IS-ENES – WP 11

D 11.3 – The e-impact portal Software Requirements/Architectural Design/ IO Specification

Abstract:

A central objective of JRA5 is to provide a prototype for a web service interface (the e-impact portal thereafter) to bridge the gap between the climate modelling community, the climate impact community and decision makers (the users or stakeholders thereafter) for developing adaptation and mitigation policies. For that purpose a number of selected and representative national Use Cases for climate data have been selected and described (deliverable D 11.1).

A detailed analysis of these national Use Cases has been performed in task 2. It brings out commonalities in term of user needs, work processes, tools and limits of data use for impact studies. The D 11.2 deliverable presents the main results resulting from this analysis. It serves as the basis to design and build the e-impact portal.

This document describes and analyzes the e-impact portal software requirements and architectural design. It provides the basis for the detailed design and the implementation of the e-impact portal. This document has been setup according to the ESA software engineering standards for small software projects and makes use of the ESA's software specification document template.

Grant Agreement Number:	**228203**	**Proposal Number:**	**FP7-INFRA-2008-1.1.2.21**
Project Acronym:	**IS-ENES**		
Project Co-ordinator:	**Dr Sylvie JOUSSAUME**		

Document Title:	The e-impact portal Software Requirements/Architectural Design/ IO Specification			**Deliverable:**	11.3
Document Id N°:	D11.3	**Version:**	12	**Date:**	2011-07-08
Status:	Final				

Filename:	IS-ENES_11_DI_11-3_Software_specification_A12.doc

Project Classification:	Public

Approval Status		
Document Manager	**Verification Authority**	**Project Approval**
Wim Som de Cerff, KNMI	Martin Juckes, STFC	
Maarten Plieger, KNMI	Joachim Biercamp, DKRZ	

REVISION TABLE

Version	Date	Modified Pages	Modified Sections	Comments
A02	15-11-2010	All	All	First draft. This document has been setup according to the ESA software engineering standards to small software projects. This document provides the base document on which the JRA5 community can work and provide input.
A04	28-12-2010	All	All	Comments accepted and worked into the document. Chapter 3 requirements specification written.
A05	25-02-2011	All	All	Comments from CERFACS and SMHI accepted and worked into the document.
A06	25-02-2011	All	2	Added Diagrams, use case analysis
A08	19-04-11	All	All	Additions and changes according to JRA5 teleconferences (March 28), JRA4/JRA3 meeting in Paris (March 29,30), discussions at EGU in Vienna (April 4-8)
A09	29-04-11	All	All	Review comments from CERFACS and CMCC worked into the document
A10	02-05-11	-	-	Release for IS-ENES internal review
A11	07-07-11	All	All	Processed the review comments from the internal reviewers Joachim Biercamp, Petra Nerge and Martin Juckes
A12	08-07-11	All	All	Finalize document.

Table of Contents

Executive Summary

IS-ENES is the infrastructure project of the European Network for Earth System Modelling (ENES). IS-ENES combines expertise in Earth system modelling, in computational science, and in studies of climate change impacts. IS-ENES will provide a service on models and model results both to modelling groups and to the users of model results, especially the impact community. Joint research activities will improve the efficient use of high-performance computers, model evaluation tool sets, access to model results, and prototype climate services for the impact community.

The work package WP11/JRA5 is aimed at Bridging Climate Research Data and the Needs of the Impact Community. The central objective of JRA5 is thus to provide a prototype for a web service interface (the e-impact portal thereafter) to bridge the gap between the climate modelling community, the climate impact community, the climate effect community, practitioners and decision makers (the users or stakeholders thereafter) for developing adaptation and mitigation policies. Building this service requires in a first time to precise user needs and determines the best way to provide the information.

The methodology adopted in JRA5 has consisted in gathering several case studies performed by the partner institutes. 17 national and representative use cases have been selected. They give a more precise idea of user request in terms of objectives, geographical coverage, time periods, parameters, horizontal and temporal resolutions, and final data format. They inform about the ways to deliver climate information to different type of users, covering a range of applications, such as water resources, forestry, agro and ecosystems, etc. Use cases documentation provides information about the main phases of the work that has been done to answer user requests. It includes workflows, data features, data processing, user/provider interaction, user documentation needs, and uncertainties assessment. This information has been synthesized in the IS-ENES deliverable D11.1.

The deliverable D11.2 presents a finer analysis of these 17 use cases. Despite the small size of the sample (17 experiences done by 8 research institutes from 5 countries); there is a large diversity in the use cases features: diversity in objectives, users, final products, data features, methods, workflows and tools. A first task was to classify this information in order to have a better overview of the request and the available resources to answer.

First key classification concerns the type of users. It influences the types of final products, users/providers interactions, types of documentation and also processing steps. The first group of users is composed of scientists and engineers that are used to work with climate data, scientific file formats, high data volumes, analysis methods and uncertainties. They ask generally for raw data they can process themselves. The second group of users comprises practitioners, members and deciders of different organisations (e.g. NGO, firms, and government). They require “elaborated products” like maps, plots or statistics to elaborate their decisions.

Second classification focuses on data. Many different parameters (final climate parameters, sectorial parameters, auxiliary parameters, indices specifically developed to answer user needs), geographical coverage (points, cities, rivers, countries, regions of all continents), time periods (different time slices of the 20th and 21st centuries), horizontal resolution (global, regional and very local resolution); temporal resolution (from hourly time step to decadal one) and data format (model native formats, scientific formats, users specific format) are needed to fulfill all user requests. Implementing this diversity in the e-impact portal will be a major challenge.

In D11.2 the workflows described in the 17 use cases were analyzed. It seemed that they are very different and focus on specific steps or actions (data processing, data delivering, users/providers interaction…). They have been analyzed and combined in order to propose a first draft of an “ideal process”. It includes 7 steps: collection, analysis and validation of the requests; definition of proceeding instructions; planning; data processing; quality assessment; documentation; data packaging and data release.

Beyond the inherent climate uncertainty sources, one must also take into account the errors due to misuse methods and tools. Ignoring these factors leads to erroneous interpretation of the results. Sources of errors are numerous as well as the methods to tackle them. Quantifying and reducing uncertainties and biases require a good knowledge of datasets, methods and model skills. It needs also complex analyses that are time expensive. Use cases show that few users are able to evaluate uncertainties by themselves. This introduces a double question: What is the best way to introduce the quality/uncertainty assessment in the e-portal? How do we deal with data that has only a low reliability? In D11.2 a set of recommendations are defined:

- The e-impact portal should be flexibly implemented because it is a prototype (it will be improved with time), methods, dataset, and tools change rapidly and because the interaction with users about the climate questions is not well known.
- The e-impact portal should include documentation, guidance and illustration to help users starting to use its functionalities and to become familiarized with the large volume of information.
- Special interest will be given to user feedback in order to improve our knowledge of user needs, adapt the e-portal and improve the service.

In D11.3 the selected use cases are analyzed from a technical point of view. A set of user requirements has been specified to cover the user's needs. A conceptual model of the e-impact portal is derived from the use cases; both a class diagram and a sequence diagram are modelled, which respectively represents the structure and the behavior of the e-impact portal. Both are modelled using Unified Modelling Language (UML).

During the analysis it became apparent that impact users want to have (1) both the ability to interact with the climate scientists -and- (2) have the ability to explore and do simple operations on the climate model data. Besides being a data portal, the portal will also be a contact point for the impact community. Typical use cases show that the impact community has needs for highly customized and tailored data, which can hardly be compiled by an automated process. The e-impact portal can play a central role in being the information broker where the impact community and climate modelling community can share their knowledge and expertise, collaborate and develop new initiatives. Other use cases show that there are users who want to have a data portal where they can search, visualize, explore and process climate model data. This part can be performed by a system and is therefore suited to be implemented by the e-impact portal.

The e--impact portal will both have a role in data distribution and processing, but also in being the information broker for the impact community to handle specific requests which need expert knowledge. The e-impact portal will be an integral part of the main ENES portal (previously known as the vERC). It must be clear that the portal does not aim at replacing national data portals, but rather to build a prototype to propose a structure, standard workflows, common tools, documentation, support system, etc. to access decentralized federated data repositories using standardized technologies such as METAFOR metadata, OGC, ESG, THREDDS, among others.

1. INTRODUCTION

The central objective of JRA5 is to provide a prototype for a web service interface (the e-impact portal thereafter) to bridge the gap between the climate modelling community, the climate impact community, and decision makers (the users or stakeholders thereafter) for developing adaptation and mitigation policies. To narrow the scope for developing the e-impact portal, the focus will be on providing a e-impact portal for the impact research community, being a narrower and well defined group, as well as having the most frequent direct interactions with the climate modelling community without any intermediate user. Also, it must be clear that this portal does not aim at replacing national or specific existing portals, but rather to define and develop a framework or structure, web services, metadata, workflows, documentation, common tools for data processing, decentralized data repositories, etc. based on existing standards.

Building this service requires to clearly specify user needs and determination of what is the best way to provide information. This requires also to really identifying what are really the users, in this present case being the impact research community. Usually, users work directly with a given institute. This bilateral mode results in an important number of methods and tools, and also to closely related data requests that could better be shared with others because of commonalities in requests. The e-impact portal will deal with this diversity in order to bring a comprehensive overview of the possible use of climate information and data.

The methodology developed in IS-ENES consists in gathering several case studies, use cases or business cases, performed into the partner institutes and making a more precise analysis of the needs, methods and tools using these examples. For that purpose, 17 national and representative use cases have been selected. These use cases describe how to deliver climate information to different type of users, covering a range of applications, such as water resources, forestry, agro- and eco-systems, hydro-geological and landslide risks. Use case documentation provides information on workflows, data features, data processing, user/provider interaction and uncertainties assessment. They have been synthesized in the IS-ENES deliverable D11.1.

The 17 IS-ENES use cases show wide ranges of objectives, products, methods, datasets and tools to process and deliver final products to users. JRA5 covers only a small sample and this diversity would be likely higher considering more cases. However the detailed analysis of these use cases points out the existence of commonalities that will facilitate their implementation in the e-impact portal. The main differences and commonalities of the use cases have been described analyzed in deliverable D11.2.

This document (D11.3) is built on the previous D11.1 and D11.2 documents.

1.1. PURPOSE OF THE DOCUMENT

This document is a software specification document (SSD) that specifies the software requirements for the e-impact portal based on user requirements. The SSD is an explicit set of requirements to be satisfied by the product or service. This document consists of two parts, first the software requirements for the e-impact portal are analyzed and second the architectural design of the e-impact portal is given.

1.2. DEFINITIONS, ACRONYMS AND ABBREVIATIONS

See appendix A.

1.3. REFERENCES

Nr	Title	URL

1	D11.1 Final description of selected Use Cases including user requirements specification	https://is.enes.org/the-project/the-workpackages/IS-ENES_11_DI_11-1_USECASES_Final.pdf/at_download/file
2	D11.2 Baseline documents on e-resources/tools and transverse themes	https://is.enes.org/the-project/the-workpackages/IS-ENES_11_DI_11-2_common_tools_final.pdf/at_download/file
3	Guide to applying the ESA software engineering standards to small software projects, BSSC(96)2 Issue 1 May 1996	http://emits.esa.int/emits-doc/e_support/Bssc962.pdf
4	Mastering the requirements process, Robertson and Robertson	ISBN-10: 0-321-41949-9
5	TGICA report	http://www.ipcc-data.org/docs/tgica14/DOC14_AR5vars.pdf

1.4. OVERVIEW OF THE DOCUMENT

In chapter 2 the general concept of the e-impact portal is described using a context diagram and a model description. The model description is detailed in chapter 3, where the requirements for each part of the model are stated in functional and non-functional requirements. Also the design constraints are listed here. Chapter 4 describes the physical model, which described the design of the e-impact portal using implementation terminology (UML diagrams). Chapter 5 provides the feasibility and resource estimates. Chapter 6 and chapter 7 provide traceability matrices for tracing the user requirements to the designed components.

2. MODEL DESCRIPTION

The e-impact portal

IS-ENES JRA5 deals with the common practice to deliver climate information to many different users of both impact and decision-maker communities in the climate modeling community. The objective is to provide services to users and help data providers in this mission. The e-impact portal will serve as a support for this purpose. There is a growing interest to use climate information, but delivery of climate variables as model outputs to users for impact studies is in general not well understood and underestimated. There is also a large heterogeneous number of climate scenarios available, with no central point access and a large diversity of datasets. Process design depends on the main objective. From the user point of view, the objective is, first, to retrieve data and visualizations of different parameters in order to perform vulnerability assessment and, second, to get quality information and guides for using these products. From the provider point of view, the objective is to give easy access to these products whilst minimizing their intervention and ensuring that the products will be used with adapted cautions and for adapted purposes.
Potential actors for such a system are scientists from climate and meteorological research institutes and people from impact and decision-making communities (engineers, scientists, and decision makers). In the context of JRA5, the portal is a prototype and hence, the types of user that are addressed are impact research scientists. However, the portal is designed so that it is extensible and can be extended to other types of users in future implementation using the same structure and standards.
There are three macro-processes identified (from D11.2):

1) Processed data, indices, and visualization responding to their initial request;
2) Guide for using information and recommendations depending on data quality;
3) Information on methods and tools used for data processing.

For D11.3 the scope of implementation will focus on the impact research community. Both experience and the use cases show that the climate change information has to be translated to impacts, which is something typically done by impact scientists. And then they become the providers of decision support material for the decision-making communities.

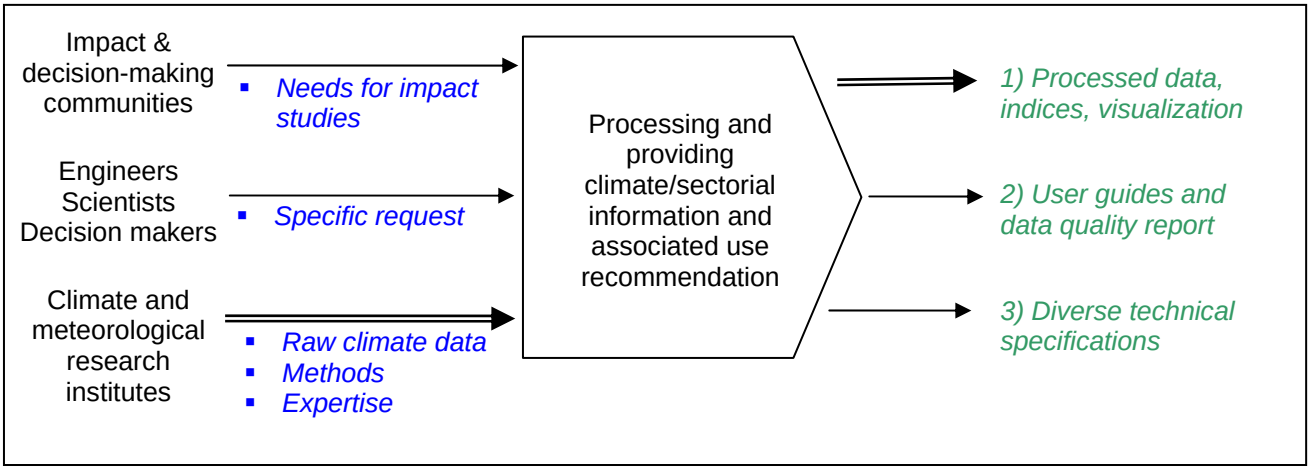

Figure 1: *Macro-process related to climate information delivering for users of both impact and decision maker communities. Data provider point of view is adopted.*

2.1. ANALYSIS OF D11.1 AND D11.2 DOCUMENTS

During the definition of the use cases, it came apparent that documentation and user guidance would be a very important part of the e-impact portal. Beyond inherent climate uncertainty sources, errors due to misuse of tools and methods must be taken into account. Quantifying and reducing uncertainties and biases require a good knowledge of datasets, methods and model skills. It also needs complex analyses that are time expensive. From the use cases it became evident that only few users are able to evaluate uncertainties by themselves. The e-impact portal must deal with this problem without the continuous need of climate modelling expert intervention. One method to achieve this goal is by providing user guidance, tool descriptions and data documentation in the e-impact portal. It is also important that the links between data, tools and documentation is strong. The goal is to minimize the chance that data or tools are being misused or misinterpreted.

