEE	Knowledge Elicitation Engine
OA	Open Access
OGC	Open Geospatial Consortium
ORD	Open Research Data Pilot
RESTful	Representational State Transfer
SDI	Spatial Data Infrastructure
WCS	Web Coverage Service
WMS	Web Map Service
WFS	Web Feature Service
WP	Work Package
XML	eXtensible Markup Language

# Executive Summary

This document presents Data Management Plan (DMP) on open access data handling (see box 1) by SIM4NEXUS. The aim is to consider the many aspects of data management, data and metadata generation, data preservation- maintenance- and analysis, whilst ensuring that data is well managed at present and prepared for preservation in the future. This Data Management Plan is compiled according to the [Guidelines on FAIR Data Management in H2020](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)<sup>1</sup>, and the Guidelines to the Rules on the [Open Access to Scientific Publications and Open Data Access to Research Data in H2020](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)<sup>2</sup>.

Thus, the sections below present the lifecycle, responsibilities and review processes and management policies of research data, produced by SIM4NEXUS. The DMP reflects the status of discussion within the Consortium on data to be produced. It is a *Living Document* with iterations (M12, M30 and M48) as SIM4NEXUS evolves. An updated version of this document will be delivered together with each reporting period, and whenever significant changes related to DMP would occur.

This document is the **update of the second release** of the DMP. In the future, it will be delivered together with the progress reports as it provides an overview of available research data, access and the data management terms of use at this stage.

For SIM4NEXUS, the DMP is defined as “the development, execution and supervision of plans, policies, programmes and practices that control, protect, deliver and enhance the value of data and information assets” obtained. In this regard, at the start of the project, the following processes and procedures for data management procedures are established:

- Data governance, such as standards management and guidelines
- Data architecture, analysis, and design including data modelling
- Data maintenance, administration, and data mapping across building blocks and solution modules
- Data security management including data access, archiving, privacy, and security
- Data quality management including query management, data integrity, data quality, and quality assurance
- Reference and master data management including data integration, external data transfer, master data management, reference data
- Document, record, and content management
- Metadata management, i.e., metadata definition, discovery, publishing, metrics, and standardization.

Readers are the project members and research institutions using the data collected and produced during the project period.

## Box 1: Open Access

**Open access (OA)** refers to the practice of providing online access to **scientific information** that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research & innovation, 'scientific information' can mean: (1) peer-reviewed scientific research articles (published in scholarly journals) or (2) research data (data underlying publications, curated and raw data)

---

<sup>1</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

<sup>2</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

## Changes with respect to the DoA

No changes with respect to the DoA

## Dissemination and uptake

The deliverable is publicly available, based on the participation of SIM4NEXUS to the Pilot on Open Research Data in Horizon 2020<sup>3</sup>. Special attention will be paid to how personal data will be properly catered together with other important data and/or scientific information.

## Short Summary of results (<250 words)

As SIM4NEXUS participates in the Pilot on Open Research Data in Horizon 2020, it is required to submit in the first 6 months of the project the Data Management Plan. This document aims to improve and maximise access and re-use of research data generated by project actions. Participating in the Open Research Data Pilot does not necessarily mean opening up all research data. In a sense, the document determines and explains which of the research data generated and/or collected will be made open. Several iterations of this document will be released as project evolves.

## Evidence of accomplishment

The deliverable itself can act as the evidence of accomplishment. Also, communication (Teleconferences, emails) between EURECAT, EPSILON and the project Coordinator (WUR) can be revealed as evidence.

---

<sup>3</sup> According to article 43.2 of Regulation (EU) No 1290/2013 of the European Parliament and of the Council, of 11 December 2013, laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" and repealing Regulation (EC) No 1906/2006.



# 1. Introduction

## 1.1 Scope

This document describes the SIM4NEXUS Data Management Plan (DMP, see box) that corresponds to Deliverable D4.2 of the SIM4NEXUS Technical Annex. The DMP:

- Provides a description of how the research data collected, processed, and generated will be handled during and after the SIM4NEXUS.
- Describes which standards and methodology for data collection and generation will be followed, how data will be shared and be curated and preserved.

The document follows the template provided by the European Commission on DMP<sup>4</sup>. The DMP is prerequisite for SIM4NEXUS as it participates in the Open Research Data Pilot<sup>5</sup>, thus first version was delivered at an early stage of the project (Month 6).

An updated version of this document has been, and will be provided together with the first two progress reports (M12, M30) and whenever significant changes occur. At month 48 the final version will be provided as Appendix to deliverable D4.7 Data Management Report.

## 1.2 Structure of the document

DMP deliverable is organized as follows:

- Section 1 is the introductory chapter, which provides the scope of the deliverable
- Section 2 presents the key questions that DMP addresses as tailored for SIM4NEXUS
- Section 3 contains information on digital data sets generated or collected in SIM4NEXUS for each Work Package. It will be updated per reporting period until the end of the project
- Section 4 contains a data summary for all Work Package data sets. It will be updated per reporting period until the end of the project
- Section 5 contains information of FAIR data for SIM4NEXUS and will be updated as the project evolves
- Sections 6 address issues related to data security & ethical aspects
- Section 7 answers FAIR data key questions related to all datasets produced or gathered in.

### Box 2: Data Management Plan

**A Data Management Plan (DMP)** is a *key element* of good data management; it describes the data management life cycle for the data to be collected, processed, and generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable, and re-usable (FAIR), a DMP should include information on: (i) the handling of research data during and after the end of the project, (ii) what data will be collected, processed, and generated, (iii) which methodology and standards will be applied, (iv) whether data will be shared/made open access, and (v) how data will be curated and preserved (including after the end of the project). A DMP is required for all projects participating in the extended ORD pilot unless they opt out of the ORD pilot; however, projects that opt are encouraged to submit a DMP on voluntary basis.

---

<sup>4</sup> 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](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

<sup>5</sup> Open Access to Scientific Publications and Research Data in Horizon 2020 Guidelines, [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

# 2 Background on SIM4NEXUS DMP

## 2.1 General

It is a well-known phenomenon that data is increasing exponentially, while the use and re-use of data that derives new scientific findings is stable. This does not imply, that data currently being unused is useless, as it can be of high value in future. Prerequisite for meaningful use, re-use or recombination of data is to be well documented according to accepted and trusted standards. These standards form a key pillar of science because they enable the recognition of suitable data. To ensure this, agreements on standards, quality level and sharing practices are to be discussed. Strategies should be fixed to preserve and store data over a defined period to ensure its availability and re-usability after the end of SIM4NEXUS. Based on the EU guidelines and because SIM4NEXUS utilizes data for various pilots, the H2020 Programme requires for such projects to participate in the Open Research Data Pilot<sup>6</sup> (ORD Pilot).

