SIM4NEXUS

Horizon 2020 Societal challenge 5

Climate action, environment, resource

Efficiency and raw materials

D3.2: ARCHITECTURE OF THE DATA BASE, THE DATA FLOW SEMANTICS, THE METADATA AND DATA MANAGEMENT DEFINED

LEAD AUTHORS: Lydia S. Vamvakeridou-Lyroudia & Xavier Domingo Albin

OTHER AUTHORS: Lluis Echeverria (EURECAT), Elena Krikigianni (EPSILON), Barry Evans (UNEXE), Maria Blanco (UPM)

DATE: 30 November 2017

Revised version 14 May 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement NO 689150 SIM4NEXUS

PROJECT	Sustainable Integrated Management FOR the NEXUS of water-land-food-energy-climate for a resource-efficient Europe (SIM4NEXUS)
PROJECT NUMBER	689150
TYPE OF FUNDING	RIA
DELIVERABLE	D3.2: ARCHITECTURE OF THE DATA BASE, THE DATA FLOW SEMANTICS, THE METADATA AND DATA MANAGEMENT DEFINED
WP NAME/WP NUMBER	Thematic Models and Integration / WP3
TASK	Task 3.1
VERSION	2.0
DISSEMINATION LEVEL	Public
DATE	07/11/2017(Date of this version) – 30/11/2017 (Due date)
LEAD BENEFICIARY	UNEXE
RESPONSIBLE AUTHOR	Lydia S. Vamvakeridou-Lyroudia (UNEXE)
AUTHOR(S)	Xavier Domingo Albin (EURECAT), Lluis Echeverria (EURECAT), Elena Krikigianni (EPSILON), Barry Evans (UNEXE), Maria Blanco (UPM)
INTERNAL REVIEWER	Floor Brouwer (WUR)

DOCUMENT HISTORY

VERSION	INITIALS/NAME	DATE	COMMENTS-DESCRIPTION OF ACTIONS
1.0	LVL (UNEXE)	07/11/2017	FIRST OUTLINE AND VERSION
1.1	XDA, LER (EUT)	20/11/2017	CONTRIBUTIONS AND REVIEW
1.2	LVL (UNEXE)	29/11/2017	REVIEW, ADDITIONS
2.0	LVL(UNEXE)	13/05/2019	REVISION AFTER THE PROJECT REVIEW

REVIEWERS' COMMENTS

D3.2	Architecture of the data base, the data flow semantics, the metadata ontology and data management defined	Request for revision	The deliverable is accepted after these minor editorial adjustments are done: - Unit of energy is strange (MV/month). - Table 3 does not seem to be finished. The following recommendations can be integrated in D4.9: - The overall architecture of the data base and data flow is appropriate. The use of one example (Sardinia) helps to better understand the process. - However, the heterogeneous nature of the source data, and the variable levels of availability and completeness for different regions, raises some questions about minimum requirement to ensure the necessary level of quality and completeness to give confidence for real use. It is recommended that before a simulation can be deemed ready for use, for both the serious game and real simulations, the minimum requirements for data and policies, including ontologies, should be clarified. - The machine learning algorithms, and more globally the learning from the players and decisions, need more explanations. - Same with the final list of inputs and outputs.

RESPONSE

1. The units for energy have been corrected in Table 2B

2. The missing info in Table 3 has been added, as well as an explanatory note.

The remaining comments are related to another Deliverable from WP4 that has not yet been submitted. They will be taken into account in time.

Table of Contents

Figures

Tables

Executive summary

This deliverable reports on the data flow between WP3 and WP4 for the development of the System Dynamics Models for each project study in WP3 and the information exchange for the development of the Serious Game in WP4. It is an intermediate report, showing that at the end of M18 the Architecture of the data base, the data flow, the metadata ontology, and data management have been defined.

More concretely, it firstly define the Semantic Repository, which is the storage where all information related to the Nexus, the specific Case Studies, and the Serious Game, among others, is stored in a standardised manner, making use of semantic technologies.

However, not all information is needed to be stored in the Semantic Repository, but is really important for the construction of the System Dynamic Models. These files, the folder structure where they are stored, the way they are shared, and accessed is provided in this report.

Finally, all this information will be put in production in the Serious Game User Interface. For doing that, web services are used to interchange information between the different modules in the SIM4NEXUS architecture. The current version of the architecture, the modules, and the web service is also provided in this report.

