018530 - SWITCH

Sustainable Water Management in the City of the Future

Integrated Project
Global Change and Ecosystems

Deliverable D1.4.4 (Includes D1.4.10 - D1.4.13 from original DoW)

The City Water Information System (CWIS) - User Manual

Due date of deliverable: 31/07/06
Actual submission date: 30/04/11

Start date of project: 1 February 2006 Duration: 63 months

Organisation name and lead contractor for this deliverable: Swiss Federal Institute of Technology, Lausanne (EPFL)

Revision [final]

Project co-funded by the European Commission within the Sixth Framework Programme (2006-2011)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Swiss Federal Institute of Technology, Lausanne (EPFL)

Authors: Bastien Roquier
Contact: marc.soutter@epfl.ch

Address:
Laboratoire d'Ecohydrologie (ECHO)
Faculté ENAC (Environnement naturel, architectural et construit)
Bâtiment GR, EPFL
Station 2
CH-1015 Lausanne

Web: http://echo.epfl.ch

SWITCH Deliverable Briefing Note

SWITCH Document: Performance assessment and scenario evaluation
Deliverable reference: D1.4.4 (includes D1.4.10, D1.4.11, D1.4.12 and D1.4.13 from original DOW)
Author(s) and Institution(s): Colin Schenk, Bastien Roquier, Marc Soutter / EPFL
Publication date: 30 Apr. 2011
Audience This deliverable is targeted to persons, mainly in the field of research, concerned by scenario planning related issues, especially in terms of data organisation and management.
Purpose To expose the language developments that were required to handle adequately the various aspects of integrated water management and especially the extensions needed to cope with scenario data, as well as provenance and uncertainty information.
Background At some point integrated water resources management (IWRM) requires strategic or scenario planning, i.e. the definition a various strategies for the future. These strategies need to be evaluated in order to make decisions. Evaluation should be based on performance indicators, some of which being evaluated with the help of simulation models. The City Water Information System, as a tool to back these processes, needs to be able to handle not only the logics of interrelations among the elements forming the water system, but also their time context (start and end of life) and situation context (base case and alternatives). These 3 dimensions (interrelation, time and situation) need to be carefully addressed when setting up the database structure, requiring additional developments to address specifically the situation/scenario related issues and their evaluation, including provenance and uncertainties issues as important dimensions in a decision making process.
Potential Impact These conceptual developments do form the basis to the latest revision of the database structure and related tools. As a consequence handling scenario data could be implemented and is now effective (see D1.4.3 and D.1.4.5).
Recommendations

A systems modelling framework to manage environmental information and scenario data in the field or integrated resources management

PhD Thesis

Bastien ROQUIER

Thursday, 24 March 2011

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Thesis advisors: Prof. A. Mermoud and Dr. M. Soutter

Environmental Engineering Institute

Laboratory of Ecohydrology (ECHO)

EPF Lausanne

Acknowledgements

This Thesis describes the results of the research I performed at EPFL in the Laboratory of Ecohydrology, with funding from the European Union through the SWITCH project (018530-2).

I first want to express my gratitude to my advisors Dr Marc Soutter and Prof. André Mermoud for the supervision and the opportunity to work in this exciting research topic.

My special thanks to Colin Schenk for the long and fruitful discussions about the concepts developed in this work, as well as for his expertise in the field of computer programming. I also acknowledge the work of Dr Marc Soutter, Philippe Brandenberg and Joël Alloh at EPFL, and Francois Vanderseypen (programmer at Orbifold), in developping the information system presented in this Thesis.

This work could not have been accomplished without the precious contributions of our scientific partners in the SWITCH project: I am particularly grateful to the colleagues of our working group: Prof. Rae Mackay and Ewan Last at the University of Birmingham, for the work carried out to integrate the model "City Water Balance" with our work; Dionysis Assimacopoulos and Elina Manoli at the University of Athens; Chris Duffy and Alison Jeffries at the University of Abertay, and Peter van der Steen, Zoran Vojinovic, Solomon Seyoum and Arlex Sanchez at UNESCO-IHE Delft. I also thank the members of the learning alliances of Alexandria, Birmingham and Belo Horizonte for the data sets.

I am also very grateful to all the members of the Laboratory of Ecohydrology, and especially to my office mate Xavier Beuchat, for their friendship and the pleasant work environment.

Finally, my gratitude goes to my wife, family and friends for their unconditional support and encouragement during these last years.

Abstract

Integrated natural resources management is an emerging discipline that aims to move beyond piecemeal approaches and promote integrated actions over a wide range of disciplines and actors. One of the technical challenges of this new discipline is the integration of information from different sources and with different levels of detail. Geographic information systems (GIS) have often been used for this purpose, but they are often inadequate to represent the structure and dynamics of human and natural systems.

This thesis aims to provide a new framework for modelling information in the field of integrated natural resources management. This framework relies on the development of a modelling language (called SYSMOD) that encodes the information required by the integrated management process. This language is based on the principles of systems approach and semantic integration that allow integrating, organising and sharing data from multiple knowledge domains.

To use the SYSMOD language, an information system was developed in the lab of EPFL Ecohydrology. This information system - known as "Combined Water Information System (CWIS)" because of its application to water management - consists of a database and a Web application. Its user interface provides a set of tools to create: (1) systemic views based on the SYSMOD language, (2) geographical views similar to those of GIS and (3) "report" views used to display and edit data such as text, files, images or numerical values.

CWIS was subsequently enhanced to manage scenario data and therefore provide support for scenario planning. Scenarios in strategic planning help to identify the impact of the most uncertain and important factors, and thus promotes the development of strategies adapted to several possible futures. In this context, CWIS has been extended to integrate the concepts of uncertainty and data provenance, to characterise the information validity and trace the sources of data.