### Box 3: Research Data

Research data Refers to information, facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge (Figure 1: Open Access and dissemination of scientific data (EC, 2017)).

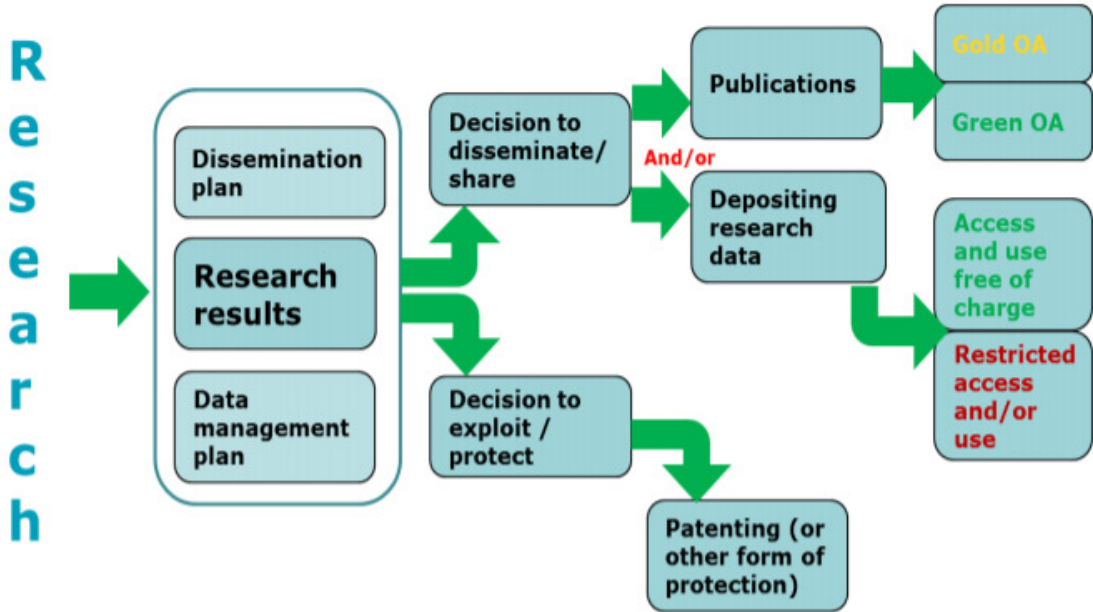


Figure 1: Open Access and dissemination of scientific data (EC, 2017)

<sup>6</sup> Open Research Data Pilot – ORD pilot: <https://www.openaire.eu/opendatapilot>

## 2.2 Why is a Data Management Plan needed?

SIM4NEXUS participates in the Open Access and the Open Research Data Pilot of the European Research Council (ERC). The DMP specifies the implementation of the pilot for: data generated and collected, standards in use, workflow to make data accessible for use, reuse and verification by the community, and definition of a strategy of curation and preservation of the data. Therefore, we refer to the SIM4NEXUS Grant Agreement (GA), Article 29.3 on “Open Access to research data”:

*'Regarding the digital research data generated in the action ('data'), the beneficiaries must:*  
*(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce, and disseminate - free of charge for any user - the following:*  
*(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible; (ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan;*  
*(b) provide information - via the repository - about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and - where possible - provide the tools and instruments themselves).*

The data management policy described in this document reflects the current state of consortium agreements<sup>7</sup> on data management. Collecting data sets on environmental systems at scale can be an expensive and time-consuming process. By making a large part of them available, this project will make a long-lasting and significant contribution to the research and industrial communities.

All datasets will be released in open formats (e.g., JSON, XML) respecting available standards, with proper documentation supporting their use by other researches. Since we will engage end-users in the development of the SIM4NEXUS platform, it is imperative to carefully address data management issues connected to the use of external data; whilst we will also deal with highly sensitive data, which may be confidential. In the development of the SIM4NEXUS platform we will explicitly deal with security issues from the technical perspective and in the DMP, delivered in WP4 with security and privacy issues arranged from the management perspective.

The following categories of data are considered within the project:

- personal status (age, gender, etc.),
- socio-economic (city of residence, social status, marital status, and income category),
- social network, and
- domain related.

Data will be stored in a project database developed by EPSILON and managed by the project coordinator. All data and especially personal data will be securely stored in a project server. The final server specifications and the system architecture may change during the execution of the project. However, a first definition made by EURECAT and DHI (Serious Game needs) can be checked in “D4.3 Game land systems requirements”. EPSILON is the responsible partner to provide the cloud server, which will be scalable when required in terms of processing power and storage space. SIM4NEXUS will use cloud server solution for maximum performance (e.g. quick access for the user to the content, quick processing and communication with SIM4NEXUS services/modules).

---

<sup>7</sup> SIM4NEXUS Grant agreement and SIM4NEXUS consortium agreement

Project participants will have secured web access to the previously anonymized data, which will have been automatically checked for consistency, homogeneity and completeness, whilst manual audits will be performed as per the standard operation procedures previously defined.

After project completion, and in case of no objection by project partners and user, anonymization is preserved (i.e., a user cannot be identified from their data) and data may be published in an Open Data portal (e.g., <http://open-data.europa.eu>) for future research always consistent with exploitation and Intellectual Property Rights (IPR) requirements.

## 2.3 Who is responsible for the implementation of the DMP?

The responsible partner is EPSILON, co-Leader in WP4, led by EURECAT, though all WP Leaders and co-Leaders shall be involved in the compliance of the DMP. The partners agreed to deliver datasets and metadata produced or collected in SIM4NEXUS according to the signed GA (Article 29.3) which is in line with rules described in the DMP. The Project and the Scientific Officer are also central players in the implementation of the DMP and will track the compliance of the rules agreed.

## 2.4 What kind of data will be affected by DMP?

The main purpose of a DMP is to describe Research Data with the metadata attached to make them discoverable, accessible, assessable, usable beyond the original purpose and exchangeable between researchers.

Thus, SIM4NEXUS focuses more on the production process and tools than on production of research or observation data and so the amount of Research Data which SIM4NEXUS intend to produce is limited, at least at this stage of the project.

As already presented the Open Research Data Pilot applies to two types of data:

1. “the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.”
1. “other data, including associated metadata, as specified and within the deadlines laid down in the data management plan—that is, per the individual judgement by each project.”

According to the “Guidelines on Data Management in Horizon 2020” (2015) the DMP describes the handling of *numerical datasets* processed or collected during SIM4NEXUS lifetime.