The e-impact portal will therefore have two main features: 1) providing climate model data, and 2) providing information and documentation in the form of tool descriptions, user guides and guidance (e.g. uncertainties, proper data uses, etc.), climate model and downscaling methodologies documentation (including limitations, hypotheses, etc.) and dataset metadata. The links between data and documentation is important and must be stressed during the use of the e-impact portal by the user.

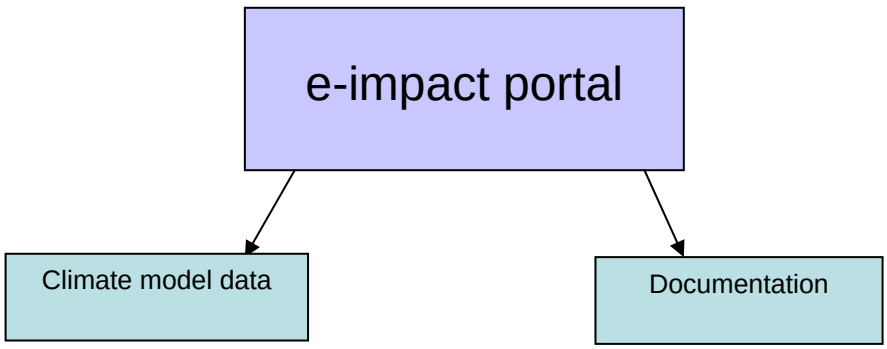

Figure 2: *The e-impact portal will have two main features: providing climate data and providing documentation. The link between data and documentation must be stressed during use of the e-impact portal. The user will also be able to ask for expert knowledge from climate scientist and is therefore a contact point.*

Climate model data:

- Raw data: model output from global climate models (GCM) and regional climate models (RCM).
- Post-processed data: raw data from GCM or RCM, which has been post-processed: e.g. averaged, statistically downscaled and bias-corrected GCM data, calibrated and bias-corrected RCM data, indices, and visualization responding to their initial request.

Documentation:

- General guide for using information and recommendations about basic use of climate scenarios and related to data quality and proper use. This section of the documentation needs to deal with the following topics (not ordered):
  - Emission scenarios and an overview of the SRES (and RCP) process.
  - Detection/attribution studies (IPCC reports).
  - Brief summary of the climate system, biogeochemical cycles, the carbon cycle (can be based on the IPCC reports, and mainly provide links to the relevant chapters).
  - Explanation of the principles of GCMs, scale, parameterizations, explain principal sources

of biases (links to spatial resolution section, e.g. downscaling which removes some biases), spin-up, trends, initial conditions, main limitations, evolution since the 1990's. Reusing work done by METAFOR and other IS-ENES work packages.

- Climate scenarios (centennial) vs. climate prediction (decadal), where the latter use properly initialized GCM integrations where the decadal variability is in phase with what is observed. Difference between weather forecasts and climate scenarios. Differences between present climate simulations by climate models and reanalysis (temporal aspects).
- Time scale (less difference between scenarios at the short/intermediate time-scale (up to ~2050). Limitations in climate scenarios for the short term (upcoming decades). Inertia of the climate system and already committed warming. SRES scenarios which are significant only past 2050.
- Spatial resolution, statistical and dynamical downscaling, principles of RCM, examples using precipitation, how gridded data is related to the point measurements? Spatial and temporal representativity of observations. Links to downscaling methods.
- Explanation of downscaling proper (i.e. scale transition), bias correction (which also should be applicable to RCM data) and the problem of mixing out-of-phase natural variability for a bias and thus 'over-adjusting' for bias.

- Guidance about cascade of uncertainties. Types of uncertainties.
- Guide about downscaling methods.
- Information on methods and tools used for data processing, including uncertainty information.
- Model and data metadata (METAFOR).
- User feedback mechanism. There are four main categories of user feedback that needs to be handled:
  - Feedback on the web portal as such regarding technical or functional aspects, layout and structure of pages, navigation, etc (i.e. related to the website structure).
  - Suggestions regarding the information content (i.e. regarding the written content and associated illustrations).
  - Request for expert support for solving specific needs.
  - Suggestions and requests regarding the data available through the portal.

  The first two are closely related to the maintenance of the web portal as such. The third one is linked to the second one in that frequently recurring requests for expert support point towards the need to expand the information content regarding these specific areas. It is essential that the limitations to available expert support is explicitly stated in relation the committed resources. Otherwise, users of the portal may likely get unrealistic expectations.
- FAQ & Do's and don'ts with climate scenarios.
- Climate Scientist Expert Knowledge Database and contact points.
- Guidance about raw data manipulation, in order to describe the procedures for the conversion of the raw data into data usable by end users. A typical example is the problem of the conversion from NetCDF to XLM or GML formats (used for programs developed in GIS environment)

2.1.1. The climate data that the e-impact portal provides or requires

This chapter provides an analysis of the data that is used by the impact and climate community. Climate model data is usually structured by parameters or variables. The data can be categorized as climate data and sectorial data. The first category refers just to the climate, as the second one is specific to a studied sector. Examples for sectorial data are river flows for the hydrological sector; fish

occurrence or population viability for ecological cases, etc.

There are 2 types of parameters: raw data and indices. Raw data are direct output of a climate model; generally they are not directly used to assess impact but used to run impact models. However this raw data should be first bias-corrected to provide data with proper quality. Indices are post-processed variables. They are developed to bring specific information on vulnerabilities or impact. Indices can be computed from simple operations or can be determined using an impact model based on statistical or dynamical methods.

There is a strong need for data at the 2 meters or surface levels. The requested parameters are: liquid precipitation, mean temperature, minimal temperature, maximal temperature, short wave radiation, solid precipitation, wind (at 10 meters), infrared radiation, relative humidity and potential evapotranspiration. Parameters that were requested occasionally were: pressure, cloud cover, diurnal temperature range, geopotential, soil temperature, soil moisture and gust wind speed. This has been described extensively in D11.2.

Geographical area and horizontal resolution:
Geographical areas extend from points to any square selected from latitude/longitude (lat/lon) or other projections. Use cases include also specific areas like river basins or countries. They can be selected using masking techniques based on EU NUTS2 (see 3.1) region. Few cases consider vertical profiles (WUR, INHGA, IPSL-4 use cases). Data are used on different type of grids: regular lat/lon; Lambert; various national grids that often implies regridding from the model grids. Models grids can be Gaussian reduced, curvilinear, non-regular.
Three kinds of spatial resolutions appear in the use cases: global (100-300km), regional (7-50 km) and very local (< 3km) ones. Regional data can be divided in two categories following modelling consideration: resolution of 25-50 km for the first generation of regional models and resolution of about 10 km for the new generation of regional models and as well as for statistical methods.

Temporal resolution
In term of temporal resolution, daily data are requested in most cases following by “sub-daily” data that includes data at 6 hourly or hourly time interval. Then, there are also requests for yearly and monthly data, and also some interest in decadal data. It’s worth noting the high rate of sub-daily data request that will surely still increase in the next few years. In fact, hourly data are often used to run impact models in hydrology and geo-technics. It will probably be extended to other sectors eventually.

In 2007 the IPCC task Group on Scenario Support for Impacts and Climate Analysis (TGICA) [R5] put in a request to WCRP for data to be included in the CMIP5 archive for the benefit of the impacts community. This request included a section of variables at 3 hourly resolution, with the consequence that the archives at BADC and DKRZ are expecting to receive around 400Tb of 3 hourly data as part of CMIP5.

Actually, confidence in this temporal resolution is very weak, but it is often necessary because of technical or modelling constraints. It is important to ensure that hourly data are used with care and with suitable uncertainty assessment. The confidence in this data ought to be dependent on application. High temporal resolution data from GCM should be treated great care. But for RCM and downscaled data, the situation is somewhat different. Basically temporal and spatial resolution goes together with the numerical approximations that restrict useful information to either average over several (say 3x3) grid cells or several time-steps. It is good practice to never use full temporal and spatial resolution at the same time.

By example, for the GCM EC-Earth with an internal time-step of 1 hour, one should not analyse data at a higher resolution than roughly 6h. For an RCM using an internal time-step of 5 minutes, one should not use data at a resolution higher than roughly 30 min. This should be always taken into account when providing data to users.

Data formats
The data file format used depends strongly on user own practice and flexibility. Most often ASCII-formatted data are provided because many users do not know how to use more specific formats, even if this conversion could cause a loss of information. Other users request NetCDF or GRIB data that

are input format of their impact models. In practice these formats have been promoted by the climate and numerical weather prediction communities, respectively. Moreover, GIS compatible formats are often requested.

Processing capabilities

In the prototype, the e-impact portal will not support all of the desired processes identified in D11.2. It is currently not feasible yet to perform downscaling online with such diversified use cases, some involving complex downscaling methodologies or regional climate models. The complexity of the process is not easy to catch in an automated process and requires expert knowledge to come to meaningful results. Therefore these products will be pre-cooked by the climate scientist and stored on the portal. The capacity of the e-impact portal will be extended with time and include more and more options.

The following processing capabilities will be part of the prototype:

- Extraction
  - For a certain time period, geographical coverage, different parameters, scenarios
  - Spatially according to a geographical mask (river basin)
    - Masks available from NUTS-2 regions
    - Discussion going on in CF conventions about large scale masks
  - with modification of data type (output can be point data)
- Subsetting
- Simple operation like averaging over 30 years
- Aggregation
- Output reformatting
  - Precision of data
  - For example limit the number of digits in ASCII files.
- Reprojection
- Regridding
  - Gaussian reduced → regular grids
- Indices calculation using existing tools like CDO
- Comparison
  - Compare parameters from different scenarios / ensembles
  - Compare parameters from simulation to those from observation ones.
- Graphical visualization

2.2. CONTEXT DIAGRAM

The context diagram gives insight in the objects that exist outside the e-impact portal and with which the e-impact portal needs to interface. The context diagram is the highest level view of the system. Every object interfacing with the e-impact portal is described in this chapter.

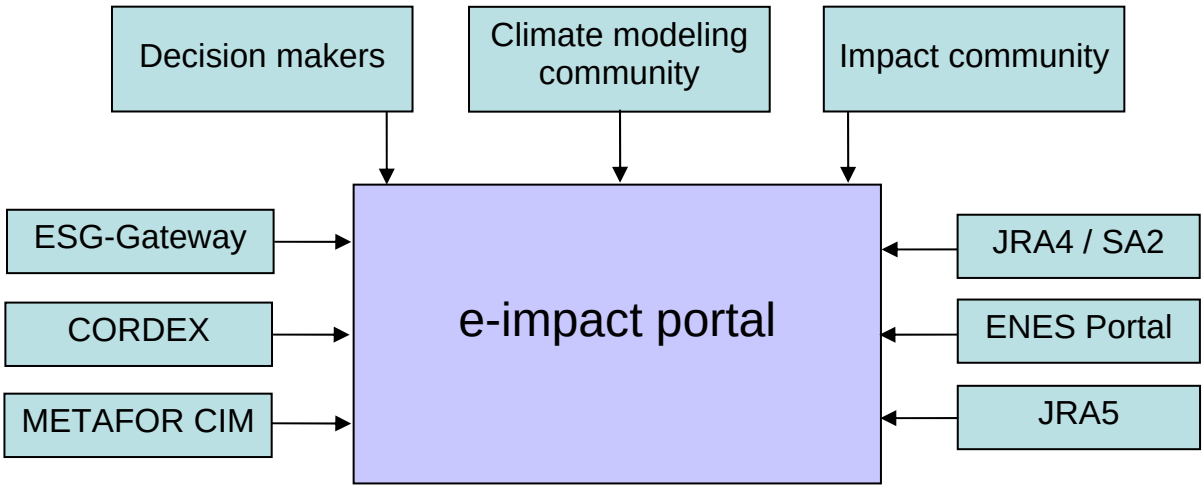

Figure 3: *The context diagram, displaying the objects with which the e-impact portal needs to interface.*

ESG Gateway – Earth Science Grid Gateway

The ESG will be a software infrastructure and a management framework to create a distributed archive capability and enhance the services beyond those provided by existing archives. The distributed archive will benefit from the resources of different and diverse institutions (e.g. BADC, WDCC, IPSL, KNMI) but will provide a service that is transparent to the user.

CORDEX

Cordex is a COordinated Regional climate Downscaling Experiment. Regional climate downscaling (RCD) techniques, including both dynamical and statistical methods, are increasingly being utilized to produce regional climate information for impact and adaptation studies. It is thus critical that the potentials and limitations of RCD-based information, along with the related uncertainties, are well understood by the modeling and user communities. The e-impact portal should be able to access CORDEX data, but will be based on JRA4 tasks.

The main goals of CORDEX are: (From CORDEX website):

- Provide a quality-controlled data set of RCD-based information for the recent historical past and 21st century projections, covering the majority of populated land regions on the globe. The RCD information will sample uncertainties in Regional Climate Change associated with (i) varying Global Climate Model (GCM) simulations; (ii) varying greenhouse gas (GHG) concentration scenarios; (iii) natural climate variability; and (iv) different downscaling methods. The CORDEX downscaling activities will be based on the latest set of GCM climate scenarios and predictions produced within the 5th Coupled Model Intercomparison Project (CMIP5).
- Define a common set of Regional Climate Model (RCM) domains for dynamical downscaling and define a standard set of variables, frequency and format for output and archival at a number of CORDEX data centres.
- Coordinate a range of RCM simulations for the defined domains, forced by analyses of observations (currently ERA-Interim) to provide a benchmark framework for model evaluation and

assessment. This exercise should include also statistical downscaling (SD) methods.

- Encourage and coordinate the development of Regional Analysis and Evaluation Teams to; (i) Evaluate the ensemble of RCD simulations, (ii) Develop a suitable set of regionally-specific metrics for RCD evaluation, (iii) Collect suitable observational data to evaluate high-resolution RCD simulations and (iv) Design experiments to investigate the added-value of RCDs and target future priorities in RCD research.
- Engage the broad RCD community in its activities and discussions.
- Provide support and information to climate impact assessment and adaptation groups interested in utilizing CORDEX RCD material in their research.

The METAFOR CIM Portal

The Common Information Model (CIM) provides a standard for metadata related to GCM models and model runs. This is being developed within the METAFOR project (started in 2008). The software suite will depend on submitted data conforming to the Climate and Forecast (CF) Metadata Convention (http://cf-pcmdi.llnl.gov/). A comprehensive list of ESM variables, which EU modeling groups would like to have, will be collected and stored on the METAFOR CIM portal. The METAFOR CIM portal will provide a search web service to find CIM instances. The CIM is a Common Information Model which describes models and experiments in detail. The METAFOR CIM portal will have a search web service, which can be used by the e-impact portal.

Decision makers

The decision makers draw their conclusions on the results that the impact community achieves. It is important that they can easily reach the documentation and guidance available about climate models, data and uncertainties. The decision makers will mainly use the documentation part of the e-impact portal to gain insight regarding the data products.

Climate modelling community

The climate modelling community plays an important role in the e-impact portal: they are the experts with the knowledge about the climate model data and the climate model limitations. They know how to use the data and should be able to provide needed feedbacks to the users. They will also use the portal to provide expertise to impact users about climate scenarios, and to download data from other institutes in support of research projects.

Impact community

The impact community is typically represented by so-called damage functions that state a statistical relationship between changes in for example global mean temperature and the associated climate damage. The impacts of climate change are usually simulated by process-based, sectorial impact models for a limited set of climate-change scenarios.

JRA4 / SA2

JRA4 will deliver the software infrastructure and a management framework to create a distributed archive capability and enhance the services beyond those provided by existing archives. The distributed archive will benefit from the resources of diverse institutions and will provide a service that is transparent to the user. The key issues to be reviewed will be data discovery, data security, performance monitoring and data handling environments. This archive is using the ESG software developed at PCMDI, and will be deployed into a federate network of main gateways and data nodes, gateways replicating the archive core of CMIP5, the forthcoming climate model intercomparison exercise. JRA4 will also prepare a list of dependencies and ensure that the METAFOR data discovery portal interfaces cleanly with the data archive. The e-impact portal will use the METAFOR data discovery portal to find its data resources. Furthermore, the authentication of the e-impact portal will be in harmony with the METAFOR data discovery portal which will be achieved with OpenID. Close cooperation with JRA4 is required to make sure that the e-impact portal authenticates and

interfaces successfully with the archives, CORDEX data and the METAFOR discovery portal. The e-impact portal will be integrated in the ENES Portal.