Most of the time, the creation of scenarios and strategies requires data that can be provided only through simulations. Consequently, a modelling approach has been developed to enable the integration of mathematical models with CWIS. The particularity of this approach is to combine in a single method the concepts of model integration, uncertainty and data provenance to provide detailed results that highlight the risks and uncertainties associated with the scenarios and strategies.

CWIS was applied in a series of case studies, in particular the management of water resources in Alexandria (Egypt). CWIS, and in general the framework that was used in its development, have proved to be powerful tools to handle the different types of information in a holistic and transdisciplinary way. This thesis proved that scenario planning, systems thinking, semantic integration and model integration are essential aspects to be considered in the development of tools for integrated natural resources management.

Keywords: integrated natural resources management, scenario planning, systems thinking, model integration, semantic integration, ontology

Résumé

La gestion intégrée des ressources naturelles est une discipline récente qui vise à s'affranchir des limites des approches sectorielles en proposant des actions transversales couvrant un large champ de disciplines et d'acteurs. Un des enjeux techniques de cette nouvelle discipline est l'intégration d'informations provenant de différentes sources et caractérisées par des niveaux de détails et de qualité variables. Les systèmes d'information géographique (SIG) ont souvent été utilisés dans ce contexte, mais ils se révèlent parfois insuffisants pour représenter la structure et la dynamique des systèmes naturels et humains.

Cette thèse a pour but de fournir un nouveau cadre conceptuel pour la modélisation de l'information dans le domaine de la gestion intégrée des ressources naturelles. Elle s'appuie sur le développement d'un langage de modélisation (nommé SysMod) qui sert à d'encoder l'information nécessaire au processus de gestion intégrée. Ce langage est basé sur les principes d'approche systémique et d'intégration sémantique qui lui permettent de structurer et partager des données provenant de multiples domaines de connaissance.

Pour permettre l'utilisation du langage SysMod, un système d'information a été développé au sein du laboratoire d'Ecohydrologie de l'EPFL. Ce système d'information – dénommé "Combined Water Information System (CWIS)" en raison de son application au domaine de l'eau – est composé d'une base de données et d'une application Web. Son interface utilisateur propose une série d'outils pour créer : (1) des vues systémiques basées sur le langage SysMod, (2) des vues géographiques similaires à celles des SIG et (3) des vues "rapport" dédiées à la présentation et à l'édition de données telles que des textes, fichiers, images ou valeurs numériques.

CWIS a ensuite été perfectionné pour gérer les données de scenarios et ainsi fournir un support au "scenario planning" (planification par scénarios). L'utilisation de scénarios dans le processus de planification permet de prendre en compte l'effet des principaux facteurs d'incertitudes, et donc de favoriser la conception de stratégies adaptées à plusieurs futurs possibles. Dans cette optique, les fonctionnalités de CWIS ont également été étendues aux concepts de d'incertitude et de provenance, pour documenter respectivement la fiabilité et l'origine des données.

La création de scénarios et l'évaluation de stratégies nécessitent la plupart du temps des données ne pouvant être obtenues que par simulation. En conséquence, une approche a été

développée pour permettre l’intégration de modèles mathématiques avec CWIS. La particularité de cette approche est de réunir en une seule méthode les concepts d’intégration des modèles, d’incertitude et de provenance des données, afin de fournir des résultats documentés permettant de mettre en évidence les risques et incertitudes liés aux scénarios et stratégies étudiés.

L’outil CWIS a été appliqué dans une série d’études de cas, en particulier la gestion des ressources en eau à Alexandrie (Egypte). CWIS, ainsi que de manière générale le cadre conceptuel qui a servi à son développement, se sont révélés des outils performants pour traiter les différents types d’information de manière holistique et transdisciplinaire. Il ressort de la thèse que le scenario planning, la pensée systémique, l’intégration sémantique des données et le couplage de modèles de simulation sont des axes incontournables pour le développement d’outils dans le domaine de la gestion intégrée des ressources naturelles.

Mots-clés: gestion intégrée des ressources naturelles, scenario planning, pensée systémique, intégration de modèles, intégration sémantique, ontologie

Contents

1. Introduction

In the early 1990's, with the Dublin Statement on Water and Sustainable Development and the Rio Earth Summit, it became widely accepted that holistic, integrated approaches were necessary to head towards more sustainable practices in environmental management in general, and in natural resources management in particular. Water management was among the first and privileged domain where such integrated approaches and their implications were largely debated. The main principle of these approaches is to avoid scattered actions and measures, by applying concerted and consistent actions over a wide range of disciplines and policies. Because of the variety of the issues and the diversity of the methods that can be considered, it is difficult to give an exhaustive definition of integrated resources management (IRM). Theoreticians and practitioners generally agree on three main aspects (Lal et al., 2001): (1) the integration of multiple disciplines, (2) the integration of various spatial and temporal scales and (3) the involvement of multiple stakeholders.

While describing the issue of integrated water resources management (IWRM), a sub-branch of IRM, McDonnel (2008) states that: "At present the possibilities for truly integrated water resources management are limited, not by a conceptual framework, but by the ability to really represent the full dimensions of variables, interactions and complexity that come into play in any water management project or policy." Therefore IWRM, as well as IRM in general, require tools and methods to adequately represent and integrate environmental information. Geographic information systems (GIS) have often been used to fulfil this task. However, as mentioned by McDonnel (2008), "whilst GIS allow an integrated approach to visualisation and basic analysis of geographical data, more complex methods are needed to understand the feedback, interactions and dynamics of water resources systems." In this context, this thesis offers a new systems modelling framework for IRM, which consists of a set of concepts, tools and methods to organise and share environmental information, to represent the complexity and dynamics of the coupled human and natural systems and to support scenario planning.

The research is founded on four research areas which are detailed below.