*“Research data refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.”*

The DMP includes clear descriptions and rationale for the access regimes that are foreseen for collected data sets. Therefore, the DMP leaves explicitly open the handling, use and curation of products like tools, software and written documents. Thus, we restrict the focus of our DMP to numerical data products like produced model data or observation data.

### 3 Digital data sets generated or collected in SIM4NEXUS

The intention of the DMP is to present data management plans of the Work Packages. The information listed below reflects the conception and design of the individual Work Packages at the beginning of the project, updated accordingly with current status. As the operational phase of the project started in June 2016, the data collection and generation is at a very first state and is rather restricted to the ongoing fast track for the Sardinia area. The objective of the fast track aims to:

1. Identify difficulties related to datasets collection from the different stakeholders and projects to run models and have a comprehensive view of all nexus components in the area
2. Identify the obstacle and find solution to harmonize data at level of scale and spatial distribution

Thus, the data register will deliver information according to Annex 1 of the H2020 guidelines (2015):

1. **Data set reference and name:** Identifier for the data set to be produced.
2. **Data set description:** Descriptions of data that will be generated or collected, 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.
3. **Standards and metadata:** Reference to existing suitable standards, in case they do not exist, an outline on how and what metadata will be created.
4. **Data 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. Identification of the repository where data will be stored, if already existing and identified, indicating the type of repository (institutional, standard repository for the discipline, etc.). 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).
5. **Archiving and preservation** (including storage and backup): Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

FAIR data of the ORD Pilot, stands as an acronym for Findable, Accessible, Interoperable, and Reused data. The following table presents the template to be used to report datasets related to each WP.

Table 1. WPXX datasets

WPXX – Dataset YY	Description
Data set reference and name	
Data set description	
Standards and metadata	
Data sharing	
Archiving and Preservation (including storage and back-up)	
Reported by	(WP leader) / (WP co-leader)

EPSILON is responsible to communicate with the WP Leaders and co-Leaders and collect the required information; assigned people for reporting and updating the above template are shown in Table 2. Assign persons.

Table 2. Assign persons

WP	Assigned person
WP1	Chrysi Laspidou (WP leader) / Mark Howells (WP co-leader)
WP2	Maria Witmer (WP leader) / Janez Sušnik (WP co-leader)
WP3	Dragan Savic (WP leader) / Maria Blanco (WP co-leader)
WP4	Xavier Domingo (WP leader) / Marc Bonazountas (WP co-leader)
WP5	Floor Brouwer (WP leader) / Maité Fournier (WP co-leader)
WP6	Alexandre Bredimas (WP leader) / Chengzi Chew (WP co-leader)
WP7	Guido Schmidt (WP leader) / Frank Wechsung (WP co-leader)
WP8	George Beers (WP leader)

The time plan (Figure 2: DMP Time plan) for the next 6 months asks to consider input from both the Sardinia pilot case and from each WP. As soon as simulations are completed more input will become available in terms of data and data ontologies that will be stored in the database and semantic repository. Additionally, bilateral teleconferences will take place between EPSILON and WP Leaders are regularly discussing in available and created data and information and terms of use (i.e. whether data will be publicly available or not).

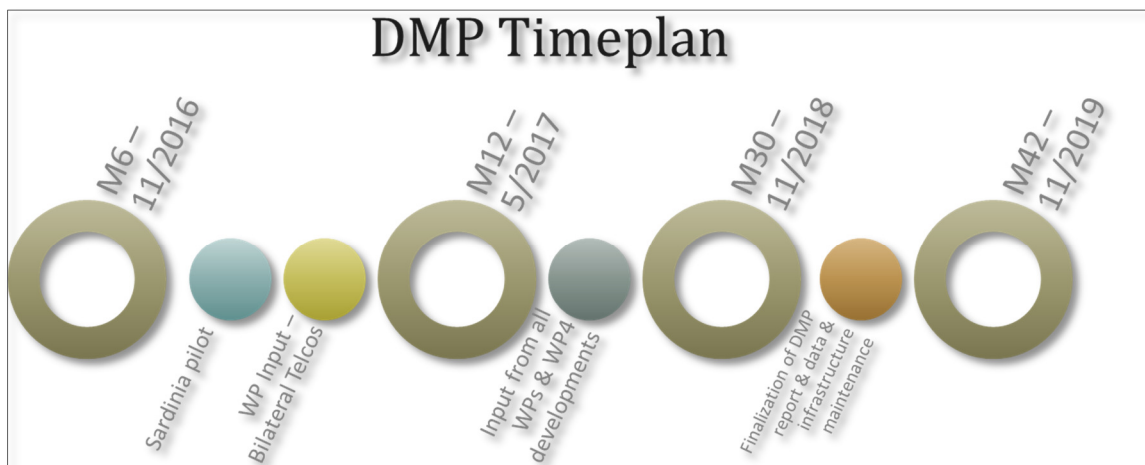


Figure 2: DMP Time plan

### 3.1 Data set reference and name

Data set reference and naming will be implemented to employ a standard identification mechanism for each data set according the metadata standard implemented. Zenodo<sup>8</sup> (a popular repository for research data, will be extensively exploited throughout the project) assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. Zenodo supports harvesting of all content via the OAI-PMH protocol<sup>9</sup>.

<sup>8</sup> <http://help.zenodo.org/features/>

<sup>9</sup> <https://www.openarchives.org/OAI/openarchivesprotocol.html>

## 3.2 Data set description

Data sets related to the project's DMP: (i) data sets referred to project publications (deliverables and papers) (ii) curated and/or raw data collected produced during the project.

SIM4NEXUS data sets collection and production is mainly linked to WP3 applying the thematic models (i.e. E3ME-FTT, MAGNET, CAPRI, IMAGE-GLOBIO, OSeMOSYS, SWIM MAGPIE) selected for the project within the individual Case Studies (as specified in the Task 3.3 of WP3) to realise a partial simulation of the Nexus components under different scenarios, feeding into the development of the Serious Game. Based on the testing, the partners will then collect and organize the data into a semantic database that houses the complexity science tools (WP1). It will also review and select the most suitable integration methodologies for the Case Studies and the Serious Game (for WP4). Thereby, integrated complexity science models will be developed for all the Case Studies. These complexity science models will then be used to run many scenarios.

Thus, SIM4NEXUS produces raw data with some parts summarized in deliverables and scientific publications. This raw data, underpinning the published work, constitute the main research data sets that will be made publicly available if the WP leaders give their permission. In cases where release of complete raw data sets is impossible due to, for example, privacy or personal data concerns (such as packet traces involving networking usage of trial participants), we will strive to find *data sanitation* and *anonymization* approaches that enable publishing as large parts of the data as possible. Any scripts used for post-processing the raw data will also be shared.