ENES Portal

IS-ENES will develop a web portal, integrating the European Earth system models (ESMs) and their hardware, software, and data environments. The overarching goal of this e-infrastructure is to further integrate the European climate modelling community, to help the definition of a common future strategy, to ease the development of full ESMs, to foster the execution and exploitation of high-end simulations, and to support the dissemination of model results and the interaction with the climate change impact community. The ENES Portal encompasses models, the tools to prepare, evaluate, run, store and exploit model simulations, the access to model results and to the European high-performance computing ecosystem – in particular the EU large infrastructures DEISA2 and PRACE. The ENES Portal developed by IS-ENES is based on generic ICT, Grid technology and subject-specific simulation codes and software environments.

JRA5

JRA5 deals with the common practice to deliver climate information to many different users of both impact and decision-maker communities in the climate modelling community. The objective is to provide services to users and help data providers in this mission. JRA5 will provide a prototype for a web service interface, the e-impact portal, to bridge the gap between the climate modelling community, the climate impact community, practitioners and decision makers (the users or stakeholders) for developing adaptation and mitigation policies.

2.3. MODEL DESCRIPTION

The e-impact portal needs to comply with the EU INSPIRE directives, indicating that data must be described, discoverable, viewable and downloadable according to INSPIRE standards.
Therefore, there are five main components identified: discovery services, view service, download service, transformation service and a documentation component.

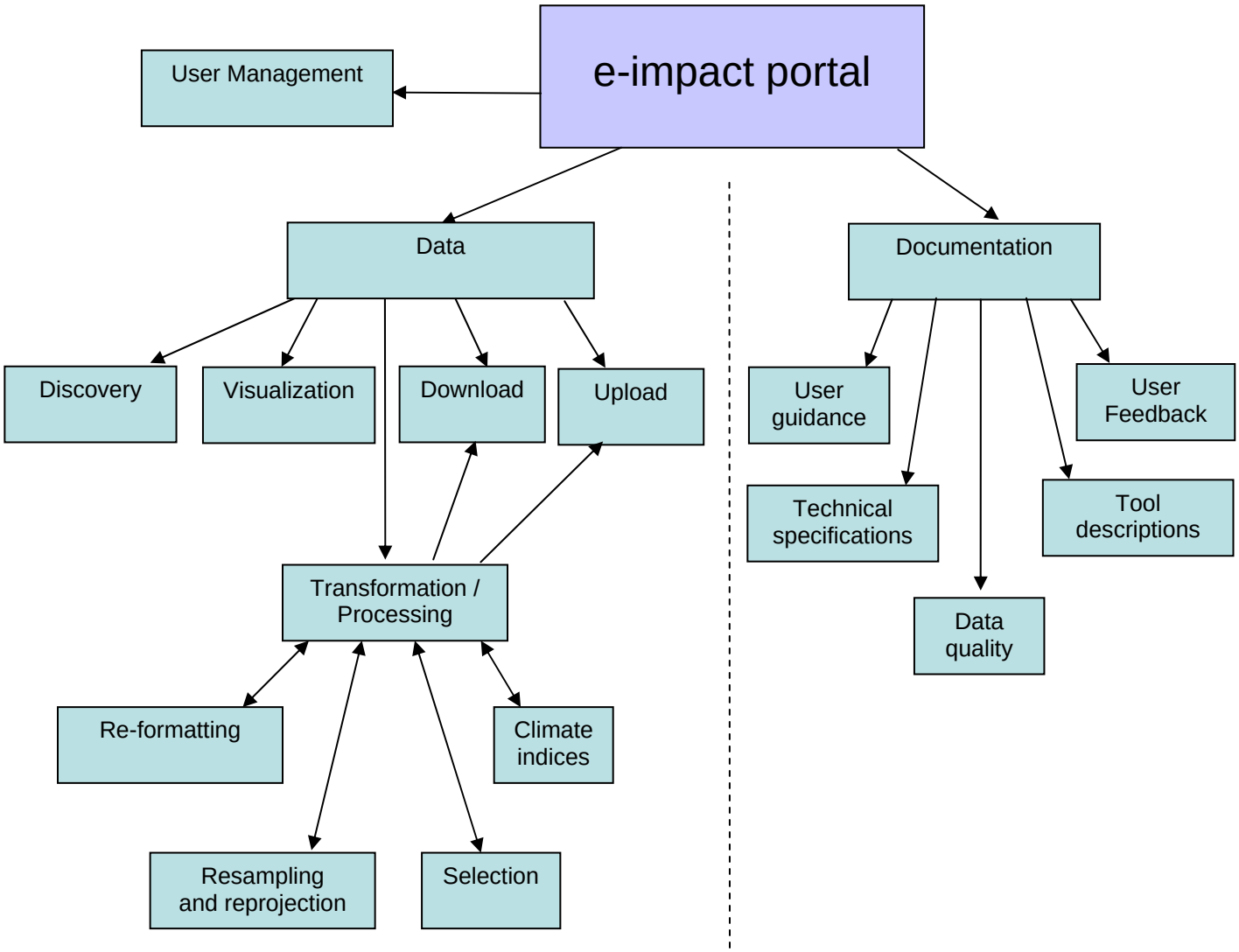

Figure 4: *e-impact portal decomposition into components.*

2.3.1. Components

This chapter describes the components that are required by the e-impact portal. This chapter will not yet focus on how these requirements are achieved or implemented, as this will be done at a later stage.

2.3.1.1. Discovery service

There are two types of discovery: 1) the e-impact portal needs to discover other services; 2) the e-impact portal offers services that need to be discoverable. Service discovery protocols are network protocols that allow automatic detection of services and data. The OGC Catalog Service defines common interfaces to

discover, browse, and query metadata about data, services, and other potential resources. Service discovery is an essential ingredient, since the e-impact portal needs to make use of one another's services without the need for continuous user intervention and the services offered by the e-impact portal need to be discoverable. The METAFOR CIM portal will cleanly interfaces the data archive, the e-impact portal should interface cleanly with the METAFOR CIM portal. The e-impact portal services can be made discoverable by making use of the OGC catalog service (CSW).

2.3.1.2. View service

A view service allows for navigating, browsing and graphically displaying the climate data and associated metadata. The e-impact portal will provide visualization in the form of 2D maps and in the form of visualizations like timeseries, delta plots and box plots. Delta plots are 2D XY graphs of DELTA temperature VS DELTA precipitation for a 30-year future period compared to a reference 30-year climate period. Usually several scenarios are plotted as scatter onto the same graph. The interface to the 2D maps shall be according to the OGC Web Map Service standard. Parts of the image generation, like creating box plots or time series graphs shall be accomplished by using the data processing services. Direct visualization of climate model data in the form of 2D maps will be accomplished by the OGC WMS service.

2.3.1.3. Download service

Allow the user to obtain the data in various formats according to their specifications and using standard protocols. Required protocols are: OGC WMS, OGC WCS, OPeNDAP, FTP and HTTP. Required formats are: ASCII, CSV, NetCDF, GRIB, XML, and GML.

2.3.1.4. Data upload service

The data upload service will be used by climate researchers to upload specific tailored datasets (e.g. statistical or dynamical downscaled data). Also, the data processors will use this service to store resulting datasets. Besides dataset data, the dataset metadata needs also to be uploaded. The dataset metadata will be used to make the uploaded data available (and findable) for the Discovery service. For using the upload service the user has to provide credentials and needs to have the appropriate access rights.

2.3.1.5. Data processing and transformation services

Goal of the data processing and transformation services is to perform pre- or post-processing tasks on climate data. Transformations will be non-interactive, e.g. processing results will always be stored on the portal temporarily. To allow the user to quickly view the results of his processing job, the e-impact portal should provide a link to the results as quickly as possible in order to minimize users' waiting time.
Required data processing services are: extraction, selection, subsetting, averaging, aggregation, output reformatting, reprojection, regridding, indices calculation, comparison and graphical visualization.
These processing services will be implemented as individual processing components. They can work independently of each other and should have well defined in and output interfaces. The e-impact portal will implement these components or “processing bricks” in a modular way: individual processing bricks can be added in the future to extend functionality of the e-impact portal.

2.3.1.6. User Management

User management encompasses all services related to user authentication, authorization, and accounting. For authentication and authorization the ESG services shall be used. Accounting will be responsible for monitoring data and site usage for providing statistics on these aspects.

2.3.1.7. Documentation and guidance

Documentation is one of the most important parts of the e-impact portal. The documentation part must be community driven. Users must be able to leave comments and provide information about the datasets. It should also be possible to annotate datasets. The content management system must enable users to add and edit documentation, in a similar way as in Wikipedia. Specific description of datasets, downscaling methods, tools, should be available and it will rely on the wiki system for partners to add relevant description. At the basis there will be only generalities not oriented toward specific tools used in the implementation of the Use Cases.

Most of the documentation should be synthetic, with a collection of links to external material for more in-depth information. Pictures should be used to shorten the text length whenever possible. Documentation should be interlinked to other parts of the documentation. The navigation among the documentation structure should always be visible. For documentation specific subjects which are rather organized as lists and database (FAQ, Do's and Don'ts, Knowledge Database), they should be structured by subject, have keywords identified so that they are searchable efficiently, and have a browsable index.

General list/database type of sections that should be available directly throughout all documentation sections

- Do's and Don'ts (Do/Don't)
- Frequently Asked Questions (FAQ)
- Contacts

User Guidance

- The Climate System
  - Generalities
  - Specific cycles (carbon, chemistry)
- Principles of Climate Numerical Models (Global Circulation Models)
  - Generalities
    - Basics
    - Grid
  - Parameterizations
  - Resolution
  - History
  - Summary of limitation
- Producing a climate numerical simulation
  - Initial conditions
  - Spin-up and trends
  - Main biases
- Centennial/decadal climate scenarios and seasonal/weekly weather forecasts
  - System forgings
  - Uncertainties
- IPCC, CMIPs and scenarios
- Detection and attribution
- Risk Assessment

Workflows: Use Cases examples (D11.3 Implemented Use Cases)

- Use Cases descriptions
- Tutorials based on the Use Cases

Bias correction methods

- Bias correction
  - Downscaling methods
    - Dynamical
    - Statistical
    - Hybrid
  - Quantile-Quantile

Uncertainties and Scenario selection

This will help the user to select which model, dataset, scenario fits to the users problem.

- Sources and Types of uncertainties
- Quantification of uncertainties
- Communicating and tackle climate change uncertainties (scenario selection)

Technical Specifications

In the e-impact portal document section, links to more data format related links and very brief explanations must be provided, on the following subjects.

- Account (OpenID) and Registration
- Data access and Authorized use

- Data Format (CF, CMOR2)
- Experiments information (METAFOR)

Dataset description
This section in the e-impact portal will be mainly brief descriptions for climate scenarios and reanalysis, with external links.

- CMIP3, AR4
- CMIP5, AR5
- CORDEX
- ENSEMBLES (GCM, RCM, Observations)
- NCEP Reanalysis
- ECMWF Reanalysis (ERA40, ERA-Interim)
- Others??

Tools descriptions
In the e-impact portal there will be a summary page with all tools (ref. appendix D11.2) briefly described, sorted by categories as in the appendix of D11.2. When clicking on the tool name, there should be a link to specific internal webpage which will include links to full documentation, external if possible, or internal for multi-purpose tools such as nco along with examples on use). These specific internal pages will enable users to leave comments on tools. Tools should be listed in multiple categories if applicable to help users. Tools description should be searchable.

Parameters and climate indices description
This section in the e-impact portal will describe parameters and climate indices identified in D11.3 (appendix).

User Feedback
User feedback is an important aspect of the e-impact portal. It is not restricted to the documentation section of the portal, but will also be available throughout the portal.

- General user feedback should be done through a discussion wiki page (or discussion forum).
- Comments must be possible on wiki pages (documentation).
- In a wiki, it is always possible to display all modifications to the wiki, hence having an overview of modifications, comments, etc., left by users. A special page should do this automatically.
- Climate Scientist should be reachable through a special contact form (hiding emails), and only on the Contacts and FAQ sections.
- Each dataset will have a specific documentation wiki page, hence enabling users to annotate and leave comments.
- An experiments and analysis section should be available for users to present data analyses.

Interactions with the Data part of the e-impact portal

- Context sensitive help should be available through pop-ups (can be on or off) in the quicklook interface. Also, a contextual menu of possible actions to apply on data should be available.
    - Interactive buttons
    - Parameters and climate indices descriptions.
    - Data processing tools
    - Data selectors
    - Workflow steps viewer (as in navigation)

2.3.2. A common use case for: Searching, processing, visualizing and downloading climate model data.

To illustrate how a user will work with the e-impact portal, a common use case is provided in which the user retrieves data. The user can optionally process data, for example to calculate climate indices. Below a use case diagram is given for this situation. The use case provides insight in the workflow and actions the user has to take to accomplish his goal.

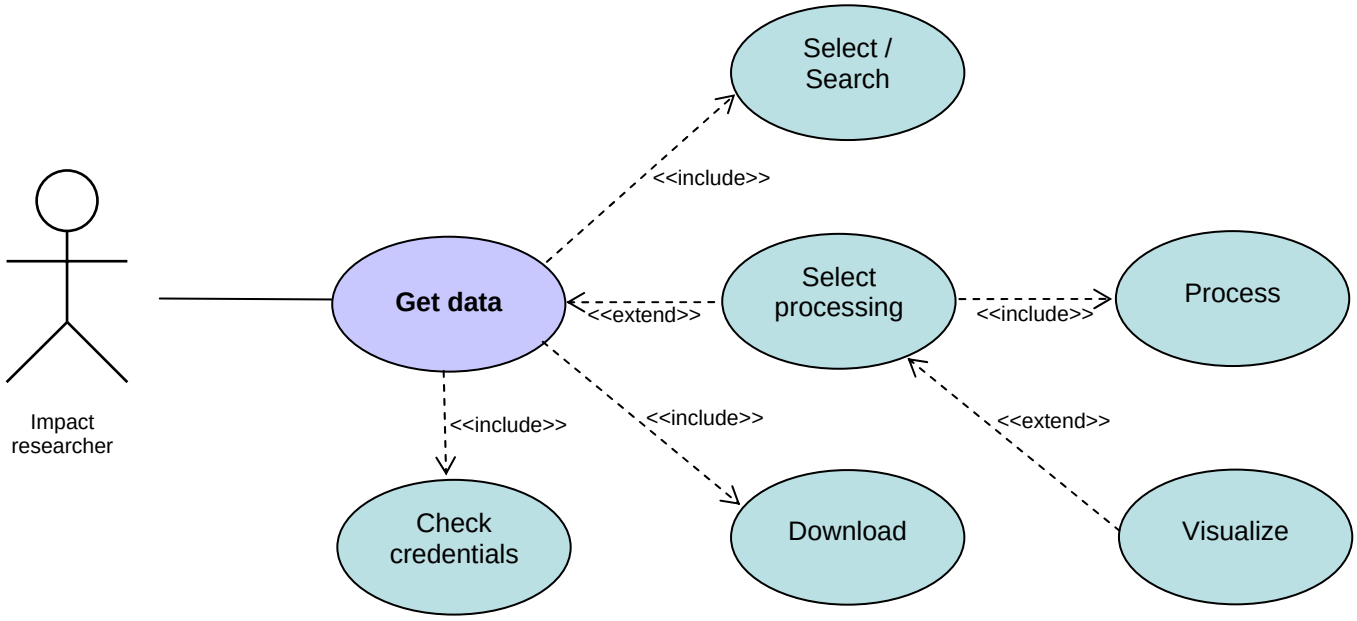

Figure 5: *A common use case for the action "get data", showing the functionality of the system with respect to the actor. This is a use case for searching, retrieving data and optionally processing and visualizing.*

Select data:

(1) Data selection should be as user friendly as possible: Area, time, (derived) parameter selection, realm (atmosphere, ocean), and emission scenario. Fields may be unspecified, and not all combinations are valid.
   a. AREA selection:
      i. Select a box on a map,
      ii. Select by country
      iii. Select by NUTS2 regions
      iv. Select by drainage basin (Elbe river, Rhine)
      v. Select by city
      vi. Select by individual points
   b. Usage of GeoNames database for searching

(2) A list with available data is presented

(3) Select or search the proper datasets from the list using:
   a. meta information
   b. viewing images, displaying quick looks of the data

(4) Go to step 1 and refine search

(5) Confirm the selected desired data

Note: For derived parameters, the datasets containing the necessary parameters will be returned. (Some derived parameters are not available in the source data)

Select processing:

Prerequisites: Data is selected

(1) Select one of the suitable data processors (e.g. calculate number of warmest days in summer)
(2) Configure the data processor by specific parameters (e.g. temperature)
(3) Add the configured data processor to the workflow
(4) Optionally go back to step 1 to add more data processors
(5) Confirm workflow

Process

Prerequisites: A workflow is defined

(1) The system should validate whether this workflow can be done on the system (Processing time, data volume, correct input)
(2) Provide the user with statistics about the workflow job (processing time and data volume)
(3) Based on these statistics the user can choose to submit, save, delete, edit the workflow
(4) The system executes the workflow job
(5) The system provides status of the workflow job (progress)
(6) Signals the user on completion (finished/error occurred)

Download

Prerequisites: Data has been selected or processing has been successfully finished

(1) The system provides an URL pointing to the data (through email also if processing takes longer than a specified time)
(2) The user uses the URL to download the data, using the desired download mechanism
(3) The system prune old temporary data files after defined pruning period

Visualize data

Prerequisites: Data has been selected or processing has been successfully finished

(1) The user selects the type of plot he wants (map, box plot, delta plot or timeseries)
(2) The system generates and displays the image

2.4. MODEL VERIFICATION USING THE BUSINESS USE CASES

The use cases provided in D11.1 are used to validate the problem domain model. For each of the use case scenarios we will check which parts can be implemented by the e-impact portal and which parts are to be solved in procedures or actions from climate scientists. The use cases as defined in D11.1 are considered as 'business use cases'. From the selected business use cases (D11.2) we can define the product use case: the work the e-impact portal can do in the business use case. [Mastering the requirements process, Robertson and Robertson, ISBN-10: 0-321-41949-9].