Changes with respect to the DoA

Not applicable.

Dissemination and uptake

This report will be released on the project website. The deliverable has been written to support the development of the SIM4NEXUS project and is open to all stakeholders, including the case study leaders and researchers contributing to the case studies.

Short Summary of results (<250 words)

This deliverable reports on the data flow between WP3 and WP4 for the development of the System Dynamics Models for each project study in WP3 and the information exchange for the development of the Serious Game in WP4. It is an intermediate report, showing that at the end of M18 the Architecture of the data base, the data flow, the metadata ontology, and data management have been defined.

Evidence of accomplishment

Submission of report.

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
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
CSV	Comma Separated Values
GDP	Gross Domestic Product
GIS	Geographic Information System
JSON	JavaScript Object Notation
KEE	Knowledge Elicitation Engine
OWL	Web Ontology Language
R	Programming language
SDM	System Dynamics Modelling
STELLA	Graphics and development software for SDM
WPS	Web Processing Service
XML	eXtensible Markup Language

1 Introduction

In SIM4NEXUS (WP3) different types of thematic models will be applied for the case studies (T3.3), each with their own data type format, under different scenarios, as detailed in this Deliverable 3.1, and Deliverable 4.2. Also the outputs from the thematic models, the downscaling (T3.2) and the data coming from freely accessible sources need to be organized in a convenient way for the complexity science models (T3.4 and T3.5) and compatible for WP4, including additional and non-numerical information, as needed. Task 3.1 coordinates all these actions and activities for the smooth execution of all the tasks in WP3 and the compatibility of data flow for smooth cooperation with the other related WP, namely WP4.

This Deliverable is an output related to the work carried out within Task 3.1. It is an intermediate report, at the end of M18, showing that the data flow and structure, for the development of complexity science models and the Serious Game for each Case Study have been defined. It covers work related to both WP3 and WP4, covering activities in both Work Packages and coordinating data share and data flow between them. It also details data types used, and shared folder structure to better understand how the information is accessed and used. Consequently the partners involved in this report are the two WP Leaders and Co-Leaders: UNEXE and UPM (WP3), EURECAT and EPSILON (WP4).

The work detailed in this document is strongly linked to another Deliverable (D4.2), where the Data Management Plan in full is detailed, together with the definition of existing datasets.

Structure of this document:

1. Introduction: This section.
2. Semantic Repository: Current version of the Semantic Repository is depicted, entities, relationships, formats, and storage is defined.
3. Metadata and other data sources: Datasets and other data sources used for building the SDM are depicted. Also how these datasets are stored, shared, and accessed.
4. Architecture and data flow between modules: this section previews the SIM4NEXUS architecture and data flow between modules which gives support to all computation needs in this project, focused in the Sardinia’s Fast track, but intended to be the guide and basis for the others.
5. Data Management: Refers to Deliverable 4.2, where it is deeply explained.

2 Semantic Repository

The Semantic Repository is designed 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.