### Data layer

The data layer is responsible for storing the SIM4NEXUS data according to different strategies:

- As files: For example, raster data can be stored in raster files formats such as JPEG, TIFF or GeoTiff. Vector data can be stored in various formats as ESRI Shapefile, MapInfo, DXF, etc.
- Within a database: This is managed by a (spatial) database management system such as Oracle Spatial, ESRI ARCSDE or Postgres/PostGIS. Spatial Databases are used for higher data volumes or when data must be accessed and updated by many users, when security policy is required, when complex spatial and non-spatial operation must be performed on the database and when the data must be integrated with non-spatial data.

## 3.3 Standards and Metadata

As mentioned data will be shared in relation to (i) publications (deliverables and papers) and (ii) curated and/or raw data. For the data linked to scientific publication, the publication will serve as the main piece of metadata documentation for the shared data. When this is not seen as being adequate for the comprehension of the raw data, a report will be shared along with the data explaining their meaning and methods of acquisition. However, for both data categories the metadata standard structure of Zenodo repository.

*For Public Availability of Data*, data will be shared when the related deliverable or paper or data set has been made available at an open access (OA) repository from the responsible partner/owner of the data. It is expected that data related to a publication will be openly shared. However, to allow the exploitation of any opportunities arising from the raw data and tools, data sharing will proceed only if all co-authors of the related publication agree. The Lead author is responsible for getting approvals and then sharing the data and metadata on Zenodo. The Lead Author will also create an entry on



OpenAIRE<sup>10</sup> to link the publication to the data. OpenAIRE is a service built to offer this functionality and may be used to reference both the publication and the data. A link to the OpenAIRE entry will then be submitted to the SIM4NEXUS Website Administrator (FT) by the Lead Author.

In view of the precautions for protection of personal data, it is explicitly confirmed that the data collected will be publicly available, after care is taken with regards to rules of confidentiality, anonymity, and protection. Anonymized final data sets will be open access and procedures are set as to how data will be preserved and archived in the repository. We are aware of post-publication risks to local researchers and end-users in our research sites and will mitigate all reasonable risk before publication according to the ethical and IPR requirements set.

However, “Opting Out” remains a choice for data owners, as it is possible that even though comprehensive measures are taken to ensure the safety of participants, researchers and their environment, it is only after a SIM4NEXUS report or peer reviewed article is published and generation of data sets is realized, that the question of open access arises. Open access does not entail an absolute obligation to publish all data, and it is up to researchers and associated organization to decide whether data is suitable and ethical to be published or not.

## 3.4 Archiving and Preservation

To ensure archiving and preservation of long-tail research data during the project, a repository with a web catalogue service will be built and maintained after the project competition. The Web Catalogue Services provides the system with a smarter interface to the SIM4NEXUS repository (geo-database). There are many technologies that can be exploited and adopted to perform this function.

### 3.4.1 Web catalogue service description

The CSW interface is based on the OGC™ Catalog Services Specification Version 2.0.2. The interaction between a client and a server is accomplished using a standard request-response model of the HTTP protocol. That is, a client sends a request to the server using HTTP, and expects to receive a response to the request or an exception message.

Repository service access is based upon the HTTP protocol with client and server requests and responses using XML or JSON. Client applications can use this interface for executing service repository queries and receiving service repository metadata results. Basically, the essential purpose of a Catalog Service is to enable a user to locate, access, and make use of resources in an open, distributed system by providing facilities for retrieving, storing, and managing many kinds of resource descriptions.

The metadata repository managed by the catalog can store a multitude of resource descriptions that conform to any standard Internet media type, as:

- XML schemas
- Audio annotations
- Specification documents
- Style sheets for generating detailed topographic maps.

Furthermore, arbitrary relationships among catalog items can be expressed by creating links between any two resource descriptions. For example, a service offer may be associated with descriptions of the

---

<sup>10</sup> <https://www.openaire.eu/>



data sets that can be acquired using the service. A catalogue can function as a stand-alone service or it can interact with other affiliated catalogues within a federation that spans multiple administrative domains. The federation then effectively enlarges the total search space within which resource descriptions may be discovered.

When a catalog is linked to a peer catalog, it makes the resource descriptions managed by the peer implicitly available to its own clients. Each catalog client connects to a single catalog service as its main point of contact with the federation. This is the agent node; the propagation of request messages to neighboring nodes is invisible to the client. It is not necessary to know where the metadata repositories are located or how they are accessed.

The CSW catalogue profile is intended to provide a flexible, general-purpose catalogue service that can be adapted to meet the needs of diverse communities of practice within the geospatial domain. In the SIM4NEXUS framework we try to respect the following rules:

1. Communicate information adopting standard protocols (e.g. XML/JSON, OGC standards, etc.).
2. Try to adopt a solution that allows for the maximum interoperability among actors who will process the data stored.
3. State of the art technologies that will be used in the context of the Web Catalogue Service include:
  - pyCSW (<http://pycsw.org/>)
  - GeoNetwork (<http://geonetwork-opensource.org/>)
  - Micka (<http://micka.bnhelp.cz/>)
  - CKAN (<https://ckan.org/>)

### **pyCSW**

The pyCSW is OGC compliant python implementation that permits clients to make searches on metadata. The CSW interface is based upon the OGC™ Catalog Services Specification Version 2.0.2. The interaction between a client and a server is accomplished using a standard request-response model of the HTTP protocol.

### **GeoNetwork**

GeoNetwork open source is a standard based and decentralized spatial information management system, designed to enable access to geo-referenced databases and cartographic products from a variety of data providers through descriptive metadata, enhancing the spatial information exchange and sharing between organizations and their audience, using the capacities and the power of the Internet. The system provides a broad community of users with easy and timely access to available spatial data and thematic maps from multidisciplinary sources, that may in the end support informed decision making. The main goal of the software is to increase collaboration within and between organizations for reducing duplication and enhancing information consistency and quality and to improve the accessibility of a wide variety of geographic information along with the associated information, organized and documented in a standard and consistent way. Main Features:

- Instant search on local and distributed geospatial catalogues
- Uploading and downloading of data, documents, PDF's, and any other content
- An interactive Web map viewer that combines Web Map Services from distributed servers around the world
- Online map layout generation and export in PDF format
- Online editing of metadata with a powerful template system
- Scheduled harvesting and synchronization of metadata between distributed catalogues
- Groups and user's management
- Fine grained access control