Clearly not all parts of the D11.1 use cases can or should be automated using the e-impact portal. This analysis shows what parts can be automated and in what parts the climate scientist has to provide expertise. The analysis also shows how the climate scientist can use the e-impact portal for parts of the process.

At each product use case a reference is made to the business use case defined in D11.1. Also at the product use case steps, the steps of the business case are mentioned to enable traceability. Note that at some use cases figures are used to describe the flow of information in the business use case. These figures also show the limitations for the e-impact portal in certain business use cases.

Within the use cases the following stakeholder acronyms are used:
ST: Stakeholder
DC: e-impact portal
CS: Climate Scientist
HS: hydrologist
GE: Geophysical Engineer

2.4.1. CERFACS (D11.1 Appendix D2)

Use case: Providing high-resolution climate change scenarios and evaluating uncertainties in the France Seine and Somme river basins

Purpose: Providing climate change scenarios for evaluating impacts on France Seine and Somme river basins, as well as evaluating changes in hydrological extremes and uncertainties using several downscaling methodologies, climate and hydrological models.

The business case of D11.1 was split into four product use cases:

1. UC Data request (step 1-5 business use case)
2. UC Data upload by the CS
3. UC Data extraction (step 8-10 business use case)
4. UC Publish report (step 14-15 business use case)

These four product use cases are also used to describe other business use cases.

Product use case success scenario for: Data Request:

1. ST: Visits site
2. ST: Fills request form with information needed for CS
3. DC: Passes the information to the CS
4. CS: Evaluates request and composes a study expert team
5. CS: Starts documentation on DC on the work to be done
6. CS: provides information back to the ST and starts interaction with ST until agreement on the data request is established.

Product use case success scenario for: Data upload:
Note: this step is optional since downscaled data could already be available on the portal data archive.

1. CS: Provides credentials
2. CS: Uploads downscaled data
3. DC: notifies upload complete

Product use case success scenario for Data Extraction

1. CS/ST: Selects downscaled dataset on DC
2. CS/ST: Selects data points of interest
3. CS/ST: Selects extraction processor
4. CS/ST: Selects ASCII formatting processor
5. CS/ST: Starts workflow
6. DC: Executes workflow
7. DC: Returns results
8. ST downloads the result data

Product use case success scenario for Publish Report

1. CS: provides credentials to DC
2. CS: Selects area on DC to place report
3. CS: Uploads report to DC to selected area on the DC
4. DC: Notifies upload report completed
5. CS: provides pointer to the report on the DC to ST

2.4.2. INHGA (D11.1 Appendix E)

Use case: Post-processing climate change scenarios and uncertainty evaluation in the Danube middle and lower basin.

Purpose: Use of climate model data by UMP_IS-ENES / INHGA for evaluating changes in the hydro meteorological extremes and uncertainties in the Danube basin using statistical downscaling methods

In the business use case the hydrologist (HS) wants downscaled data for the Donabe lower basis region for risk assessment. This business use case can be split over four product use cases:

1. UC Data request
2. UC Data extraction and correction
3. UC Data download
4. UC Data upload

As these product use cases are previously described, the business scenario will be described using the defined product use cases.

Business use case success scenario:

1. HS: Contacts CS using e-impact portal (UC Data request)
2. CS: Searches and downloads data, representing the requested area (UC Data extraction)
3. CS: Applies bias corrections to the data
4. CS +HS: Reveal extreme events from the data
5. CS: Applies bias corrections to the data
6. CS: Performs statistical Downscaling
7. CS: Uploads downscaled data to the e-impact portal (UC Data upload)

Note the steps 3-6 are not done using the e-impact portal.

2.4.3. CMCC-2 (D11.1 Appendix F)

Use case: Evaluation of the effects of climate changes on landslides phenomena triggered by precipitation events (clay strain)

Purpose: In the business use case, the Geophysical Engineer (GE) wants to evaluate the effect of climate changes on landslides phenomena in the AMRA/CMCC landslide framework triggered by precipitation events.

The business use case will be described using the previously defined product use cases where possible.

Business use case success scenario:

1. GE: Contacts CS using the e-impact portal (UC Data request)
2. CS: Prepares the dataset (UC Data extraction)
3. CS: Uploads dataset (UC Data upload)
4. CS: Notifies GE of the prepared dataset

The next steps of the business use case can be performed by the GE using the e-impact portal:

5. GE: Visit e-impact portal and search for the prepared dataset
6. GE: Selects period of interest
7. GE: Selects area of interest by using NUTS sub setting
8. GE: Selects averaging method for averaging grid points to produce a single value
9. GE: Submits workflow
10. DC: Executes workflow
11. DC: returns result and signals GE
12. GE: Downloads data (UC Download)

The uses of the finite elements model of GE is out of scope of the business use case.

2.4.4. Meteo France (D11.1 Appendix G2)

Use case: Scenarios for adaptation studies to climate change over the Loire Basin

Purpose: Providing climate change scenarios for vulnerability studies of human activities and environment to the effects of climate change on flood regimes and droughts in the Loire Basin.

The business case is described below. It was chosen to put this business case complete into this document to better explain how it can be supported by the designed e-impact portal.

Business use case success scenario:

1) Production by the climate research group of climate scenario datasets within the context of research projects: simulations of climate scenarios using global low resolution atmosphere/ocean/sea-ice climate models, downscaling of climate scenarios using a statistical method and a variable resolution climate model combined with a correction method.
2) Workshop associating climatologists and Loire Basin managers to exchange knowledge on climate change scenarios and selecting the scenarios and variables that will be provided in the context of the call of opportunity. The importance of uncertainty in the scenarios is re-assessed.
3) Data extraction (regional re-analysis and downscaled scenario data) over the domain of interest and calculation of complementary variables (evapotranspiration, minimum and maximum daily temperature, relative humidity).
4) Preparation of a documentation describing the data, the methods for data production, the data access methods, the uncertainty evaluation and recommendations for data use.
5) Data delivery by the climate research groups in a data repository located in a centre in contact with the selected research teams (provisional before the settlement of a web data portal).
6) Launch of the call of opportunity including documentation on the available climate change scenarios and selected of variables.
7) Selection of the research teams by the scientific committee of the program.
8) Workshop associating the climatologists, the Loire Basin managers and scientists from the selected research teams to present the research project and the data made available for the studies, along with recommendations for data use (uncertainty).
9) Data recovery by the selected research teams from the data repository with the associated documentation.
10) Support from the climatologists to the data users during the duration of their projects through e-mails, phone or meetings (practical use, scientific aspects), initiated by means of a unique e-mail contact point.

In Figure 6 "Event flow diagram for Meteo France G2 use case" the flow of information and the role of the e-impact portal in the business use case is shown.

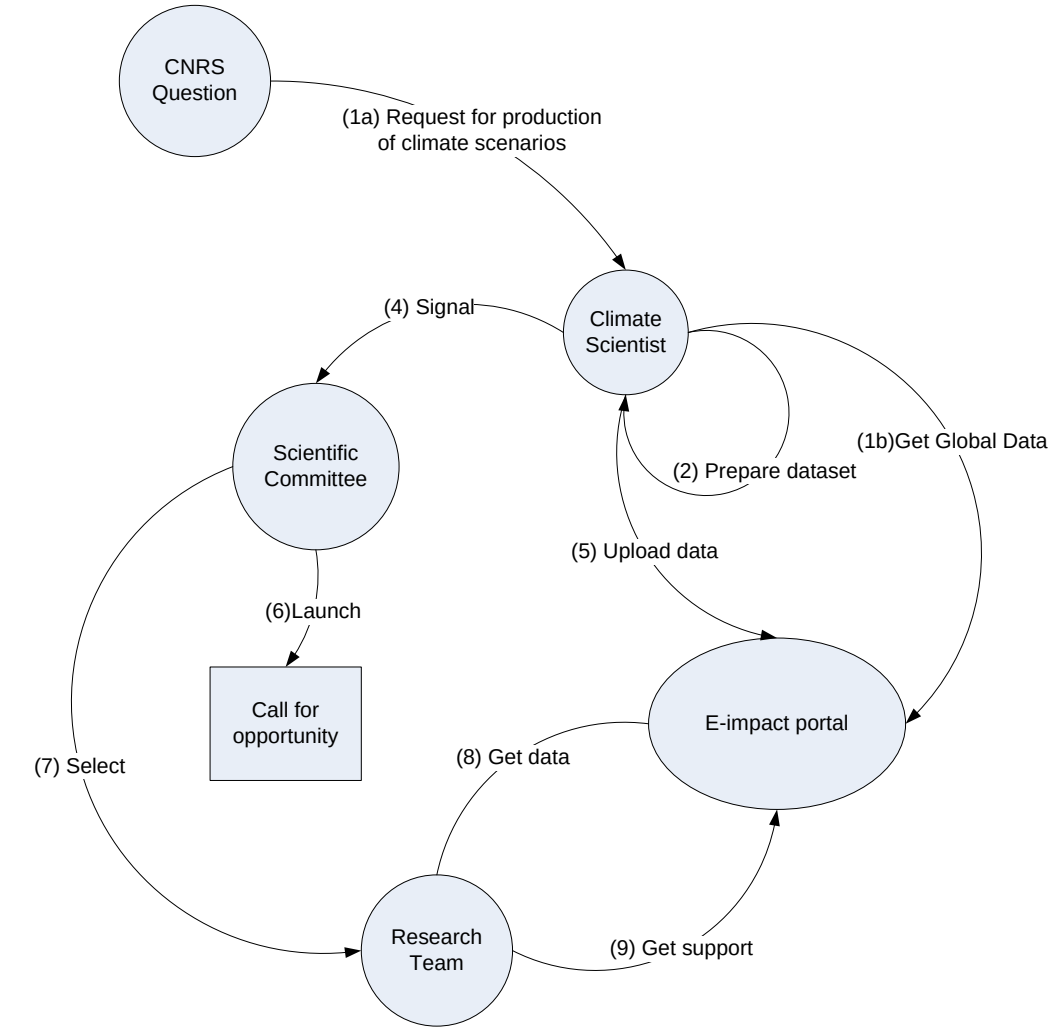

Figure 6 "Event flow diagram for Meteo France G2 use case"

As can be observed from the figure, the business case can be composed using previously defined product use cases:

- UC Download data in step 1b and step 8
- UC Data upload in step 5
- UC Data request in step 9

The figure shows the e-impact portal only plays a small role in the business use case.

2.4.5. WUR (D11.1 Appendix H)

Use case: Regional climate modelling using PRECIS

Purpose: Dynamical downscaling of climate scenarios using the Hadley Centre PRECIS model requires forcing data to run. Downscaled data are used as input for impact models and assessments on biodiversity and carbon stocks.

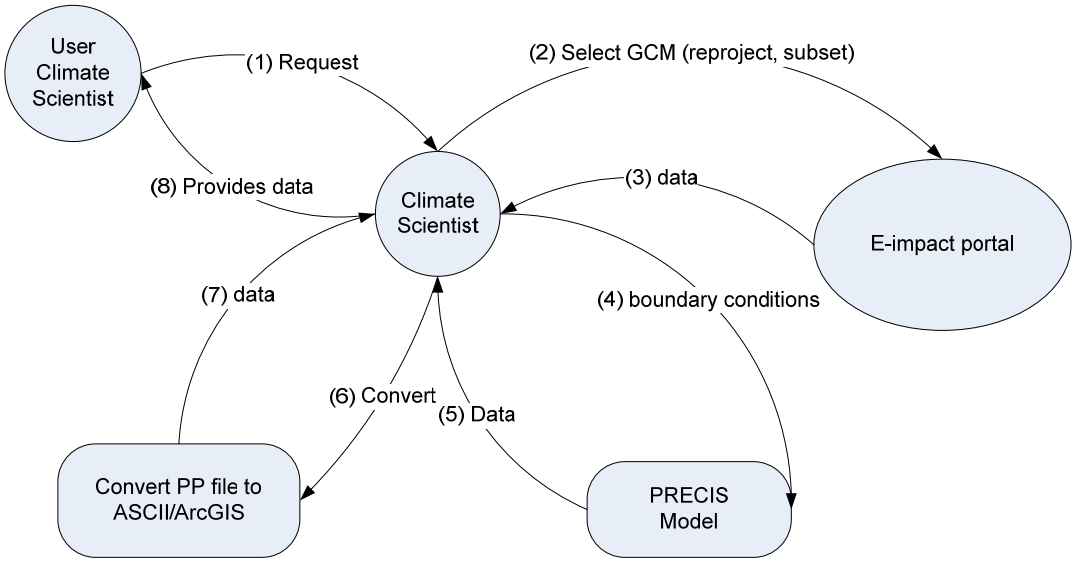

Figure 7 "Event flow diagram for WUR use case"

In Figure 7 the WUR business case is shown. In this business case the role of the e-impact portal is limited to assisting the WUR climate scientist getting the boundary conditions for the PRECIS regional downscaling model. The specific format conversion of PP files to ASCII or Arc GIS formats are out of scope for the e-impact portal prototype.

