D11.4 | *Data Management Plan*

Deliverable:	*Data Management Plan*
Author(s):	*Loredana Marmora – ISIS, Bart Verspagen - UN-MERIT, Boris Le Hir, Pierre Le Mouël - SEURECO*
Version:	Final
Quality review:	*Andrea Ricci - ISIS*
Date:	04/10/2015
Grant Agreement N°:	645884
Starting Date:	01/03/2015
Duration:	36 months
Coordinator:	Bart Verspagen, UN-MERIT
E-mail:	b.verspagen@maastrichtuniversity.nl

Table of Contents

Introduction

The purpose of the I3U Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used in the project with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document, but evolves during the lifespan of the project.
This first version contains only a short overview of the methodology applied in the project to collect data and an overview of the inputs and outputs of the NEMESIS model.
At M10 (december 2015) a first round of deliverables will be produced by the partners: they will concern the literature review and data collection for each technical workpackages. By february 2016 (M12) an updated version of the DMP will be submitted to the Commission. The document will follow the template provided by the Commission.

1. Data exchange protocol I3U project

1.1 Introduction

This section describes the data exchange protocol used in the I3U project. The purpose of the protocol is to describe the computer format in which data is made available for use within the project. The data exchange protocol is binding for all partners. It must be used to submit all data in a to the project database in a common format. The common project database will consists of a number of files organized according to this protocol.

1.2 File format and file names

The file format is Microsoft Excel binary file format, version 2007 or later. This file format is the default Excel file format and has file extension .xlsx. (also OpenOffice can also produce this file format).
The files will be named as follows: *"WPXXnameYYYY.xlsx"*
where *"XX"* is the number of the workpackage that produced the file, *"name"* is the name of a group of indicators, and *"YYYY"* is the version number of the data, where all digits are used. Version numbers is not consecutive but indicate major steps in data construction process.
We expect several updates during the project's lifetime, so several version numbers will be saved.

1.3 Data documentation

Use Calibri font throughout the file, point 11.
The leftmost worksheet in the file should be named documentation, and should contain a description of all data contained in the file. Column A in this sheet should be set to column width 100 and text wrapping on. Cell A1 contains a general description of the data contained in this file. It should cover approximate definitions (what phenomenon is measured by these indicators?). If this description needs several paragraphs, use one cell per paragraph, and continue using as many rows as are needed. Leave one row empty after the general description is completed.
Start reporting formal variable definitions on the next row, starting with the variable name, in bold, followed by the formal definition. Use one cell per variable, and leave one row empty after the last definition.
Start reporting sources on the next row, starting with text "source for variable name", in bold, followed by a description of the source. Use one cell per variable. Leave one row empty after the last

variable.

Write any messages about permissions for data use and/or attribution of efforts in collecting the data in this cell. Mention the I3U project in the attribution.

1.4 Data presentation

The worksheets following the documentation sheets contain the actual data. Use one worksheet per variable, and name the worksheet by the exact variable name (used in the documentation sheet).

The top row of a worksheet containing data documents the units to which the data refer (countries, sector, regions, etc.; we refer to these as labels in this document), and the years for which data is available. Start with label country in column A, and use subsequent columns for additional labels in the database (such as sector or region). Use as many columns as there are label types (e.g., 3 columns if there are countries, regions and sectors). Document the first year for which data are available in the column following this, and continue years after this. Freeze panes at the 2nd row below the first year.

Adjust column width according to the data format and labels, but do not make columns any smaller than width 3, nor wider than width 15 (including columns for labels). Left align label columns, right align data columns.

Always provide text for any label column that is used (do not leave any cells empty below a label), and set the cell format to General for all labels. Use full country names as used on the Eurostat website (see below for selected countries). For any other labels than countries, provide a separate worksheet explaining the labels used (see below).

Provide the data below the years, and set the cell format to Number for all cells containing data. Use an appropriate fixed number of decimals throughout the worksheet for a single variable, but implement this as a display format, not as actual rounding (provide full decimals in the actual writing of variables). Use two dots (..) for missing data (also right align these), and 0 for values that are actually 0.

1.5 Notes to individual datapoints

In case your data has any notes (e.g., to indicate exceptions to definitions, breaks in definitions or sources, etc.), include a separate sheet for every variable for which such notes exist, and name this sheet "variable name – notes". Insert the sheet to the immediate right of the sheet with data.

