

UNIVERSITÀ DELLA CALABRIA



University of Calabria

Department of Physics

Doctoral School

Scuola in Scienza e Tecnica "Bernardino Telesio"

Cycle XXVIII

**Integration of mercury atmospheric models in a framework for
interoperable workflow execution**

(Integrazione di modelli atmosferici del mercurio in un framework per l'esecuzione di
workflow interoperabili)

PhD Thesis

Scientific Sector FIS/06

Coordinator: Prof. Vincenzo Carbone (University of Calabria)

Signature Vincenzo Carbone

Supervisors: Dr. Nicola Pirrone (CNR - IIA)

Signature Nicola Pirrone

Dr. Stefano Nativi (CNR - IIA)

Signature Stefano Nativi

Prof. Vincenzo Carbone (University of Calabria)

Signature Vincenzo Carbone

Candidate: Enrico Boldrini

Signature Enrico Boldrini

1	Introduction.....	4
1.1	International programs, initiatives.....	4
1.1.1	GEOSS.....	5
1.1.2	GMOS.....	6
1.1.3	GEO Model Web.....	6
1.1.4	Brokering Approach.....	7
2	Model Access and Interoperability.....	11
2.1	From offline model applications to MaaS.....	12
2.2	Business Processes and Workflows.....	13
2.3	The BP-Broker.....	13
2.3.1	System Architecture.....	14
3	Mercury atmospheric models.....	16
3.1	WRF model.....	16
3.2	WRF Chem model.....	18
3.3	WRF/Chem-Hg model developed at CNR/IIA.....	19
4	Implementing WRF/Chem-Hg as a MaaS.....	21
4.1	Publication of the model as a WPS in the cloud.....	21
4.2	BPMN description.....	25
5	The e-WRF/Chem-Hg system.....	27
5.1	BP-Broker and Registry.....	28
5.2	Model Adaptor realization.....	28
5.3	Publication of the business process.....	29
5.4	BP-Broker Portal.....	29
6	System tests.....	30
6.1	The basic workflow.....	30
6.1.1	Discovery and evaluation of the mercury model.....	30
6.1.2	Configuring and Running the model.....	31
6.1.3	Execution.....	34
6.1.4	Working with the result.....	34
6.2	A complex workflow.....	35
6.2.1	A Hg Post Processing BP.....	35
6.2.2	Composition of models.....	36
7	Discussion and Challenges.....	38
	Appendix A: UML quick reference.....	39
7.1	Deployment Diagrams.....	39
7.1.1	Node.....	39

7.2	Node Stereotypes	39
7.3	Node as Container	39
7.4	Component Diagrams	40
7.4.1	Representing Components	40
7.4.2	Assembly Connector	40
Appendix B. BPMN quick reference		41
7.5	Events	41
7.5.1	Start event	41
7.5.2	End event	41
7.6	Activity	41
7.6.1	Task	41
7.7	Artifacts	42
7.7.1	Data objects	42
7.7.2	Group	42
7.7.3	Annotation	42
7.8	Connections	42
7.8.1	Sequence Flow	42
7.8.2	Association	42
List of figures		43
Glossary		45
Bibliography		47
Acknowledgements		53

1 INTRODUCTION

Earth and Space Science (ESS) role is fundamental to perform analysis and prediction of a complex and changing environment. Informatics importantly contributes to many aspects of the scientist working in this field, easing the automatic processing of information. Example given big quantities of Earth related data (Baumann, et al., 2015) can now be acquired through automatic systems, such as satellites and observing stations. Web services such as inventory, catalog and data access services are used to publish and effectively share it amongst different research groups. Finally, data can be further viewed, analyzed and processed, e.g. using statistical applications and software models to produce forecast and reanalysis.

In the ESS global scenario, it appears clear that many actors need to interoperate, such as research groups, decision makers and private companies along with their many heterogeneous software systems and information models: it's the interoperability challenge.

International programs and organizations have the role of coordinating and easing such interoperability efforts, e.g. drafting guidelines, promoting standardization and implementing infrastructures:

- The GEO Model Web initiative envisions an increased interoperability of ESS computational models, which will become online available and interoperable, allowing chaining and complex workflows, as well as multidisciplinary applications.
- The Brokering approach emerges in the ESS context with the aim of easing interoperability between heterogeneous parties (Boldrini, Craglia, Mazzetti, & Nativi, 2014): an important implementation of this approach is to be found in the Group on Earth Observation Discovery and Access Broker (GEO-DAB) which takes charge of discovering and accessing geospatial resources around the globe and presenting them in a user friendly way.

The subject of the present work conducted at the National Research Council of Italy, Institute of Atmospheric Pollution Research (CNR/IIA) is to apply both Model Web and Brokering approach to the context of Atmospheric Pollution research.

CNR/IIA modelers working at the Rende division have produced mercury atmospheric models in the context of the Global Mercury Observation System (GMOS) based on the Weather Research and Forecasting Model (WRF) to gain deeper knowledge about this pollutant cycle (Gencarelli, De Simone, Hedgecock, Sprovieri, & Pirrone, Development and application of a regional-scale atmospheric mercury model based on WRF/Chem: a Mediterranean area investigation, 2014).

In particular by adopting and enhancing the Business Process-Broker framework developed by researchers of the CNR/IIA Florence division, it has been possible to make the mercury models online available and more interoperable. The BP-Broker is capable of interpreting custom workflows described using the Business Process Model and Notation (BPMN) and executing them after resolving the needed resources (e.g. models, input data) by leveraging the brokering approach.

Once integrated in the BP-Broker framework, mercury models can be remotely configured and executed by scientists using a user friendly web client. Moreover, the models become readily available as building blocks for realizing and executing complex workflows designed using standard BPMN diagrams as well.

1.1 INTERNATIONAL PROGRAMS, INITIATIVES

International programs and initiatives in the context of Earth and Space Science infrastructures have the role to coordinate scientific and technological efforts at a European and global level, promoting interoperability, good practices, and discussion forums open to the community.

In the following sections GEOSS program, GMOS project, and the GEO Model Web initiative will be presented for their relevance to the scope of this work. An overview of the Brokering approach to support multidisciplinary applications is provided, too.

1.1.1 GEOSS

In 2002 the World Summit on Sustainable Development and the Group of Eight (G8) governmental political forum recognized that international collaboration on Earth Observation (EO) was essential. Subsequently, the Group on Earth Observations (GEO) has been established as an intergovernmental organization comprising 96 member countries and the European Commission, as well as 68 participating organizations.

GEO coordinates the efforts to develop the Global Earth Observation System of Systems (GEOSS), a global scale network of interconnected systems, interoperating to collect and share Earth Observation data. A 10 years implementation plan was drafted (Group on Earth Observations, 2005), defining the expected benefits the final system shall provide to 9 identified Societal Benefit Areas (SBA): disasters, health, energy, climate, water, meteorology, ecosystems, agriculture and biodiversity.

In 2013, 65 million Earth Science Earth Observation elements/files (pertaining to all the SBAs), were discoverable through the GEOSS Portal (Group on Earth Observations, 2014). Of these discoverable resources, more than 50 million (data products, files, images) can be accessed without restriction and free of charge. Presently (Oct 2015), about 200 million elements are discoverable and potentially accessible, shared by more than 120 autonomous and disparate data infrastructures.



Figure 1 The GEOSS Portal enables discovery and access to the comprehensive data produced by the GEOSS community by connecting to the GEO-DAB, the broker (mediator) component in the GEOSS architecture.

The Discovery & Access Broker (DAB) developed by CNR/IIA is the enabler of this major result. GEO-DAB is the core component of the GEOSS architecture, acting as a broker between the GEO Portal and the many and heterogeneous Data Providers maintained by the GEOSS community members (Figure 1). Through the GEO Portal users can thus discover and access resources of interest in a friendly way. Client applications can also benefit of APIs and standard service interfaces to leverage the functionalities of GEO-DAB (e.g. to the aim of building community or thematic portals).

1.1.2 GMOS

The Global Mercury Observation System (GMOS) is a global monitoring network collecting near-real-time data from ground-based sites (see Figure 2), shipboard measurements over the Pacific and Atlantic Oceans and European Seas, as well as aircraft-based measurements (Cinnirella, D'Amore, Bencardino, Sprovieri, & Pirrone, 2014) (Pirrone & Cinnirella, 2012).

The system comprises tools (i.e. databases, catalogs, services) (Cinnirella, D'Amore, Mazzetti, Nativi, & Pirrone, 2013) used to collect datasets, harvest mercury databases, and a data portal (CNR-Institute of Atmospheric Pollution Research, 2015) providing user friendly search capabilities. Global and Regional Models are developed in the context of GMOS to provide assessment of Hg pollution levels and evaluation of future pollution changes (Gencarelli, et al., European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux, 2015) according to different emission control scenarios. A specific GMOS Mercury Modelling Task Force (MMTF) has been established in 2014 (Global Mercury Observation System, 2014).

GMOS has been developed by CNR/IIA and other partners in the context of the European Union 7th Framework Program (FP7) and includes contributions from Canada (CAMNet), European Commission (EC-FP7), United States (NADP/AMNet), the European Monitoring and Evaluation Program (EMEP), the Task Force on Hemispheric Transport of Air Pollutants (TFHTAP), UNEP Mercury Programme and other international monitoring and modeling efforts.

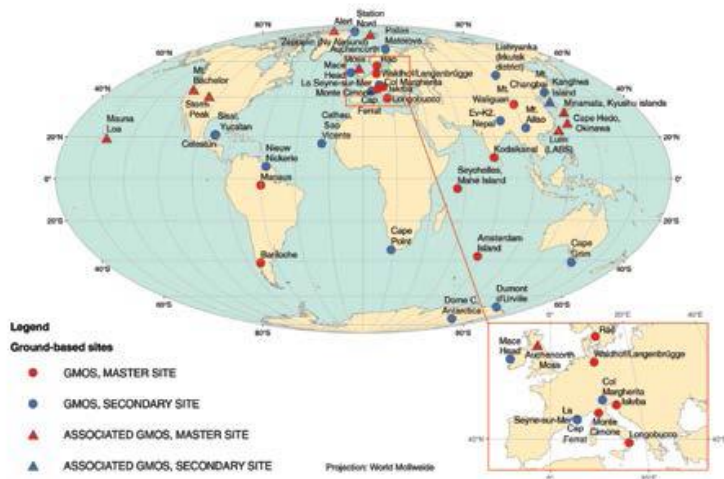


Figure 2 The GMOS ground-based monitoring network. (CNR-Institute of Atmospheric Pollution Research, 2015)

1.1.3 GEO Model Web

The GEO Model Web initiative envisions the adoption of the Model as a Service (MaaS) approach to increase model access and sharing. Four basic principles underlie the GEO Model Web (Nativi, Mazzetti, & Geller, Environmental model access and interoperability: The GEO Model Web initiative, Environmental Modelling & Software, 2013):

- open access: models should be published online with no constraints on the access
- minimal barriers to entry: accessing or publishing a model should require a low effort
- service-driven: models should be available online as Web services
- scalability: scalability shall and is inherently assured for Model Web, being based on Web services

Any implementation approach meeting these principles is a step towards the Model Web long term vision: a global scale network of independent but interacting models.

Given the conceptual nature of the Model Web vision, no architectural approach and interoperability solution is mandatory for its implementation, provided that they satisfy the Model Web principles. In (Nativi, Mazzetti, & Geller, Environmental model access and interoperability: The GEO Model Web initiative, Environmental Modelling & Software, 2013) five characteristics were identified to facilitate Model Web development and its compliance with GEOSS: Distributed System, System of Systems, World Wide Web, Semantic Web and Web 2.0.

1.1.4 Brokering Approach

The figure of brokers (or mediators) is long standing in the broad social-organizational and technological contexts (Boldrini, Craglia, Mazzetti, & Nativi, 2014). Gathering from these experiences the Brokering Approach has been conceived to realize multidisciplinary interoperability in the Geoscience by Nativi et al. (Nativi, Craglia, & Pearlman, The brokering approach for multidisciplinary interoperability: a position paper, 2012). The main principles of brokering are:

- **Autonomy:** Keep the existing disciplinary infrastructures as autonomous as possible, not asking community members to implement any “more general” service bus.
- **Subsidiarity:** Supplement but not supplant disciplinary infrastructure mandates and governance arrangements by interconnecting and mediating their service buses.
- **Interconnection:** Build incrementally on existing infrastructures and introduce distribution and mediation functionalities to interconnect the heterogeneous service buses characterizing any domain specific or other infrastructure.
- **Low Entry Barrier:** Assure a low entry barrier for both users and resource providers of any disciplinary infrastructure.
- **Flexibility:** Be flexible enough to accommodate existing and future information systems and information technologies that will augment the service bus implemented by any discipline.
- **Effectiveness:** Address the full range of information exchange needs (discovery, access, semantics, workflow, etc.).

The resulting brokering architecture enables multidisciplinary applications, allowing clients and servers traditionally conceived and confined in a specific community to interoperate with the ones pertaining to different disciplines, thanks to the Brokering mediation services (Figure 3).

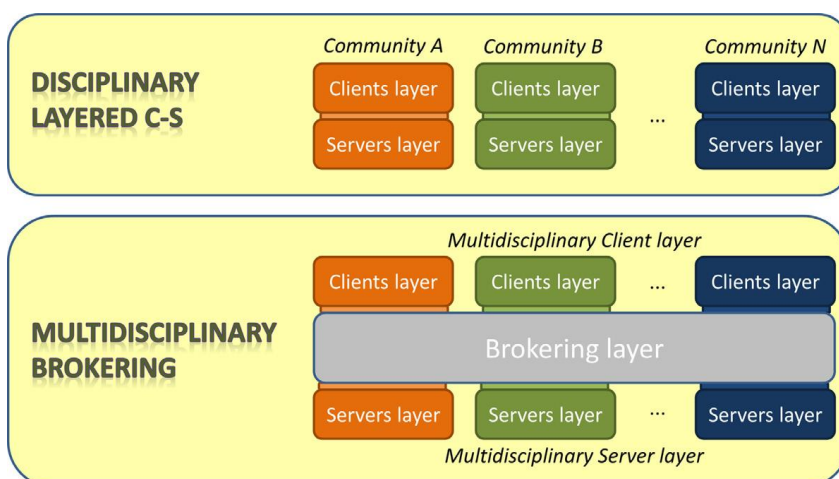


Figure 3 Disciplinary Layered Client-Server architectures and the Brokering oriented architecture for multi-disciplinary environments (Nativi, Craglia, & Pearlman, The brokering approach for multidisciplinary interoperability: a position paper, 2012).

Following the Brokering principles, the GI-suite brokering framework co-designed and implemented by the candidate (Boldrini, Papeschi, Santoro, & Nativi, Extending the GI Brokering Suite to Support New Interoperability Specifications, 2014) (Boldrini, Papeschi, Santoro, & Nativi, Enabling interoperability in

Geoscience with GI-suite, 2015), is an extensible and modular software framework composed by different mediation services, each one specialized in brokering a particular capability:

- **GI-cat** (Nativi, Bigagli, Mazzetti, Boldrini, & Papeschi, 2009) (Boldrini, E. & the GI-cat Team, 2010) discovery broker enables discovery of geoscientific resources across a series of heterogeneous data providers (e.g. catalog services publishing CSW/ISO and OpenSearch, inventories publishing THREDDS and OAI-PMH) to a heterogeneous set of data consumers (e.g. client applications, web portals)
- **GI-axe** (Santoro, Papeschi, Boldrini, & Bigagli, 2011) (Boldrini, Nativi, Santoro, Papeschi, & Mazzetti, 2012) access broker enables access of geoscientific resources across a series of heterogeneous data providers (e.g. data access services publishing OGC WCS and OGC WMS, data repositories publishing FTP) to a heterogeneous set of data consumers (e.g. OpenLayers libraries, ArcGIS web client)
- **The semantic broker (GI-sem)** is used within the suite to obtain semantically expansions of a given term from a set of heterogeneous semantic services. The user can benefit this way example given of semantic expanded queries.

GI-suite is a configurable and extensible framework, indeed it supports an always increasing number of community and international standard service interfaces and data models. It has been tailored and successfully deployed in many international projects and initiatives, such as:

- **FP7 SeaDataNet:** SeaDataNet II was launched on October 1st 2011 for a duration of 4 years as the second phase of the EC research project SeaDataNet (FP7 SeaDataNet, 2015). The GI-suite discovery broker is used in the SeaDataNet Pan-European marine data infrastructure participated by 35 countries to provide a standardized INSPIRE compliant service interface to international programs interested in SeaDataNet data, such as GEOSS (Boldrini, Nativi, & Schaap, INSPIRE compliant international standards for the SeaDataNet marine metadata, 2013).
- **FP7 EuroGEOSS:** The EuroGEOSS research project was launched on May 1st 2009 for a 3 year period with 20 partners at regional, national, European, and international levels with the main aim of making a long-term contribution to the development of GEOSS. As a result, the EuroGEOSS broker based on GI-cat has been adopted in the GEOSS Common Infrastructure and its successful deployment was recognized at the GEO plenary in November 2011 (Serrano, et al., 2012) (Santoro, Boldrini, Papeschi, Baldini, & Nativi, 2010).
- **FP7 MEDINA:** Medina research project was launched on October 1st 2011 for a duration of 3 years with the aim of enhancing Northern African Countries monitoring capacity of their coastal areas ecosystems (FP7 Medina, 2015). Both GEO-DAB and BP-Broker have been successfully experimented and applied in this context (Bigagli, Santoro, Mazzetti, & Nativi, 2015).
- **FP7 ODIP:** The Ocean Data Interoperability Platform (ODIP) coordination & support action was launched on October 1st 2012 for a duration of 3 years with the aim of facilitating and harmonizing Europe, USA and Australia efforts of building regional infrastructures to discovery and access marine data (FP7 ODIP, 2015) (Glaves & Schaap, 2015). The action has been particularly aimed at promotion of international workshops to foster development of common standard and interoperability solution prototypes. The SeaDataNet broker based on GI-suite has been successfully experimented in this context.
- **FP7 GEOWOW:** The GEOWOW research project was launched on September 1st 2011 for a 3 year period with 15 partners from 7 countries (FP7 GEOWOW, 2012). GEOWOS has developed a set of multidisciplinary use scenarios to advance the GEOSS Common Infrastructure. The middleware services provided by the GEO-DAB were employed to successfully address multidisciplinary interoperability challenges (Santoro, Sorichetta, Roglia, Craglia, & Nativi, 2013).
- **GEO GEOSS:** the GEO system of systems enables user friendly discovery and access of millions resources from the major ESS data providers at a global scale. The GEO-DAB, based on GI-suite and managed by CNR/IIA, is presently the core component of the GEOSS architecture.