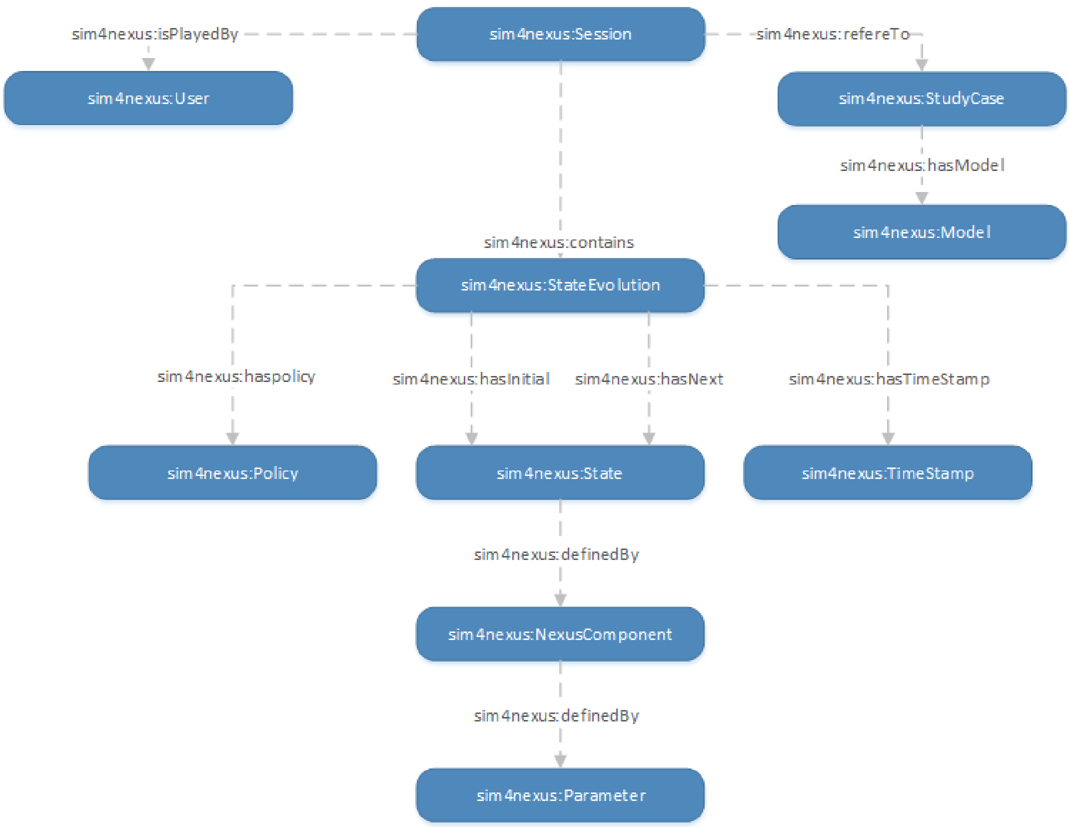

Figure 1: SIM4NEXUS Ontology summarised

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.

3 Metadata and other data sources

Several datasets are used as input for the System Dynamics Models, basically coming from the thematic models. The SIM4NEXUS repository (see D4.2) provides cloud storage and file sharing (among others), and stores the different necessary datasets and other files to build the SDM. These files can be divided in 4 main categories: i) Model Data; ii) Arbitrary Data; iii) Thematic Data; and iv) Climate Data.

The present report focuses on the data used and/or generated for the Sardinia's Case Study, which has been set as the Sardinia's Fast Track, a simple case study taken to the end to provide an example and guidance to the other case studies.

3.1 Case study related data: model, arbitrary, thematic, and climate data

The model data contains all the files necessary to run each thematic or conceptual model involved in the case study. The Table 1: Data used in each Case Study – Sardinia Fast Track current status summarises which types of data have been used for the Sardinia's Fast Track. When necessary, some metadata is added to each file to better describe its contents, units, etc. An example of this can be checked in Figure 2: Example of metadata information.

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^2
rsds = Short-wave downward solar radiation at the ground in W/m^2
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 2: Example of metadata information

This metadata is only added inside the file when the file format allows it. Otherwise, a small text file is added in the folder containing the related files to simplify understanding of contents. At the moment, only Sardinia's datasets and other sources are defined. It is expected that these efforts will be useful for other case studies, but is also known that new data formats, and file types will appear, and may require modify the folder structure, and of course, add more file descriptions.

Table 1: Data used in each Case Study – Sardinia Fast Track current status

Thematic model, concept, or other sources	Files and datasets
CAPRI (model, thematic)	Excel/CSV files for the baselines
E3ME (model, thematic)	Excel/CSV files for electricity, oil, gas, coal, biomass, and heat demand. Also GDP, investment, employment, among others
GTAP (model, thematic)	Excel/CSV files for GDP, etc.
SDM (model)	Stella files for the Sardinia's SDM
Climate data (climate)	DAT files with historical data about 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

Thematic model, concept, or other sources	Files and datasets
Conceptual (model, thematic, arbitrary, climate)	ASCII/DAT files for monthly climate inputs GIS layers for dams location GIS layers for digital elevation model Excel/CSV files for outputs PowerPoint file for conceptual model representation

These files are clearly organised in a shared data folder, which structure is being replicated while the other case studies are being studied and evolved. So, it is not only needed for current research on Sardinia's Case Study, but it is being used as example and guide for the other case studies.