2.4.6. IPSL (D11.1 Appendix I)

Use case: Impact of global changes on freshwater fish biodiversity at large scale

Purpose: Developing a hydro-climatic-ecological (HCE) model-chain to link large scale climate data to local scale occurrence of aquatic biodiversity

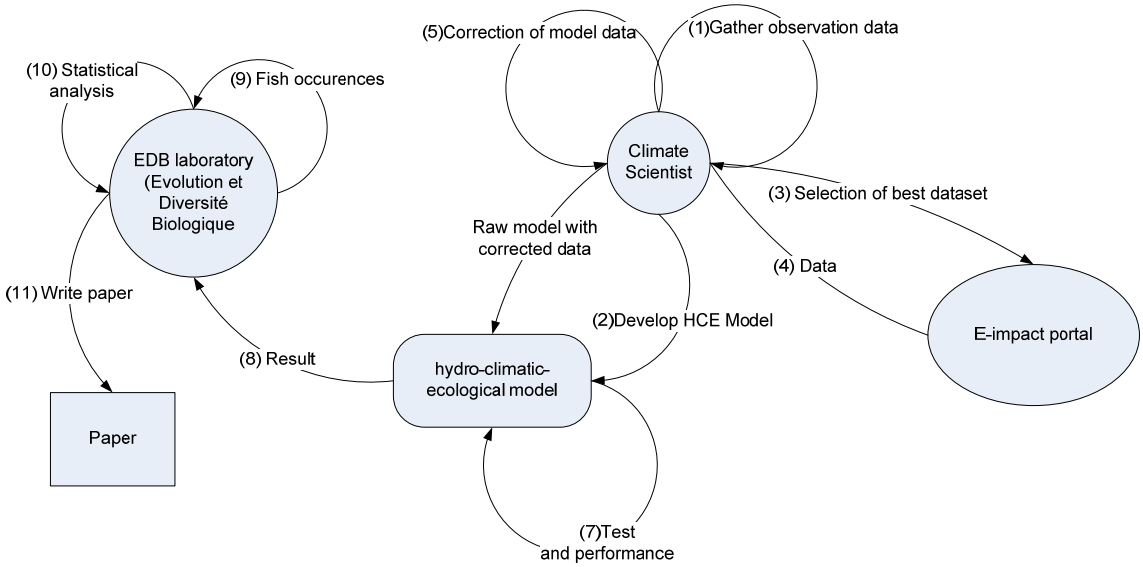

Figure 8 "Event flow diagram for IPSL I use case"

Figure 8 shows the event flow diagram of the HDE model chain development. The figure clearly shows a limited role for the e-impact portal in this specific use case: its role is limited to the GCM model data selection by the climate scientist. For this purpose the UC Download data can be used.

Not described in the use case is how the EDB contacts the climate scientist. For getting this contact the e-impact portal can be used, using the UC Data request product use case.

2.5. CONSTRAINTS FOR THE E-IMPACT PORTAL

- **Processing:**

In the prototype, the following action will not be supported:

- Dynamic downscaling is very computer intensive and some use cases will start directly from climate output of regional models.
- Statistical downscaling requires expert knowledge. Support cannot be done in an automated manner.

- **Processing limit**

The e-impact portal shall calculate the processing time and data storage it requires to execute a job, in order to prevent misuse of the e-impact portal's processing and storage resources and to prevent an overload of the systems capabilities.
The impact portal shall keep track of user activity in order to prioritize requests and be able to identify 'heavy users'.

- **Portal**

The e-impact portal will interface with the ESG-gateway and follow its interface requirements. The portal will also be a part of the ENES Portal and will follow its interface requirements and design choices (Plone, OpenID, Redmine). For metadata, the e-impact portal will use the CIM model as defined by METAFOR.

- **Data store**

The e-impact portal will connect to the ESG gateway for the climate data, and will also store pre-cooked data on local data storages at the partner sites. This distributed data store will be accessible using OPeNDAP.

3. SPECIFIC REQUIREMENTS

3.1. DEFINITIONS LIST

Besides the generic list of acronyms specified in Appendix A, the table below lists the definitions used in the requirements specified in the following chapters.

Definition	Explanation
7 main variables	The 7 main variables the e-impact portal shall provide are: precipitation, temperature, wind, radiation, humidity, pressure and orography.
ASCII format	American Standard Code for Information Interchange
CDF transformation	Cumulative Distribution Function transformation
CF	Climate and Forecast convention
CIM	Common Information Model - The FP7 METAFOR project develops this standard.
Climate indices	By their very nature, indices are simple, and combine many details into a generalized, overall description of the atmosphere or ocean which can be used to characterize the factors which impact the global climate system.
Climate parameters	Parameter- is a computation made from a population (rain free days, for example) it is a computation from data values recorded- but it is not actually a data value recorded from a subject
CMOR	Climate Model Output Rewriter
Common Tools	Tools as defined in appendix C of D11.2
CORDEX	A COordinated Regional climate Downscaling Experiment
downscaling	In numerical modeling, downscaling refers to techniques that take output from the model and add information at scales smaller than the grid spacing
e-impact portal	A central objective of JRA5 is to provide a prototype for a web service interface (the e-impact portal thereafter) to bridge the gap between the climate modelling community, the climate impact community and decision makers (the users or stakeholders thereafter) for developing adaptation and mitigation policies.
ESG Gateway	The Earth System Grid (ESG) integrates supercomputers with large-scale data and analysis servers located at numerous national labs and research centers to create a powerful environment for next generation climate research.
GRIB-1/2	Gridded Binary data format
ISO 8601	Numeric representation of dates and time
METAFOR	Common Metadata for Climate Modelling Digital repositories (http://ncascms.nerc.ac.uk/METAFOR/) - FP7 infrastructure project under ENES, which focuses on developing common standards for data and model information exchange that will be implemented in IS-ENES
Model data uncertainty information	Climate modelling output data holds uncertainties because of many sources, such as internal climate variability, model errors and biases, etc.
NetCDF-3 or 4	Network Common Data Format
NUTS regions /NUTS-2 regions	Nomenclature of Territorial Units for Statistics
OGC CSW	Catalogue Service for web
OGC services	webservices for geographical data compliant with standards defined by the Open Geospatial Consortium
OGC WMS	Web Map Service
OPeNDAP	Open Data Access Protocol
OpenID	OpenID is an open standard that describes how users can be authenticated in a decentralized manner, obviating the need for services to provide their own ad hoc systems and allowing users to consolidate their digital identities.

Pre-cooked data	Datasets derived from Climate Model data, containing the commonly used data (meteorological parameters) and data retrieved by downscaling
Quantile-quantile	In statistics, a Q-Q plot ("Q" stands for quantile) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. It is also known as a data correction method which implements a correction given each quantile in a probability density function.
Scenario	Coherent, internally consistent and plausible descriptions of possible future states of the world (IPCC, 1994)
Sectorial parameters	Climate parameters focused on a certain sector of impact users (e.g. agriculture)
STARDEX	Statistical and Regional dynamical Downscaling of Extremes for European regions

3.2. GENERAL REQUIREMENTS

Functional requirements

ID	*Stakeholder*	*Description*	*Importance*	*Trace*
GEN-19	JRA5	The user management system provides logging of user activity	High	
GEN-20	JRA5	The user management system provides user account management	High	
GEN-21	JRA5	The user management system provides self service facilities for users (passwords, registration)	High	
GEN-22	JRA5	The user management system provides monitoring of user activity	High	
GEN-23	JRA5	The e-impact portal shall be web based	High	
GEN-24	impact community	The e-impact portal will provide model data uncertainty information	High	
GEN-25	JRA5	The e-impact portal shall provide a data upload facility to upload pre-cooked data by JRA5 partners	High	
GEN-26	impact community	The 7 main variables the e-impact portal shall provide are: precipitation, temperature, wind, radiation, humidity, pressure and orography.	High	
GEN-27	impact community	The e-impact portal shall be able to process with the 7 main variables.	High	
GEN-28	impact community	The e-impact portal shall provide access to derived parameters (diagnosed and stored during climate simulations)	High	
GEN-29	impact community	The e-impact portal shall provide access to climate indices (calculated from the 7 main variables and/or derived parameters)	High	
GEN-30	impact community	The e-impact portal shall support the following temporal resolutions: hourly, 6 hourly, daily, monthly, yearly, decadal	High	
GEN-31	impact community	The e-impact portal shall support the regular lat lon projection (EPSG 4326)	High	
GEN-32	impact community	The e-impact portal shall support the Lambert projection	High	
GEN-33	impact community	The e-impact portal shall support the EPSG 28992 projection (rijksdriehoeks)	High	
GEN-34	impact community	The e-impact portal shall support Gaussian grid	High	
GEN-35	impact community	The e-impact portal shall support Gaussian reduced grid	High	
GEN-36	JRA5	The e-impact portal shall only provide data to authorized	High	

		users		
GEN-37	JRA5	The e-impact portal shall only give access to processing capabilities to authorized users	High	
GEN-38	impact community	The upload facility shall support FTP	High	
GEN-39	impact community	The e-impact portal shall provide documentation related to the products provided on the portal	Medium	
GEN-40	JRA5	The e-impact portal shall provide a request form for off line processing by climate scientists (e.g. downscaling)	Medium	
GEN-41	JRA5	The e-impact portal will provide basic operations on pre-cooked data	Medium	
GEN-42	impact community	The e-impact portal shall be flexible to support other temporal resolutions	Medium	
GEN-43	JRA5	The e-impact portal shall notify the user when using a sub-daily temporal resolution that this data should be used with care with suitable uncertainty assessment	Medium	

Non-functional requirements

ID	*Stakeholder*	*Description*	*Importance*	*Trace*
GEN-44	impact community	The e-impact portal shall be public accessible	High	
GEN-45	JRA5	For using the processing and data download facilities a user will login using his/her OpenID account	High	
GEN-46	impact community	Provided documentation shall be understandable for a large number of communities (The working group considered nine different domains: agriculture, forest / health / tourism / biodiversity / water / risk / transport infrastructure and building / energy / territories)	Medium	

Design constraints

ID	**Stakeholder**	**Description**	**Importance**	*Trace*
GEN-01	CIM	The e-impact portal shall use the following metadata description: METAFOR+CIM	High	
GEN-02	ESG-Gateway	The e-impact portal shall be able to retrieve data from the ESG gateway	High	
GEN-03	ESG-Gateway	The e-impact portal shall be able to authenticate users using OpenID.	High	
GEN-04	JRA5	The e-impact portal shall provide a user management system	High	
GEN-05	ENES Portal	The e-impact portal shall be integrated in the main ENES Portal	High	
GEN-06	JRA5	The use cases of D11.1 will provide the base for implementation of the e-impact portal	High	
GEN-07	JRA5	The e-impact portal will not provide downscaling services (neither statistical nor dynamic)	High	
~~GEN-08~~	~~JRA5~~	~~The use cases of D11.1 will provide the base for implementation of the e-impact portal~~ Removed: same requirement as GEN-06	~~High~~	GEN-06
~~GEN-09~~	~~JRA5~~	~~The e-impact portal shall only cover a limited set of use cases (D11.1)~~ Removed: is replaced by GEN-13 which is more specific	~~High~~	GEN-13

GEN-10	impact community	The e-impact portal shall at least provide access to the final climate parameters: liquid precipitation, mean temperature, minimal temperature, maximal temperature and short wave radiation, solid precipitation, wind (at 10 meters), IR radiation; relative humidity and potential evaporation.	High	
GEN-11	JRA5	The e-impact portal shall comply to the ISO 8601 date time standard	High	
GEN-12	JRA4	The e-impact portal shall make use of PLONE and REDMINE for implementation	High	
GEN-13	JRA5	The selected use cases of D11.1 will be implemented in the e-impact portal	High	
GEN-14	ESG-Gateway, JRA5,	The e-impact portal shall have a distributed data store	Medium	
GEN-15	JRA5	The e-impact portal should be flexibly implemented because it is a prototype (it will be improved with time), methods dataset and tools change rapidly and because the interaction with users about the climate questions is not well-known	Medium	
GEN-16	JRA5	The e-impact portal shall implement a limited number of selected tools [D11.2], based of the strength and weaknesses of each tool.	Medium	
GEN-17	JRA5	The e-impact portal shall use the defined 7 steps: collection, analysis and validation of the requests; definition of proceeding instructions; planning; data processing; quality assessment; data packaging and data release, where possible	Medium	
GEN-18	impact community	The e-impact portal will provide special interest to user feedback in order to improve knowledge on user needs, adapt the e-portal and improve the service.	Medium	
GEN-19	CORDEX	The e-impact portal shall be able to retrieve data from the CORDEX data portal	Medium	

3.3. DATA REQUIREMENTS

3.3.1. Data Discovery

Functional requirements

ID	Stakeholder	Description	Importance	Trace
DISC-01	METAFOR CIM portal	The portal shall interface with the METAFOR CIM portal.	High	
DISC-02	JRA5	OGC services offered by the impact portal shall be made visible by using OGC CSW services	High	
DISC-03	impact community	The portal provides selection options on different parameters, area, periods, horizontal resolution, and temporal resolution	High	
DISC-04	impact community	parameters provided include final climate parameters, sectorial parameters, auxiliary parameters, indices specifically developed to answer user needs	High	
DISC-05	impact community	Area include points, cities, rivers, countries, regions of all continents	medium	DISC-04
DISC-06	impact community	Periods include different time slices of the 20th and 21st centuries	medium	DISC-04
DISC-07	impact community	Horizontal resolution is global, regional and local resolution	medium	DISC-04

DISC-08	impact community	Temporal resolution ranges from hourly time step to decadal one	High	GEN-30, DISC-04
DISC-09	CORDEX	The portal shall interface with the CORDEX gateway data discovery service.	High	

Non-functional requirements
Not applicable

Design constraint

Not applicable

3.3.2. Data Visualization

Functional requirements

ID	Stakeholder	Description	Importance	Trace
VIS-01	JRA5	The portal shall visualize climate model data in the form of 2D maps using an OGC WMS service.	medium	
VIS-02	impact community	The portal shall be able to generate timeseries, delta plots and box plot graphs	high	

Non-functional requirements
Not applicable

Design constraints
Not applicable

3.3.3. Data Download

Functional requirements

ID	Stakeholder	Description	Importance	Trace
DOW-02	JRA4	Data shall be accessible in it's original form by using an OPENDAP service	High	
DOW-03	impact community	Data shall be accessible by using HTTP	High	
DOW-04	impact community	Data shall be accessible by using FTP	High	
DOW-05	impact community	Data can be provided in ASCII format by the e-impact portal	High	
DOW-06	impact community	Data can be provided in NetCDF 3 or 4 by the e-impact portal	High	
DOW-07	JRA5	Data shall be accessible in the form of graphical images by using an OGC WMS service	Medium	
DOW-08	JRA5	Data shall be accessible in the form of geographical referenced data in GML format by using an OGC WCS service	Medium	
DOW-09	impact community	Data can be provided in GRIB1 and GRIB2 data format by the e-impact portal	Medium	
DOW-10	JRA5	At download, the user will receive: 1) guidance for using data 2) recommendations depending on data quality and uncertainties	Medium	