- **NSF ACADIS:** The Advanced Cooperative Arctic Data and Information Service (ACADIS) is a joint effort by the National Snow and Ice Data Center (NSIDC), the University Corporation for Atmospheric Research (UCAR), UNIDATA, and the National Center for Atmospheric Research (NCAR). Its purpose is to provide data support, preservation and access services for all projects funded by NSF's Arctic Science Program (ARC). GI-suite is a key component of the architecture, providing search capabilities among heterogeneous resource types (Khalsa, et al., 2013).
- **NSF BCube:** Funded by the National Science Foundation (NSF) and titled "A Broker Framework for the Next Generation Geoscience", the BCube project was launched on September 1st 2013 for a 2 year period (BCube, 2015). GI-suite has been deployed as the BCube broker to ease discovery and access to geoscience resources by scientists coming from different domains: hydrology, oceans, polar, and weather (Boldrini, Easton, Fuka, Pearlman, & Nativi, 2015). It is part of NSF EarthCube program "A Community-Driven Data and Knowledge Environment for the Geosciences." (Khalsa, et al., 2014).
- **CEOS Water Portal:** CEOS Water Portal led by Japan Aerospace Exploration Agency (JAXA) is a project of the Applications Subgroup of the Committee on Earth Observation Satellites (CEOS). The purpose of the CEOS Water Portal Project is to provide assistance to the water relevant scientists and general users (or non-researchers) in the development of data services associated with data integration and distribution. The system architecture is based on GI-suite (Miura, Sekioka, Kuroiwa, & Kudo, 2015).
- **MIUR NextData:** The Italian initiative NextData (NextData, 2015) was launched in 2012 to build a national system for the retrieval, storage, access and diffusion of environmental and climate data from mountain and marine areas (Nativi, et al., 2014). The brokering technology first developed by FP7 EuroGEOSS project was successfully applied in this context.
- **FP7 MED-SUV:** MED-SUV was launched on June 1st 2013 for a duration of 3 years (FP7 MED-SUV, 2015). It aims at improving the consortium capacity of assessment of volcanic hazards in Supersites of Southern Italy (e.g. by optimizing and integrating existing and new observation/monitoring systems). The GI-suite is deployed to the aim of accessing and making available resources of interest from the both the project partners and international data providers.
- **FP7 GeoViQua:** QUALity aware Visualization for the Global Earth Observation System of Systems (GeoViQua) was launched on February 1st 2011 for a duration of 3 years (FP7 GeoViQua, 2015). Its aim was to complement GEOSS with rigorous data quality specifications and quality-aware capabilities, in order to improve reliability in scientific studies and policy decision-making. The GeoViQua quality-augmentation broker based on GI-suite was designed and deployed in this context (Bigagli, Papeschi, Mazzetti, & Nativi, 2012).
- The GI-suite brokering services have been tested in a collaboration between CNR/IIA and **Oak Ridge National Laboratory** (ORNL) to normalize and compare Carbon Cycle model outputs and observations. The work was carried on between the National NSF EarthCube Brokering Concept Award project, the NSF DataONE Exploration, Visualization, and Analysis (EVA) working group, the two NASA-funded projects Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP), and Modeling and Synthesis Thematic Data Center (Santoro, et al., 2013).
- **ICSU PANGAEA** digital library: The International Council for Science (ICSU) information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. GI-cat is providing a CSW interface to access the PANGAEA OAI-PMH service.
- **WIS:** The WMO Information System is deployed with the aim of managing and transferring weather, water and climate information at a global scale (WMO, 2015). WIS is amongst GEOSS sources brokered by GEO-DAB.
- **H2020 ECOPotential:** ECOPotential was launched on June 1st 2015 for a duration of 4 years, funded under the Horizon 2020 (H2020) EU research and innovation programme (H2020 EcoPotential, 2015). The GI-suite brokering framework is going to be tested to realize open and interoperable access to terrestrial and marine ecosystems data.

- A **CERIF** accessor component has recently been added to GI-suite (Boldrini, Luzi, Nativi, & Pecoraro, 2015): it is in charge of mediation between CERIF data model and ISO 19115 (ISO TC211, 2003).
- **ESRI ArcGIS Online**: A Memorandum of Understanding (MoU) between the Environmental Systems Research Institute (ESRI) and Earth and Space Science Informatics Laboratory of CNR/IIA has created a partnership between the two organizations. Per the MOU, CNR-IIA and ESRI are designing brokering arrangements and direct dataset access technologies as well as open standards for data interoperability and cataloging (ESRI, 2014). GI-suite has been improved and tested during the collaboration.

2 MODEL ACCESS AND INTEROPERABILITY

A scientific model aims to represent empirical objects, phenomena, and physical processes in a logical and objective way. It necessarily abstracts the modelling object by capturing its salient features, possibly producing a useful simplification of reality (Box & Draper, 1987).

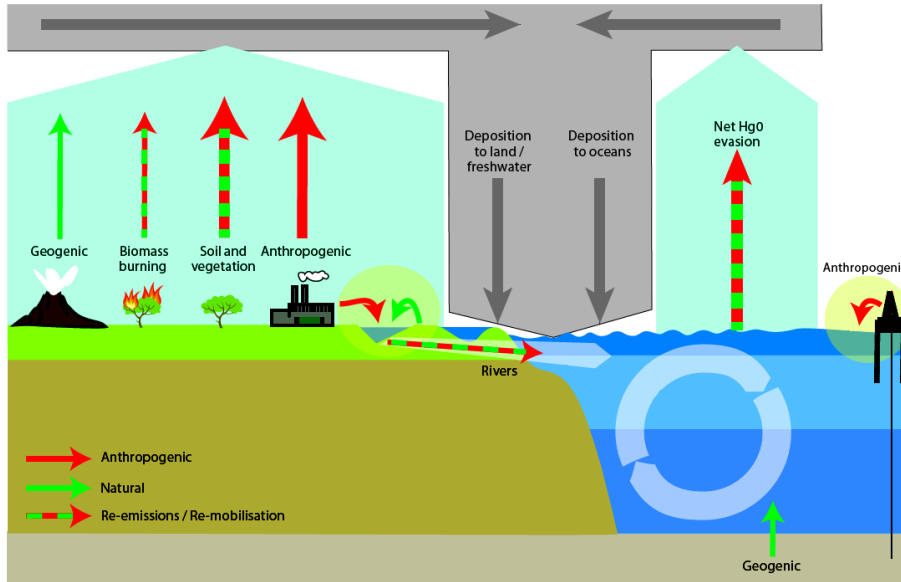


Figure 4 Scientific modelling of components related to global mercury cycle (Arctic Monitoring Assessment Programme, UNEP Chemicals Branch, 2013)

Figure 4 shows the modelling of the global mercury cycle, in which various components of both anthropogenic and natural emissions and re-emissions/re-mobilizations are highlighted (Arctic Monitoring Assessment Programme, UNEP Chemicals Branch, 2013).

Computational models (or just models in the following) are currently used in many scientific disciplines, such as weather forecasting, earth simulators, protein folding, to study complex systems that require extensive computational and data resources.

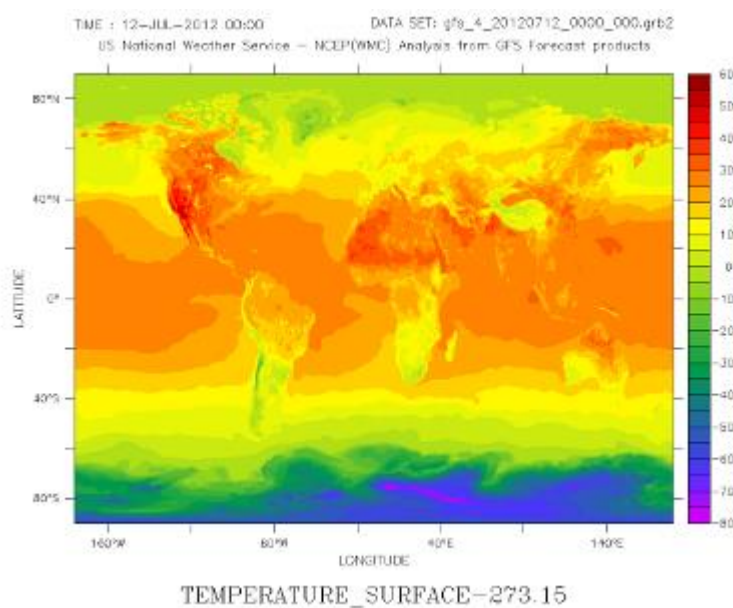


Figure 5 A global analysis plot of GFS surface temperatures, in degrees Celsius, for July 12, 2012 at 00UTC. (National Oceanic Atmospheric Administration, National Centers for Environmental Information, 2015)

Example given, the Global Forecast System (GFS) is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP). Dozens of atmospheric and land-soil variables are available through this dataset, from temperatures (Figure 5), winds, and precipitation to soil moisture and atmospheric ozone concentration. The entire globe is covered by the GFS at a base horizontal resolution of 28 kilometers between grid points, which is used by the operational forecasters who predict weather out to 16 days in the future (National Oceanic Atmospheric Administration, National Centers for Environmental Information, 2015).

In the context of ESS research infrastructures, realization of computational models that are able to interoperate (e.g. where the output of model A is the input for model B) represents a great scientific and technological challenge, to the aim of developing multidisciplinary applications.

2.1 FROM OFFLINE MODEL APPLICATIONS TO MAAS

The simpler (and less interoperable) approach to modelling is to realize software applications specific for a given model; through an interface or using command line the scientist can select algorithms and parameters of the simulation, including input data. At simulation end, user can access and visualize the simulation output. This is the case e.g. for the Weather Research and Forecasting Model (WRF) (National Center for Atmospheric Research, 2015), used for weather forecasting.

An advanced step in model interoperability is realized by specific modelling frameworks, such as Object Modelling System (David, Markstrom, Rojas, Ahuja, & Schneider, 2002), ModCom (Hillyer, Bolte, van Evert, & Lamaker, 2003), The Invisible Modelling Environment (Rahman, et al., 2003), OpenMI (Gregersen, Gijbers, & Westen, 2007), Spatial Modeling Environment (Maxwell, Villa, & Costanza, 2015), Tarsier (Watson & Rahman, 2004), Interactive Component Modelling System (Gilmour, Croke, & Watson, 2000), Earth System Modelling Framework (Hill, DeLuca, Balaji, Suarez, & da Silva, 2004), SEAMLESS (Van Ittersum, et al., 2008), etc. These frameworks differ by many aspects, including the domain (i.e. disciplinary or multidisciplinary), functionalities (e.g. single execution of a model or chaining of multiple models) and technology (e.g. modelling language, supported platforms): these differences reflect the specific requirements of the community they address. By adopting a modelling framework, users can thus design and implement models and in some cases workflows of connected models that are interoperable inside a specific framework; interoperability with models developed outside the framework is however not ensured.

The paradigm of Model as a Service (MaaS) is an important advancement that has its roots in the context of Service Oriented Architecture (SOA): it enables model functionalities through a Web service interface (a set of operations that characterize the entity behavior). MaaS allows the remote execution of a model; a necessary condition for inserting the model in a more complex scenario which may include service chaining and workflows.

A MaaS implementation is provided by OpenModeller, a tool used by the biology community to model ecological niches that can be remotely operated through its Web interface (Muñoz, et al., 2009).

In the MaaS context interoperability conventions are shifted from the specific technological level agreements to the level of service interface definitions. International specifications such as the Web Service Description Language (WSDL) (World Wide Web Consortium, 2001) and Open Geospatial Consortium (OGC) Web Processing Service (WPS) (Open Geospatial Consortium, 2007) are drafted to support Web implementations of models.

Formal definitions of workflows and business processes are currently in use. Example given, in the e-Commerce and e-Government contexts the Business Process Execution Language (BPEL) is used for describing workflows and orchestrating web services (OASIS Web Services Business Process Execution Language TC, 2007).

2.2 BUSINESS PROCESSES AND WORKFLOWS

A Business Process (BP) is a sequence or flow of related, structured activities or tasks in an organization that produce a specific service or product (serve a particular goal) for a particular customer (Object Management Group, 2010).

BPMN is a standard for business process modeling that provides a graphical notation for specifying business processes in a Business Process Diagram, based on a flowcharting technique very similar to activity diagrams from UML. The objective of BPMN is to support business process management, for both technical users and business users, by providing a notation that is intuitive to business users, yet able to represent complex process semantics. The BPMN specification also provides a mapping between the graphics of the notation and the underlying constructs of execution languages, particularly BPEL (Object Management Group, 2010).

BPMN can describe both:

- Abstract Business Processes (ABP): partially specified processes that are not intended to be executed. An Abstract Process may hide some of the required concrete operational details. Conceptual tasks and people can be represented in ABP. Example given a task that may appear in ABP is “meteorology data processing”.
- Executable Business Processes (EBP): define a portable execution format for business processes that rely exclusively on Web Service resources and XML data. Moreover, such processes execute and interact with their partners in a consistent way regardless of the supporting platform or programming model used by the implementation of the hosting environment.
 - Model Workflows are a subset of Executable Business Process, in which web models are described and chained to produce a complex output. Workflows can be automatically run by a business process execution engine.

2.3 THE BP-BROKER

The BP-Broker is introduced and experimented (Santoro, Environmental Model Composition in the GEO Model Web, 2013) (Bigagli, Santoro, Mazzetti, & Nativi, 2015) in the context of the FP7 project UncertWeb (Bigagli, Santoro, Angelini, Mazzetti, & Nativi, 2011) for the composition and execution of model workflows. In this solution the advantages of a SOA approach (abstraction of functionalities to the service level) are summed to the ones of Brokering approach (managing of heterogeneity: of data resources, of models, of business processes, of workflow engines).

BP-Broker is a service having the task of easing the implementation and execution of workflows based on existing SOA services. BP-Broker is in charge of typical tasks of workflow definition (matching of types, identification of service endpoints, etc.), moreover taking care of existing heterogeneity, both of functionalities and technologies; the BP-Broker is also able to perform advanced functionalities in model composition particularly useful in multidisciplinary applications (such as perfecting an abstract workflow through a service registry to make it executable) (Santoro, Nativi, & Mazzetti, Towards GEO Model Web implementation: a Brokering Service for Business Processes, 2015).

At a conceptual level, the BP is the abstract functionality that users want to utilize. The Business Process is modeled during the composition phase, in which the available (atomic or composite) processes are automatically retrieved and assembled (each process is implemented as a service). The Business Process model is then published by the BP-Broker as EBP, so that it can be run by a specific Execution Engine; the execution result is represented by a Run, containing the execution outputs.

2.3.1 System Architecture

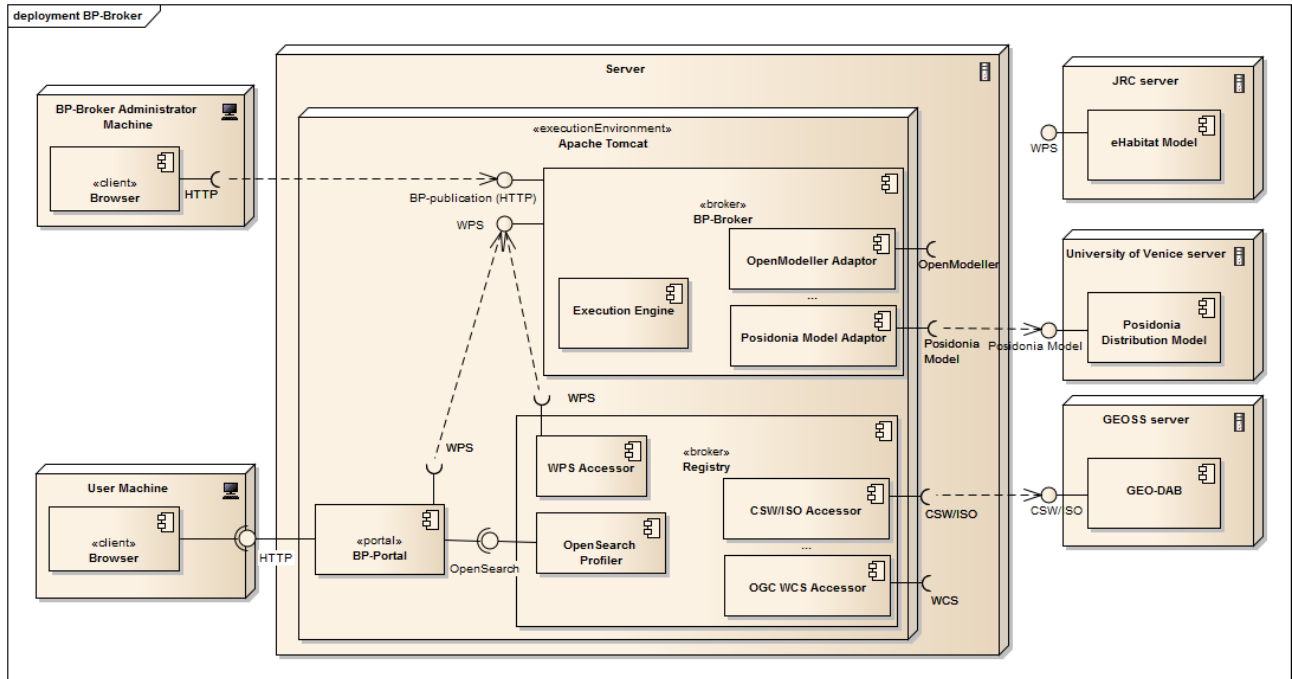


Figure 6 Possible deployment of the BP-Broker. Users and the administrator access the BP-Broker functionalities through the user friendly web interfaces. The BP-Broker is able to access functionalities from different remote heterogeneous services.

Figure 6 shows a typical deployment of the BP-Broker. In this case the BP-Broker is hosted on an Apache Tomcat Web container, along with BP-Portal and a Registry.