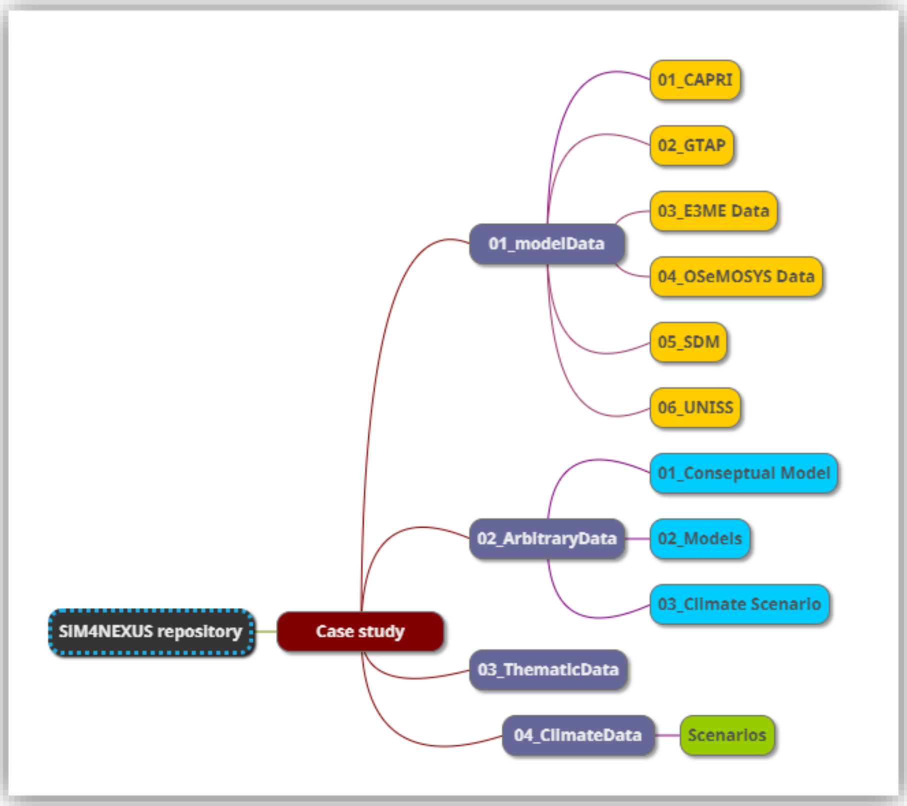

Figure 3: Case Study's folder structure

4 Architecture and data flow between modules

4.1 SIM4NEXUS Architecture

Figure 4: SIM4NEXUS Architecture depicts the different modules which make the Serious Game developed under SIM4NEXUS project a reality. At the top of the solution schema there is the **Serious Game** Graphic User Interface, which allows the users to interact with the system, creating an immersive environment, where they can take different decisions and actions, giving them the corresponding responses and, finally, allowing the fact "to learn by doing".