Non-functional requirements
Not applicable

Design constraints

ID	Stakeholder	Description	Importance	Trace

DOW-01	JRA4	Delivered data should adhere to the CF and CMOR conventions	Medium	

3.3.4. Data Processing

Functional requirements

ID	Stakeholder	Description	Importance	Trace
~~PROC-01~~	~~JRA5~~	~~The e-impact portal shall use the Common Tools as defined in D11.2 [R2] to provide data processing capabilities~~ Removed: is design constraint and mentioned there	~~High~~	
PROC-02	impact community	Data shall be extractable by time period, geographical coverage, parameter and scenario	High	
PROC-03	impact community	Data shall be extractable at requested point coordinates, modifying the dataset into a point dataset.	High	
PROC-04	impact community	The e-impact portal shall provide the option to do the operation Averaging on the data	High	
PROC-05	impact community	The e-impact portal shall be possible to re-project the data to other projections (as defined in Generic requirements)	High	
PROC-06	impact community	The e-impact portal provides functionality to do Regridding from Gaussian reduced grids to regular grids	High	
PROC-07	impact community	The e-impact portal provides functionality to calculate indices based on the ECA&D system's indices list	High	
PROC-08	impact community	The e-impact portal provides functionality to compare different scenarios/ensembles	High	
PROC-09	impact community	The processing system shall provide workflow support	High	
PROC-10	impact community	The e-impact portal shall support the data extraction method: region/point selection	High	
PROC-11	impact community	The e-impact portal shall support the data extraction method: period selection	High	
PROC-12	impact community	The e-impact portal shall support the data extraction method: parameter selection	High	
PROC-13	impact community	The e-impact portal shall support the interpolation method: Nearest neighbour	High	
PROC-14	impact community	The e-impact portal shall support the Common data manipulation method: mean	High	
PROC-15	impact community	The e-impact portal shall support the Common data manipulation method: max	High	
PROC-16	impact community	The e-impact portal shall support the Common data manipulation method: min	High	
PROC-17	impact community	The e-impact portal shall support indices calculations	High	
PROC-18	impact community	The indices calculations in the e-impact portal shall be flexible to incorporate tailored indices calculations	High	
PROC-19	impact community	The e-impact portal shall provide statistical analysis methods for comparison of model results	High	
PROC-20	impact community	The e-impact portal shall support the data correction method: Quantile-quantile correction	Low	
PROC-21	impact community	The e-impact portal shall support the data correction method: CDF transformation	Low	
PROC-22	impact community	Data shall be extractable according to a geographical mask like river basins or NUTS-2 regions.	Medium	
PROC-23	impact community	It shall be possible to do subsetting on the data	Medium	
PROC-24	impact	The e-impact portal shall provide the option to do the	Medium	

	community	operation Aggregation in time on the data		
PROC-25	impact community	The e-impact portal shall provide the option to reformat the output of ASCII files, by reducing the number of digits.	Medium	
PROC-26	impact community	The e-impact portal provides functionality to generate boxplots of datasets	Medium	
PROC-27	impact community	The e-impact portal provides functionality to generate timeseries plots of the datasets	Medium	
PROC-28	impact community	The e-impact portal provides functionality to generate delta plots of the datasets	Medium	
PROC-29	impact community	The e-impact portal shall provide support to integrate scripts provided by the climate scientist (e.g. R scripts)	Medium	
PROC-30	impact community	The e-impact portal shall support the data extraction method: according to NUTS regions	Medium	
PROC-31	impact community	The e-impact portal shall support the interpolation method: bilinear	Medium	
PROC-32	impact community	The e-impact portal shall support the interpolation method: bi-cubic	Medium	
PROC-33	impact community	The e-impact portal shall support the interpolation method: spline	Medium	
PROC-34	impact community	The e-impact portal shall support the data correction method Mean correction	Medium	
PROC-35	impact community	For indices calculations the e-impact portal shall use the STARDEX list	Medium	
PROC-36	impact community	The e-impact portal shall provide the statistical analysis method: Significativity test (kendall, kolmogorov-smirnov)	Medium	
PROC-37	impact community	The e-impact portal shall provide statistical analysis methods for trend analysis	Medium	
PROC-38	impact community	The e-impact portal shall provide statistical analysis methods for extreme events analysis	Medium	
PROC-39	impact community	The e-impact portal shall provide statistical analysis methods for validation	Medium	
PROC-40	impact community	The e-impact portal shall support batch processing for long running jobs	Medium	
PROC-41	impact community	The e-impact portal shall provide aggregation functionality	Medium	
PROC-42	impact community	The e-impact portal shall provide user feedback on data processing errors	High	

Non-functional requirements
Not applicable

Design constraints

ID	Stakeholder	Description	Importance	Trace
PROC-01	JRA5	The e-impact portal shall use the Common Tools (reference) to provide data processing capabilities	High	
PROC-42	impact community	The e-impact portal shall provide user feedback on data processing errors	High	

3.4. DOCUMENTATION REQUIREMENTS

Functional requirements

ID	Stakeholder	Description	Importance	Trace
DOC-01	impact community	The e-impact portal shall provide information on climate system and climate modelling	high	

DOC-02	impact community	The e-impact portal shall provide information on products: dataset, area, period, resolution parameters, indices, analysis, visualization	high	
DOC-03	impact community	The e-impact portal shall provide information on methods and tools to process data	high	
DOC-04	JRA5	The e-impact portal shall provide information concerning job status and completion	high	
DOC-05	impact community	The e-impact portal shall provide guidance for data use and vulnerability/risk assessment	high	
DOC-06	impact community	The e-impact portal shall provide an electronic help desk	high	
DOC-07	impact community	The e-impact portal shall provide contact points to get specific expertise from climate experts	high	
DOC-08	impact community	The e-impact portal shall provide general documentation on: • climate impact studies, uncertainties, models • link between model runs from available global climate projections and national ones. • models (see CIM developed in METAFOR) methods available for data processing, packaging and data analysis • data processing tools including function/functionality/use information • data packaging tools • data delivering way • uncertainty types and methods to deal with. • limitations on usability of the data	high	
DOC-09	impact community	The e-impact portal shall provide a list of do's and dont's for using climate data, along with a FAQ	high	
DOC-10	JRA5	The e-impact portal provides Common Tools documentation	high	
DOC-11	JRA5	The e-impact portal shall provide workflow examples	low	
DOC-12	JRA5	The e-impact portal shall provide technical information like data volume, delivering methods/tools, format	medium	
DOC-13	impact community	The e-impact portal shall provide uncertainties, quality and limits of the final products	medium	
DOC-14	impact community	The e-impact portal shall provide use case descriptions as illustration for inexperienced users	medium	
DOC-15	JRA5	The e-impact portal shall provide advices on how to make workflows	medium	
DOC-16	impact community	Provided documentation shall be understandable for a large number of communities (The working group considered nine different domains: agriculture, forest / health / tourism / biodiversity / water / risk / transport infrastructure and building / energy / territories)	medium	

Non-functional requirements
Not applicable

Design constraints
Not applicable

4. SYSTEM DESIGN

4.1. DESIGN METHOD

The Unified Modeling Language (UML) has been used to make the conceptual design of the e-impact portal. In the following chapter four diagram types are used and explained for describing the e-impact portal: the component decomposition (non-UML), a class diagram, sequence diagrams and a package diagram.

The component decomposition describes the logical structure of the e-impact portal. The class diagram is the main building block in object oriented modeling. It is used both for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. The classes in a class diagram represent both the main objects and or interactions in the application and the objects to be programmed. The sequence diagram shows different processes/objects that live simultaneously and the messages that are exchanged between them. The package diagram shows the organization of the classes into packages.

4.2. DECOMPOSITION DESCRIPTION

The decomposition was already shown in Figure 4: ***e-impact portal decomposition into components.*** Figure **Erreur ! Source du renvoi introuvable.**9 shows the main components only.

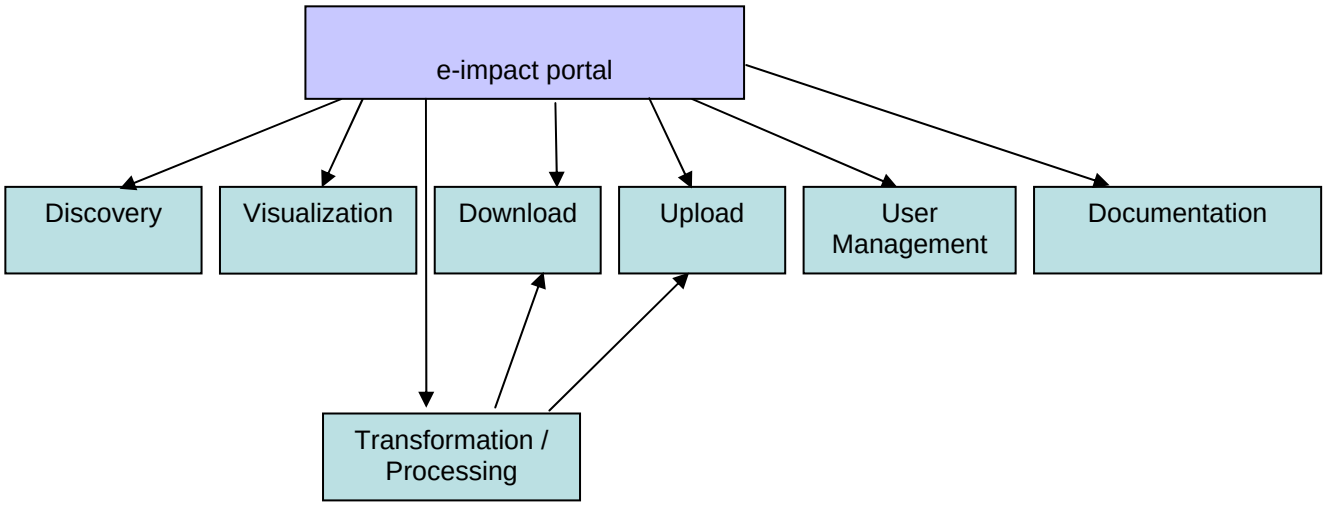

Figure 9 e-impact portal component decomposition

The main components of the e-impact portal are:

- Data Discovery
- Data Visualization
- Data Download
- Data Upload
- Data Transformation/Processing
- User Management
- Documentation

The envisaged components are described in detail in chapter 2.3.1.

4.3. UML CLASS DIAGRAM OF THE E-IMPACT PORTAL

The decomposition description is provided by a class diagram which models the structure of the e-impact portal and by a sequence diagram which models the behavior of the e-impact portal.

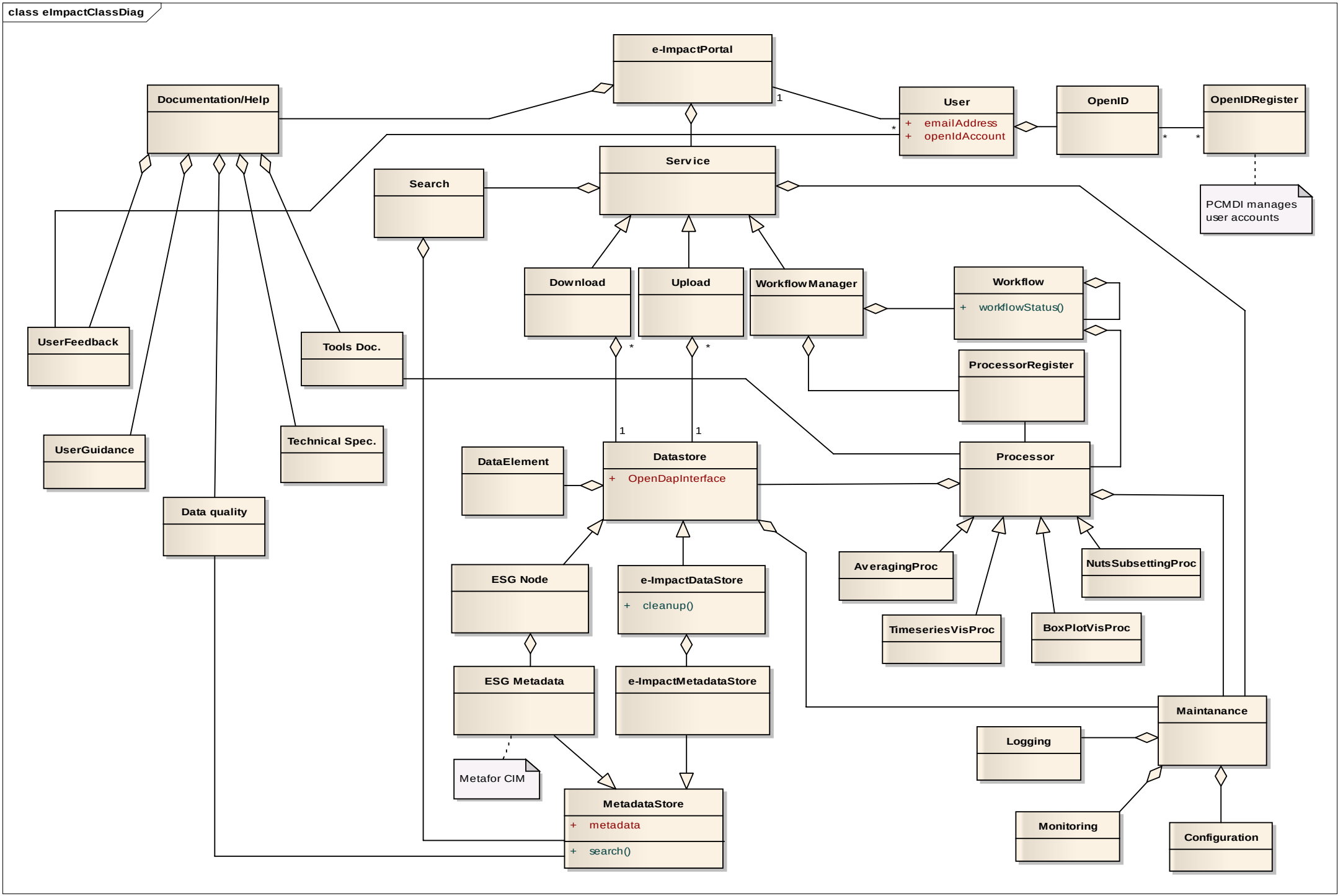

Figure 10 UML Class diagram of e-impact portal

Class descriptions:

AveragingProc
The averaging processor is a Processor which calculates averages of data in space or time. This can for example be used to generate timeseries data of global averaged temperature. In this case the input data is selected as a series of 2D fields, while the output data is a series of single values.

BoxplotVisProc
The box plot visualization Processor generates a box plot visualization in the form of an image. The input data can be any collection of data. The Processor shall be configurable in such a way that multiple box plots over a time period can be processed.

Configuration
Configuration is an aggregation of Maintenance and holds settings configured by the Maintainer. Configuration is usually represented by configuration files / database settings.

Data quality
Data quality is a specialization of documentation and holds information about the quality of data, how quality of data can be monitored/ checked. Data quality has an association with the MetaDataStore, the MetaDataStore provides information about requested data, including data quality.

DataStore
The DataStore is the source of the data and it keeps all the data. The DataStore provides an interface to the data which allows for selection of variables/parameters and the selection of this data along dimension axes. The DataStore will basically have the Common Data Model and is implemented using OPeNDAP.

Documentation/Help
Documentation/Help is the part of the e-impact Portal where the user can find documentation topics like the common tools description, technical specifications, data quality, guidance and get feedback from climate scientists. Documentation can be implemented by using wiki pages or using a CMS. Certain documentation parts have links with the services, as they provide information about the data, data search and data quality.

ESG Metadata
ESG Metadata relies heavily on METAFOR/CIM as these actually develop the metadata standard for climate model data. For the e-impact portal ESG Metadata is a specialization of MetaDataStore.

ESG Node
The ESG Node is a DataStore which contains the initial bulk of climate model data. The e-impactPortal uses this DataStore as its data source, all initial data is retrieved from this data store. The ESG Node is a specialization of DataStore.

Logging
Logging is aggregated by Maintenance and provides the general mechanism to keep track of the different e-impact portal statuses, activity and running processes.

Maintenance
Maintenance aggregates Logging, Monitoring and Configuration. Maintenance is used by DataStore, Service and Processor to provide general support for these functionalities.

MetaDataStore
MetaDataStore provides an endpoint for Search to provide search functionality. The MetaDataStore executes the actual store and keeps a database with information about the data. MetaDataStore specializes MetaDataStores like ESG Metadata and e-impactMetadataStore.

Monitoring
Monitoring is aggregated by Maintenance and provides support for several classes for monitoring. Monitoring

signals the system administrator when a problem or malfunction occurs. Monitoring provides also support to keep track of statistics, e.g. number of visits, what users do on the e-impact portal,

NutSubsettingProc
NutSubsettingProc is a processor which enables the selection of data according to specified regions, the NUTS regions. NutSubsettingProc is a specialization of Processor.

OpenID
OpenID is the digital identity of the User.

OpenIDRegister
OpenIDRegister keeps track of all registered and authorized OpenID accounts of the e-impact portal. This will probably be managed by PCMDI.

Processor
Processor is the generalization class of all processors in the e-impact portal. These processors can be data processing processors or visualization processors. Processor is controlled by Workflow and takes it data from DataStore. All available processors are registered in the ProcessorRegister.

ProcessorRegister
ProcessorRegister is a list of all available Processors.

Search
Search provides a mechanism for the user where he can search information about available data by using search terms and or keywords. Because the data is ordered according to multiple classifications the search should be faceted. Search itself does not execute the process of searching in a database, it delegates commands to the MetaDataStore which has support to do the actual searching through the data. Search connects to MetaDataStore and delegates the searching to the store.

Service
Service is a class which specializes the classes Download, Upload and WorkflowManager. It provides a common interface to these classes. Service aggregates search, and so do download, upload and the workflowmanager.

Technical Specifications
Technical Specifications is a specialization of documentation. It documents standards and interfaces like Metafor/CIM, ESG nodes, data formats.