A BP-publication interface is available to allow the administrator to add business processes to the system. This is done by uploading BPMN diagrams, along with metadata fields to allow discovery (e.g. title, abstract, keywords). Publication entails the resolution of all possibly ambiguous and incomplete aspects (e.g., missing endpoints to concrete service instances, binding issues, type mismatches), as well as BP validation. Upon successful publication, an EBP is produced: this can be interpreted and run by the internal Execution Engine. Moreover, the EBP is also published as a process through the WPS interface published by the BP-Broker, and can hence become part of other BPs. In other words, the newly published BP is immediately available for reuse.

The BP-Portal provides a web user friendly interface to discover resources of interest: available business processes and allowed data inputs. The Process parameters can be configured and the process can be run directly through the portal.

The discovery is performed by the Registry, which is a broker based on the GI-suite framework (e.g. GI-cat and GI-axe brokers). Leveraging its functionalities, users can search a wide range of geoscience services for processes and datasets (e.g. the GEO-DAB). Moreover, the Registry periodically harvest the BP-Broker WPS interface, to learn about new EBP that have been published. The components that are responsible to contact the remote sources and mediate access to their information content using specific service interfaces are called accessors: there exists a different accessor for each service type.

The Execution Engine of the BP- Broker is based on the JBoss Business Process Management technology (Red Hat, Inc., 2015). The core of jBPM is a lightweight, extensible workflow engine that allows executing annotated BPs using the latest BPMN 2.0 specification.

The Execution Engine makes use of the Adaptor components to invoke remote processing services (e.g. Figure 6 shows BP-Broker connection to the eHabitat model (Dubois, et al., 2011), hosted by JRC and to the

Posidonia SeaGrass Model (Santoro, et al., 2014), hosted by University of Venice). There exists a different adapter for each different service type. It's the adapter task to perform functional and data model mediation in order to provide the remote processing service with the needed inputs in the appropriate format, start the execution, collect feedback and retrieve the results once ready.

A Run represents the outcome of an EBP execution, including the resulting output, as well as provenance and additional ancillary information (execution logs, etc.)

3 MERCURY ATMOSPHERIC MODELS

The problem posed by mercury pollution has increased its relevance over the recent years (Arctic Monitoring Assessment Programme, UNEP Chemicals Branch, 2013) (UNEP, 2015): the concerns are particularly related to the high concentration of this pollutant in fishes at the top of the food chain (e.g. tuna and swordfish) (Kraepiel, Keller, Chin, Malcolm, & Morel, 2003).

Chemical Transport Models (CTMs) are software tools which are useful to gain a deeper understanding about the mercury cycle, including Hg dispersion over long distances and mercury pollution (Ariya, et al., 2015).

CNR/IIA modelers have implemented a regional high resolution Mercury Model (WRF/Chem-Hg) consisting of an online CTM based on the WRF/Chem model (Gencarelli, 2013). This model can take in account both environmental and anthropogenic emissions, as well as changes in Hg species concentrations which occur as a result of chemical transformations. WRF/Chem-Hg model has been used to evaluate atmospheric Hg deposition over Europe and the Mediterranean: anthropogenic and natural emissions, transport and removal processes were simulated with two different oxidation mechanisms (Gencarelli, et al., European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux, 2015).

In the next section an overview about the WRF and WRF/Chem models is provided, the focus of the last section is finally moved on the description of the WRF/Chem-Hg model.

3.1 WRF MODEL

The Weather Research and Forecasting (WRF) Model is a mesoscale numerical weather prediction system designed for both atmospheric research and operational forecasting needs (National Center for Atmospheric Research, 2015). It features two dynamical cores, a data assimilation system, and a software architecture facilitating parallel computation and system extensibility. The model serves a wide range of meteorological applications across scales from tens of meters to thousands of kilometers. The effort to develop WRF began in the latter part of the 1990's and was a collaborative partnership principally among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (represented by the National Centers for Environmental Prediction (NCEP) and the (then) Forecast Systems Laboratory (FSL)), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA).

WRF is composed by the following major components:

- The WRF Preprocessing System
- WRF Data Assimilator
- ARW solver
- Post-processing & Visualization tools

The WRF Preprocessing System is used to define simulation domains (e.g. spatial and temporal extents), interpolating terrestrial data (such as terrain, land use, and soil types) to the simulation domain and interpolating GRIB meteorological data from another model (i.e. ECMWF or GFS) to the simulation domain. Its main features include (National Center for Atmospheric Research, 2015):

- GRIB (World Meteorological Organization, 2003) 0.5 degree meteorological data from various centers around the world
- USGS 24 category and MODIS 20 category land datasets; USGS GTOPO30 elevation dataset; Global 5-minutes United Nation FAO, and North-America STATSGO 30 sec soil category dataset; 10-min greenness fraction data based on AVHRR and 30-sec greenness fraction data based on 10 years

MODIS; MODIS-based leaf-area index; 0.15 degree monthly albedo and snow albedo data; and 1-degree deep soil temperature data; plus a few specialized datasets

- Map projections for polar stereographic, Lambert-Conformal, Mercator and latitude-longitude
- Nested domains
- User-interfaces to input other static data as well as internal format data

WRF Data Assimilator is an optional WRF component used to ingest observations into the interpolated analyses created by WRF Preprocessing System.

The ARW solver is the key component of the modeling system, which is composed of several initialization programs for idealized, and real-data simulations, and the numerical integration program. The key features of the WRF model include (National Center for Atmospheric Research, 2015):

- Fully compressible non-hydrostatic equations with hydrostatic option
- Regional and global applications
- Complete Coriolis and curvature terms
- Advanced nesting management
- Mass-based terrain-following coordinate
- Vertical grid-spacing can vary with height
- Map-scale factors for these projections:
 - polar stereographic (conformal)
 - Lambert-conformal
 - Mercator (conformal)
 - Latitude and longitude, which can be rotated
- Arakawa C-grid staggering
- Runge-Kutta 2nd and 3rd order time integration options
- Scalar-conserving flux form for prognostic variables
- 2nd to 6th order advection options (horizontal and vertical)
- Monotonic transport and positive-definite advection option for moisture, scalar, tracer, and TKE
- Weighted Essentially Non-Oscillatory (WENO) advection option
- Time-split small step for acoustic and gravity-wave modes:
 - no small step horizontally explicit, vertically implicit
 - divergence damping option and vertical time off-centering
 - external-mode filtering option
- Upper boundary absorption and Rayleigh damping
- Lateral boundary conditions
 - idealized cases: periodic, symmetric, and open radiative
 - real cases: specified with relaxation zone
- Full physics options for land-surface, planetary boundary layer, atmospheric and surface radiation, microphysics and cumulus convection
- Ocean models
- Grid analysis nudging using separate upper-air and surface data, and observation nudging
- Spectral nudging
- Digital filter initialization
- Adaptive time stepping
- Orographic gravity wave drag
- Stochastic kinetic-energy backscatter scheme
- A number of idealized examples

Finally, several programs shipped with the WRF model package are used for post-processing and visualization of model outputs, including RIP4 (Stoelinga, 2009), NCAR Graphics Command Language (NCL) (University Corporation for Atmospheric Research, 2015), and conversion programs for other readily available graphics packages like GrADS (Doty, 1995).

3.2 WRF CHEM MODEL

WRF/Chem model is modularly built on top of WRF and maintains consistency with the WRF model I/O Application Program Interface (I/O API). WRF-Chem package adds the following chemistry functionalities to WRF (in addition to resolved and non-resolved transport) (Department of Commerce-National Oceanic Atmospheric Administration, The Cooperative Institute for Research in Environmental Sciences, The University Corporation for Atmospheric Research, The National Center for Atmospheric Research, The Max Planck Institute, The University of Chile, Centro de Previsão de Tempo e Estudos Climáticos, 2015):

- Dry deposition, coupled with the soil/vegetation scheme
- Four choices for biogenic emissions
- Three choices for anthropogenic emissions
- Several choices for gas-phase chemical mechanisms
- Three choices for photolysis schemes
- Five choices for aerosol schemes
- Aerosol direct/indirect effect
- Passive tracer transport of greenhouse gases
- Volcanic ash aerosol scheme
- A plume rise model for emissions of wildfires

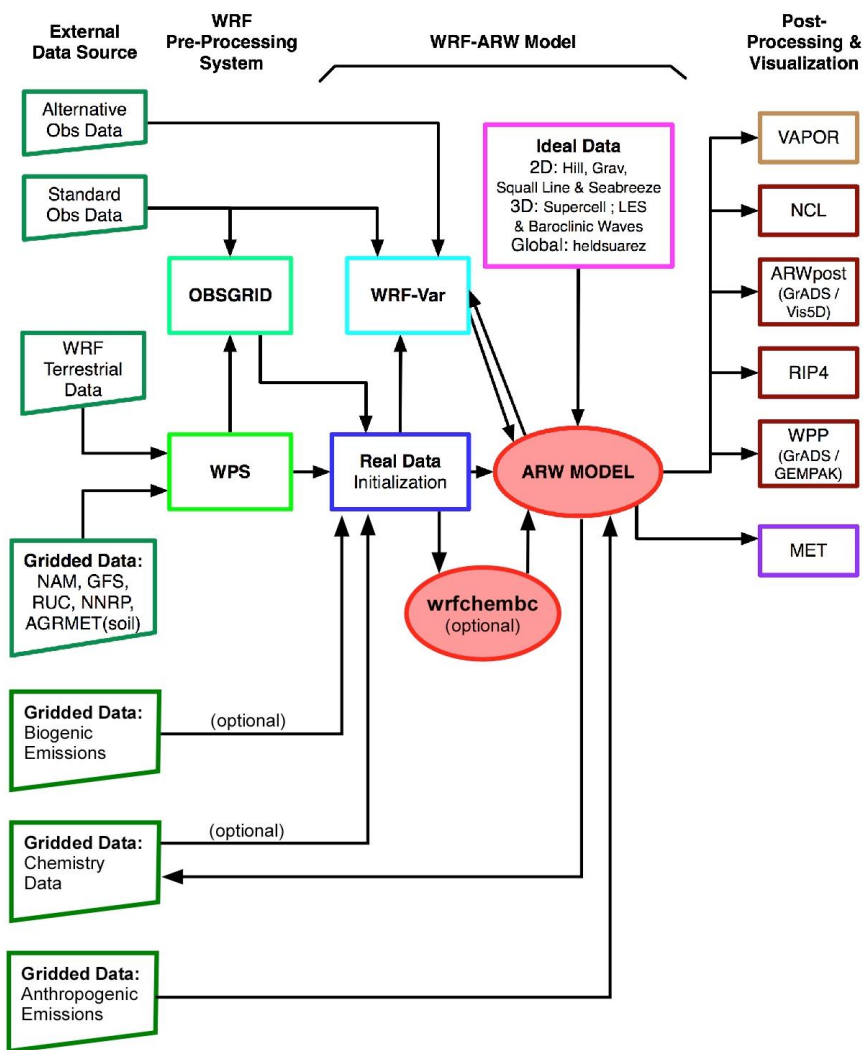


Figure 7 WRF/Chem modelling system main components (Department of Commerce-National Oceanic Atmospheric Administration, The Cooperative Institute for Research in Environmental Sciences, The University Corporation for Atmospheric Research, The National Center for Atmospheric Research, The Max Plank Institute, The University of Chile, Centro de Previsão de Tempo e Estudos Climáticos, 2015)

As shown in Figure 7, the WRF-Chem modeling system follows the same structure as the WRF model by consisting of these major components:

- The WRF Pre-Processing System
- WRF-Var data assimilation system
- WRF solver (ARW core only) including chemistry
- Post-processing and visualization tools

The difference with regular WRF comes from the chemistry part of the model needing to be provided additional gridded input data related to emissions. This additional input data is provided either by the WRF Pre-Processing System (dust emission fields), or read in during the real.exe initialization (e.g., biomass burning, biogenic emissions, GOCART background fields, etc.), or read in during the execution of the WRF solver (e.g., anthropogenic emissions, boundary conditions, volcanic emissions, etc.).

3.3 WRF/CHEM-HG MODEL DEVELOPED AT CNR/IIA

The modelling group at CNR/IIA has developed the WRF/Chem-Hg model to perform simulations of the Atmospheric chemistry of Hg, used mainly over Europe and the Mediterranean area with different oxidation

mechanisms (Gencarelli, et al., European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux, 2015) (Žagar, et al., 2014).

This section provides an overview of the WRF/Chem-Hg model. Further details of the model, and the various physics options employed can be found in the literature (Gencarelli, et al., European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux, 2015) (Gencarelli, De Simone, Hedgecock, Sprovieri, & Pirrone, Development and application of a regional-scale atmospheric mercury model based on WRF/Chem: a Mediterranean area investigation, 2014) (Žagar, et al., 2014).

WRF/Chem-Hg is built on top of WRF/Chem (version 3.4). Gas phase chemistry of Hg and a parametrized representation of atmospheric Hg aqueous chemistry have been added to the RADM2 chemical mechanism using KPP (Sandu & Sander, 2006) and the WKC coupler (Salzmann & Lawrence, 2006). Anthropogenic and natural Hg emissions have also been included and dry and wet deposition processes for Hg have been implemented.

Meteorological input was from the Global Forecast System (National Oceanic Atmospheric Administration, National Centers for Environmental Information, 2015) (1° by 1°) at six hourly intervals.

Both anthropogenic and natural emissions have been considered as input to the model:

- AMAP/UNEP 2010 (Arctic Monitoring Assessment Programme, UNEP Chemicals Branch, 2013) inventory as input dataset about Hg Speciation
- RETRO (ENES, 2005) and EDGAR v4 (PBL Netherlands Environmental Assessment Agency, 2001) datasets for non-mercury emissions.
- Hg emissions from biomass burning, from FINNV1 (NCAR, 2015) emissions inventory, making use of the Hg/CO enhancement ratio
- The WRF/Chem built in plume rise module
- WRF/Chem built in Guenther scheme for online calculation of non-Hg species biogenic emission
- Evasion of Hg⁰ from the sea surface, as described in Gencarelli et al, 2014 (Gencarelli, De Simone, Hedgecock, Sprovieri, & Pirrone, Development and application of a regional-scale atmospheric mercury model based on WRF/Chem: a Mediterranean area investigation, 2014) and in Žagar et al, 2014 (Žagar, et al., 2014)

The model was implemented on a dedicated workstation, simulations were configured and started by a local operator.

- Simulations can be configured to run using different parameters; for a discussion of obtained results the reader is referred to Gencarelli et al. 2015 (Gencarelli, et al., European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux, 2015) and in Žagar et al, 2014 (Žagar, et al., 2014)

4 IMPLEMENTING WRF/CHEM-HG AS A MAAS

The first required step to make the WRF/Chem-Hg model interoperable according to the Model Web initiative, is to have it published as a web service. According to the MaaS paradigm, once completed this step, a web service interface is published in front of the model and it becomes possible to remotely start a run through an appropriate web service request.

The OGC WPS service has been chosen as the Web service interface to be implemented. The following section describes its cloud based implementation; in the last section the correspondent BPMN description of the model will be presented.

4.1 PUBLICATION OF THE MODEL AS A WPS IN THE CLOUD

Cloud computing is a model for enabling ubiquitous, convenient, on-demand access to a shared pool of configurable computing resources. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centers.

The WRF/Chem-Hg model deployment has been made as a virtual machine instance on the Amazon Web Services (AWS) Cloud Computing Infrastructure (CCI), to leverage its convenient cloud capabilities. Example given, by using a CCI, it is possible to setup automatic allocation of new instances that are exact copies of the original virtual machine (i.e. auto-scaling) in case of an increased need of computational power. This could be the case of multiple simulations to be run at the same time (e.g. issued by different user requesting the service from different locations). A load balancer is in charge of delegating incoming requests to the available virtual machine instances.

A new virtual machine, based on Amazon Linux, has been instantiated on the CNR/IIA AWS cloud. The WRF/Chem-Hg model, has been compiled for the virtual machine environment, configured and installed on the instance.

At this point some offline tests have been run (e.g. launching locally the model with local input datasets and analyzing the outputs for correctness).

Following, the WPS interface has been chosen to be implemented as the web service frontend to the model.

WPS includes three fundamental and mandatory operations (Open Geospatial Consortium, 2007):

- **GetCapabilities:** used to retrieve the service capabilities (e.g. available processes, responsible organization)
- **DescribeProcess:** used to retrieve a full description of the process of interest (e.g. admissible inputs parameters, outputs and formats)
- **Execute:** used to run a specific process implemented by the server, using the provided input parameters. Inputs can be included directly in the Execute request, or reference web accessible resources. The outputs can be returned in the form of an XML response document, either embedded within the response document or stored as web accessible resources. If the outputs are stored, the Execute response shall consist of a XML document that includes a URL for each stored output, which the client can use to retrieve those outputs. Alternatively, for a single output, the server can be directed to return that output in its raw form without being wrapped in an XML response document. Normally, the response document is returned only after process execution is completed. However, a client can instruct the server to return the Execute response document immediately following

acceptance by the server of the Execute request. In this case, the Execute response includes a URL from which the response document can later be retrieved during and after process execution. The server can be instructed to provide regular updates to a measure of the amount of processing remaining if the process is not complete. This allows the client to determine the process status by polling this URL.

Apart from the technological difficulties that existed to correctly implement the WPS standard protocol and data model specification, the primary challenge of this work was perhaps here to define allowed and semantically meaningful input and output parameters to characterize the WRF/Chem-Hg model.

After discussing with the mercury model experts, the following parameters have been identified as the more useful to be selected to run a WRF/Chem-Hg simulation.

Input parameters:

- **Meteo data** (mandatory): a reference to an online dataset containing observations or simulations of meteorological data (e.g. pressure, snow, temperature, ...).
- **Anthropogenic emissions data** (optional): a reference to an online dataset containing observations or simulations of emissions related to human activity (e.g. Hg, Isoprene, PM₁₀, ...).
- **Bounding Box** (optional): used to select the spatial extent of the simulation. It shall coincide or be a subset of the spatial extent of the input datasets. By default, it's coincident with the input data extent.
- **Time Begin + Time End** (optional): used to select the temporal extent of the simulation. It shall coincide or be a subset of the temporal extent of the input datasets. By default, Time Begin is equal to the oldest time available in the input data and Time End is set to Time Begin + 6hrs.
- **Emission Scenario** (optional): used to select the mercury emissions to be activated in the simulation: anthropogenic emissions, marine emissions or both of them. By default, is set to marine emissions only.
- **Oxidation mechanism** (optional): used to select one of the two available oxidation mechanisms to be activated in the simulation (i.e. Br/BrO and O₃/OH). By default is set to Br/BrO.