The notes sheet has exactly the same format as the actual datasheet, except that the cells where the data are in the data sheet will contain the notes. Set the format to General for these cells, but keep them right aligned.

1.6 Aggregations for sectors and EU

When possible, provide EU totals for all variables that you supply. When appropriate, these totals are weighted averages, using the natural weights that lead to a value that spans the entire country set.

1.7 Labels for countries, sectors, regions and other dimensions

1.7.1 Countries

Use full country names (as specified below, or for non EU countries, use official country names as specified in this UN document.

The following table provides country memberships of the EU12, EU15, EU25, EU27 and EU28 groups:

Country	Remarks	EU12	EU15	EU25	EU27	EU28
Belgium		Yes	Yes	Yes	Yes	Yes
Bulgaria					Yes	Yes
Czech Republic				Yes	Yes	Yes
Denmark		Yes	Yes	Yes	Yes	Yes
Germany	For data until 1990 use former territory of the FRG, indicate this in notes if any data for 1990 or before are included	Yes	Yes	Yes	Yes	Yes
Estonia				Yes	Yes	Yes
Ireland		Yes	Yes	Yes	Yes	Yes
Greece		Yes	Yes	Yes	Yes	Yes
Spain		Yes	Yes	Yes	Yes	Yes
France		Yes	Yes	Yes	Yes	Yes
Croatia						Yes
Italy		Yes	Yes	Yes	Yes	Yes
Cyprus				Yes	Yes	Yes
Latvia				Yes	Yes	Yes
Lithuania				Yes	Yes	Yes
Luxembourg		Yes	Yes	Yes	Yes	Yes
Hungary				Yes	Yes	Yes
Malta				Yes	Yes	Yes
Netherlands		Yes	Yes	Yes	Yes	Yes
Austria			Yes	Yes	Yes	Yes
Poland				Yes	Yes	Yes
Portugal		Yes	Yes	Yes	Yes	Yes
Romania					Yes	Yes
Slovenia				Yes	Yes	Yes
Slovakia				Yes	Yes	Yes
Finland			Yes	Yes	Yes	Yes
Sweden			Yes	Yes	Yes	Yes
United Kingdom		Yes	Yes	Yes	Yes	Yes

Table 1: country memberships of the EU12, EU15, EU25, EU27 and EU28 groups

The official membership countries are: Iceland, Montenegro, the former Yugoslav Republic of Macedonia, Albania, Serbia, Turkey.

1.7.2 Regions
The project uses the NUTS 2013 classification, with NUTS-2 as the default level of disaggregation. Whenever NUTS-3 data exist, these can be provided, but in any case NUTS-2 (when available) must be provided. Use NUTS codes to indicate regions.

1.7.3 Sectors
The project uses the NACE classification, Rev. 2. Data availability will determine the level of disaggregation. Use NACE codes to indicate sectors.

1.7.4 Other labels
When other labels are necessary, use an official classification, and provide details of this classification by referencing an official document.

2. Inputs and outputs of the NEMESIS model

This section does not deal with the management of dataset strictly speaking but with the information that will be used to run the NEMESIS model and the different results that the model will be able to provide.

The first subsection briefly explains the general variables used as exogenous inputs in the model, the second subsection gives a first idea of the inputs that will be provided by the WP1 to 8 allowing the simulation of the implementation of the Innovation Union Commitments with NEMESIS. The last part describes succinctly the nature of the outputs provided by the model.

2.1 General inputs used to run the model

The usefulness of a model relies on its ability to describe the impacts of the variations of exogenous variables (inputs of the model) on endogenous variables (outputs of the model), in a theoretically and empirically consistent way. The exogenous variables are of two types: those resulting from the incapacity of the model to represent every phenomenon and that reflect the limitations of the model, and those that are by nature exogenous. The first type of exogenous variables regroups notably a set of assumptions related to interest rates, exchange rates, activity proxies for the rest of the world, prices of wholesale commodities and especially oil; demographic assumptions by country such as total population, population and participation rates to labour force by gender, by skill and by age per 5 years cohorts. The second type of exogenous variables is mainly composed of government-related decisions: assumptions on national policies and notably fiscal policies (indirect and direct taxes, social security benefits and contributions); government expenditures (defence, health, education, infrastructures, others expenditures) and investments; and energy and environment assumptions as excises duties and other energy tax rates, CO_2 taxation, etc.