1.1 Research areas

1.1.1 Integrated approaches to resources management

Integrated resources management advocates the use of holistic, interdisciplinary and participative processes. The paradigm of integrated resources management (IRM) mainly results from the Integrated Water Resources Management (IWRM) concept. IWRM emerged subsequently to the 1992 International Conference on Water and Environment in Dublin. It is inspired from the first Dublin principle for freshwater management and has been mainly popularised by the Global Water Partnership (GWP, 2000; Mitchell, 2005). It is now recognised as one of the most pertinent strategy to overcome the shortcomings of fragmented management approaches, as shown by the abundant literature on the topic (NRC 1999; Gleick 2003; Letcher 2005; Barreira 2006; Liu 2008; Savenije 2008).

According to the GWP, IWRM is “a process which promotes the coordinated development and management of water, land and related resources, in order to maximize the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems” (GWP, 2000). Although this definition seems quite simple, it does not provide further information about what aspects should be considered, as well as how to integrate them into the water management process.

To address the first issue “What aspects should be considered?”, the management process should basically include all the concerned stakeholders, and integrate information from the various knowledge domains and spatial-temporal scales involved in the problematic. The second issue “How to integrate them?” can be addressed by systems approaches (Mitchell, 2005; Saravanan, 2008), which provide a transdisciplinary framework.

1.1.2 Systems thinking

Whereas information systems typically used in IRM are centred on the spatial dimension of information (geographic information systems), systems approaches allow describing the real world complexity in terms of elements and interactions, including non-spatial components such as laws and policies.

Systems thinking provides an alternative approach to problem solving. It considers the problems as parts of an overall system, rather than focusing on individual outcomes. Systems thinking is therefore coherent with the philosophy of IRM which aims at going beyond the usual fragmented management of natural resources.

Systems thinking is the basis of several theories, such as systems theory (Cabrera et al., 2008), complexity theory (Max-Neef, 2005), chaos theory or system dynamics (Winz et al., 2009). These systems theories have been applied successfully in many scientific disciplines, e.g. biology, sociology, ecosystem science, political science, psychology, etc.

Recent studies based on systems approaches such as Bayesian belief networks or system dynamics to natural resources management obtained encouraging results, as illustrated by the work of Bellamy and Smith (Bellamy et al., 2001; Smith et al., 2007). In the same way, applications specific to water management succeeded in conducting integrated assessments (Bosch et al., 2007; Liu et al., 2008; Saravanan, 2008).

From a qualitative perspective, the soft system methodology introduced by Checkland (1981) is applied to systems with incomplete or conflicting knowledge. Its use is particularly adapted to the framework of participatory management, providing a systemic framework with quantitative assessment. In some cases, the soft system methodology has been combined with other approaches, such as systems dynamics, to tackle complex social issues (Mendoza and Prabhu, 2006; Paucar-Caceres and Rodriguez-Ulloa, 2007).

In addition, several methods, also based on a systems perspective, may provide relevant contributions to water management. Mental mapping is for instance a relevant method for participatory management and decision making (Kolkman et al., 2005). On the other hand, material flow analysis (Huang et al., 2007) and life cycle assessment (Lundie et al., 2004) are relevant methods to assess environmental (and sometimes economic) issues.

1.1.3 Scenario planning

Based on the development of scenarios that explore a wide range of possible futures and the creation of strategies that can be adapted accordingly, scenario planning is the most adequate decision-making process to handle uncertainties linked to the natural and human systems complexity.

As a key process in water management, scenario planning involves the development of scenario and strategies to achieve a vision produced by the decision makers. A strategy is elaborated to face a variety of scenarios, and should therefore be flexible and robust as regards the risks and uncertainties of the future. In that sense the definition of scenario planning is opposed to the notion of programme, which consists of a sequence of predetermined actions without flexibility for adjustments (Morin, 1990).

Scenario planning and strategic planning are concepts extensively explored in the literature (Peterson et al., 2003; Ralston and Wilson, 2006). Many methods have been developed

specifically for the context of natural resources management (Dewhurst and Kessler, 1999; Groves and Lempert, 2007; Mahmoud et al., 2009; Makropoulos et al., 2008). Scenario planning is generally defined as a stepwise approach involving scientists, policymakers, managers, and other stakeholders that work together to consensually develop strategies. The process starts with the identification of the problems and/or objectives. Then, based on the identified issues and the understanding of the system to be studied, the most uncertain and important factors are selected to build scenarios that represent divergent but plausible evolution of the system. Scenarios, as well as the strategic actions defined by the working group, can be tested in a variety of ways, possibly using simulation models to quantitatively assess the effects of both scenarios and strategies.

In order to select the most appropriate strategy, an assessment of the strategies' performance is required. Many tools are available for this purpose, as stated by the abundant literature on the topic. The first are the multi-criteria assessment methods (Hajkowicz and Collins, 2007; Hajkowicz and Higgins, 2008; Matos, 2007; Scott, 2007; Srdjevic et al., 2004). The multi-criteria approach is particularly interesting for the integration of decision makers' expectations through indicators. However, a recurrent issue of this approach is the weighting of the indicators. Optimisation procedures based on mathematical algorithms are also methods frequently used for strategies assessment and ranking (Liberatore et al., 2006; Maia and Schumann, 2007; Tan et al., 2002; Xevi and Khan, 2005; Zitzler and Künzli, 2004). However, these methods are mainly determinist and often present limitations in choosing flexible and robust strategies. In other words, they usually provide a single best solution, while a range of potential alternatives would be more relevant, with details about their uncertainty, flexibility and resilience.