TimeseriesVisProc
The timeseries visualization processor can generate a timeseries plot in the form of an image. For example an image of globally averaged temperature in the period 1981-2010 can be generated with this Processor. The input data must be a set of values over a time period which can be generated by the AveragingProcessor.

Tools Doc.
Tools Documentations provide information and search functionality of the common tools. It enables searching tools for certain criteria. For example a user has to be able to search what tools are available for the processing he wants to accomplish.

Upload
Upload is a specialization of Service and provides functionality for the Climate Scientists to store customized data specifically intended for the impact community. It enables Climate Scientists to store output of their downscaling methods and data which does not belong to the ESG nodes.

User
User is the e-impact portal User. The user is always free to browse through the documentation, but the user will need a valid OpenID account when he/she wants to search, download or process data.

UserFeedBack

UserFeedBack is a specialization of Documentation and provides contact forms and tracking mechanisms for Impact Users who have questions about climate topics. It should also provide functionality for Impact Users and Climate Scientists to come to new initiatives about climate data. When a specific dataset is required for a specific topic which requires expert knowledge, the e-impactPortal can help with the UserFeedBack mechanism to be a point of contact.

UserGuidance
UserGuidance is a specialization of Documentation and provides information about the Do's and Don't that a impact user can do with climate modelling data. It also contains the FAQ, information about used scenarios, etc.

Workflow
A workflow is an ordered sequence of Processors defined by the User in the WorkflowManager. In a workflow one or several Processors can be chained. For example, a timeseries plot can be generated by using the averaging Processor and timeseries visualization Processor together in a single workflow. When for a certain job just one Processor is needed, a Workflow still needs to be defined.

WorkflowManager
The workflow manager allows the user to specify workflows and keeps track of the running workflows.

e-impactDataStore
The e-impactDataStore is the DataStore owned by the e-impactPortal. It enables Climate Scientists and Processors to store their custom data.

e-impactMetadataStore
The e-impactDataStore is the MetaDataStore owned by the e-impactPortal.

e-impactPortal
The e-impact portal is the central access point for the impact user. It consists of two main parts, the service part and the documentation part.

4.4. UML SEQUENCE DIAGRAMS

Sequence diagrams show how objects communicate with each other in terms of a sequence of messages. They also indicate the lifespan of objects relative to those messages.

4.4.1. UML sequence diagram of use case "process data"

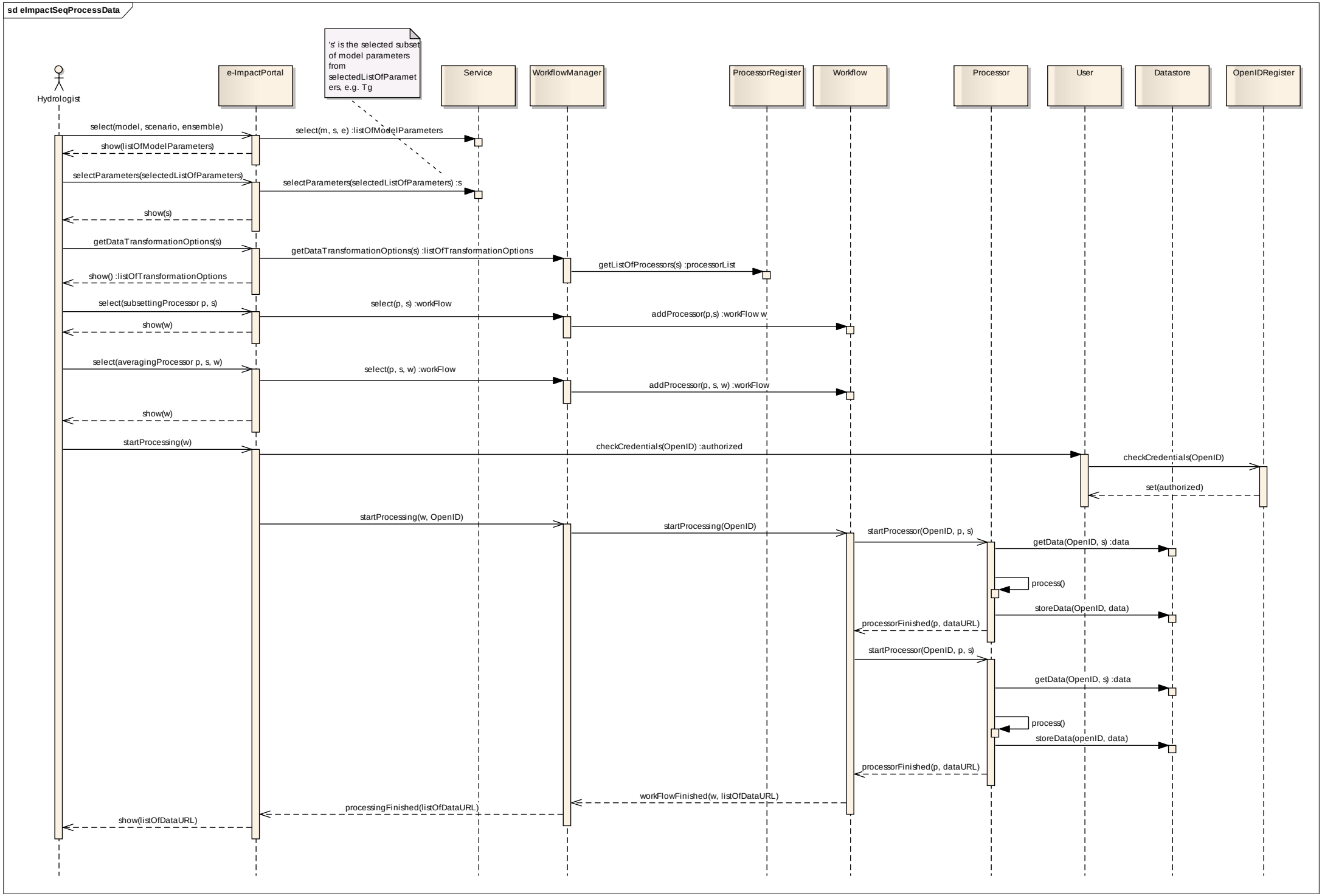

Figure 11 UML sequence diagram 'process data'

Erreur ! Source du renvoi introuvable. shows the sequence of messages for a use case where a user first selects data (e.g. Tg) from a GCM model, and processes this in a workflow where Tg is first sub-setted and then averaged in order to obtain a global averaged timeseries. After processing is finished, the user receives a URL to the result dataset.

4.4.2. UML sequence diagram of use case “upload data”

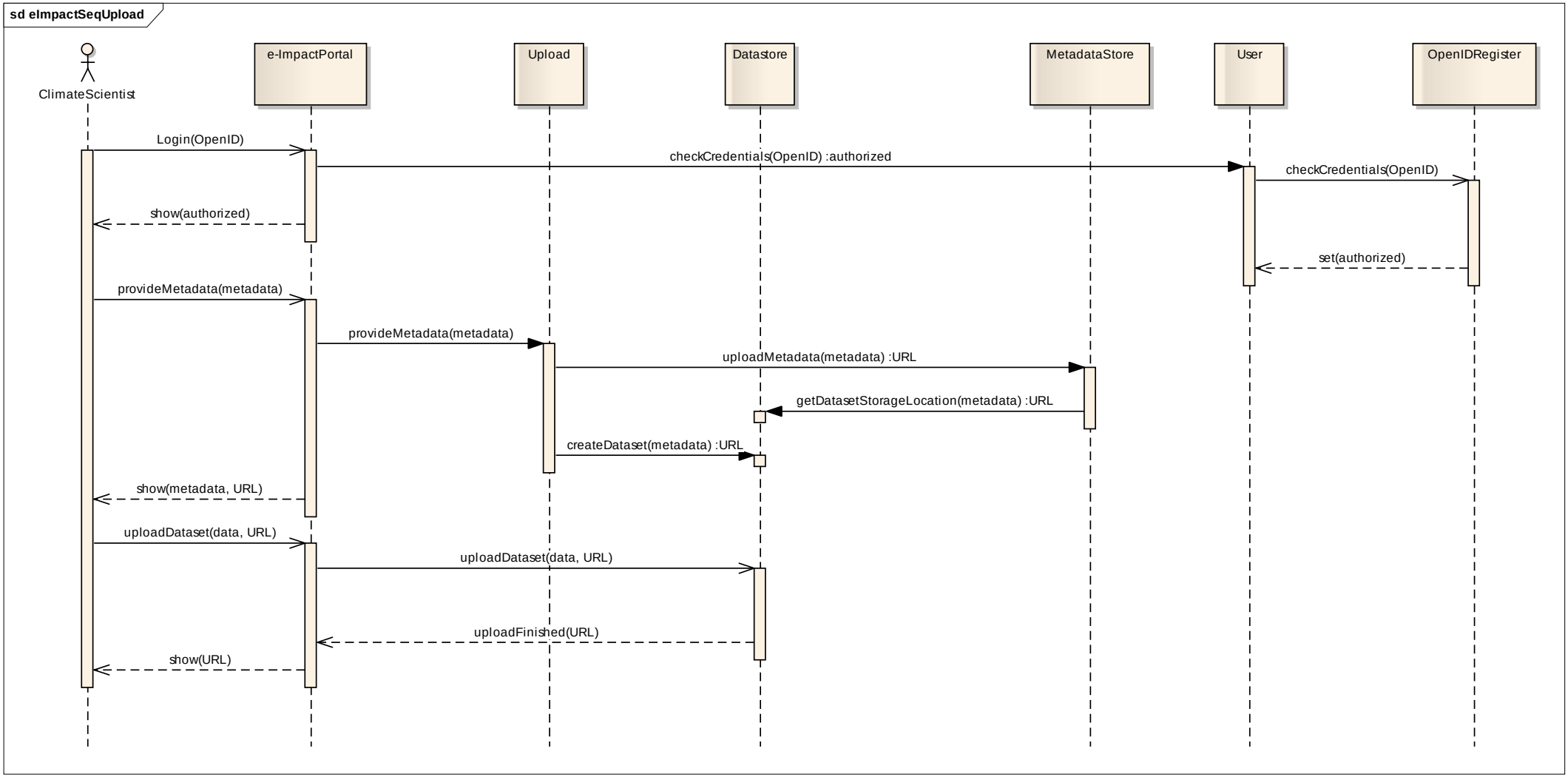

Figure 12 "UML sequence diagram 'upload data' "

Erreur ! Source du renvoi introuvable. shows the sequence of messages for dataset upload. First the climate scientist needs to provide credentials. Then the metadata of the dataset must be provided. A URL for uploading the data is returned as a location to upload the data. When uploading is finished, the e-impact portal will show the URL where the uploaded data (and metadata) are located.

4.5. PACKAGE DIAGRAM E-IMPACT PORTAL

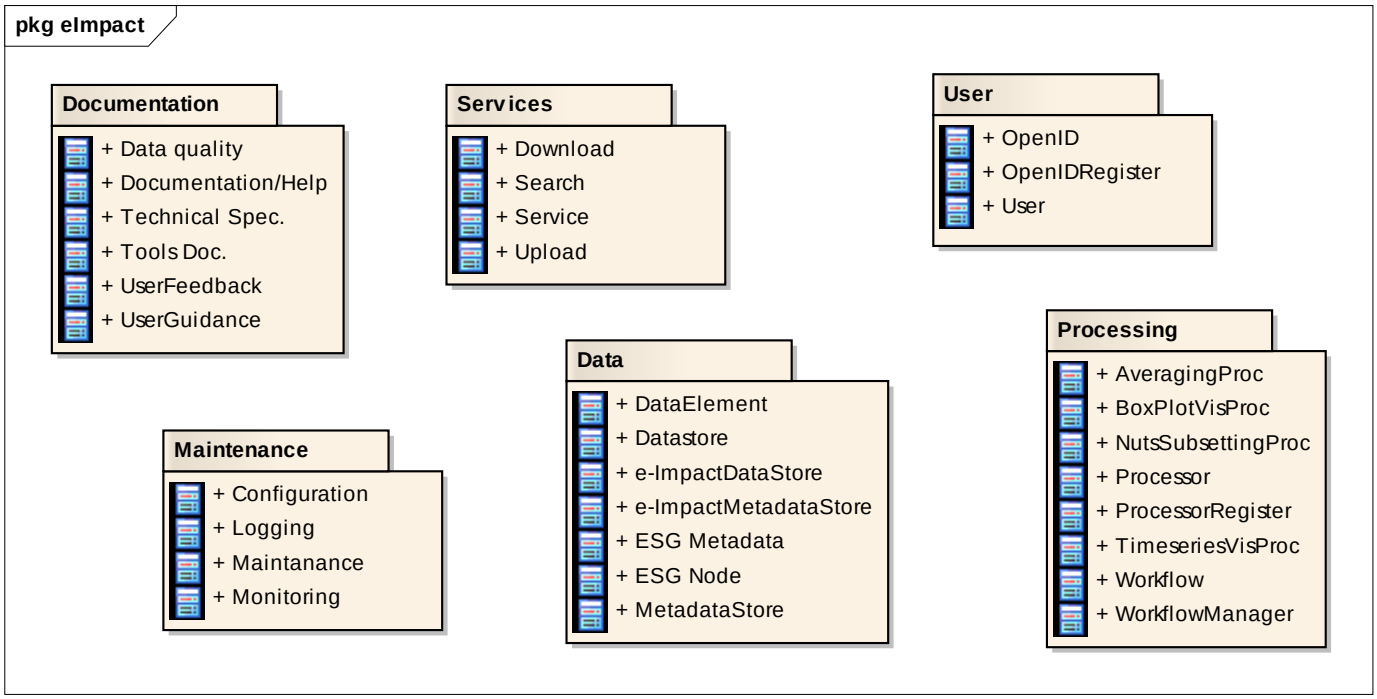

Figure 13 "Package diagram e-impact portal"

The packages diagram describes the logical organization of classes into packages. The interfaces between these packages will be defined in the detailed design in the next phase of the e-impact portal development.

5. FEASIBILITY AND RESOURCE ESTIMATES

The e-impact portal prototype will consist of 4 main components

1. e-impact portal server, containing the e-impact software and documentation
2. dataset storage, ESG nodes and local e-impact data storage
3. metadata storage. ESG catalogue and e-impact catalogue
4. processing server, e-impact workflow processing

For the e-impact portal the following resources are available:

- Data storage: 8 TB storage at KNMI
- Data storage: 1 TB storage at CERFACS
- Server: 12 Intel core 24 Gb memory HP blade server at KNMI for hosting the e-impact portal and processing services
- Server: 4 cores AMD Opteron 2.4 GHz 32 Gb memory at CERFACS for hosting final version of e-impact portal prototype

Note the e-impact storage will only store pre-cooked data, processed data and intermediate data results. The GCM CMIP-5 and CORDEX data is stored at the ESG nodes.

Note the documentation part of the e-impact portal will be integral part of the ENES Portal and stored in its CMS system.

For each of the components we will list the feasibility and remaining issues to be solved in the following chapters.

5.1. E-IMPACT PORTAL SERVER:

Integration of Impact Portal and ENES Portal:
The ENES Portal has a Content Managed part (CMS) for documentation and a dynamical part for functionality. The CMS is achieved by using Plone, the dynamical part is achieved by using Pylons and other techniques. JRA4 discourages the use of Plone for a dynamical website (JavaScript, AJAX, etc...). This division of documentation and functionality can also be applied for the impact portal: Use the Plone CMS for the documentation part, and use more dynamical techniques for the functional/data part (ExtJS for example). The dynamical part can than be integrated in Plone by using a HTML iframe.

Security:
ESG OPeNDAP authorization is achieved by using PKI X.509. MyProxy can be used to delegate a credential. How this exactly works needs to be discussed with JRA4. JRA4 managed to get PKI X.509 working with pydap. Pydap is currently being used in the Web Processing Service of the impact portal.

5.2. DATA STORAGE:

Downscaled data / regional climate model data
CORDEX core data will probably be made available on ESG. The ESG infrastructure does not depend on CMOR2, so CMOR2 is not a requirement for CORDEX. CORDEX is CF compliant. It is not yet known whether the data-quality checks from CMIP5 can be applied on CORDEX data, because CORDEX is not CMOR compliant.