The WRF/Chem-Hg WPS web service has been implemented using Java technologies (e.g. Java servlets, Apache CXF, JAXB). A Basic Access Authentication method (Network Working Group, 1999) has been setup to prevent access to unauthorized users and automatic crawlers.

The WPS has been deployed on the Apache Tomcat servlet container (version 7) on the same virtual machine instance of the model, in order to ease the run of the model and access to the results upon user client requests are issued (see Figure 8).

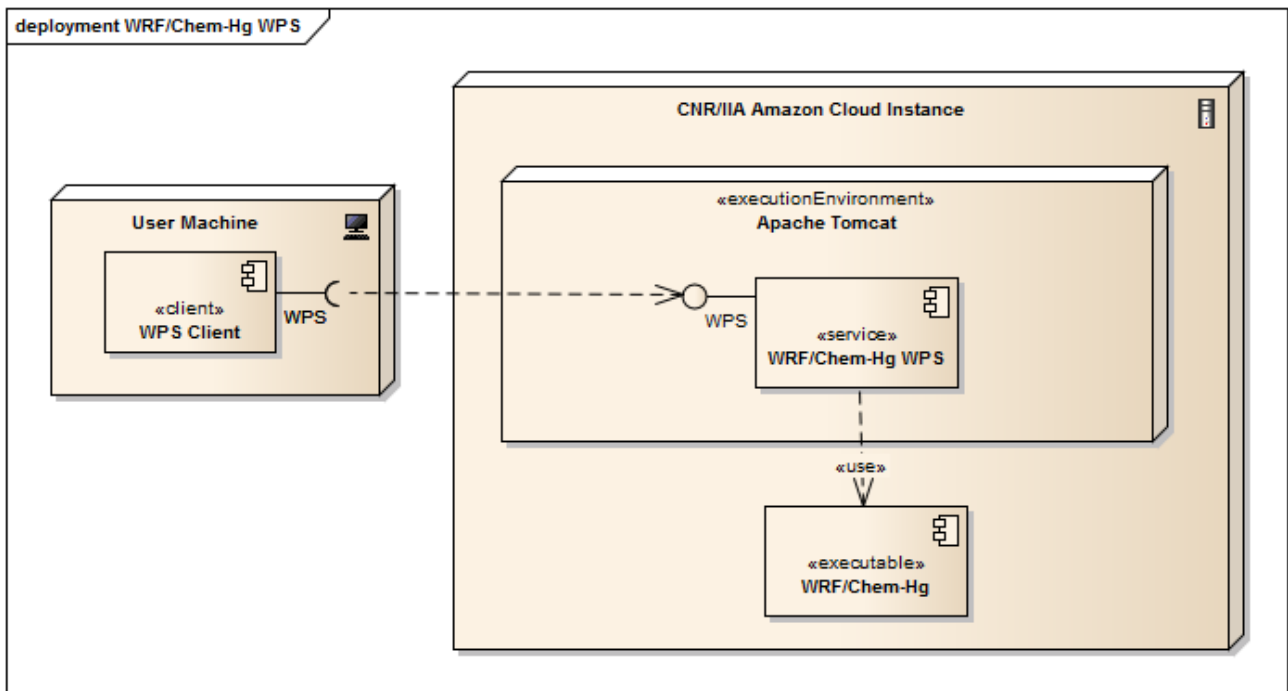


Figure 8 WRF/Chem-Hg deployment as a WPS in the CNR/IIA Amazon based cloud.

Figure 9 shows the sequence diagram of a typical interaction with WRF/Chem-Hg WPS started by a WPS user client, which include asynchronous execution of the model and retrieval of results. The diagram shows the interaction of the WPS service with the WRF/Chem-Hg model as well.

The main steps of the sequence are:

1. The user client issues a **GetCapabilities** request to the WRF/Chem-Hg WPS service.
 - 1.1. WPS replies with a Capabilities document version 1.0.0, advertising the WRF/Chem-Hg process among his available processes.
2. The user client issues a **DescribeProcess** request to WPS, requesting the description of the WRF/Chem-Hg process.
 - 2.1. WPS replies with a Process Description document, detailing the available model input parameters described above, including their types and obligation (optional/mandatory). Output parameters is the result of the simulation, encoded as a network Common Data Form (NetCDF) grid coverage.
 - 2.2. The user client is now able to visualize a form to be filled by the user with the parameters required in order to start the run; in case of a non-assisted client this step may be skipped and parameters be automatically filled in by the client itself.
3. The user client issues an **Execute** request to the WPS, including the parameters required to configure and start the simulation.
 - 3.1. WPS checks the input parameters. If some parameter is missing or incorrectly specified, then an explicative error message is provided. Otherwise the run can (asynchronously) start and a reply is immediately returned to the client. The execute response is an acknowledgment of having accepted a process in the execution queue. The “status location” attribute contained in the response is a link that can be further followed by the client to retrieve the running status.
 - 3.2. Model running is started by WPS after the process has been accepted and the queue of processes is free. As a first step, WPS reads the input parameters and resolve those who are expressed as references (e.g. input data is downloaded from the provided links). Then the WRF local environment is configured (e.g. appropriate parameters are set into the WRF namelist.input, according to user choices). The domain and input data is subsequently prepared for the run (programs *real.exe* and *echmeritbc.exe*). Finally, *wrf.exe* is run to start the simulation. After simulation is completed, the

resulting NetCDF output file containing the mercury emission variables is published on the WPS by making use of a HTTP publishing service.

3.3. The client polls at regular time intervals (e.g. 5 seconds) the “status location” link to be informed of the current status of the process execution. The XML response describe the process execution status, along with an information message. Process status can be one of the following:

- *Accepted*: The process has been accepted and is in the execution queue. In case another simulation is running, the process will wait until its termination.
- *Processing*: The simulation is being run. A textual message informs about the status of the simulation.
- *Completed*: The simulation has completed successfully. The response attribute “resource location” provides a link to the simulation output. The client can fetch it and visualize the NetCDF coverage.
- *Error*: An error has occurred during the simulation. A textual message in the XML response informs about the problem.

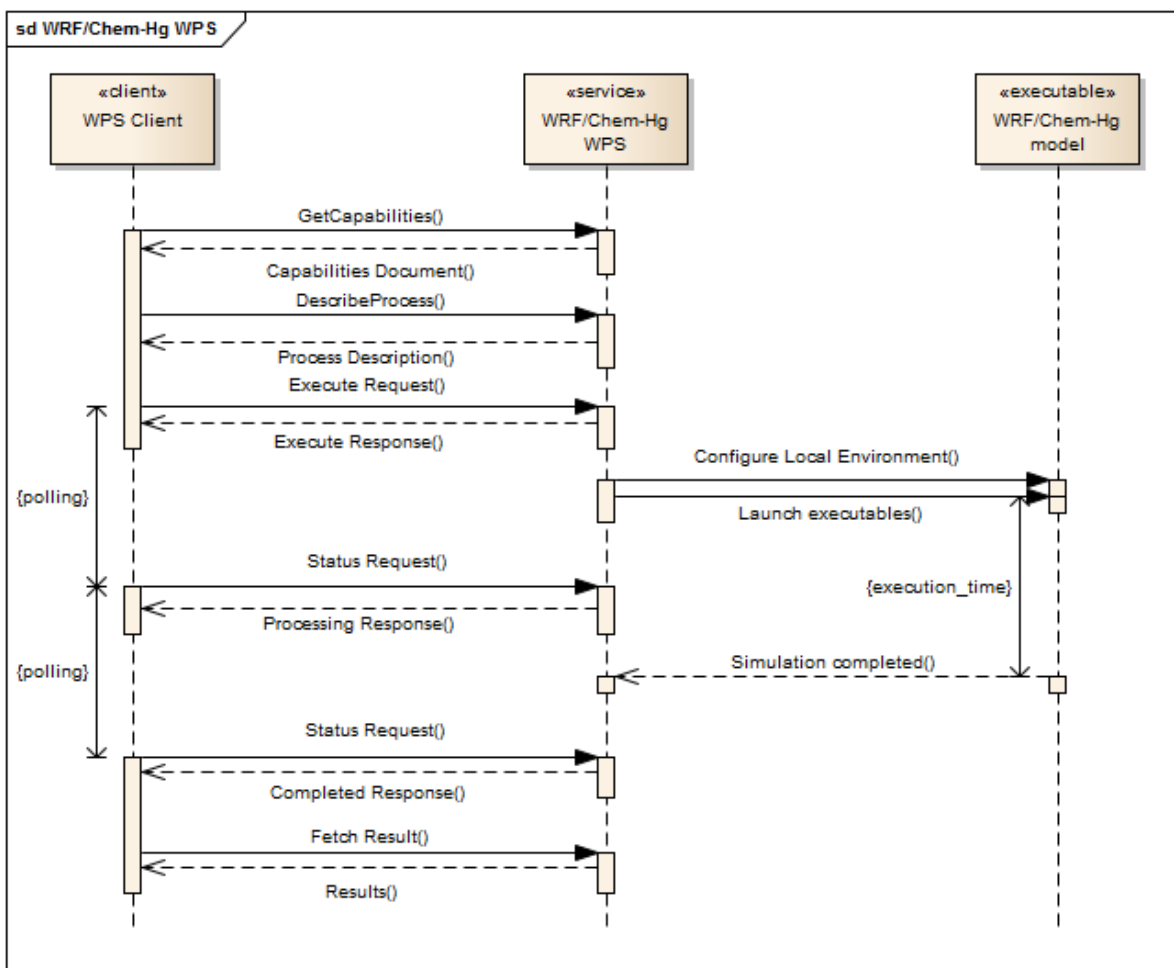


Figure 9 WRF/Chem-Hg WPS sequence diagram showing a typical interaction of an asynchronous execution of the model, started by a WPS client.

Publishing the model on the web immediately increases its interoperability: at this stage the model becomes usable by all web client applications, provided they have support for OGC WPS standard.

4.2 BPMN DESCRIPTION

Figure 10 depicts a BPMN description of the published WRF/Chem-Hg WPS service, realized using the Yaoqiang BPMN editor (Yaoqiang, 2009).

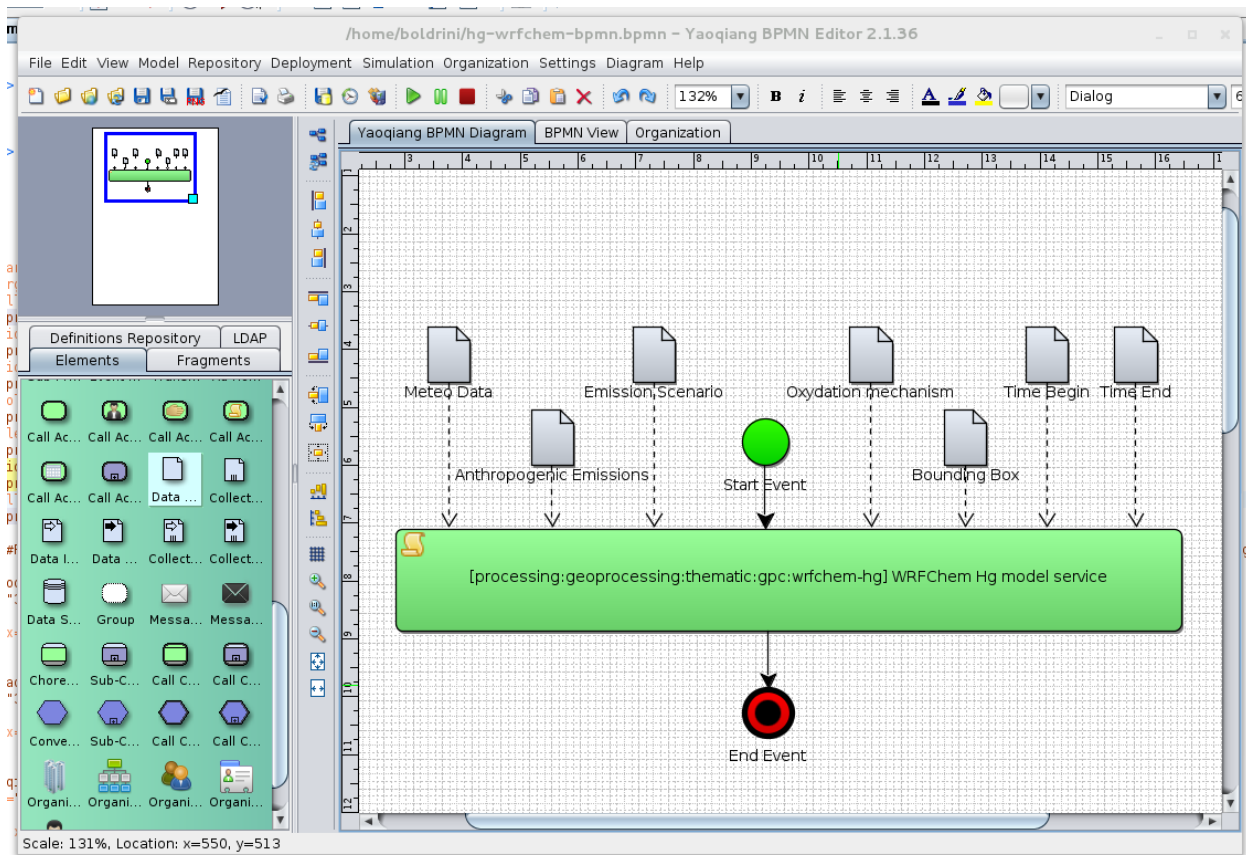


Figure 10 Editing the BPMN description of WRF/Chem-Hg WPS service with the Yaoqiang BPMN editor.

The process main task is labeled WRFChem Hg model service, and is tagged with the convention *processing:geoprocessing:thematic:gpc:wrfchem-hg*. The tag will be useful during the publication phase, to retrieve a matching adaptor capable to make the BP executable.

This simple Business Process provides a high level view of the process which both omits the technological detail and highlights the WRF/Chem-Hg input parameters. Represented as data inputs linked to the model, they are labeled as:

- Meteo data
- Anthropogenic emissions data
- Emission scenario
- Oxidation mechanism
- Bounding box
- Time begin
- Time end

Two more elements are present in the diagram, they are the start event and the end event. Representing the start and the end of model invocation, they can be used to connect this simple business process to other processes, realizing process chaining.

The abstraction represented by the BPMN makes possible for the scientists to concentrate on model aspects that are central for their work, such as multidisciplinary modelling and comparing simulation outcomes in case of different input parameters (such as different observation data).

5 THE E-WRF/CHEM-HG SYSTEM

The next step towards the Model Web approach is to plug the WRF/Chem-Hg model into the BP-broker architecture. The resulting *e-WRF/Chem-Hg* system increases the mercury model interoperability by making it available as a building block to compose complex workflows.

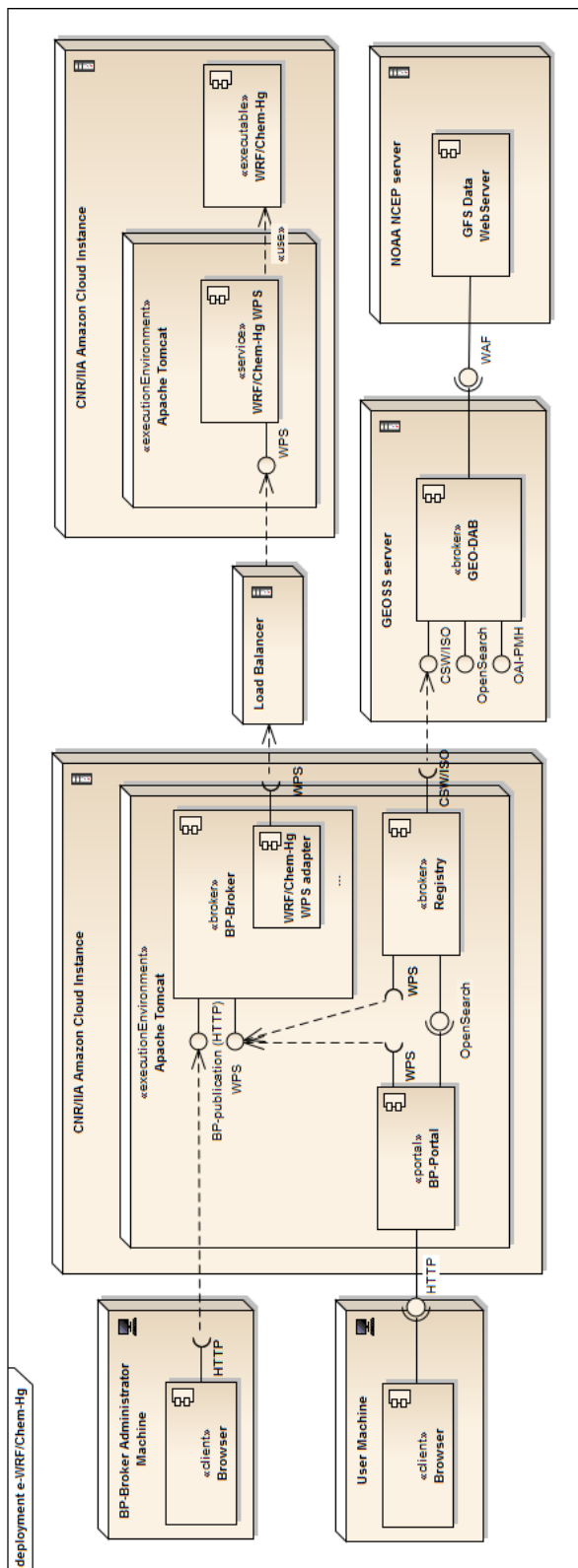


Figure 11 *e-WRF/Chem-Hg* system: deployment diagram.