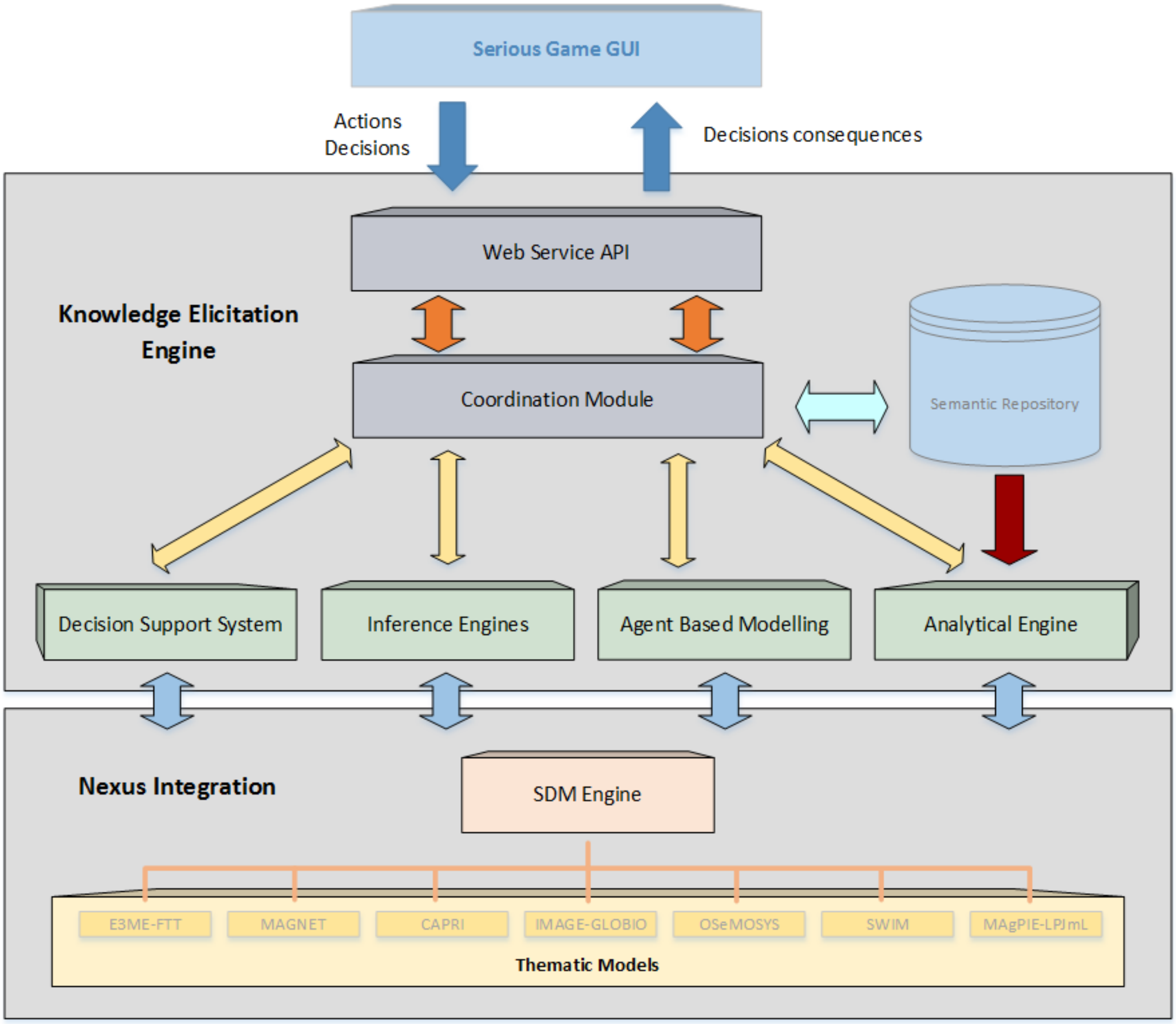

Figure 4: SIM4NEXUS Architecture

In the next level, there is the **Knowledge Elicitation Engine** (KEE) (WP4), the core of the system, which englobes the Web Service API (WS), the Coordination Module (CM), the Semantic Repository (SR), the Decision Support System (DSS), the Inference Engines (IE), the Agent Based Modelling (ABM) and the Analytical Engine (AE).

The Web Service API establishes the communication point between the SG and the KEE, dealing with all the requests and responses. Next, the Coordination Module, which manages all the logics in the system and monitors all the infrastructure status. The Semantic Repository works as a knowledge base, where the generated data is stored following the defined ontology (see Section 2 in this report), allowing the Analytical Engine to be able to learn from these data through machine learning algorithms. The Decision Support System tries to provide recommendations and feedback to users in each step of the SG. Finally, the Agent Based Modelling implements intelligent software agents, based on the acquired knowledge.

At the bottom of the schema there is the **Nexus Integration** layer (WP3), which provide the system with the basic Nexus knowledge to operate with, based on the Thematic Models, through the SDM Engine.

Web Service API

The Web Service (WS) follows the provided rules by the OpenGIS® Web Processing Service (WPS) Interface Standard for standardizing how have to be the inputs and outputs for geospatial processing services. The standard also defines how an execution of a process has to be requested by a client, and how the process output is handled. It defines an interface that facilitates the publishing of geospatial processes and clients discovery of and binding to those process. More details in the next section.

Coordination module

The Coordination Module (CM) works as a central core of the KEE, managing and monitoring all the processes that are involved in the system and ensuring the correct behaviour. It is the responsible for storing all the data in the semantic repository following the SIM4NEXUS ontology.

Analytical Engine

The Analytical Engine (AE) uses different techniques, from the machine learning scope (supervised and unsupervised learning), to learn from the data stored in the semantic repository and the decisions taken by users.

Inference Engine

The Inference Engine (IE) analyse in real time each decision taken by the users to obtain new knowledge from them.