From technical point of view, GeoNetwork has been developed following the principles of a Free and Open Source Software (FOSS) and is based on International and Open Standards for services and protocols, like the ISO-TC211 and the Open Geospatial Consortium (OGC™) specifications. The architecture is largely compatible with the OGC™ Portal Reference Architecture, i.e. the OGC™ guide for implementing standardized geospatial portals. Indeed, the structure relies on the same three main modules identified by the OGC Portal Reference Architecture that are focused on spatial data, metadata, and interactive map visualization. The system:

- Is fully compliant with the OGC™ specifications for querying and retrieving information from Web Catalog (CS-W)
- It supports the most common standards to specifically describe geographic data (ISO19139 and FGDC) and the international standard for general documents (Dublin Core)
- It uses standards (OGS WMS) also for visualizing maps through the Internet

A common use case is the harvesting of geospatial data in a shared environment. In fact, within the geographic information environment, the increased collaboration between data providers and their efforts to reduce duplication have stimulated the development of tools and systems to significantly improve the information sharing and guarantee an easier and quicker access of data from a variety of sources without undermining the ownership of the information.

The harvesting functionality in GeoNetwork is a mechanism of data collection in perfect accordance with both rights to data access and data ownership protection. Through the harvesting functionality it is possible to collect public information from the different GeoNetwork nodes installed around the world and to copy and store periodically this information locally. In this way, a user from a single-entry point can get information also from distributed catalogues.

## **MICKA**

MICKA is a meta-information catalogue that fully complies with the ISO 19115 standard and is fully compliant with the INSPIRE principles. It can be integrated with map applications. It is multilingual. The web catalogue service uses OGC specifications.

MICKA is a complex system for metadata management used for building Spatial Data Infrastructure (SDI) and geo-portal solutions. It contains tools for editing and management of metadata for spatial information, web services and other sources (documents, web sites, etc.). It includes online metadata search engine, portrayal of spatial information and download of spatial data to local computer.

MICKA is compatible with obligatory standards for European SDI building (INSPIRE). Therefore, it is ready to be connected with other nodes of prepared networked metadata catalogues (its compatibility with pilot European geo-portal is continuously being tested). Functions include:

- Spatial data metadata (ISO 19115)
- Spatial services metadata (ISO 19119)
- Dublin Core metadata (ISO 15836)
- Feature catalogue support (ISO 19110)
- OGC CSW 2.0.2 support (catalogue service)
- User defined metadata profiles
- INSPIRE metadata profile
- Web interface for metadata editing

Multilingual (both user interface and metadata records). Currently 16 languages are supported. It is possible to dynamically extend the system for other languages. Context help is multilingual and import from the following metadata formats is supported:

- ESRI ArcCatalog
- ISO 19139
- OGC services (WMS, WFS, WCS, CSW)
- Feature catalogue XML
- Export – ISO 19139, GeoRSS
- Support of thesauri and gazetteers.
- Display of changes with GeoRSS

Template base interface with possibilities to change according to user requirements is available with possibility in deep cooperation with any map clients for display of on-line map services.

MICKA stores metadata in a relational database and it is edited by dynamically generated forms. Therefore, it is possible to amend other standards or profiles. It is possible to switch between profiles while editing. Individual profiles can be distributed into sections. With the help of control elements, it is possible to duplicate individual items, select from code lists or connects to supporting applications. Checking of mandatory items is enabled while editing.

The MICKA integrated application is divided into 3 independent components:

- Metadata creation
- Metadata importing
- Metadata Management

## CKAN

CKAN is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding, and using data. It is aimed at data publishers (national and regional governments, companies, and organizations) wanting to make their data open and available. CKAN implements several of the administrative services that are described in the data management plan and provides both an attractive end-user client as well as a Web Service API.

CKAN is built with Python on the backend and JavaScript on the frontend, and uses The Pylons web framework and SQLAlchemy as its ORM. Its database engine is PostgreSQL and its search is powered by SOLR. It has a modular architecture that allows extensions to be developed to provide additional features such as harvesting or data upload.

CKAN uses its internal model to store metadata about the different records, and presents it on a web interface that allows users to browse and search this metadata. It also offers a powerful API that allows third-party applications and services to be built.

### 3.4.2 Zenodo Repository

After the project completion of the project, the final dataset will be transferred to the Zenodo repository which ensures sustainable archiving of the final research data sets and publications produced.

Zenodo is built and developed by researchers, in the context of The OpenAIRE project, that in the vanguard of the open access and open data movements in Europe, commissioned by the EC to support their nascent Open Data policy by providing a catch-all repository for EC funded research. One of its major advantages is it works closely with GitHub, enabling users to make the work they share on GitHub<sup>11</sup> citable by archiving one of your GitHub repositories and assigning a DOI with the data archiving tool Zenodo.

---

<sup>11</sup> <https://github.com/>

## 4 Data Summary – Specifics

This section must be understood as a living section that will be further updated in future iteration as the project evolves, and more input in terms of data and data reporting comes from the SIM4NEXUS partners. The purpose of this section is to provide an executive summary of the different SIM4NEXUS data addressing the following issues:

Provide a summary addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful
- Defines how this data is going to be accessible both for internal or public use

### 4.1 System Dynamics models data sets

EPSILON has implemented the SIM4NEXUS repository (currently using Dropbox infrastructure) for the efficient management of the SIM4NEXUS datasets. The service supports cloud storage, file sharing and synchronization and client software. The file organization structure adopted in the context of the project, supports the easy identification and revisions of the datasets provided by each partner. For each of the SIM4NEXUS cases, a specific folder has been created in the file hosting service. The folder contains the following sub-folders: 01-ModelData, 02-ArbitraryData, 03-ThematicData, 04-ClimateData. The 1<sup>st</sup> folder provides the outputs from the selected models applied in each case study. The file type of the outputs is either Microsoft Excel Open XML Spreadsheet (.XLSX File Extension) or Comma Separated Values files (.CSV File Extension). The sub-folder 02-ArbitraryData contains information about the relevant case Study. These include the conceptual model of the case, the concept harmonization process etc. The 3<sup>rd</sup> sub-folder contains the thematic datasets of each case study along with its metadata. Finally, the sub-folder 04-ClimateData contains the climate datasets such as precipitation, relative humidity, long-wave downward solar radiation at the ground, long-wave downward solar radiation at the ground, daily maximum air temperature, daily minimum air temperature, and wind speed at 10m height. The datasets in this folder are of a generic file type with .DAT file extension. Each dataset of this type may contain data in binary or text format. A standardised name has been assigned to each dataset in the following format:

*Country code\_Earth System Model\_Simulation Method\_Period\_Time frequency.dat*

In this way, all the necessary elements of each dataset such as the way that each dataset has been produced (i.e. model and simulation methods), the addressed area, and the duration and time frequency, are provided. Each subfolder in the file hosting service contains files with descriptive information about the available datasets.