To assess and plan strategies that cope with uncertain future evolutions, scenario planning need to monitor the system's evolution and adjust the strategies accordingly. Scenario planning is in that sense an extension of the concept of adaptive management (AM), which aims at dealing with complex uncertain systems. AM is an iterative process of decision making under uncertainty. In this regards, Medema (2008) reveals that "Often people think of AM as learning by doing, but this simplifies and misses the essential goal of needing to experiment with complex systems to learn from them." The concept of adaptive management emphasizes the fact that knowledge management and exploration tools are more relevant than optimization methods.

1.1.4 Integration of environmental data and models

Getting a complete knowledge of the coupled human and natural systems is not feasible, and would provide systems representations too complex for an effective management. Therefore 'integrated' approaches should focus "on the key components and relationships accounting for the greatest variability in system behaviour'' (Mitchell, 2005). Hence to ensure the adoption of a holistic framework, reference to a "generic model" which identifies the key components of the systems under consideration, is a prerequisite. Such a comprehensive and generic system model was developed by Schenk et al. (2009) for the field of water management.

In order to create a knowledge base for natural resources management, and given the fragmentary state of knowledge, the management of data faces the problem of semantic integration. This issue is frequently addressed by ontology languages, such as the Web Ontology Language (OWL) (Smith et al., 2004). Ontologies are usually saved in a file format, and some developments may be required to store them in a database.

Evaluating scenarios/strategies and making decisions relies or should rely on quantifiable criteria. Some of these may be advantageously estimated by running simulation tools. The simulation of future scenarios in IRM can be achieved by the integration of mathematical models into some chains of connected models. This approach, named integrated modelling, is supported by a series of tools referred to as integrated modelling frameworks (IMFs), such as OpenMI (Gregersen et al., 2007), TIME (Rahman et al., 2003) or OMS (David, 1997).

1.2 Thesis objectives

The object of this thesis is the development of a systems modelling framework to support collaborative scenario planning in the field of IRM. As stated earlier in this introduction, the ability to tackle the complexity and dynamics of coupled human and natural systems is an essential quality, which has oriented this thesis towards the development of an information sharing platform based on systems thinking. The main objectives of the research are to:

- Develop a system modelling language, to handle information from various disciplines, stakeholders and spatial-temporal scales.
- Implement the developed system modelling language into a relational database and a suite of software tools to form an information system.
- Extend these tools and concepts to support scenario data.
- Develop a modelling approach to assess scenarios and strategies.

1.3 Thesis outline

The first four chapters address the objectives listed above and are followed by a chapter devoted to the application of the research and a general conclusion.

Chapter 2 introduces the new modelling language, SysMod, which is designed to organise and share data used in IRM. SysMod integrates a series of principles drawn from systems theories and allows defining ontologies to semantically integrate data from various sources and knowledge domains.

Chapter 3 presents an information system called the Combine Water Information System (CWIS) developed on the basis of the SysMod language. CWIS consists of a relational database that implements the SysMod constructs (i.e. the conceptual elements that form the language) and a modular software application containing several tools to manage and visualise environmental information.

Chapter 4 provides an extension of SysMod (and CWIS) to cope with scenario data. Developments focus on the management of data representing various possible futures and integrate the concepts of provenance and uncertainty to provide meta-information about the origin and quality of the data.

Chapter 5 defines a modelling approach to assess scenarios and strategies. This approach makes use of the SysMod formalism and its extensions (thematic groups, data provenance and uncertainty) to create some workflows that combine data, mathematical models and possibly other kind of data processes such as measurements, expert judgments or data transformations. Workflows provide an overview of the processes involved in the assessment of scenarios and strategies. They are particularly useful to plan and check the validity of the modelling process, and, retrospectively, to give information about the sources of data and associated uncertainties. The chapter presents different techniques to integrate mathematical models with CWIS, possibly using existing integrated modelling frameworks.

Chapter 6 explores the application of the tools and methodological concepts. The aptitude of CWIS to support the decision making and scenario planning process is demonstrated with a case study on integrated urban water management (IUWM) in the city of Alexandria, Egypt.

1.4 Thesis framework

This thesis was performed within the framework of the SWITCH project (Sustainable Water Improves Tomorrows Cities' Health), funded by the European Union. The SWITCH project aims at promoting the concept of integrated urban water management (IUWM). It involves 33

groups of researchers and practitioners from 15 countries. The Ecohydrology laboratory of EPFL was leading the work package on “Strategic planning and performance assessment” and played an important role in developing City Water, a suite of tools and models including CWIS as a central database and data visualisation component.

SWITCH involves thirteen demonstration cities all over the world: Accra, Alexandria, Beijing, Belo Horizonte, Birmingham, Cali, Chongqing, Emscher Region, Hamburg, Lima, Lodz, Tel Aviv and Zaragoza. The tools developed at EPFL have been tested through case studies in the cities of Alexandria, Birmingham, and Belo Horizonte.

2. SysMod: A system modelling language[1]

2.1 Introduction

Systems thinking provides concepts and methods to better understand complex issues. While working in multidisciplinary contexts, systems approaches are of great interest to deal with complexity and to bring the various fields of science together (Midgley, 1992; Mulej et al., 2004). This is particularly true in the frame of integrated resources management (IRM) and its related fields, where modelling of the ecological system as well as the social, technical, political and economic systems are key issues towards sustainable planning (Bellamy et al., 2001; Rammel et al., 2007). To date, IRM requires new tools and methods to represent the complexity of the human and natural systems (McDonnell, 2008).