2.2 Inputs provided by the other partners to simulate the commitments

The simulation of the commitments with the NEMESIS model can imply two main ways of implementation in the model. The first consists in modifying exogenous variables such as public investment in R&D for instance; the second way consists in modifying parameters such as productivity coefficient or elasticities.

In addition, if a commitment involves mechanisms that are not represented in the model, modifications of equations of the model may also be suggested. In that case, the results of the WP 1 to 8 should provide enough information for the modification of the equations (the value of an additional parameter for instance).

Regarding the coverage of the data, the exogenous variables, in the ideal situation, must cover the 28 countries and the 30 sectors of the model over the simulation period (from the start of implementation of commitments to 2050). However, because of their non-directly observable nature, most of these variables will be calculated thanks to the direct impact assessments realized in WP 1 to 8. If robust calculus is not possible for the requested coverage, assumptions will have to be done to overcome this limitation.

Concerning the modification of parameters, their value will, in most cases, remain constant over the simulation period, so that only the value of their variation due to the implementation of the concerned commitment is required.

The table 2 below gives a first list of exogenous variables and parameters needed for the implementation of commitments (in each case, we need the deviation of the exogenous variable or of the parameter due to the implementation of the commitment with respect to the situation in which the commitment would not have been implemented). This list is obviously still subject to discussion and is potentially modifiable depending on the findings of the WP 1 to 8. In addition to this list of data, the question of the modification of equations is raised in some cases.

Finally, we should keep in mind that the simulation of every commitment in NEMESIS will not be systematic: (i) some of them will be either grouped together because it will be not possible or not relevant to disentangle their effects and (ii) some of them will not be simulated because no significant economic impact are expected from these commitment (ranking university or the development of the innovation scoreboard for instance, these commitments are indicated with a grey line in the table).

2.3 Outputs that will be provided by the model

On the output side, NEMESIS can deliver results at EU27, country and regional NUTS2 levels for key economic indicators. There are three main layers of economic indicators: (i) macro-economic, such as GDP (European, National or Regional) and its counterparts (final consumption, investment, exports, imports, etc.), unemployment rates, ...; (ii) sectoral, such as production, value added and employment per NACE economic sector or sector clusters, and (iii) those related to agent accounts with five categories: Government, Non-Financial Corporations, Financial Corporations, Households including NPIH, and outside.

The inclusion in the model of detailed data on population and working force, allows finally NEMESIS to deliver many social indicators, such as employment, unemployment and labour force participation rates by gender, GINI coefficients for wages and earnings, and a set of indicators dedicated to measure inequalities between European countries and regions for key variables as GDP and final consumption per capita. All these indicators result from the mechanisms incorporated in NEMESIS.

Com. Nbr	Exogenous variables							Parameters				Equations			
	Public R&D	Employment of researcher in public sector	Wages of researchers in public sector	High skill labour supply	Subside for innovation	Leverage effect on private R&D	investments in ICT and OI	Productivity of innovation activities: λ	Economic performance of innovation: α and α'	Private knowledge diffusion parameter: delay or ψ or θ	Public knowledge diffusion parameter: delay or $\propto$	Explicit public research production function	Cost of use of innovation investments	finance and innovation	heterogeneity of firms
1	X	X	X	X											
2.1															
2.2											X				
3				X			X								
4				X						X	X				
5	X				X						X				
6	X				X										
7					X			X	X						X
8	X				X										
9	X				X	X									
10													X		
11													X		X
12													X		X
13	X												X		
14						X		X					X		
15						X		X		X			X		
16						X		X		X			X		
17	X					X									X
18	X					X	X								X
19						X	X								X
20							X			X	X				
21										X	X				
22								X		X					
23								X							
24	X					X	X								X
25	X					X	X								X
26	X					X	X								
27	X										X				
28															
29										X	X				
30	X	X	X	X											
31										X	X				
32	X					X					X				
33															
34															

Table 2 Expected information as inputs for modelling commitments