At the moment, none of this datasets are intended for public availability, as they are only useful in the context of building System Dynamics Models, and specific for the different scenarios of the case studies (only Sardinia so far).

It has been decided internally that only people involved on each case study will have access to the specific case study folder. This is a default initial status, which can be revoked if the Case Study leader considers it is convenient. The access is by default read-only, in the same way, if there is a need for edition access, this can be requested to EPSILON by the Case Study leader. As this datasets are only useful for the specific case studies' System Dynamic Models generation, it has been decided to not make them available publicly. At this stage of the project, this decision only affects to Sardinia's Case Study.

However, some baselines for some case studies are being developed and may be of interest for the scientific community. In that case, SIM4NEXUS project will make these baselines available under Open Access.

The value depends on the variable coded in the individual file name. These are the standardised variable acronyms used in climatology:

- pr = Precipitation, originally given as m/s or mm/s, converted to mm/d
- rhs = Relative humidity in %, 2 m above ground
- rlds = Long-wave downward solar radiation at the ground in W/m<sup>2</sup>
- rsds = Short-wave downward solar radiation at the ground in W/m<sup>2</sup>
- tas = Average air temperature 2 m above ground, usually given in K, but converted to °C
- tasmax = Daily maximum air temperature
- tasmin = Daily minimum air temperature
- wind = Wind speed at 10 m height, given in m/s

Figure 3: Example of metadata information in each subfolder

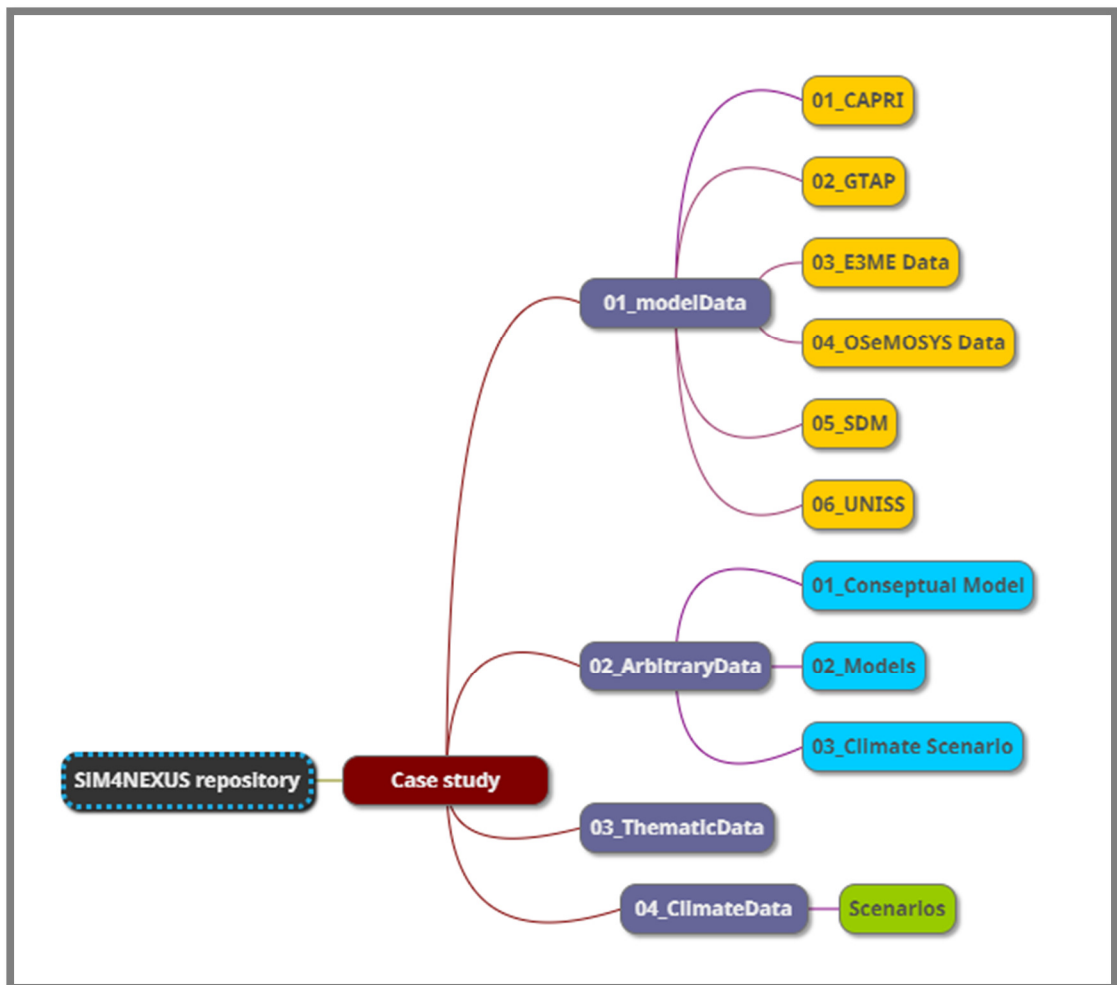


Figure 4 Structure of the SIM4NEXUS repository

## 4.2 Semantic repository

A Semantic Repository is being developed to store information related to the concepts, properties and restrictions from the Nexus procedures; to improve the data integration of diverse sources and, finally, to give a better analytical power. This repository, which is currently focused in provide support to the information flow between the System Dynamics Models and the Serious Game User interface, will also allow for knowledge storage related to the Nexus, policies, etc., coming from WP1 and WP2.

A triple-store is being used to be the base of the repository and an ontology to semantically represent the stored data. The ontology is still under development and represents the SIM4NEXUS knowledge and scope. Currently, the top concept is the 'Session', which represent a game or session in the Serious Game. Linked to the 'Session' there are a 'User', representing the player, and a 'StudyCase' which, in turn, it is related with a specific 'Model', a SDM. In order to represent the Nexus state through the game, the 'Session' has a list of 'StateEvolution', which represent the way from one state ('State') to the new state ('State'), applying certain polices ('Policy'). The 'State' is defined by 'NexusComponent' (for instance 'Climate and environment' or 'Water') and these components have a specifics parameters ('Parameter').