Multiple steps are required to compose the complete system (see deployment diagram in Figure 11):

- **BP-Broker** and **Registry**: A BP-broker and a Registry instances are configured and deployed on the CNR/IIA cloud.
- **Model adaptor** realization: A new component with the role of accessing and mediating the WRF/Chem-Hg WPS functionalities is designed and developed. The component is then added to the modular and flexible architecture of the BP-Broker.
- **Publication** of the business process: This step consists in the publication of the BPMN diagram on the BP-broker. As a result, a new process related to this BP is created in the BP-broker, and the Registry is populated with a correspondent entry.
- The **BP-Broker Portal** is deployed as a user friendly frontend to the BP-Broker.

5.1 BP-BROKER AND REGISTRY

An additional CNR/IIA Amazon Cloud Instance is setup with an Apache Tomcat environment. A BP-Broker and a Registry are deployed and configured on it. The Registry (discovery broker) is configured with three accessors:

- **WPS accessor**, to periodically harvest the BP-Broker WPS interface, with the aim of finding new BP that may have been published
- **CSW/ISO Accessor**, configured with the GEO-DAB CSW/ISO endpoint, with the aim of discovering Geoscience resources coming from GEOSS that may be useful to the user
- **WAF Accessor**, configured with a local web folder, with the aim of discovering additional resources that may be useful to the WRF/Chem-Hg model and are not yet published in GEOSS (e.g. Anthropogenic emissions).

A MarkLogic database instance (MarkLogic, 2015) is configured as the storage backend for the Registry.

5.2 MODEL ADAPTOR REALIZATION

The WRF/Chem-Hg WPS adapter is designed and implemented. The adapter is a Java module which fits in the BP-Broker extensible architecture, to the aim of making accessible new service types: in this case new WRF/Chem-Hg WPS instances that may as well be deployed on different machines.

To realize a new adaptor five main methods of the Java abstract interface must be implemented:

- **getComponentType()**, it returns the string *processing:geoprocessing:thematic:gpc:wrfchem-hg*, advertising to the framework its ability to support (making executable) processes that are tagged with the same string
- **getInputMap()**, it must return an hash map describing the allowed inputs and their types. To implement this method the adaptor issues a DescribeProcess request to the WPS to receive the available inputs from the model. Default values are provided, for recognized input types.
- **getOutputMap()**, it must return an hash map describing the model outputs. It has been implemented to return a single output of type NetCDF coverage.
- **executeDry()**, the method used during execution simulation, that simulates the consumption of inputs in the stack before actual execution.
- **executeInternal()**, the actual method invoked during run. Its responsibility includes issuing the Execute request to the WPS, and subsequently checking the execution feedback at regular time intervals by issuing requests at the appropriate status location URL. Upon execution completion, it acquires the NetCDF result from the WPS and sets it as the Run output.

After implementation, the module has been packaged and deployed along with the BP-Broker instance already deployed.

5.3 PUBLICATION OF THE BUSINESS PROCESS

After the adapter has been realized and integrated in the BP-Broker, the administrator can publish the WRF/Chem-Hg WPS BP, described as the BPMN diagram of section 4.2, directly through the BP-Broker publication interface.

The BP-Broker is able at this point to obtain an EBP, by matching the annotated BPMN diagram with the new model adaptor.

5.4 BP-BROKER PORTAL

The BP-Broker Portal is deployed locally to the BP-Broker, although it is a separate component that communicates using standard service interfaces:

- **OpenSearch**: to execute queries to the Registry, and retrieve resources of interest for the user (BPS and their inputs)
- **WPS**: to request the BP-Broker to launch the BP of interest (including users provided by the user) and to retrieve the execution status and the result once the execution is finished.

The BP-Broker Portal represents the user entry point to access e-WRF/Chem-Hg functionalities. User can access its web interface using a common browser.

6 SYSTEM TESTS

This section is dedicated to the description of evaluation tests of the e-WRF/Chem-Hg system after its deployment, to assert and demonstrate the implemented functionalities.

6.1 THE BASIC WORKFLOW

The first section is dedicated to the description of the basic workflow, which is a business process composed by the mercury Hg model alone (including its inputs and outputs).

6.1.1 Discovery and evaluation of the mercury model

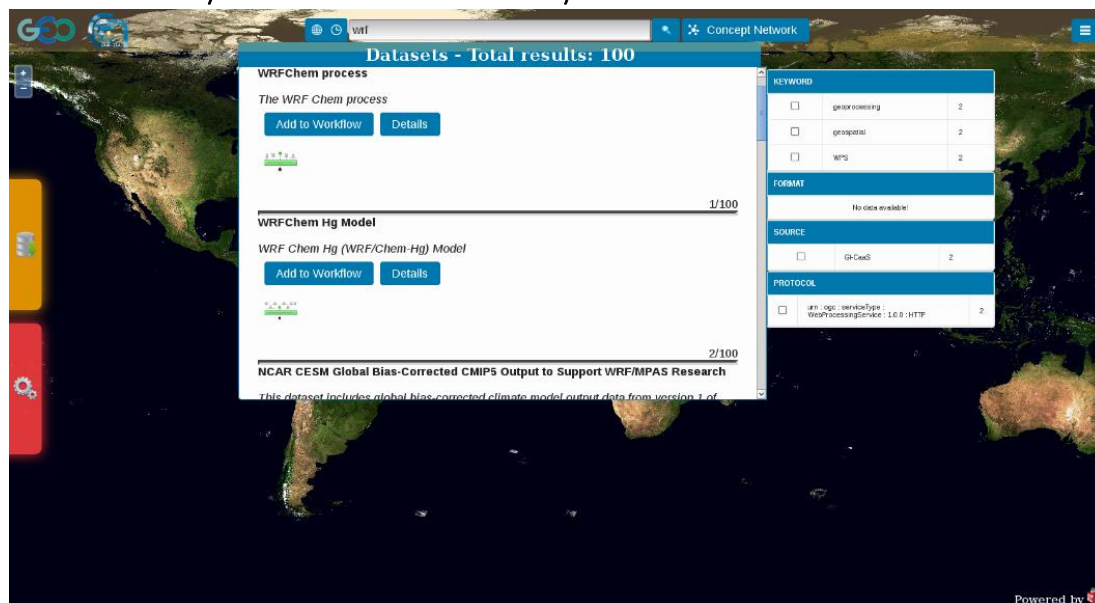


Figure 12 Search with keyword "WRF" for WRF based models and WRF related resources (e.g. simulation outputs).

The portal shows a user friendly web interface to search geoscience resources against the Registry discovery broker (see Figure 12). Several basic discovery criteria can be selected by user:

- The currently shown map represents the **spatial extent** that will be used for the query. If user is interested in local (e.g. regional) resources, he or she shall zoom on the interested area before performing the query, to exclude results from other regions.
- A **temporal extent** can be selected as well by clicking on the "clock" icon on the top panel. In this case only resources which represent observations included in the time range are returned.
- The text field on the top of the portal allows **text searches**: resources are accompanied with a metadata description, that (depending from the data provider) can be very thorough with more than 300 ISO 19115 (ISO TC211, 2003) metadata fields available. The text search for matching resources is exercised on all the metadata fields; a ranking is in place to put the more relevant and more accessible resources on top of the result set.
- A last tool that can improve discovery is the **semantic search** capability (see Figure 13): user can input a keyword (e.g. "temperature") and obtain a graph of related terms (e.g. "atmospheric temperature", "ocean temperature"). The terms are obtained by querying semantic services (i.e. web enabled ontologies) that can be explored by the broker by means of semantic queries for terms that are in a relation (e.g. more general, more specific, related, and corresponding) with the input term. The suggested terms can finally be included as query terms.

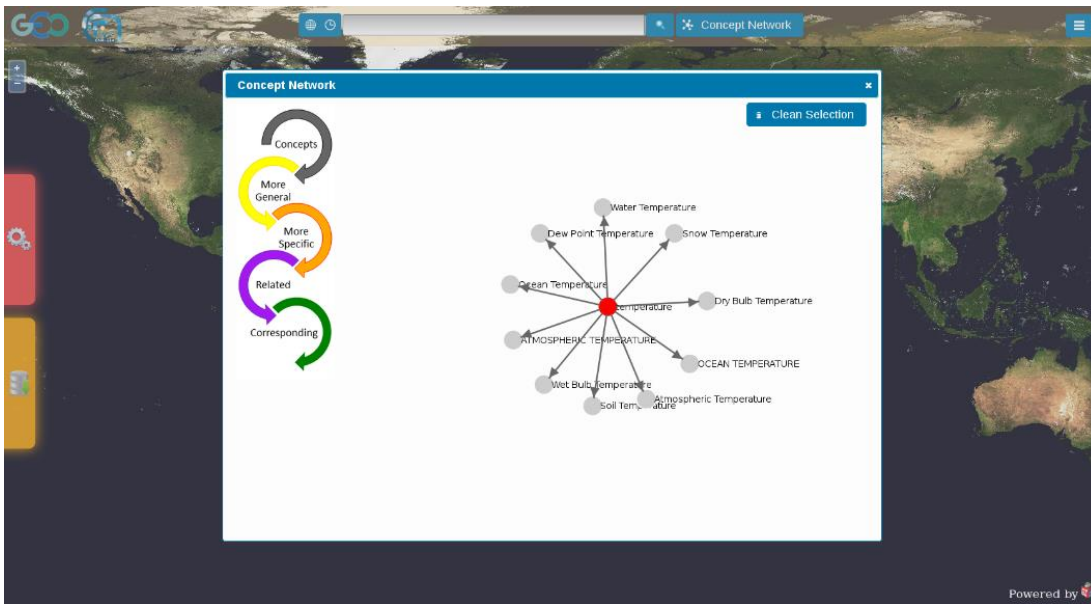


Figure 13 Semantic search capability enables user with semantic expanded terms.

Figure 12 shows a search made with keyword “WRF” on a global spatial extent. 100 paginated results are provided back by the discovery broker. The first ones are business processes found in the BP-Broker, WRF related resources coming from GEOSS are subsequently listed (the most of them are output of WRF simulations that have been shared to the GEO community).

By clicking on the “Details” button accompanying each resource, an overview of resource metadata is provided along with a preview (where available) to support resource evaluation.

It is possible to **download dataset resources**, provided an online endpoint is available in the metadata description and the service type is supported. In that case the Access Broker can be contacted to perform the download, according to the Common Grid Environment (CGE) specified by the user. CGE is a combination of format, grid resolution, spatial-temporal subset and Coordinate Reference System (CRS). The Access Broker can supplement the remote data provider if a required transformation (e.g. interpolation) is not provided by the source data provider.

It is also possible to perform **process execution**, in case a BP is returned as a result of the query. In this case the “Add to Workflow” button is clicked and the WRF/Chem-Hg BP is added to the workflow panel on the left making it available for configuration and execution.

6.1.2 Configuring and Running the model

The BPMN diagram is shown to the user upon clicking the saved workflow (Figure 14). At this stage user can proceed to configuration of input parameters.

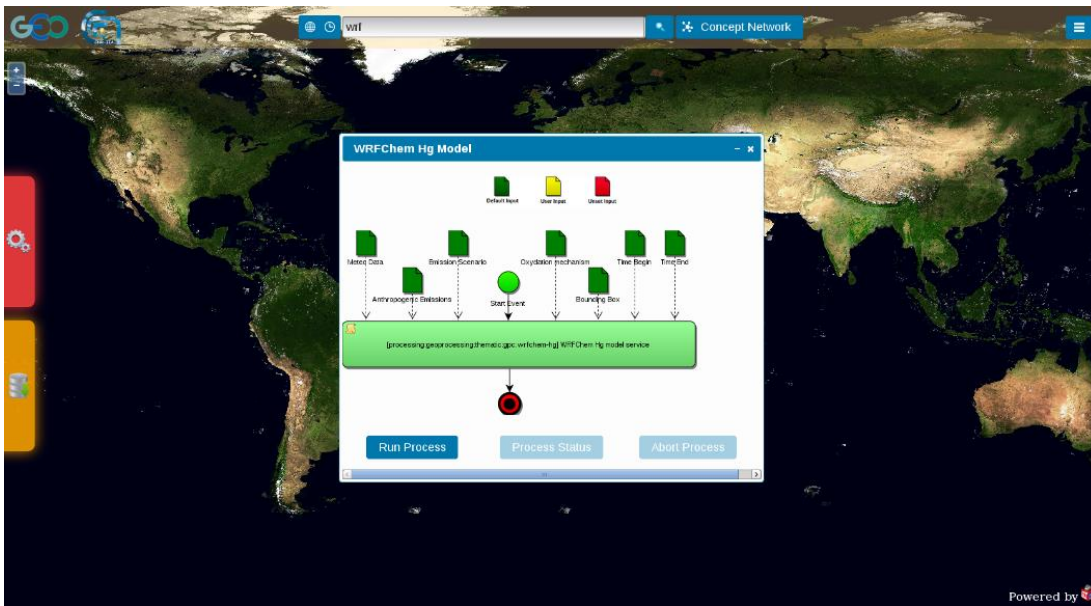


Figure 14 BP configuration of parameters prior to running.

User can directly click on the BPMN Data Objects to set their values. Data objects are displayed in three colors:

- Green data objects: they are already set with a value by default. User can still change them.
- Red data objects: they require user input because they are mandatory inputs with no default value set.
- Yellow data objects: after user has set a value for a data object, it becomes yellow.

Different inputs are differently rendered by the graphical interface:

- Emission scenario and Oxidation mechanism presents the user with a drop-down list to select from the finite set of allowed values.
- Bounding box input enables a selection of the area of interest directly on the map, by displaying a scalable rectangle to the user.
- Time begin and Time end presents a calendar widget to the user.
- Meteo data and anthropogenic emissions represents dataset inputs to the model. By clicking on them a search against the Registry broker is issued with the associated keyword.

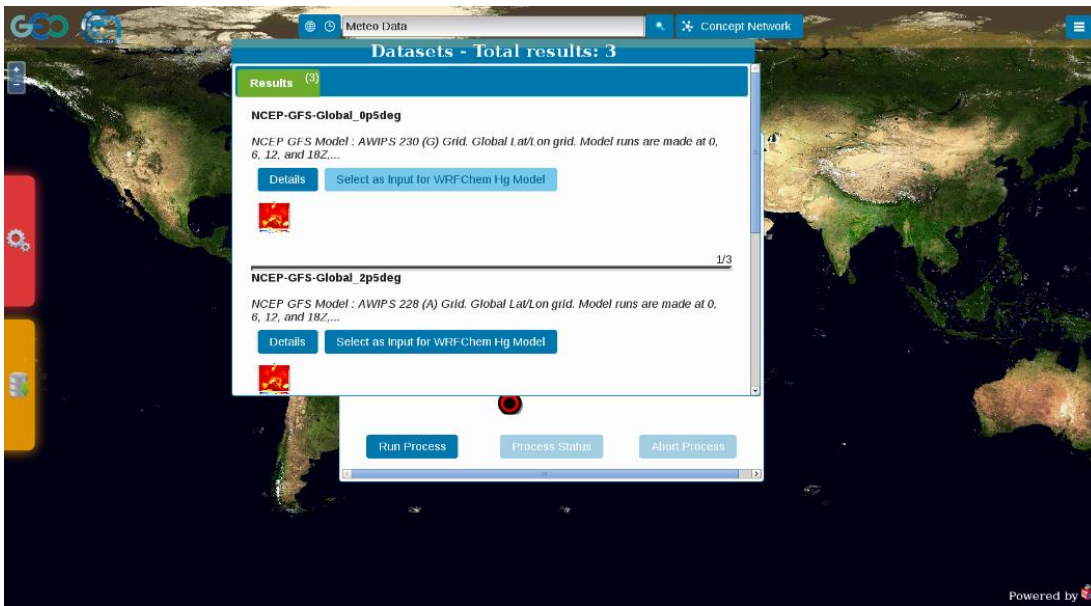


Figure 15 Discovery of required model input: GFS meteo data

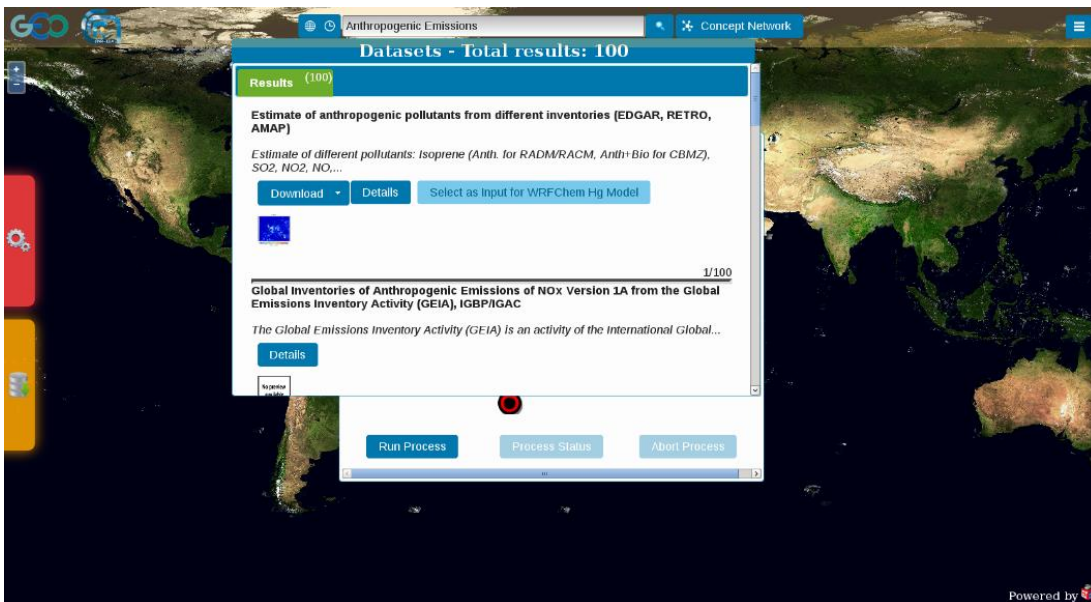


Figure 16 Discovery of required model input: Anthropogenic Emissions

Figure 15 and Figure 16 show the search results for model data inputs (Meteo data and anthropogenic emissions respectively), issued by clicking on the correspondent BPMN data objects. Results that can be used as inputs for the current BP present an additional button (reading “Select as input for WRF/Chem-Hg Model”).