This chapter deals with the choice of a modelling language to support the development of tools for IRM. The term “modelling” refers here specifically to the descriptive representation of objects or phenomena, and therefore differs from the term “mathematical modelling” which implies the use of mathematical equations to simulate the response of the system being modelled. Because of the diversity and the complexity of the IRM issues, tools need to handle and integrate information from various sources, multiple scales (e.g. spatial, temporal and organisational) and various formats, whether in the form of numeric values, texts, spatial geometries or files such as documents and images (Schenk, 2010). To handle the aforementioned aspects, the language should match three essential requirements. The first one is domain-independence, as the language should model any kind of system without relying on a specific knowledge domain. The second feature is semantic integration, which relies on ontologies (see Section 2.2.2). Ontologies allow sharing unambiguous meaning about the concepts used in models. This is particularly important in the context of IRM, where models may integrate information from various disciplines, be the results of participative processes that involve many stakeholders with a variety of knowledge background. As highlighted by Villa (2009), the integration of ontologies in the field of systems modelling will

[1] Based on a paper by Roquier, B. et al., "SYSMOD: A systems modelling language for environmental information." Submitted to Ecological Modelling, 2010

be an important step towards the reusability and the potential coupling of systems models. The third feature is the ability of the language to model open systems. In contrast to closed systems, open systems interact with their environment through the transfer of energy, material or information. As a result, any open system may be considered as part of larger systems.

Next paragraphs provide details about the language's requirements to model complex human/natural systems in the framework of IRM, along with a review of existing languages. Since, to our knowledge, no existing language meet simultaneously the criteria of domain-independence, semantic integration and ability to model open systems, a new modelling language, called SysMod, has been developed, The features of the SyMod language are described and its use is illustrated with short examples.

2.2 Language requirements

The next sections detail the language requirements, with an overview of the pros and cons of two categories of existing languages: the ontology languages and the systems modelling languages.

2.2.1 Domain-independence and holism

It is generally admitted that IRM requires a holistic framework; the definition of holism being summarised by the principle of "The whole is more than the sum of its parts" (Aristotle, Metaphysics, 10f-1045a). One example advocating this approach resides in the first principle of the "Dublin Statement on Water and Sustainable Development" which relates to the integrated management of water resources (ICWE, 1992): *"Since water sustains life, effective management of water resources demands a holistic approach, linking social and economic development with protection of natural ecosystems. Effective management links land and water uses across the whole of a catchment area or groundwater aquifer."*

As a consequence, the holistic framework involves taking into account any significant element that is directly or indirectly connected to the IRM issues. The types of elements and information to be considered depend on the situations. They are various and unpredictable, that is why the domain-independence of the modelling language is a prerequisite. Examples of domain independent languages are the data modelling language EXPRESS (Schenck and Wilson, 1994) or the Unified Modelling Language (UML) (Rumbaugh et al., 2005). When dealing with multiple fields of knowledge, such generic languages must tackle the problems of semantic interoperability, as mixing terminologies of various domains usually results in more ambiguity.

2.2.2 Semantic integration and ontologies

In information science, the term "ontology" refers to a set of concepts and relationships between these concepts used for describing a field of knowledge. As noted by Villa et al. (2009), "ontologies can be used at different degrees of internal complexity and expressive power". At a basic level, ontologies are simple inventories of concepts that are not or loosely connected, such as basic thesauri or vocabulary indexes. At a more complex level, ontologies consist of classes (types of objects) hierarchically organised with a range of attributes (often named properties) and relationships.

Ontology languages provide a solution to formalise and structure knowledge domains, and therefore allow clarifying semantic issues associated with the integration of multiple knowledge sources. In 2003, a review of the various ontology languages has been carried out by Corcho et al. (2003). Since then new developments and new languages gave rise to an abundant literature. Examples of such languages are the Web Ontology Language (OWL) (W3C, 2009) or the Integrated Definition for Ontology Description Capture Method (IDEF5) (KBSI, 2010).

One of the major contributions towards the foundations of an ontological framework has been carried out by Bunge (1977, 1979) and extended by Wand and Weber (1988; 1990). Known as the Bunge-Wand-Weber (BWW) model, this framework proposes "a set of high-level, abstract constructs that are intended to be a means of representing all real-world phenomena". These ontological constructs are often used as references to evaluate the grammar constructs of any modelling language (Opdahl and Henderson-Sellers, 2001; Opdahl and Henderson-Sellers, 2004).

The ontologies allow users to agree on a common "vocabulary" to describe a knowledge domain. Several examples of ontologies for environmental management can be found in the literature. For instance, the Ontology-based Environmental Decision Support System for Wastewater (OntoWEDSS), which is based on an ontology of microbiologic knowledge to model the wastewater treatment process (Ceccaroni et al., 2004), the "system model for water management" of Schenk et al. (2009) that provides an ontology of the water system and its related elements to support the process of integrated water resources management or the Extensible Observation Ontology (OBOE), a formal ontology to "capture the semantics of generic scientific observation and measurement" (Madin, 2007).

Systems modelling languages are generally used in computer science, systems engineering or project management. Families of languages such as UML or IDEF can cover a wide range of applications, principally in order to follow all the steps of the software development

lifecycle. Other languages such as the System Modelling Language (SysML) (Willard, 2007), Modelica (Mattsson et al., 1998) or the Universal Systems Language (USL) (Hamilton and Hackler, 2008) are specially adapted for advanced systems engineering, and therefore are domain-specific. It is interesting to note that these few examples are only a sample of the many languages available in these fields, whereas languages dedicated to environmental modelling are very rare.

One of the few examples related to environmental modelling is the Energy System Language (ESL) developed by Howard T. Odum (1960), which models ecosystems through energy flow diagrams. ESL is based on the analogies between energy flows within the ecosystems and those in electronic circuits and has been applied successfully in a variety of situations from ecological to economic modelling (Brown, 2004b; Odum and Odum, 2000; Rivera et al., 2007). Although it is a powerful language, ESL remains too domain-specific according to the requirements of IRM.

Finally, in the frame of systems dynamics, which involves causal loop diagrams and stock and flow diagrams, some approaches such as SIMILE, the Semantic Interoperability of Metadata and Information in unLike Environments, (Muetzelfeldt and Massheder, 2003) and STELLA (Costanza and Gottlieb, 1998) incorporate their own graphical modelling language. These languages remain unfortunately specific to their modelling task, and therefore difficult to apply in other contexts.