The ontology is defined using the Web Ontology Language (OWL), a Semantic Web language designed to represent rich and complex knowledge about things, and relations between things. Some existing ontologies, related to the Nexus, have been analysed to be involved in the SIM4NEXUS context:

- WatERP ontology, which reflects the water manager's expertise to manage water supply and demand. The novelty of WatERP ontology lies on including man interactions with the natural paths as a mechanism to understand how affect into the water resources management with the objective to match supply with demand, these interactions could range from infrastructures to management decisions.
- WEFNexus ontology, which concern Water, Energy and Food derived by the European Directives: Article 2 of EU Directive 98/83/EC that defines the water intended for human consumption; Article 2 of EU Directive 2003/30/EC that defines bio-fuels; Article 2 of EU Regulation 178/2002/EC that defines food.

At the moment, the ontology is not available publicly, as it is in an initial state. As soon as it gets a stable version, will be made available broadly.

## 4.3 3D map and terrain data in serious game

The 3D map available in current Serious Game user interface prototype is rendered from 2 sources – a height map and a texture overlay:

- The height map data comes from NASA Shuttle Radar Topographic Mission (SRTM) v4.1 and is distributed freely by USGS. The SRTM data is available with a 90m resolution. Not all data from SRTM is being used in the serious game, only data related to the geographical regions of the 12 case studies are used. This data is downloaded from the USGS web and then stored as mesh elements within the serious game client.
- The texture overlay is from Google Earth, it is downloaded from Google and then stored within the serious game client.

Consequently, we are not generating this data, but reusing existing publicly available datasets.

## 4.4 Other considerations

As stated in D.9.1, SIM4NEXUS project, foresee to provide both short and long-term benefits for the involved decision-makers and their associated networks. Given the above, *«the only ethics issues involved in the SIM4NEXUS study concern general ethical issues of informed consent, anonymity and confidentiality associated with the voluntary involvement of human participants in the European Union»*.

SIM4NEXUS will not involve types of data related to sensitive topics (well described in D.9.1), which might generate uncomfortable situations such as psychological stress, any kind of anxiety or humiliation, deception, or any potential increased danger to participants, or gathering of personal data from participants. Thus, SIM4NEXUS, has defined approaches for the following issues:

1. Collection and processing of personal data (described in D.9.1)
2. General ethics commitments (described in D.9.1)
3. Storing and sharing information (described in D.9.1)
4. Accessing and using of information (described in D.9.1)
5. Protection of Information (described in D.9.1)

In addition, deliverable D9.2 provides templates of the informed consent forms and information sheets. This information is confidential, and only for the consortium members (including the Commission Services). Also, SIM4NEXUS consortium, agreed on signing an ethics agreement based on the European Code of Conduct for Research Integrity, published by the European Science Foundation ([http://www.esf.org/fileadmin/Public\\_documents/Publications/Code\\_Conduct\\_ResearchIntegrity.pdf](http://www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf)) and the ethical principles for conducting community-based participatory research, as defined by the National Co-ordinating Centre for Public Engagement of Durham University, UK ([www.publicengagement.ac.uk](http://www.publicengagement.ac.uk)). This will ensure the fair and equal power relationships between researchers and participants.

Last but not least, Deliverable 9.3 stipulates a keen awareness of most ethical issues, as they were presented to the Social Sciences Ethics Committee (SEC). *«The SEC is convinced that fair and respectful treatment in terms of inconvenience, consent and privacy is assured»*.

## 5 FAIR Data – Specifics

This section must be understood as a living section that will be further updated in future iteration as the project evolves, and more input in terms of data and data reporting comes from the SIM4NEXUS partners.

Intellectual property rights (IPR) management in SIM4NEXUS project, is a substantial part of its data management plan. Usually data content and their system are treated as one parameter, but when the matter comes to IPR a distinction between the databases and data content is of utmost importance. It is imperative for other users to know how they can reuse both the data collected, assembled, or generated and the databases where these are included.

The Open Data Commons group (<http://opendatacommons.org>) developed the following tools to govern the use of data sets. The three ODC licenses are:

- Public Domain Dedication and License (PDDL): This makes the use of the database and its content free to the public domain.
- Attribution License (ODC-By): Users can make use of the database and its content in new and different ways, but they need to provide an attribution to the source of the data and/or the database.
- Open Database License (ODC-ODbL): ODbL stipulates that any use of the database must provide attribution, and any new outcomes must use the same terms of licensing (also an unrestricted version of the new product must always be accessible).

In addition, it is acceptable to articulate for SIM4NEXUS project a set of “community norms” that can be used complementary to the use of formal licenses.

At this version of the D4.2 this section is briefly answered in the section 9 Questions & Answers on FAIR data, and thus is structured here to give the outline of future DMP iterations.

### 5.1 Making data findable, including provisions for metadata

This section will be updated on next iterations to provide detailed information on how data will be made discoverable, and more specifically:

- Discoverability of data (metadata provision)
- Identifiability of data and refer to standard identification mechanism
- Use of persistent and unique identifiers such as Digital Object Identifiers
- Naming and conventions used
- Approach towards search keyword
- Approach for clear versioning
- Specify standards for metadata creation
- Type of metadata created and how

As detailed in section 4, all data, information, and knowledge considered relevant for the scientific community will be made accessible under Open Access. At the moment, all datasets, and other means of information storage are in draft status, and this information has not been defined for each dataset. When a dataset is set to be accessible publicly, this information will be fulfilled and the DMP updated accordingly.



## 5.2 Making data openly accessible

This section will be updated on next iterations to provide detailed information on how data will be made accessible, assessable and intelligible more specifically:

- Specifics on which data will be made openly available
- Which data is kept closed and provide the rationale?
- How the data will be made available
- What methods and software tools are used to access the data
- Documentation of software needed to access the data included
- Inclusion of relevant software
- Data and associated metadata, documentation and code deposit
- Provision of access provided in case of restrictions

As detailed in section 4, and 5.1, all data, information, and knowledge considered relevant for the scientific community will be made accessible under Open Access. At the moment, all datasets, and other means of information storage are in draft status, and this information has not been defined for each dataset. When a dataset is set to be accessible publicly, this information will be fulfilled and the DMP updated accordingly.

## 5.3 Data interoperability

This section will be updated on next iterations to provide detailed information on how data will be made interoperable to specific quality standards and more in detail:

- Assess the interoperability of project data
- Specifics on data/metadata vocabularies, standards, methodologies followed
- Use of standard vocabulary for all data types present to allow inter-disciplinary interoperability
- Provision of mapping to more commonly used ontologies

To assure data interoperability, SIM4NEXUS project will follow state of the art ontologies and standards. The two main elements which will store and make available information and services in SIM4NEXUS publicly are the Knowledge Elicitation Engine, and the Semantic Repository. The Knowledge Elicitation Engine is being implemented under Service Oriented Architecture principles, following the OGC standards and services defined for information publication, discovery, exchange, etc. (WPS, WFS, GML...). Please refer to deliverable D4.3, section 3.2 for more information. In regards to the Semantic Repository, please refer to section 4.2 in this document for more information.