In this case the portal sends the dataset distribution information (comprising the endpoint, service type and resource name) to the BP-Broker. It is model adaptor task to retrieve the actual data by possibly recurring to the Access Broker functionalities (if data transformations such as reformat or resampling are needed).

6.1.3 Execution

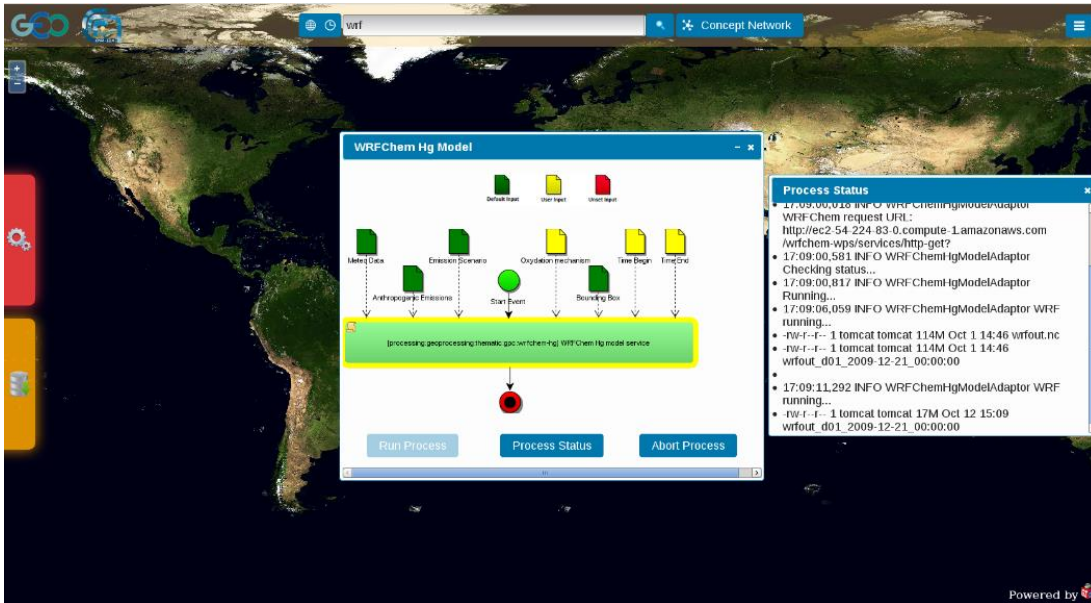


Figure 17 BP execution, highlighting the WRF-Chem-Hg model service, and showing status feedback on the right.

Finally, execution can be started. The component highlighting is switched from the start event to the BP main task to the end event, once finished.

The BP status can be regularly checked, by clicking on the central button “Process Status”. This will open a new dialog window, where the status information is updated as it is being received from the WPS.

At process end, the BP output represented by a NetCDF coverage is ready to be downloaded through an informative window.

6.1.4 Working with the result

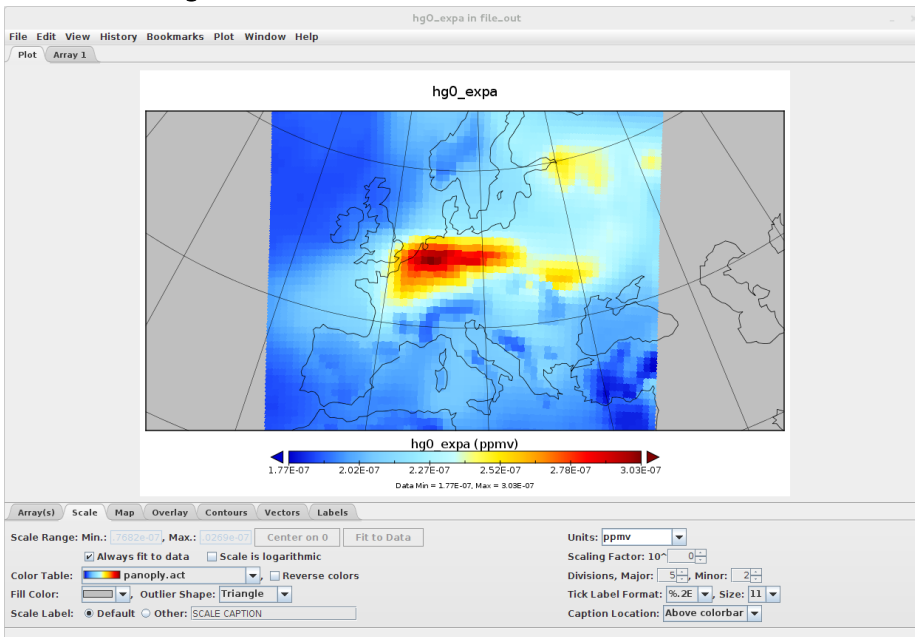


Figure 18 Download and evaluation of the results (Hg₀ concentration variable) with the NASA tool Panoply. (National Aeronautics Space Administration-Goddard Institute for Space Studies, 2015)

The NetCDF output can be viewed with any NetCDF enabled tool, such as Panoply (National Aeronautics Space Administration-Goddard Institute for Space Studies, 2015). Figure 18 shows the estimated

concentration of mercury, a variable that is included amongst the results of a simulation over Europe for December 2009, with the base oxidation mechanism.

6.2 A COMPLEX WORKFLOW

Aim of this section is to demonstrate the possibility to include the just realized BP in a complex workflow, which includes chaining of two BP.

6.2.1 A Hg Post Processing BP

To this aim a second process has been published as a WPS. Task of this service is to calculate dependent variables, starting from the output of the described mercury model. It takes as input a NetCDF file (named "Mercury Data" for simplicity) containing the following variables:

- perturbation pressure (P)
- base state pressure (PB)
- perturbation potential temperature (T)
- Hg0 concentration (hg0)
- Hg2 concentration (hg2)
- HgP concentration (hgp)
- Hg2 scavenging wet deposition, accumulated (WDS_HG2)
- Hgp scavenging wet deposition, accumulated (WDS_HGP)
- Hg2aq scavenging wet deposition, accumulated (WDS_HG2AQ)
- Hg2aq surface wet deposition, accumulated (WD_HG2AQ)
- Hg2aq surface dry deposition, accumulated (DD_HG2AQ)
- Hgp surface dry deposition, accumulated (DD_HGP)
- Hg2 surface dry deposition, accumulated (DD_HG2)

It processes them to provide as an output a NetCDF with the following variables:

- gaseous elemental mercury (GEM)
- reactive gaseous mercury (RGM)
- particle bound mercury (PBM)
- total mercury deposition (dry and wet) (depHgT)

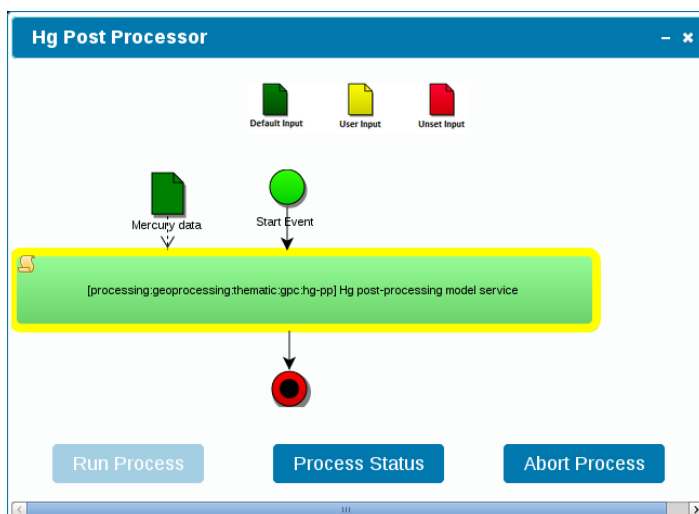


Figure 19 Hg Post Processing BP

The BP has been published on the BP-Broker, following the procedure already described for the mercury model which included creation of a new adapter for the Post and subsequently tested. Figure 19 shows the BPMN description of the Hg Post Processing service, as rendered by the BP-Portal during execution.

6.2.2 Composition of models

Thereafter it has been possible to run the WRF/Chem-Hg BP chained with the Hg Post Processing BP.

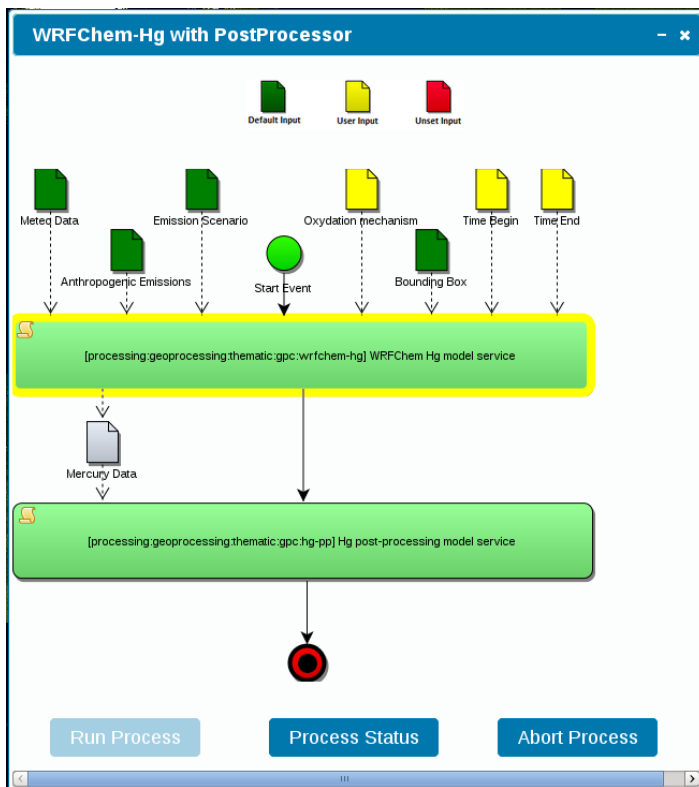


Figure 20 BP of a complex workflow including the WRF/Chem-Hg model

To this aim, it has been sufficient to draft the BPMN diagram shown in Figure 20. It is just a composition of the two BPMN diagrams described before. The output of WRF/Chem-Hg (indicated as “Mercury Data” in the diagram), acts as the input for the Hg Post Processing service.

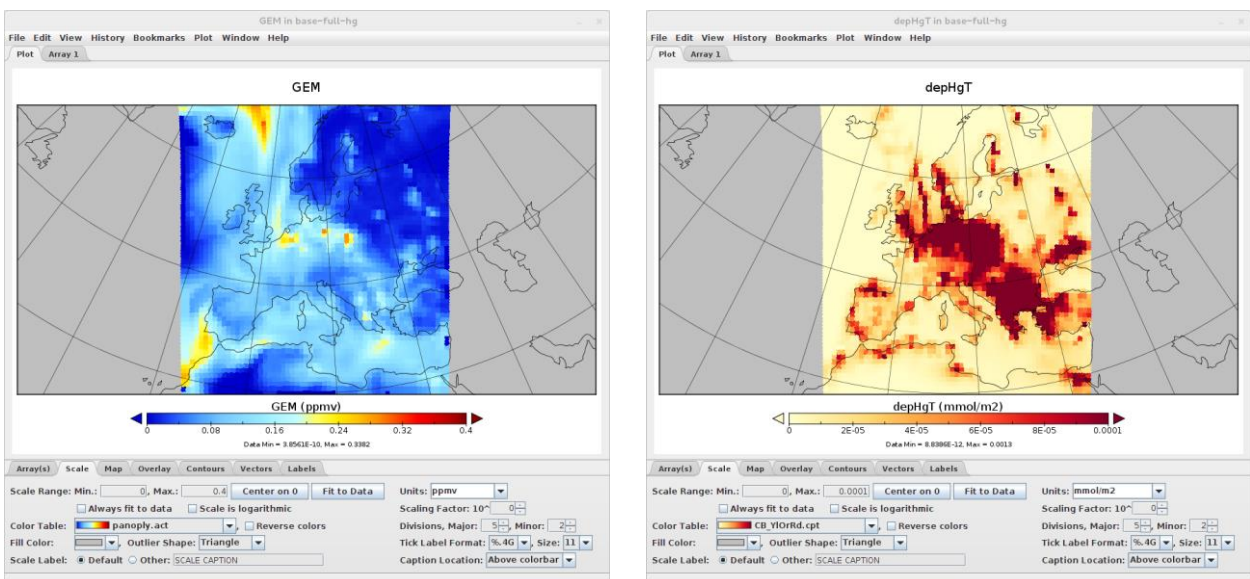


Figure 21 Gaseous Elemental Mercury (GEM) concentration and mercury deposition variables after a 1 week period simulation started on July 2005, with both marine and anthropogenic emissions activated.

The complete workflow has been run for a test over the first week of July 2005, with the Base oxidation mechanism turned on and both marine and anthropogenic emissions turned on. Two output variables which are results of the Post Processing (GEM and depHgT) are shown in Figure 21.

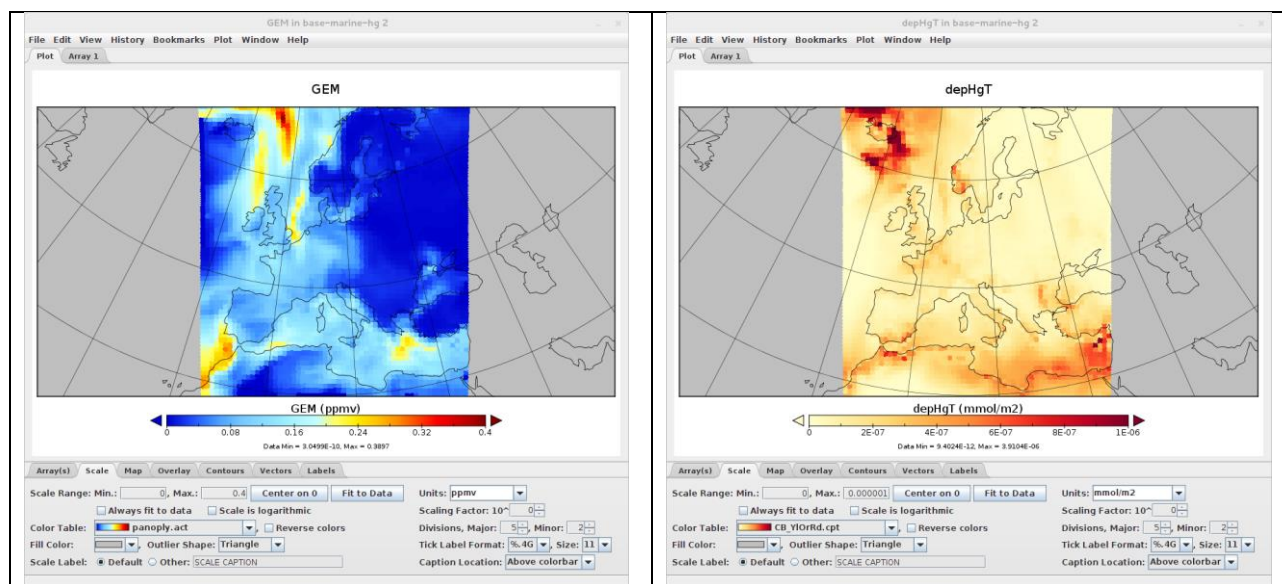


Figure 22 Gaseous Elemental Mercury (GEM) concentration and mercury deposition variables after a 1 week period simulation started on July 2005, with anthropogenic emissions switched off. Please note that the deposition map has a magnified (100x) scale with respect to the previous simulation, in order to highlight the areas where mercury is being deposited in this second case (although with a far less concentration).

Figure 22 is the result of a second run, obtained by simply switching off the anthropogenic emissions from the web interface (in this case only marine emissions contribute to the simulation). The two experiments were conducted with the aim of comparing mercury depositions derived from the two different scenarios, in particular highlighting the contributions of the anthropogenic emissions: these results successfully confirm the results already found by the CNR/IIA modelers (Gencarelli, et al., European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux, 2015).

Figure 23 A video demonstrating the e-WRF/Chem-Hg system has been published on the YouTube platform. To play it, it is sufficient to scan the QR code displayed in figure or point a browser to the URL: <https://www.youtube.com/watch?v=pY2QoibEFq>

Finally, a video demonstration of the e-WRF/Chem-Hg system has been published on the YouTube platform (see Figure 23), showing the final system functionalities.

7 DISCUSSION AND CHALLENGES

This research work has been focused on increasing interoperability of the mercury models developed at CNR/IIA, by harmonizing them according to the Model Web initiative and the Brokering approach.

The source model, developed at CNR/IIA and called WRF/Chem-Hg, ran before on a workstation system as an offline process. Input configuration, execution and output retrieval were manual operations executed by the expert modeler.

As a first step the model was moved from the offline workstation to the CNR/IIA cloud environment, which benefits of virtualization and load balancing capabilities. Then, a WPS service interface has been published as a frontend to the model, so as to be possible the remote execution and retrieval of results.

As a second step, a BPMN description of the mercury model has been drafted in coordination with the CNR/IIA modelers. Using this high level notation, scientists could concentrate on the important details relevant to their work such as characterization of input and output parameters.

The technological burden represented by the need of actually executing the model through an execution engine (including remote accessing of the required input data) is delegated to the BP-Broker, and in particular to a new component that has been added to this extensible framework called WRF/Chem-Hg WPS adapter.

The BPMN diagram describing the mercury model abstract BP could thus be published on the BP-Broker and be transformed to an EBP.

A user friendly portal has been implemented and deployed along the BP-Broker, to ease accessibility to the BP-Broker and make possible demonstration of the functionalities of the full e-WRF/Chem-Hg system.

With the BP-Portal it is possible to search for published EBP using a combination of spatial-temporal and free text (possibly semantic enhanced) constraints. It is then possible to configure the EBP prior to executing them. The needed parameters can be easily selected by the user. In case of input datasets (e.g. meteorological data and anthropogenic emissions), the search of available resources is started from the same web interface across a series of heterogeneous sources (such as the GEOSS data providers) leveraging the Brokering architecture of GEO-DAB.

By adopting Model Web and the Brokering approach, not only interoperability and accessibility is increased, but also repeatability of experiments: everyone can access and use the e-WRF/Chem-Hg system to reproduce the results found in the literature. This is in line with other use cases previously discussed which regarded increasing repeatability of computational results through a Brokering approach (Fuka, et al., 2015).