Although this is not their primary purpose, ontology languages can be used for systems modelling, as demonstrated by Tudorache (2006) who explored the application of ontologies for systems engineering. Conversely, systems modelling languages can be used (or extended, in the case of UML) to create ontologies (Kogut et al., 2002; OMG, 2009).

2.2.3 Ability to model open systems

An open system is defined as a system that interacts with its environment, for example by exchanging raw materials, energy or data. Systems inputs, outputs and interactions are therefore essential elements of any system definition and thus deserve a special attention.

Interactions can be represented by two types of relationships, the flows that carry energy, material or information between systems and the causal relationships that describe how the change of a system variable influences other variables, and vice-versa through possible feedbacks. By analogy to a system which is composed of a group of subsystems, an interaction may also be perceived as an aggregation of a multitude of elements. For instance, a flow may have a property that quantifies the flow value or lists the items being

transferred during a time period. Moreover, a flow may also have additional properties that describe some secondary elements carried by the stream, such as the concentration of pollutant in a water flow.

On the one hand, systems languages, such as UML, Petri-Net, SysML or ESL, have interactions or relationships that typically provide a single piece of information, whether a specific influence or a flow of a particular kind between two elements. With these languages, there is therefore no possibility to represent an interaction as an aggregation of multiple systems elements and/or information. On the other hand, ontology languages, such as OWL, have the ability to model interactions as aggregation of multiple systems elements; however they do not provide a standard way to classify the properties (or systems attributes) as inputs, outputs, components or other system features.

Given the restrictions imposed by the language requirements, the existing systems modelling languages have been excluded, because they are often domain-specific and none of them can be used for modelling various types of "complex" interactions. Therefore, two options remain open: (1) the use of an ontology language even if not dedicated to systems modelling and (2) the creation of a new language. The second option has been chosen.

2.3 The SysMod language

The new SysMod (System Modelling) language is at the same time an ontology and a system modelling language, i.e. the modelling of both ontologies and systems are combined in a single framework. The creation of a new language has been chosen instead of using an existing ontology language, as it offers the opportunity to explore and put into practice concepts drawn from the systems theories (Bertalanffy, 1973; Skyttner, 2005), such as:

- **Hierarchy**: Systems are complex wholes, which are made up of smaller subsystems.
- **Inputs and outputs**: Systems are open systems that interact with their environment through inputs and outputs.
- **Interrelationship and interdependence** of objects and their attributes: A system is composed of interdependent elements. The state of the subsystems and their relationships regulates the state of the system and its inputs/outputs.
- **Control and communication mechanisms**: complex systems, such as life beings (plants, animals, human, societies…) are able to process and store information, to adapt themselves to (or act on) their environment. Such capacities of control and communication generally rely on circular (feedback) mechanisms.

- **Negative and positive feedback causality**: System's equilibrium relates on the existence of negative feedbacks that tend to bring the system in its previous state (also known as the principle of equifinality). On the other hand, positive feedbacks may lead to radically different end states of the system (principle of multifinality).

2.3.1 SysMod elements constructs

The SysMod constructs are abstract concepts inferred from the description of the real world as a complex system. The constructs are divided into two groups: ten "SysMod elements" that define the lexical categories of the language and six "SysMod relationships" that describe its grammatical rules.

SYSMOD describes system elements along two axes: the instantiation levels that define whether the elements are classes, properties or instances, and the types that differentiate holons, information items, flows and causal relationships. Table 1 summarises the ten constructs that compose the SysMod language, along with the symbology that has been proposed to represent them graphically.

Table 1 List of SysMod elements constructs

SysMod elements		Types			
		Holon	**Information**	**Flow**	**Causal relationship**
Instantiation levels	**Class**	E-HC Holon class	E-IC Information class	E-FC Flow class	E-RC Causal rel. class
	Instance	E-H Holon	E-I Information	E-F Flow	E-R Causal relationship
	Property	E-HP Holon property	E-IP Information property		

The concept of "holon" has been adopted, as a mean to focus on the essential while keeping a holistic view of all potential issues. A holon is a thing that is simultaneously a whole and a part (Koestler, 1967, 1969). The concept of holon allows the modellers to describe any system as a hierarchy of wholes or, in adopting Koestler's terminology, as a "holarchy". This

hierarchical approach may be of great interest when trying to tackle the philosophical debate between the relative merits of reductionism (the system is explained by analysing its parts) and holism (the system is more than the sum of its part and is explained by the interactions with its environment) (Edwards and Jaros, 1994; Koestler, 1969; Naveh, 2000) and turns out to be particularly adapted for complex systems modelling, as demonstrated by many studies (Giampietro et al., 2006; Kay et al., 1999; Kira and van Eijnatten, 2008). In a holarchy, any system is also part of a larger system; it ensures that models can always be considered as open systems.

A complete definition of the ten SysMod elements' constructs is given thereafter:

Holon class (E-HC), information class (E-IC), flow class (E-FC) and causal relationship class (E-RC)

A class is a conceptual element that groups the objects (also named instances) sharing one or more identical properties (E-HP/E-IP). The class is a template from which new instances can be derived. SysMod distinguishes four types of classes, one per type of system element defined in SysMod:

- The holon class (E-HC) categorises the "things" of the real world and inventories their properties, e.g. the holon class "River" defines the properties common to all rivers and can be used to derive instances (E-H) from the class, such as the instance "Nile river"
- The information class (E-IC) is characterised by a data type such as text, numeric value, file, etc. and a value type, which enumerates, if required, the available units, e.g. the information class "Volume" is used to derive some instances of volume such as "20 m^3" or "15 litres",
- The flow class (E-FC) describes the types of flow and their properties, e.g. a class "Wastewater flow" is defined by the property "Flow value" and some additional properties such as the "Biochemical oxygen demand".
- The causal relationship class (E-RC) defines the types of influences among properties. For instance, a positive causal relationship indicates that the change on both sides vary in the same direction, i.e. if the value of the link origin decreases, the target also decreases, whereas if the origin increases, the target increases too. On the other hand, a negative causal relationship indicates that the change on both sides vary in opposite directions, i.e. if the value of the link origin decreases, the target increases, and vice versa.