Decision Support System

The Decision Support System (DSS) provide recommendations to users based on the Nexus knowledge, case studies affecting policies, system dynamics results and case study context information, the Analytical Engine and also uses the Inference Engine to obtain knowledge from current action analysis.

Agent Based Modelling

The intelligent software agents are used in the SG to simulate artificial players that interact with the real users. Also are used to run agent-based simulations of possible consequences of policy scenarios that directly or indirectly influence actors' behaviours.

4.2 Data flow between modules

The main interface to obtain information inside the KEE is the Semantic Repository, which interchanges information between modules directly. However, a difference exists in regards to the relation between the Serious Game User Interface, the Knowledge Elicitation Engine, and the SDM in the Nexus integration. In that case, the Web Service API is used.

The deployed web service (WS) follows the provided rules by the OpenGIS® Web Processing Service (WPS) Interface Standard for standardizing how have to be the inputs and outputs for geospatial processing services. The standard also defines how an execution of a process has to be requested by a client, and how the process output is handled. It defines an interface that facilitates the publishing of geospatial processes and clients discovery of and binding to those process.

By integrating artificial intelligence, the game will be able to learn from the players and the decisions they make, expanding its ability to identify nexus effects and improving its predictive capabilities with each interaction. So, these parameters will be stored into the knowledge base with an identifier, assigned to each session or game in the SG, and a timestamp, to be able to establish a link and an order among them and, finally, to ensure the correct learning.

When enough data will be available to provide a good source for the machine learning algorithms, the system will enable these procedures.

The current version of the WS has the input parameters which can be seen in Table 2: WS input parameters. The outputs of this WS can be checked in Table 3: WS Outputs. This is a simplified version of the WS to be able to go ahead with the Sardinia's Fast Track Case Study. The structure of the web service's inputs and outputs will vary during the project execution. However, a generalisation of them is being studied in order to simplify adaptation of the user interface to new versions.

Table 2: WS input parameters

Variable	Unit	Min	Max
Serious Game Parameters			
Game_ID	Unitless		
Timestamp	Unitless	0	
Socio economic			
Population	Million	0	2500000
Water			
Reservoir	Million m^3		
Basin_surface	Km^2		
PPtable			
Coef_defluss	Unitless		
ET_open_body			
Industrial	Million m^3/month		
Tourism	Million m^3/month		
Energy			
Energy_balance			
Solar_production	MW	0	
Wind_production	MW	0	
Other_energy	MW	0	
Agri_energy_demand	Joule	0	
Domestic_energy_demand	Joule	0	
Industry_energy_demand	Joule	0	
Service_energi_demand	Joule	0	
Percap_demand	Joule	0	
Food and Agriculture			
Fruit	Tons/km^2	0	2500
Grape	Tons/km^2	0	2500
Maize	Tons/km^2	0	2500
Oat	Tons/km^2	0	2500
Olive	Tons/km^2	0	2500
Pasture	Tons/km^2	0	2500
Potato	Tons/km^2	0	2500
Citrus	Tons/km^2	0	2500
Rice	Tons/km^2	0	2500
Rotation	Tons/km^2	0	2500
Tomato	Tons/km^2	0	2500
Veg	Tons/km^2	0	2500
Wheat	Tons/km^2	0	2500
Fruit_area	km^2	0	
Grape_area	km^2	0	
Maize_area	km^2	0	
Oat_area	km^2	0	
Olive_area	km^2	0	
Pasture_area	km^2	0	
Potato_area	km^2	0	
Citrus_area	km^2	0	
Rice_area	km^2	0	
Rotation_area	km^2	0	

Tomato_area	km^2	0	
Veg_area	km^2	0	
Wheat_area	km^2	0	

Table 3: WS Outputs

Variable	Unit	Min	Max
time	Months	1	12
Reservoir	Million m^3	1467.24	2102.76
Energy_balance	Unitless	1.0	1.0

Please note: For the Energy Balance the model adaptive and was designed so that supply always equalled demand so the energy balance in this model was always equal to 1