## 5.4 Increased data re-use

This section will be updated on next iterations to provide detailed information on how data will be made usable beyond the original purpose for which it was collected, and more in detail:

- Data licensing to permit the widest reuse possible
- Data availability for re-use
- Why and for what period a data embargo is induced
- Data useable by third parties after the end of the project
- Restriction of re-use of some data
- Data quality assurance processes
- Length of time for which the data will remain re-usable

As detailed in section 4, 5.1, and 5.2, all data, information, and knowledge considered relevant for the scientific community will be made accessible under Open Access. At the moment, all datasets, and other

means of information storage are in draft status, and this information has not been defined for each dataset. When a dataset is set to be accessible publicly, this information will be fulfilled and the DMP updated accordingly.

# 6 Ethical & Security aspects

## 6.1 General

Within the SIM4NEXUS study only general ethical issues are concerned such as informed consent, anonymity and confidentiality associated with the voluntary involvement of human participants in the European Union. Types of such data collected in SIM4NEXUS are various user interviews, opinions and reviews associated with project's components. Non-exhaustive list is as follows:

- Stored involvement of Serious Game users to gain insight into the decisions and behaviors of the players and to allow further analysis
- The visualization and interaction tool - to collect information from users so that the Knowledge Elicitation Engine (KEE) can learn from user decisions
- A series of interviews with key stakeholders and decision makers – in particular those which might be affected most by a Nexus-compliant implementation of policies, or which behavioral change is central to the achievement of a resource efficient Europe
- Planned contacts with representative persons of the targeted users. Interviews should be carried out by phone or face-to-face when convenient. Interviews should help define the expected functionalities/ services to be offered, test the price that could be acceptable and identify distribution channels to access these clients
- The end-users, potential developers, and partners, etc. will be provided the opportunity to test and review the latest products and services
- Methodology and procedures for sensitive data processing and storing will be specified as a part of the ethics Deliverable 9.1. It is important to emphasize that special efforts will be devoted to anonymize information and securing accessibility. Mechanisms to delete personal data will be provided in an easy and usable manner

## 6.2 Intellectual Property Rights (IPR)

Intellectual Property Rights (IPR) will receive special attention from the beginning. All rules regarding management of knowledge and IPR will be governed by the Consortium Agreement (CA). SIM4NEXUS was based on DESCA (Consortium Agreement Model) H2020 model for the Consortium Agreement (CA). SIM4NEXUS will not act in contradiction with the rules laid down in Annex II of the Grant Agreement. The CA will address background and foreground knowledge, ownership, protected third party components of the products, and protection, use and dissemination of results and access rights.

The following principles will be applied:

- Confidentiality: During the project duration and beyond (Section 10 of the GA – nondisclosure of the information for a period of 4 years after the end of the project), the contractors shall treat any information, which is designated as property by the disclosing contractors, as confidential. They also shall impose the same obligations to their employees and suppliers.
- Pre-existing know how: Each Contractor is and remains the sole owner of its IPR over its pre-existing know-how. The Contractors will identify and list the pre-existing know-how over which they may grant access rights for the project. The Contractors agree that the access rights to the pre-existing know-how needed for carrying out their own work under the project shall be granted on a royalty-free basis.
- Ownership and protection of knowledge: The ownership of the knowledge developed within the project will be governed by an open source license.
- Open data: Data and results obtained during the project that are based on open public sector data will be made available free of charge.

## 7 Questions & Answers on FAIR data

Furthermore, the goal of this document is to clarify a series of questions related to all datasets produced or gathered in SIM4NEXUS project. In this conclusive section, we describe how the DMP answers to these questions. In the following table, we report the characteristics of the dataset together with the questions to which the DMP should answer.

Table 3. SIM4NEXUS data characteristic description.

Data Characteristic	Question	SIM4NEXUS Answer
Discoverable	Are data and associated software produced and/or used in the SIM4NEXUS discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?	<p>Data produced within the project can be discoverable in SIM4NEXUS Database and will be uniquely identified most probably by RESTful (Representational State Transfer) service pointing to Database resource.</p> <p>Third party datasets will be referenced and identified with data name and version, thus discoverable in their original data repositories.</p> <p>Existing software that is used should be background information of a given partner and as such could be documented but not discoverable.</p> <p>This may also apply to modifications made as part of the SIM4NEXUS case.</p>
Accessible	Are data and associated software produced and/or used in SIM4NEXUS accessible and in what modalities, scope, licenses (e.g. licensing framework for research and education, embargo periods, commercial exploitation, etc.)?	<p>Intermediate data (i.e. non-final data produced during the processing chain elaboration) will be stored in the database but only Consortium Members will access to them.</p> <p>Final products data instead will be stored in the Database and freely accessible also by externals through a dedicated web-service obtaining data from the database.</p>
Assessable and intelligible	Are data and associated software produced and/or used in the project accessible for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for peer review, is data provided in a way that judgments can be made about their reliability and the competence of those who created them)?	Final products data will be freely accessible for and intelligible to third parties.

Data Characteristic	Question	SIM4NEXUS Answer
Useable beyond the original purpose for which it was collected	Are data and associated software produced and/or used in SIM4NEXUS project useable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata, and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?	Final products data will be useable by third parties even long time after their production. Thanks to infrastructure (i.e. IaaS – Infrastructure as a Service that will be applied to SIM4NEXUS) behind the database, data retention will have no physical limit about size and the time of validity (just economic limits related to maintenance). The historical data will be accessible through same interface and methods of the most recent ones also thanks to dedicated storage methodologies.
Interoperable to specific quality standards	Are data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organizations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?	The web-service (providing data) will be based on well-known protocols (i.e. Data Access Protocol - DAP 2.0, RESTful/HTTP), so all data can be accessed in a standardized way, through a compliant HTTP / DAP client. Moreover, data will be stored, when possible, with recognized state of the art standards and protocols, thus assuring interoperability and maximizing exploitation of results.

# References

- Article 43.2 of Regulation (EU) No 1290/2013 of the European Parliament and of the Council, of 11 December 2013, laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" and repealing Regulation (EC) No 1906/2006.
- 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](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- Open Access to Scientific Publications and Research Data in Horizon 2020 Guidelines, [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- Open Research Data Pilot – ORD pilot: <https://www.openaire.eu/opendatapilot>
- SIM4NEXUS Grant agreement & SIM4NEXUS consortium agreement