Holon (E-H)

In the SysMod language, the holon represents an element of the real world and can be instantiated from a holon class (E-HC). It is equivalent to the “thing” of the BWW ontological model (Opdahl and Henderson-Sellers, 2004), to the “individual thing” of Chisholm’s ontology (Chisholm, 1996) or to the “individual” in OWL.

A holon has the particularity to be simultaneously a whole and a part, i.e. a system in itself and a part of a larger system. Holons are used for modelling the real world as a set of hierarchically embedded systems. For instance, a river can be described as a system that belongs to the super-system watershed and contains subsystems such as fishes.

Information (E-I)

The information construct represents any kind of data that can describe or document a system. They can be texts (i.e. descriptions, comments or web links), a spatial geometries (points, lines, polygons or multipart geometries), values (single values, intervals, time series or fuzzy values), classifications (e.g. a temperature classification: cold -warm-hot) and files (text document, image, video, etc.).

Flow (E-F)

A flow connects two holon properties (E-HP) and can be used to model the movement a holon from a system to the other. A typical illustration of this construct is the modelling of a flow between two reservoirs or the arrivals and departures of individuals in a group (e.g. the population migration induced by climate change, the bird migration).

Causal relationship (E-R)

Causal relationships allow modelling the influences between information properties (E-IP). If they are interconnected, the causal relationships can form some “causal loop diagrams” that help visualising how the properties of the system(s) affect one another, as shown by the predator-prey model in Figure 9. The causal relationships represent either some positive influences (e.g. the lynx’s abundance has a positive effect on the hare’s death rate) or some negative influence (e.g. the number of hare’s deaths has a negative influence on the hare’s abundance).

Holon property (E-HP) and Information property (E-IP)

SysMod uses the concept of property to describe the characteristics of the classes and instances. The property is assigned to the subject (which can be any system element, not necessarily from a specific class) using the relationship “property role” (R-PR). The object of

the property is expressed through the "general instantiation" (R-GI), which defines the relationship to the class that characterise the property[2].

SysMod properties act as kinds of abstract containers that store information items or holons. The properties may also have cardinality constraints. It means that some minimum and maximum cardinality values can be specified, to define whether the property contains a fixed or a bounded number of elements.

The holon property (E-HP) is used to model systems as a hierarchy of holons, i.e. a holon can possess properties to group its subsystems. This hierarchical structure is illustrated in Figure 1, with the example of a "Watershed" that contains the subsystem "River", which itself possess subsystems such as "Fishes". The properties and holons are connected through different relationships, which will be described in Section 2.3.2.

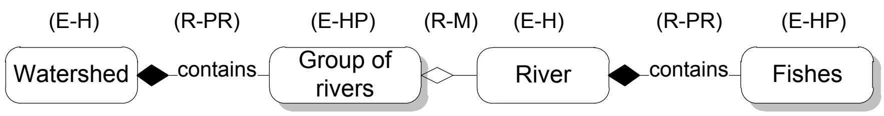


Figure 1 Example of hierarchy of holons combining holons (E-H), holon properties (E-HP), property role relationships (R-PR) and membership relationship (R-M)

In the same way than holon properties, information properties (E-IP) can host data sets. For example, a river may have a property "Water flow" that contains the instance of information "10 m^3/s" (Figure 2).

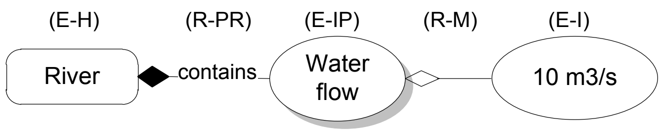


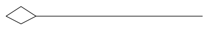
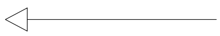
Figure 2 Example of information property (E-IP) having an instance of information (E-I) linked through a membership relationship (R-M)

[2] This definition differs from the one adopted by the popular Web Ontology Language (OWL), where the subject and the object of a property are defined by a "domain" (a relationship to the subject class) and a "range" (a relationship to the object class).

2.3.2 SysMod relationships constructs

The SysMod relationships constructs are the grammatical rules of the language. They describe how the SysMod elements can be combined to model ontologies or systems. Table 2 gives an overview of the six constructs and defines the origin and target elements of each relationship, as well as the type of arrow chosen for the graphical representation. These six constructs are explained in detail in the next paragraphs.

Table 2 List of SysMod relationship constructs

Code	**Relationship name**	**Relationship between**	**Symbol**
R-PR	Property-role	System element – Property	
R-GI	General instantiation	Class – Property	
R-PI	Particular instantiation	Class – Instance	
R-M	Membership	Property – Instance	
R-G	Generalisation/Specialisation	Class – Class	
R-MT	Mereotopology	System element – System element	

Property-role relationship (R-PR)

The property-role relationship is used to assign properties to system elements. A list of "roles" is given to characterise the function of the property from a systemic perspective. The roles are defined by the labels: "has", "contains", "produces", "consumes", "requires", "acts on" and "interprets". Figure 3 illustrates the different functions of the property roles.

The first property-role relationship "has" aims to document the system with "descriptive" properties. It means that the properties (E-HP/E-IP) linked by a relation "has" do not represent some system's features such as inputs, outputs, stocks or subsystems. These relationships are mainly dedicated to descriptive information such as numerical coefficients, textual comments, web links or images.