Currently, in order to fulfil requirements coming from both sides (Serious Game UI, and SDMs), the web service response is provided using XML format. However, we are studying whether it would be interesting to use a hybrid approach using XML and JSON format to both give support to WPS standards, and simplify data manipulation in WS clients. Figure 5: Sample WS request in XML format and Figure 6: Sample WS answer in XML format, show examples of a request and the corresponding answer performed while interacting with the Web Service.


```xml
<?xml version="1.0" encoding="utf-8"?>
<wps:Execute xmlns:ogc="http://www.opengis.net/ogc"
        xmlns:gml="http://www.opengis.net/gml" xmlns:wfs="http://www.opengis.net/wfs"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink"
        xmlns:ows="http://www.opengis.net/ows/1.1" xmlns:wps="http://www.opengis.net/wps/1.0.0"
        xmlns="http://www.opengis.net/wps/1.0.0" service="WPS" version="1.0.0"
        xsi:schemaLocation="http://www.opengis.net/wps/1.0.0
        http://schemas.opengis.net/wps/1.0.0/wpsExecute_request.xsd">
        <ows:Identifier>testcomplex</ows:Identifier>
        <wps:DataInputs>
                <wps:Input>
                        <ows:Identifier>xml_in</ows:Identifier>
                        <wps:Data>
                                <wps:ComplexData encoding="UTF-8" mimeType="text/xml">
                                        <parameters>
                                                <Game_ID>10500</Game_ID >
                                                <TimeStamp>3444</TimeStamp>
                                                <Basin_surface>10500</Basin_surface>
                                                <Citrus_area>3444</Citrus_area>
                                                <Coef_defluss>0.26</Coef_defluss>
                                                <Fruit_area>1877</Fruit_area>
                                                <Grape_area>5474</Grape_area>
                                                <Industry_energy_demand>0.522</Industry_energy_demand>
                                                <Maize_area>7311</Maize_area>
                                                <Oat_area>1222</Oat_area>
                                                <Olive_area>4102</Olive_area>
                                                <Other_energy>0.047116667</Other_energy>
                                                <Pasture_area>2903</Pasture_area>
                                                <Percap_demand>0.000007604</Percap_demand>
                                                <Population>1675411</Population>
                                                <Potato_area>326</Potato_area>
                                                <Rice_area>3503</Rice_area>
                                                <Rotation_area>18284</Rotation_area>
                                                <Service_energy_demand>0.201</Service_energy_demand>
                                                <Solar_production>0.005889583</Solar_production>
                                                <Tomato_area>2300</Tomato_area>
                                                <Veg_area>10674</Veg_area>
                                                <Wheat_area>1222</Wheat_area>
                                                <Wind_production>0.085987917</Wind_production>
                                                <Energy_balance>1</Energy_balance>
                                                <Reservoir>2000</Reservoir>
                                        </parameters>
                                </wps:ComplexData>
                        </wps:Data>
                </wps:Input>
        </wps:DataInputs>
</wps:Execute>
```


Figure 5: Sample WS request in XML format

```xml
<?xml version="1.0" encoding="UTF-8"?>
<output_values xmlns:gml="http://www.opengis.net/gml"
        xmlns:ows="http://www.opengis.net/ows/1.1"
        xmlns:wps="http://www.opengis.net/wps/1.0.0"
        xmlns:xlink="http://www.w3.org/1999/xlink"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
        <variable>
                <var_name>Energy_balance</var_name>
                <value>1,1,1,1,1,1,1,1,1,1,1,1,1</value>
        </variable>
        <variable>
                <var_name>Reservoir</var_name>
                <value>2000,2166.74872308923,2221.55748522354,2242.23711685724,2242.07134079978,2210.5449859
                9037,2112.03655032357,2008.44046009765,1990.65399889987,2035.70693065564,2131.36527277498,22
                22.89591439931,2280.28678171047</value>
        </variable>
        <variable>
                <var_name>time</var_name>
                <value>1e-
        04,1.0001,2.0001,3.0001,4.0001,5.0001,6.0001,7.0001,8.0001,9.0001,10.0001,11.0001,12.0001</value>
        </variable>
</output_values>
```


Figure 6: Sample WS answer in XML format

5 Data Management overall- Conclusions

For more information regarding the Data Management, please refer to the latest version of D4.2 Data Management Plan, a living document which 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 Deliverable has tried to avoid duplications with regard to Deliverable 4.2. It focuses explicitly on Data Flow between WP3 and WP4, as well as Ontologies, reflecting the progress achieved until M18. According to the Grant Agreement it will not be updated. Any sub updates will be detailed in D4.2, which is a living document.