The current plan is to put CORDEX in the form of tar zipped NetCDF3 files on ESG. If this plan is carried out, OPeNDAP cannot be enabled for this dataset. The ESG software itself has probably problems with serving tar zipped NetCDF data. JRA4 will look for a solution for this problem. A few solutions have been put on the table: (1) Convert the NetCDF3 data
to compressed NetCDF4, or (2) serving the 7 main variables as unzipped NetCDF3 (for JRA5). CORDEX data will also be available in the form of interpolated regular grids.

Data search:
JRA4 has made a THREDDS catalog crawler which harvests THREDDS catalogs available on the ESG nodes. The results are put in a PostGRES database, which can be queried from a web interface.

There is currently no plan for a search webservice interface for external portals, like the impact portal from JRA5. JRA4 will investigate whether it is feasible to make a search webservice interface available. An alternative is that the impact portal makes use of the
HTML search form from the ENES Portal directly. Plans are to add links to the search results from the ENES Portal which link directly to the impact portal.

OPeNDAP:
OPeNDAP is currently not enabled on every ESG by default. JRA4 will look in to this and will make sure that OPeNDAP will be enabled. There is a small possibility that some issues may arise when OPeNDAP is enabled for all datasets, this will be investigated by JRA4.

5.3. METADATA STORAGE

MetaData search:
The METAFOR CIM portal will provide a search web service to find CIM instances. The CIM is a Common Information Model which describes – in our case - models and experiments in detail. There has been a discussion on how we can find CIM instances corresponding to datasets directly by using DOI's (Digital Object Identifiers).

5.4. PROCESSING SERVER

COWS WPS for Multi model ensemble comparison: It is worth noting that JRA4 will setup a Web Processing Service based on the COWS WPS framework to make it possible to do multi-model ensemble comparisons. This setup might be co developed or reused in JRA5.

6. USER REQUIREMENTS VS SOFTWARE REQUIREMENTS TRACEABILITY MATRIX

The Software requirements stated in the document are derived from:

- The final description of selected use cases including user requirements specification (D11.1)
- The JRA5 e-impact portal design workshop meeting and design workshop minutes
- The baseline document on e-resources/tools and transverse themes (D11.2)

As there is no numbered list of user requirements to trace the software no matrix will be provided in this chapter.

7. SOFTWARE REQUIREMENTS VS COMPONENTS TRACEABILITY MATRIX

The matrix below enables the tracing of the software requirements to the envisaged components.

Component	Requirements
Data Discovery	DISC-01, DISC-02, DISC-03, DISC-04, DISC-05, DISC-06, DISC-07, DISC-08, GEN-26, GEN-30. GEN-31, GEN-32, GEN-33, GEN-34-, GEN-35, GEN-36, GEN-37, GEN-41, GEN-42, GEN-43, GEN-44
Data Visualization	VIS01, VIS02, GEN-30. GEN-31, GEN-32, GEN-33, GEN-34-, GEN-35, GEN-36, GEN-37, GEN-41, GEN-42, GEN-43, GEN-44
Data Download	DOW-01, DOW-02, DOW-03, DOW-04, DOW-05, DOW-06, DOW-07, DOW-08, DOW-09, DOW-10, GEN-28, GEN-29, GEN-30. GEN-31, GEN-32, GEN-33, GEN-34-, GEN-35, GEN-36, GEN-37, GEN-41, GEN-42, GEN-43, GEN-44, GEN-45
Data Processing	PROC-01, PROC-02, PROC-03, PROC-04, PROC-05, PROC-06, PROC-07, PROC-08, PROC-09, PROC-10, PROC-11, PROC-12, PROC-13, PROC-14, PROC-15, PROC-16, PROC-17, PROC-18, PROC-19, PROC-20, PROC-21, PROC-22, PROC-23, PROC-24, PROC-25, PROC-26, PROC-27, PROC-28, PROC-29, PROC-30, PROC-31, PROC-32, PROC-33, PROC-34, PROC-35, PROC-36, PROC-37, PROC-38, PROC-39, PROC-40, PROC-41, GEN-27, GEN-30. GEN-31, GEN-32, GEN-33, GEN-34-, GEN-35, GEN-36, GEN-37, GEN-40, GEN-41, GEN-43, GEN-44, GEN-45
Data Upload	GEN-25, GEN-38, GEN-44
Documentation	DOC-01, DOC-02, DOC-03, DOC-04, DOC-05, DOC-06, DOC-07, DOC-08, DOC-09, DOC-10, DOC-11, DOC-12, DOC-13, DOC-14, DOC-15, DOC-16, DOC-17, GEN-24, GEN-39, GEN-40, GEN-43, GEN-44
User Management	GEN-19, GEN-20, GEN-21, GEN-22, GEN-23, GEN-44, GEN-46

Table 1 Software v.s. Components matrix

Note: GEN-01 until GEN-18 are design constraints and therefore not traceable in the table

APPENDIX A : ACRONYMS

ARPEGE: regional model developed at Météo-France.

ASCII: American Standard Code for Information Interchange is a character-encoding scheme based on the ordering of the English alphabet. ASCII codes represent text in computers, communications equipment, and other devices that use text. Most modern character-encoding schemes are based on ASCII, though they support many more characters than ASCII does.

BADC: The British Atmospheric Data Centre is the Natural Environment Research Council's (NERC) Designated Data Centre for the Atmospheric Sciences. The role of the BADC is to assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by NERC projects.

C4MIP: Coupled Climate-Carbon Cycle Model Inter-comparison Project (http://c4mip.lsce.ipsl.fr/) - International project devoted to run and evaluate coupled climate-carbon models as part of ESMs.

CCMVal: Chemistry-Climate Model Validation Activity (http://www.pa.op.dlr.de/CCMVal/) - International project devoted to evaluate the chemistry component of ESMs.

CERFACS: European Centre for Research and Advanced Training in Scientific Computation

CF: Climate and Forecast Metadata Convention (http://cf-pcmdi.llnl.gov/) - International standard for model data files format. .

CGCM: Coupled global circulation model

CIM: Common Information Model - The FP7 METAFOR project develops this standard.

CMCC: Italian Ltd company realizing scientific and applied activities in the field of climate change

CMIP-5 is the Coupled Model Intercomparison Project Phase 5. CMIP is a standard experimental protocol for studying the output of coupled ocean-atmosphere general circulation models (GCMs) that will be used to analyse multi-model simulations ensemble performed in the frame of the IPCC AR5 report

CMOR: Climate Model Output Rewriter (http://www2-pcmdi.llnl.gov/cmor) – Comprises a set of FORTRAN 90 functions that can be used to produce CF-compliant NetCDF files that fulfill the requirements of many of the climate community's standard model experiments. The output resulting from CMOR is "selfdescribing" and facilitates analysis of results across models.

CORDEX: A COordinated Regional climate Downscaling EXperiment.

COSMO-LM: Members of the **C**limate**L**imited-area**M**odelling-Community (CLM-Community) are applying and developing the COSMO-CLM or CCLM, which is the COSMO model in CLimate Mode. The COSMO model is the nonhydrostatic operational weather prediction model applied and further developed by the national weather services joined in the COnsortium for SMall scale MOdelling (COSMO).

CSV: The Comma-Separated Values file format is a set of file formats used to store tabular data in which numbers and text are stored in plain textual form that can be read in a text editor.

CSW: OGC Catalogue Service for Web interface standard

DEISA: Distributed European Infrastructure for Supercomputing Applications - an EC infrastructure project to optimise the access to high-performance computers for European users (http://www.deisa.org). After a first phase during FP6, a second phase, DEISA2, is funded under FP7.

DRS: Data reference syntax

ECA&D: The European Climate assessment and Dataset project aims to present information on changes in weather and climate extremes, as well as the daily dataset needed to monitor and analyse these extremes. ECA&D is initiated by the European Climate Support Network and supported by the Network of European

Meteorological Services EUMETNET.

ENES: European Network for Earth System Modelling (http://www.enes.org) - A consortium of European institutions aiming at helping the development of use of ESMs for climate and Earth System studies.

ENSEMBLES: EU FP6 funded Integrated Project (http://ensembles-eu.metoffice.com/) -Provides future climate change projections in Europe.

ESG: Earth System Grid: http://www.earthsystemgrid.org/

ESM(s): Earth System Model(s). These models are developed to simulate the climate system in its full complexity, i.e. atmosphere, ocean and land which are the basic components included in climate models together with biogeochemical cycles, i.e., carbon cycle, vegetation, aerosol and chemistry processes.

GML: Geography Markup Language is the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. GML serves as a modeling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.

GRIB: GRIB (GRIdded Binary) is a mathematically concise data format commonly used in meteorology to store historical and forecast weather data. It is standardized by the World Meteorological Organization's Commission for Basic Systems, known under number GRIB FM 92-IX, described in WMO Manual on Codes No.306.

HCE: Hydro-climate-ecological model

ICT: Information & Communication Technology

INHGA: Romania national authority in hydrology, hydrogeology and water management

INSPIRE: Infrastructure for Spatial Information in the European Community (http://www.ecgis.org/inspire)

IPCC: Intergovernmental Panel on Climate Change (http://www.ipcc.ch) - Provides regular scientific assessments reports (AR) on climate change issue under the auspices of UNEP and ICSU. The last one is the AR4 produced in 2007; the next one is AR5 to be issued in 2013…

IPSL: Institut Pierre Simon Laplace, federation of 5 research laboratories working on global environment and climate studies.

IS-ENES: InfraStructure for the European Network for Earth System Modelling

JRA: Joint Research Activity

KNMI: Royal Netherlands Meteorological Institute

LMDZ-zoomed is the regional version of the “Laboratoire de Météorologie Dynamique” model.

METAFOR: Common Metadata for Climate Modelling Digital repositories (http://ncascms.nerc.ac.uk/METAFOR/) - FP7 infrastructure project under ENES, which focuses on developing common standards for data and model information exchange that will be implemented in IS-ENES.

MF (Météo-France): French weather Institute.

NCAR: National Center for Atmospheric Research in Boulder, USA (http://www.ncar.ucar.edu/)

NetCDF: network Common Data Form (http://www.unidata.ucar.edu/software/netcdf/) - A set of software libraries and machine-independent data formats that support the creation, access, and sharing of arrayoriented scientific data

NUTS: The Nomenclature of Units for Territorial Statistics is a hierarchical system for dividing up the economic territory of the EU for 1) the collection, development and harmonisation of EU regional statistics; 2) socio-economic analyses of the regions; 3) framing of EU regional policies. This geocode standard is developed and regulated by the European Union, and thus only covers the member states of the EU in detail. For each EU member country, a hierarchy of three NUTS levels is established. NUTS 1 corresponds to major

socio-economic regions; NUTS 2 to the basic regions for the application of regional policies and NUTS3 as small regions for specific diagnoses.

OASIS: Ocean Atmosphere Sea Ice and Soil coupler (http://www.cerfacs.fr/globc/software/oasis/) – A software component allowing synchronized exchanges of coupling information between numerical codes representing different components of the climate system. The latest versions, OASIS3 and OASIS4, were developed in the framework of the EU FP5 PRISM project, and are now supported and developed further by CERFACS (3), NEC (14) and CNRS (1) within the PRISM Support Initiative. Approximately 25 groups use the OASIS coupler internationally.

OGC: Open Geospatial Consortium - The OGC Catalog Service defines common interfaces to publish, discover, browse, and query metadata about geospatial data, services, and related resource information. It is applicable to the implementation of interfaces on catalogues of a variety of information resources.

PCMDI: The Program for Climate Model Diagnosis and Intercomparison has been established in 1989 at the Lawrence Livermore National Laboratory. The PCMDI mission is to develop improved methods and tools for the diagnosis and intercomparison of general circulation models (GCMs) that simulate the global climate.

PP file format is a record-based binary format used in a number of the BADC's datasets. It is a Met Office proprietary format so is mainly associated with Met Office products.

PRACE: Partnership for Advanced Computing in Europe (http://www.prace-project.eu/) - An FP7 infrastructure project devoted to prepare the implementation of world-class high-performance computers in Europe.

PRECIS: **P**roviding **RE**gional **C**limates for **I**mpacts **S**tudies". PRECIS is a regional climate modelling system developed at the Hadley Centre at the UK Met Office. PRECIS can be easily applied to any area of the globe to generate detailed climate change projections.

RPC: Representative Concentration Pathways. New AR5 IPCC emission scenarios. RCP8.5: Rising radiative forcing pathway leading to 8.5 W/m2 in 2100. RCP6: Stabilization without overshoot pathway to 6 W/m2 at stabilization after 2100. RCP4.5: Stabilization without overshoot pathway to 4.5 W/m2 at stabilization after 2100. RCP3-PD2: Peak in radiative forcing at ~ 3 W/m2 before 2100 and decline.

SAFRAN: Atmospheric reanalysis developed by Meteo-France at a national scale.

SMHI: Swedish institute providing operational services on climate, oceanography, meteorology and hydrology.

SRES: Special Report on Emission Scenarios IPCC for AR4 (2007).

STARDEX: STAtistical and Regional Dynamical Downscaling of EXtreme for European Region, research project supported by the European Union Framework 5 Program
http://www.cru.uea.ac.uk/projects/stardex/reports/STARDEX_FINAL_REPORT.pdf

TREDDS: Thematic Realtime Environmental Distributed Data Services, http://www.unidata.ucar.edu/projects/THREDDS/

WCRP: World Climate Research Programme (http://www.wmo.ch/pages/prog/wcrp)

WP: Work Package

WU: Wageningen University and Reasearch Center

XML: Extensible Markup Language

APPENDIX B : CLASSIFICATION OF IMPACT MODEL PARAMETERS

Main Variables	Derived parameters	Indices
Precipitation	convective large scale liquid solid total total atmosphere water vapor content	Rain free days Drought (definition) Extremes Return period Peak intensity Wettest/coldest month
Temperature	2meter and other heights, Air, soil, sea surface Avg, max, min (2meter)	Tropical days (WMO) Warm days (WMO) Frost days (WMO) Ice days Heat waves Degree days Return period Diurnal (daily/monthly) temperature range
Wind	u, v component (multiple heights) speed direction	Peak Gust (10 meter) Return period
Radiation	incoming/downwelling long & short wave outgoing/upwelling cloud fraction	Sunshine hours
Humidity	2meter (multiple heights) specific relative (incl. avg, min, max) dew point soil moisture	Potential evapotranspiration Actual evaporation
Pressure	multiple heights mean sea level geo potential height	
Orography	land sea mask (mask or fraction) Ice cover (mask or fraction) Terrain (mask or fraction)	

APPENDIX C : REFERENCES

CDO	http://www.mpimet.mpg.de/fileadmin/software/cdo/
CMOR	http://www2-pcmdi.llnl.gov/cmor
CF conv	http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.4
CERFACS' dsclim	http://www.cerfacs.fr/~page/
CORDEX	http://wcrp.ipsl.jussieu.fr/RCD_CORDEX.html
GDAL	http://www.gdal.org/
GMT	http://gmt.soest.hawaii.edu/
GNU GSL	http://www.gnu.org/software/gsl/
GNU Fortran	http://gcc.gnu.org/fortran/
GNU C	http://gcc.gnu.org/
GRIB1, GRIB2	http://www.wmo.int/pages/prog/www/WMOCodes.html
Ferret	http://ferret.pmel.noaa.gov/Ferret/
Matlab	http://www.mathworks.com
Metafor	http://metaforclimate.eu/
NCO	http://nco.sourceforge.net/
NCL	http://www.ncl.ucar.edu/
NetCDF4,NetCDF3	http://www.unidata.ucar.edu/software/netcdf/
Numpy	http://numpy.scipy.org/
OASIS3	http://www.prism.enes.org/PAEs/coupling_IO/software_OASIS3.php
Inspire	http://inspire.jrc.ec.europa.eu/
OGC	http://www.opengeospatial.org/
OPeNDAP	http://opendap.org/
OpenLayers	http://openlayers.org/
Pydap	http://pydap.org/
R	http://www.r-project.org/
stochasticTools	http://iri.columbia.edu/climate/forecast/stochasticTools/
WCS	http://www.opengeospatial.org/standards/wcs
WMS	http://www.opengeospatial.org/standards/wms
WPS	http://www.opengeospatial.org/standards/wps