Composition of WRF/Chem-Hg BP in complex workflows has been demonstrated by designing and executing an additional BPMN diagram which includes chaining of the mercury model to a post processing service. Chaining the mercury model with models from other disciplines may be the subject of further investigations aimed at realizing multidisciplinary applications.

APPENDIX A: UML QUICK REFERENCE

The Unified Modeling Language (UML) is a graphical language for visualizing, specifying, constructing, and documenting a software related system in terms of its structure, behavior, and architecture (Object Management Group, 2015).

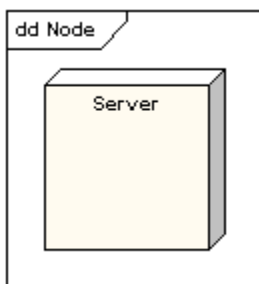
UML has been used throughout this document to describe components and deployments inherent to the e-WRF/Chem-Hg system. A quick guide to UML 2 notation is provided in the following sections; the reader can refer to the bibliography for a comprehensive UML dissertation (Object Management Group, 2015) (SPARX Systems, 2015).

7.1 DEPLOYMENT DIAGRAMS

A deployment diagram models the run-time architecture of a system. It shows the configuration of the hardware elements (nodes) and shows how software elements and artifacts are mapped onto those nodes.

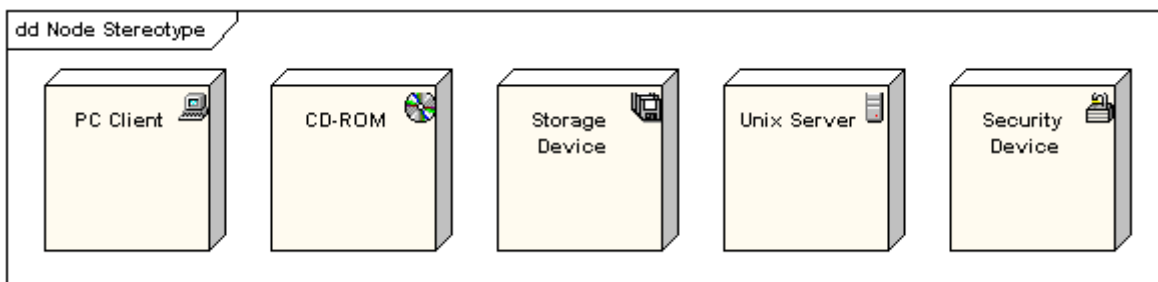
7.1.1 Node

A Node is either a hardware or software element. It is shown as a three-dimensional box shape, as shown below.



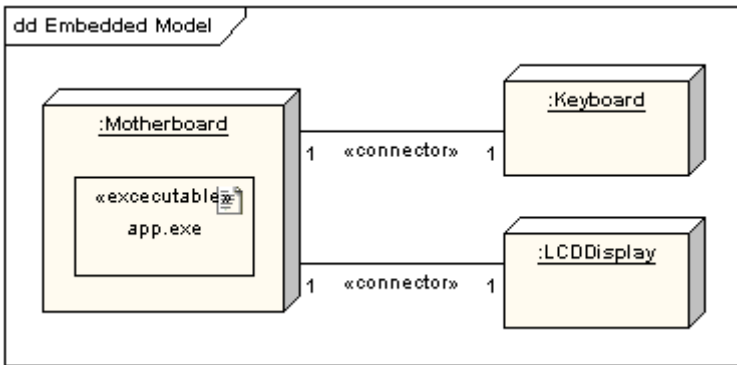
7.2 NODE STEREOTYPES

A number of standard stereotypes are provided for nodes, namely «cdrom», «cd-rom», «computer», «disk array», «pc», «pc client», «pc server», «secure», «server», «storage», «unix server», «user pc». These will display an appropriate icon in the top right corner of the node symbol



7.3 NODE AS CONTAINER

A node can contain other elements, such as components or artifacts. The following diagram shows a deployment diagram for part of an embedded system, depicting an executable artifact as being contained by the motherboard node.

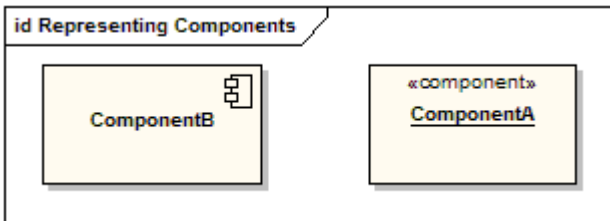


7.4 COMPONENT DIAGRAMS

Component diagrams illustrate the pieces of software, embedded controllers, etc., that will make up a system. A component diagram has a higher level of abstraction than a Class Diagram - usually a component is implemented by one or more classes (or objects) at runtime. They are building blocks so a component can eventually encompass a large portion of a system.

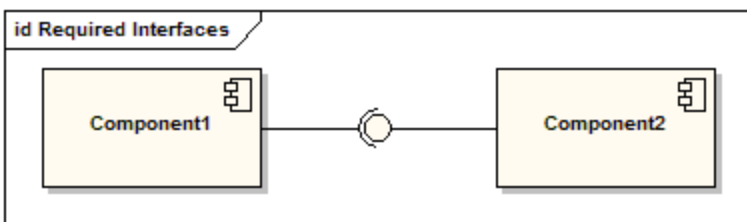
7.4.1 Representing Components

Components are represented as a rectangular classifier with the keyword «component»; optionally the component may be displayed as a rectangle with a component icon in the right-hand upper corner.



7.4.2 Assembly Connector

The assembly connector bridges a component's required interface (Component1) with the provided interface of another component (Component2); this allows one component to provide the services that another component requires.



APPENDIX B. BPMN QUICK REFERENCE

BPMN is a standard for business process modeling that provides a graphical notation for specifying business processes in a Business Process Diagram, based on a flowcharting technique very similar to activity diagrams from UML.

BPMN has been used throughout this document to describe the WRF/Chem-Hg BP (section 4.2), the Post Processor BP (section 6.2.1) and the complex workflow BP (section 6.2.2).

A quick reference to the BPMN elements used in this work follows. For a complete dissertation on BPMN diagrams the reader can refer to the bibliography (Object Management Group, 2010).

7.5 EVENTS

An Event is represented with a circle and denotes something that happens. Icons within the circle denote the type of event (e.g., an envelope representing a message, or a clock representing time).

7.5.1 Start event



Figure 24 Start event element

Acts as a process trigger; indicated by a single narrow border, is shown with an open (outline) icon.

7.5.2 End event



Figure 25 End event element

Represents the result of a process; indicated by a single thick or bold border, is shown with a solid icon.

7.6 ACTIVITY

An activity is represented with a rounded-corner rectangle and describes the kind of work which must be done.

7.6.1 Task



Figure 26 Task element.

A task represents a single unit of work that is not or cannot be broken down to a further level of business process detail without diagramming the steps in a procedure (which is not the purpose of BPMN).

7.7 ARTIFACTS

Allow developers to bring some more information into the model/diagram. There are three pre-defined Artifacts (Data Objects, Group, Annotation).

7.7.1 Data objects



Meteo Data

Figure 27 A data object element

Data objects show the reader which data is required or produced in an activity.

7.7.2 Group

A Group is represented with a rounded-corner rectangle and dashed lines. The group is used to group different activities but does not affect the flow in the diagram.

7.7.3 Annotation

An annotation is used to give the reader of the model/diagram an understandable impression.

7.8 CONNECTIONS

Flow objects are connected to each other using Connecting objects, which are of three types: sequences, messages, and associations.

7.8.1 Sequence Flow



Figure 28 Sequence flow element

A Sequence Flow is represented with a solid line and arrowhead, and shows in which order the activities are performed. The sequence flow may also have a symbol at its start, a small diamond indicates one of a number of conditional flows from an activity, while a diagonal slash indicates the default flow from a decision or activity with conditional flows.

7.8.2 Association

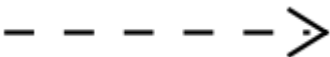


Figure 29 Association element

An Association is represented with a dotted line. It is used to associate an Artifact or text to a Flow Object, and can indicate some directionality using an open arrowhead (toward the artifact to represent a result, from the artifact to represent an input, and both to indicate it is read and updated). No directionality is used when the Artifact or text is associated with a sequence or message flow (as that flow already shows the direction).

LIST OF FIGURES

Figure 1 The GEOSS Portal enables discovery and access to the comprehensive data produced by the GEOSS community by connecting to the GEO-DAB, the broker (mediator) component in the GEOSS architecture... 5	5
Figure 2 The GMOS ground-based monitoring network. (CNR-Institute of Atmospheric Pollution Research, 2015)..... 6	6
Figure 3 Disciplinary Layered Client-Server architectures and the Brokering oriented architecture for multi-disciplinary environments (Nativi, Craglia, & Pearlman, The brokering approach for multidisciplinary interoperability: a position paper, 2012). 7	7
Figure 4 Scientific modelling of components related to global mercury cycle (Arctic Monitoring Assessment Programme, UNEP Chemicals Branch, 2013) 11	11
Figure 5 A global analysis plot of GFS surface temperatures, in degrees Celsius, for July 12, 2012 at 00UTC. (National Oceanic Atmospheric Administration, National Centers for Environmental Information, 2015) ... 11	11
Figure 6 Possible deployment of the BP-Broker. Users and the administrator access the BP-Broker functionalities through the user friendly web interfaces. The BP-Broker is able to access functionalities from different remote heterogeneous services..... 14	14
Figure 7 WRF/Chem modelling system main components (Department of Commerce-National Oceanic Atmospheric Administration, The Cooperative Institute for Research in Environmental Sciences, The University Corporation for Atmospheric Research, The National Center for Atmospheric Research, The Max Planck Institute, The University of Chile, Centro de Previsão de Tempo e Estudos Climáticos, 2015) 19	19
Figure 8 WRF/Chem-Hg deployment as a WPS in the CNR/IIA Amazon based cloud..... 23	23
Figure 9 WRF/Chem-Hg WPS sequence diagram showing a typical interaction of an asynchronous execution of the model, started by a WPS client. 24	24
Figure 10 Editing the BPMN description of WRF/Chem-Hg WPS service with the Yaoqiang BPMN editor.... 25	25
Figure 11 e-WRF/Chem-Hg system: deployment diagram..... 27	27
Figure 12 Search with keyword "WRF" for WRF based models and WRF related resources (e.g. simulation outputs). 30	30
Figure 13 Semantic search capability enables user with semantic expanded terms. 31	31
Figure 14 BP configuration of parameters prior to running..... 32	32
Figure 15 Discovery of required model input: GFS meteo data 33	33
Figure 16 Discovery of required model input: Anthropogenic Emissions 33	33
Figure 17 BP execution, highlighting the WRF-Chem-Hg model service, and showing status feedback on the right. 34	34
Figure 18 Download and evaluation of the results (Hg0 concentration variable) with the NASA tool Panoply. (National Aeronautics Space Administration-Goddard Institute for Space Studies, 2015) 34	34
Figure 19 Hg Post Processing BP 35	35
Figure 20 BP of a complex workflow including the WRF/Chem-Hg model 36	36
Figure 21 Gaseous Elemental Mercury (GEM) concentration and mercury deposition variables after a 1 week period simulation started on July 2005, with both marine and anthropogenic emissions activated. ... 36	36
Figure 22 Gaseous Elemental Mercury (GEM) concentration and mercury deposition variables after a 1 week period simulation started on July 2005, with anthropogenic emissions switched off. Please note that the deposition map has a magnified (100x) scale with respect to the previous simulation, in order to highlight the areas where mercury is being deposited in this second case (although with a far less concentration). 37	37
Figure 23 A video demonstrating the e-WRF/Chem-Hg system has been published on the YouTube platform. To play it, it is sufficient to scan the QR code displayed in figure or point a browser to the URL: https://www.youtube.com/watch?v=_pY2QoibEFg 37	37

Figure 24 Start event element.....	41
Figure 25 End event element	41
Figure 26 Task element.	41
Figure 27 A data object element	42
Figure 28 Sequence flow element	42
Figure 29 Association element	42

GLOSSARY

ABP: Abstract Business Processes	13
ACADIS: Advanced Cooperative Arctic Data and Information Service	9
AFWA: Air Force Weather Agency	16
ARC: Arctic Science Program	9
AWS: Amazon Web Services.....	21
BP: Business Process.....	13; 14; 25; 27; 28; 29; 31; 32; 33; 34; 35; 36; 38; 41
BPEL: Business Process Execution Language	12; 13
BPMN: Business Process Model and Notation	4; 13; 14; 21; 25; 26; 28; 29; 31; 32; 33; 36; 38; 41
CCI: Cloud Computing Infrastructure	21
CEOS: Committee on Earth Observation Satellites	9
CNR/IIA: National Research Council of Italy, Institute of Atmospheric Pollution Research. 4; 5; 6; 8; 9; 10; 16; 19; 21; 23; 28; 38	
CSW: Catalogue Service for the Web	8; 28
CTM: Chemical Transport Model.....	16
DAB: Discovery & Access Broker	4; 5; 8; 14; 28; 38
EBP: Executable Business Process	13; 14; 15; 29; 38
EC: European Commission.....	6
EMEP: European Monitoring and Evaluation Program	6
EO: Earth Observation	5
ESRI: Environmental Systems Research Institute	10
ESS: Earth and Space Science	4; 8; 12
EVA: Exploration, Visualization, and Analysis.....	9
FAA: Federal Aviation Administration	16
FP7: 7th Framework Program	6; 8; 9; 13
FTP: File Transfer Protocol.....	8
G8: Group of Eight	5
GEO: Group on Earth Observations.....	4; 5; 6; 8; 14; 28; 31; 38
GEOSS: Global Earth Observation System of Systems.....	5; 7; 8; 28; 31; 38
GeoViQua: QUALity aware VISualization for the Global Earth Observation System of Systems	9
GFS: Global Forecast System	12; 20; 33
GMOS: Global Mercury Observation System	4; 5; 6
H2020: Horizon 2020.....	9
ICSU: International Council for Science	9
JAXA: Japan Aerospace Exploration Agency.....	9
MaaS: Model as a Service.....	6; 12; 21
MMTF: Mercury Modelling Task Force	6
NCAR: National Center for Atmospheric Research	9; 16; 18
NCEP: National Centers for Environmental Prediction	12; 16
NetCDF: network Common Data Form.....	23; 24; 28; 34; 35
NSF: National Science Foundation	9
NSIDC: National Snow and Ice Data Center.....	9
OAI: Open Archive Initiative	8
ODIP: Ocean Data Interoperability Platform	8
OGC: Open Geospatial Consortium	8; 12; 21; 24
ORNL: Oak Ridge National Laboratory	9

PMH: Protocol for Metadata Harvesting.....	8
SBA: Societal Benefit Areas	5
SOA: Service Oriented Architecture	12; 13
TFHTAP: Task Force on Hemispheric Transport of Air Pollutants	6
UCAR: University Corporation for Atmospheric Research	9
UML: Unified Modeling Language	39; 41
WCS: Web Coverage Service	8
WMS: Web Map Service	8
WRF: Weather Research and Forecasting Model... 31; 33; 34; 36; 38; 41	4; 12; 16; 17; 18; 19; 20; 21; 22; 23; 24; 25; 27; 28; 29; 30;
WSDL: Web Service Description Language.....	12

BIBLIOGRAPHY

- Arctic Monitoring Assessment Programme, UNEP Chemicals Branch. (2013). *Technical Background Report for the Global Mercury Assessment 2013*. Tech. rep. Retrieved from <http://www.unep.org/hazardoussubstances/Mercury/Informationmaterials/ReportsandPublications/tabid/3593/Default.aspx>
- Ariya, P. A., Amyot, M., Dastoor, A., Deeds, D., Feinberg, A., Kos, G., . . . Toyota, K. (2015). Mercury Physicochemical and Biogeochemical Transformation in the Atmosphere and at Atmospheric Interfaces: A Review and Future Directions. *Chemical Reviews*, 3760-3802.
- Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., . . . Wagner, S. (2015). Big data analytics for earth sciences: the earth server approach. *International journal of digital earth*.
- BCube. (2015). *BCube Informatics Project Homepage*. Retrieved from <http://nsidc.org/research/informatics/bcube>
- Bigagli, L., Papeschi, F., Mazzetti, P., & Nativi, S. (2012). GeoViQua: quality-aware geospatial data discovery and evaluation. *Geophysical Research Abstracts*. Vienna.
- Bigagli, L., Santoro, M., Angelini, V., Mazzetti, P., & Nativi, S. (2011, 8). *UncertWeb Deliverable 2.2, Service frameworks for modelling resources*. Tech. rep. Retrieved from <http://www.uncertweb.org/documents/deliverables>
- Bigagli, L., Santoro, M., Mazzetti, P., & Nativi, S. (2015). Architecture of a Process Broker for Interoperable Geospatial Modeling on the Web. *ISPRS International Journal of Geo-Information*, 4, 647-660.
- Boldrini, E. & the GI-cat Team. (2010). A SOA broker solution for standard discovery and access services: the GI-cat framework. *Geophysical Research Abstracts*. Vienna.
- Boldrini, E., Craglia, M., Mazzetti, P., & Nativi, S. (2014). The Brokering Approach for Enabling Collaborative Scientific Research. In *Collaborative Knowledge in Scientific Research Networks* (pp. 282-302). IGI Books.
- Boldrini, E., Easton, Z., Fuka, D., Pearlman, J., & Nativi, S. (2015). Brokering technologies to realize the hydrology scenario in NSF BCube. *EGU General Assembly Conference Abstracts*. Vienna.
- Boldrini, E., Luzi, D., Nativi, S., & Pecoraro, F. (2015). Harmonising CERIF and INSPIRE metadata models to support multidisciplinary data sharing. *Int. J. Metadata, Semantics and Ontologies*, 10(2), 139-153.
- Boldrini, E., Nativi, S., & Schaap, D. (2013). INSPIRE compliant international standards for the SeaDataNet marine metadata. *Bollettino di Geofisica teorica ed applicata*, 54.
- Boldrini, E., Nativi, S., Santoro, M., Papeschi, F., & Mazzetti, P. (2012). GI-axe: an access broker framework for the geosciences. AGU Fall Meeting, San Francisco.
- Boldrini, E., Papeschi, F., Santoro, M., & Nativi, S. (2014). Extending the GI Brokering Suite to Support New Interoperability Specifications. *AGU Fall Meeting, San Francisco*.
- Boldrini, E., Papeschi, F., Santoro, M., & Nativi, S. (2015). Enabling interoperability in Geoscience with GI-suite. *EGU General Assembly Conference Abstracts*. Vienna.
- Box, G., & Draper, N. (1987). In *Empirical Model-Building and Response Surfaces* (p. 424). Wiley.

- Cinnirella, S., D'Amore, F., Bencardino, M., Sprovieri, F., & Pirrone, N. (2014, March). The GMOS cyber(e)-infrastructure: advanced services for supporting science and policy. *Environment Science and Pollution Research*, 21, 4193-4208.
- Cinnirella, S., D'Amore, F., Mazzetti, P., Nativi, S., & Pirrone, N. (2013). A Spatial Data Infrastructure for the Global Mercury Observation System. *E3S Web of Conferences*, 1, 28001.
- CNR-Institute of Atmospheric Pollution Research. (2015). *Global Mercury Observation (GMOS) System Data Portal*. Retrieved from <http://www.gmos.eu/sdi>
- CNR-Institute of Atmospheric Pollution Research. (2015). *Global Mercury Observation System (GMOS) Homepage*. Retrieved from <http://www.gmos.eu>
- David, O., Markstrom, S., Rojas, K., Ahuja, L., & Schneider, I. (2002). The Object Modeling System. In *Agricultural System Models in Field Research and Technology Transfer* (pp. 317-331). Eds., Lewis Publishers, CRC Press LLC, 2002.
- Department of Commerce-National Oceanic Atmospheric Administration, The Cooperative Institute for Research in Environmental Sciences, The University Corporation for Atmospheric Research, The National Center for Atmospheric Research, The Max Plank Institute, The University of Chile, Centro de Previsão de Tempo e Estudos Climáticos. (2015, 7). WRF-Chem Version 3.7 User's Guide. *WRF-Chem Version 3.7 User's Guide*. Retrieved from http://ruc.noaa.gov/wrf/WG11/Users_guide.pdf
- Doty, B. (1995, 9). The Grid Analysis and Display System (GrADS). *The Grid Analysis and Display System (GrADS)*. Retrieved from <http://dggf.uchile.cl/GF750/GrADS/gadoc151.pdf>
- Dubois, G., Skoien, J., Mendes de Jesus, J., Peedell, S., Hartley, A., Nativi, S., . . . Geller, G. (2011). eHabitat: A Contribution to the model web for habitat assessments and ecological forecasting. *Proceedings of the 34th International Symposium on Remote Sensing of Environment*. Sidney, Australia.
- ENES. (2005). *RETRO inventory homepage*. Retrieved from <http://retro.enes.org/>
- ESRI. (2014). *Creating Understanding Out of Measure, Knowledge, and Science*. Retrieved from <http://www.esri.com/~media/Files/Pdfs/news/arcuser/0314/creating-understanding.pdf>
- FP7 GeoViQua. (2015). *GeoViQua Homepage*. Retrieved from <http://www.geoviqua.org/>
- FP7 GEOWOW. (2012). *GEOWOW Fact Sheet*. Retrieved from http://www.geowow.eu/downloads/GEOWOW%20Project%20Fact%20Sheet_11072011.pdf
- FP7 Medina. (2015). *Medina Project Homepage*. Retrieved from <http://www.medinaproject.eu>
- FP7 MED-SUV. (2015). *MEDiterranean SUPersite Volcanoes Homepage*. Retrieved from <http://med-suv.eu/>
- FP7 ODIP. (2015). *ODIP Homepage*. Retrieved from <http://www.odip.eu/>
- FP7 SeaDataNet. (2015). *SeaDataNet Homepage*. Retrieved from <http://www.seadatanet.org>
- Fuka, D., Collick, A., MacAlister, C., Braeckel, A., Wright, D., Khalsa, S., . . . Easton, Z. (2015). Brokering as a framework for hydrological model repeatability. *EGU General Assembly Conference Abstracts*. Vienna.
- Gencarelli, C. (2013). Ph.D. dissertation, Bernardino Telesio Doctorate School of Science and Technique, University of Calabria, Cosenza, Italy.

- Gencarelli, C., De Simone, F., Hedgecock, I., Sprovieri, F., & Pirrone, N. (2014, 3). Development and application of a regional-scale atmospheric mercury model based on WRF/Chem: a Mediterranean area investigation. *Environmental Science and Pollution research*, 21(6), 4095-109.
- Gencarelli, C., De Simone, F., Hedgecock, I., Sprovieri, F., Yang, X., & Pirrone, N. (2015, 9). European and Mediterranean mercury modelling: Local and long-range contributions to the deposition flux. *Atmospheric Environment*, 117, 162-168. Retrieved from <http://dx.doi.org/10.1016/j.atmosenv.2015.07.015>
- Gilmour, J., Croke, B., & Watson, W. (2000). An Integrated Modelling Approach for Assessing Water Allocation Rules. In T. I. of Engineers ACT (Ed.), *Hydro 2000 3rd International Hydrology and Water Resources Symposium of The Institution of Engineers, Australia*. Perth, Western Australia.
- Glaves, H., & Schaap, D. (2015). Ocean Data Interoperability Platform (ODIP): developing a common framework for marine data management on a global scale. *Geophysical Research Abstracts*. Vienna.
- Global Mercury Observation System. (2014, 3). GMOS Mercury Modelling Task Force Kick-off meeting. *GMOS Mercury Modelling Task Force Kick-off meeting*. Retrieved from http://www.gmos.eu/freedocuments/gmos_mmtf_agenda.pdf
- Gregersen, J., Gijsbers, P., & Westen, S. (2007). OpenMI: Open Modelling Interface. *Journal of Hydroinformatics*, 9, 175-191.
- Group on Earth Observations. (2005, January). *The GEOSS 10-year implementation plan, February 16 2005*. Report on International Workshop on Spatial Data Infrastructures' Cost-Benefit / Return on Investment, European Commission, DG Joint Research Centre, Institute for Environment and Sustainability.
- Group on Earth Observations. (2014). *Report on Progress, Geneva Ministerial Summit, 2011-2013*. Tech. rep.
- H2020 EcoPotential. (2015). *EcoPotential Project Homepage*. Retrieved from <http://www.ecopotential-project.eu/>
- Hill, C., DeLuca, C., Balaji, V., Suarez, M., & da Silva, A. (2004). Architecture of the Earth System Modeling Framework. *Computing in Science and Engineering*, 6(1).
- Hillyer, C., Bolte, J., van Evert, F., & Lamaker, A. (2003). The ModCom modular simulation system. *Europ. J. Agronomy*, 18, 333-343.
- ISO TC211. (2003). *ISO 19115:2003, Geographic information – Metadata, ISO standard*. Tech. rep.
- Khalsa, S., Nativi, S., Pearlman, J., Pearlman, F., Santoro, M., Duerr, R., . . . Browdy, S. (2014). Achievements in Advancing the EarthCube Cyberinfrastructure Vision. *AGU Fall Meeting, San Francisco*.
- Khalsa, S., Parsons, M., Yarmey, L., Truslove, I., Pearlman, J., & Boldrini, E. (2013). The Advanced Cooperative Arctic Data and Information Service (ACADIS), Integrated Research Infrastructures and Services to users: supporting excellence in a science for society. *EGU General Assembly Conference Abstracts*. Vienna.
- Kraepiel, A. M., Keller, K., Chin, H. B., Malcolm, E. G., & Morel, F. M. (2003). Sources and Variations of Mercury in Tuna. *Environmental Science & Technology*, 37, 5551-5558. doi:DOI: 10.1021/es0340679
- MarkLogic. (2015). *Enterprise NoSQL Database MarkLogic*. Retrieved from <http://www.marklogic.com/>

- Maxwell, T., Villa, F., & Costanza, R. (2015). *Spatial Modeling Environment Homepage*. Retrieved from http://www.likbez.com/AV/Spatial_Modeling_Book/2/
- Miura, S., Sekioka, S., Kuroiwa, S., & Kudo, Y. (2015). The DIAS/CEOS Water Portal, distributed system using brokering Architecture. *General Assembly Conference Abstracts*. Vienna.
- Muñoz, M., Giovanni, R., Siqueira, M., Sutton, T., Brewer, P., Pereira, R., . . . Canhos, V. (2009). openModeller: a generic approach to species' potential distribution modelling. *GeoInformatica*, 15, 111-135.
- National Aeronautics Space Administration-Goddard Institute for Space Studies. (2015). *NASA GISS: Panoply 4, netCDF, HDF and GRIB Data Viewer Homepage*. Retrieved from <http://www.giss.nasa.gov/tools/panoply/>
- National Center for Atmospheric Research. (2015, 7). Advanced Research WRF Version 3 Modeling System User's Guide. *Advanced Research WRF Version 3 Modeling System User's Guide*. Retrieved from http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/ARWUsersGuideV3.pdf
- National Oceanic Atmospheric Administration, National Centers for Environmental Information. (2015). *Homepage of Global Forecast System*. Retrieved from <http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>
- Nativi, S., Bigagli, L., Mazzetti, P., Boldrini, E., & Papeschi, F. (2009). GI-Cat: A Mediation Solution for Building a Clearinghouse Catalog Service. *International Conference on Advanced Geographic Information Systems and Web Services*, (pp. 68-74).
- Nativi, S., Craglia, M., & Pearlman, J. (2012). The brokering approach for multidisciplinary interoperability: a position paper. *International Journal of Spatial Data Infrastructures Research*, 7, 1-15.
- Nativi, S., Mazzetti, P., & Geller, G. N. (2013). Environmental model access and interoperability: The GEO Model Web initiative, Environmental Modelling & Software. *Environmental Modelling & Software*, 39, 214-228.
- Nativi, S., Rorro, M., Mazzetti, P., Fiameni, G., Papeschi, F., & Carpenè, M. (2014). Building a multidisciplinary e-infrastructure for the NextData. *Geophysical Research Abstracts*. Vienna.
- NCAR. (2015). *Fire Inventory from NCAR (FINN) data homepage*. Retrieved from <http://bai.acom.ucar.edu/Data/fire/>
- Network Working Group. (1999, 6). *HTTP Authentication: Basic and Digest Access Authentication, RFC 2617*. Tech. rep. Retrieved from <http://tools.ietf.org/html/rfc2617>
- NextData. (2015). *NextData Homepage*. Retrieved from <http://www.nextdataproject.it>
- OASIS Web Services Business Process Execution Language TC. (2007, 4). Web Services Business Process Execution Language Version 2.0. *Web Services Business Process Execution Language Version 2.0*. Retrieved from <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>
- Object Management Group. (2010). Business Process Model and Notation (BPMN) Version 2.0. *Business Process Model and Notation (BPMN) Version 2.0*. Retrieved from <http://www.omg.org/spec/BPMN/2.0/>
- Object Management Group. (2015). *OMG Unified Modeling Language TM (OMG UML)*. Retrieved from <http://www.omg.org/spec/UML/2.5/PDF>

- Object Management Group. (2015). *Unified Modeling Language (UML) Resource Page*. Retrieved from <http://www.uml.org/>
- Open Geospatial Consortium. (2007, 6). OpenGIS Web Processing Service. OGC Implementation Specification OGC 05-007r7. *OpenGIS Web Processing Service. OGC Implementation Specification OGC 05-007r7*. Retrieved from <http://www.opengeospatial.org/standards/wps>
- PBL Netherlands Environmental Assessment Agency. (2001). Retrieved from Emision Database for Global Atmospheric Research Homepage: <http://themasites.pbl.nl/tridion/en/themasites/edgar/>
- Pirrone, N., & Cinnirella, S. (2012). The GMOS contributions to GEOSS. *Proceedings of the 16th International Conference on Heavy Metals in the Environment*. Rome. Retrieved from http://www.e3s-conferences.org/articles/e3sconf/pdf/2013/01/e3sconf_ichm13_1800c1.pdf
- Rahman, J., Seaton, S., Perraud, J.-M., Hotham, H., Verrelli, D., & Coleman, J. (2003). It's TIME for a New Environmental Modelling Framework. *Proceedings of MODSIM 2004 International Congress on Modelling and Simulation*, (pp. 1727-1732). Townsville, Australia. Retrieved from http://www.mssanz.org.au.previewdns.com/MODSIM03/Volume_04/C05/03_Rahman.pdf
- Red Hat, Inc. (2015). *JBPM – Open Source Business Process Management – Process engine Homepage*. Retrieved from <http://www.jbpm.org/>
- Salzmann, M., & Lawrence, M. (2006). Automatic coding of chemistry solvers in WRF-Chem using KPP. *7th WRF Users Workshop, Boulder, Colorado, USA*.
- Sandu, A., & Sander, R. (2006). Technical note: simulating chemical systems in FORTRAN90 and Matlab with the Kinetic PreProcessor KPP -2.1. *Atmospheric Chemistry and Physics*, 6, 187-195.
- Santoro, M. (2013, 3). *Environmental Model Composition in the GEO Model Web*. Ph.D. dissertation, Dipartimento di Ingegneria e Fisica dell'Ambiente, Facoltà di Ingegneria, Università della Basilicata, Potenza, Italy.
- Santoro, M., Boldrini, E., Papeschi, F., Baldini, A., & Nativi, S. (2010). *D.2.2.2. Specification of EUROGEOSS Initial Operating Capacity*. Retrieved from http://www.eurogeoss.eu/Documents/EuroGEOSS_D_2_2_2.pdf
- Santoro, M., Nativi, S., & Mazzetti, P. (2015). *Towards GEO Model Web implementation: a Brokering Service for Business Processes*. unpublished, Consiglio Nazionale delle Ricerche (CNR).
- Santoro, M., Papeschi, F., Boldrini, E., & Bigagli, L. (2011). GI-axe – an ICT Solution for Sensor Data Processing. *Geophysical Research Abstracts*. Vienna.
- Santoro, M., Pastres, R., Zucchetta, M., Venier, C., Roncella, R., Bigagli, L., . . . Nativi, S. (2014). Sustainable management of seagrass meadows: The GEOSS AIP-6 Pilot. *EGU General Assembly Conference Abstracts*. Vienna.
- Santoro, M., Sorichetta, A., Roglia, E., Craglia, M., & Nativi, S. (2013). GEOWOW: a drought scenario for multidisciplinary data access and use. *Geophysical Research Abstracts*. Vienna.
- Santoro, M., Wei, Y., Boldrini, E., Pearlman, J., Cook, R., & Nativi, S. (2013). Brokering Services to Evaluate, Visualize, and Analyze Terrestrial Biosphere Model Output and Observations. *EGU General Assembly Conference Abstracts*. Vienna.

- Serrano, J. J., Bertrand, F., Craglia, M., Nativi, S., Dubois, G., Vogt, J., . . . Pearlman, J. (2012). *EuroGEOSS FINAL REPORT*. Retrieved from http://www.eurogeoss.eu/Documents/EuroGEOSS_final_public_report.pdf
- SPARX Systems. (2015). *UML 2 Tutorial*. Retrieved from http://www.sparxsystems.com/resources/uml2_tutorial/
- Stoelinga, M. (2009, 3). A Users' Guide to RIP Version 4.5: A Program for Visualizing Mesoscale Model Output. *A Users' Guide to RIP Version 4.5: A Program for Visualizing Mesoscale Model Output*. Retrieved from <http://www2.mmm.ucar.edu/wrf/users/docs/ripug.htm>
- UNEP. (2015). Retrieved from Minamata Convention on Mercury: <http://www.mercuryconvention.org/>
- University Corporation for Atmospheric Research. (2015). *NCL Language Reference Guide*. Retrieved from http://www.ncl.ucar.edu/Document/Manuals/Ref_Manual/
- Van Ittersum, M. K., Ewert, F., Heckeley, T., Wery, J., Olsson, J., Andersen, E., . . . Wolf, J. (2008). Integrated assessment of agricultural systems – A component-based framework for the European Union (SEAMLESS). *Agricultural Systems*, 96, 150-165.
- Watson, F., & Rahman, J. (2004). Tarsier: a practical software framework for model development, testing and deployment. *Environmental Modelling & Software*, 19, 245–260.
- WMO. (2015). *WMO Information System*. Retrieved from http://www.wmo.int/pages/themes/wis/index_en.html
- World Meteorological Organization. (2003, 1). FM 92 GRIB Edition 2. *FM 92 GRIB Edition 2*. Retrieved from https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB2_062006.pdf
- World Wide Web Consortium. (2001). Web Services Description Language (WSDL) 1.1. *Web Services Description Language (WSDL) 1.1*. Retrieved from <http://www.w3.org/TR/wsdl>
- Yaoqiang, I. (2009). Yaoqiang BPMN Editor Homepage. *Yaoqiang BPMN Editor Homepage*. Retrieved from <http://sourceforge.net/projects/bpmn/>
- Žagar, D., Sirnik, N., Četina, M., Horvat, M., Kotnik, J., Ogrinc, N., . . . Pirrone, N. (2014). Mercury in the Mediterranean. Part 2: processes and mass balance. *Environmental Science and Pollution Research*, 21, 4081-4094.

ACKNOWLEDGEMENTS

I would like to express my special appreciation and gratitude to my supervisors Dr. Nicola Pirrone, Dr. Stefano Nativi and Prof. Vincenzo Carbone, who always provided help and guidance during the research work and introduced me to the exciting research activities going on in their groups and in the international context.

I would like to thank the colleagues of the CNR/IIA Rende Division, who provided the scientific expertise and technical support on the mercury models, essential for the thesis work, in particular Dr. Christian Natale Gencarelli, Dr. Ian Mickael Hedgecock and Francesco De Simone.

I would like to thank all the colleagues of the CNR/IIA Florence Division, who provided invaluable support on day-to-day research activities and during the thesis work, in particular Dr. Mattia Santoro, Roberto Roncella, Fabrizio Papeschi, Massimiliano Olivieri, Dr. Lorenzo Bigagli and Paolo Mazzetti.

Finally, I would like to thank my parents Maria Teresa Rossi and Giancarlo Boldrini and my wife Viola Foggi for their infinite support and patience.