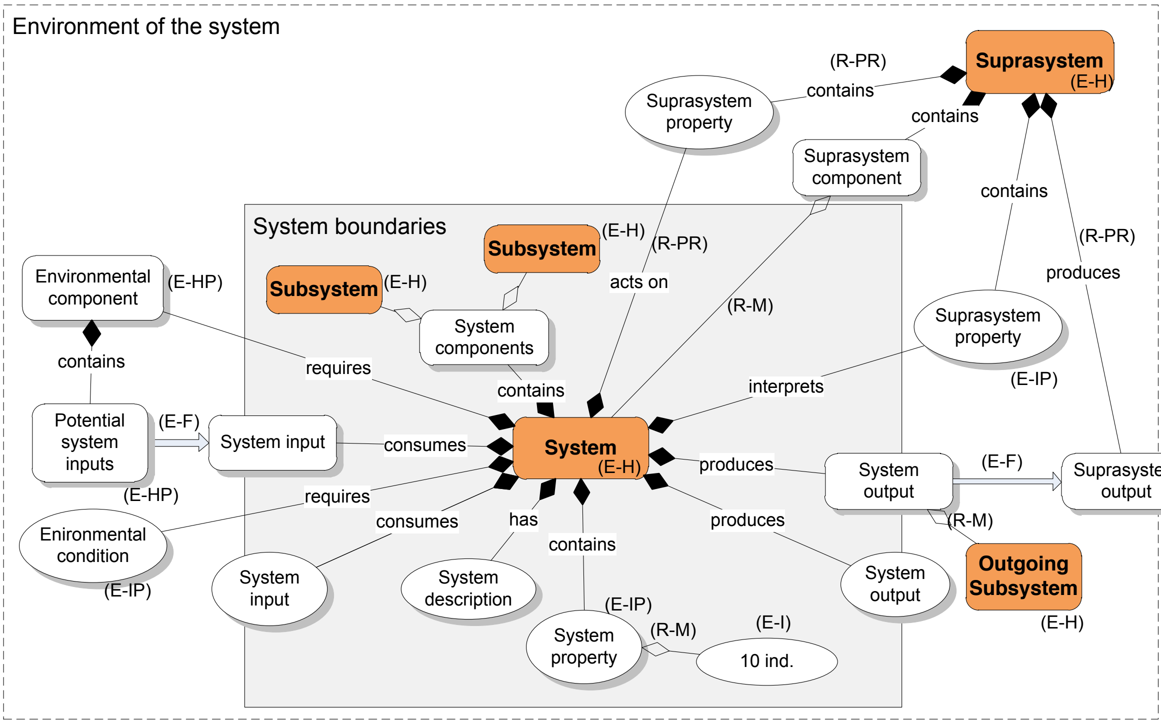


Figure 3 Example of a system model representing the system inputs, outputs, composition, environmental requirements and control over other elements

The label “contains” specifes the composition of the system. If the subsystems need to be represented in the model, the relationship “contains” must be connected to a holon property (E-HP) (e.g. the property “River of a watershed” that contains an instance of river). Otherwise, the relationship can be connected to an information property (E-IP) that will describe the subsystems (e.g. the river can have a property “Fish abundance”, which is an information group that records the number of fish in the river without considering each individual fish).

Inputs and outputs of the systems are characterised by properties with the label “consumes” and “produces” respectively. In order to keep track of the subsystems that have entered or leaved the system boundaries, the property-role relationship need to be connected to a holon property (E-HP). Otherwise, if only a quantification of the inputs and outputs is required, an information property (E-IP) should be used.

In order to exist or operate, the system may need some specific environmental conditions, which can be modelled using the relationship “requires”. Examples of such conditions could be a required interval of temperature (based on an information property), or the proximity to another element (member of a holon property) that provides the inputs needed by the system.

When there is not enough knowledge to represent the influences between systems as collections of flows (E-F) or causal relationships (E-R), the property-role “acts on” allows to indicate that a system has an effect or controls the property of another system..

Finally, the relationship “interprets” characterises the aptitude of some living beings to observe and analyse the elements of their environment, or the capacity of measuring instruments to monitor environmental phenomena (e.g. a rain gauge that determines the rainfall intensity).

General instantiation (R-GI), Particular instantiation (R-PI) and membership relationships (R-M)

The general instantiation (R-GI), particular instantiation (R-PI) and membership relationships (R-M) allow defining the relationships between the different instantiation levels of the SysMod elements (as shown in Figure 4). The general instantiation relationship is used for creating a property from a class. For instance, the property “Water flow” of a river is an information property that is instantiated from the class “Flow”. On the other hand, the particular instantiation relationship connects the instance to its corresponding class, for example the “10 m^3/s” to the class “Flow”. Then, the membership relationship links the property to its members (holons or information items), e.g. the “Water flow” to the value “10 m^3/s”.

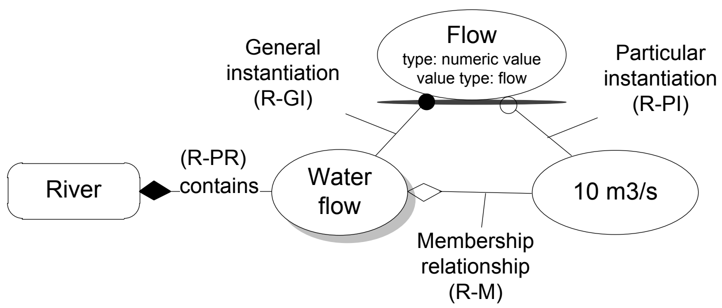


Figure 4 Example of relationships between system elements

Generalisation/specialisation relationship (R-G)

The generalisation/specialisation relationship represents a link between two classes where one class (the specialisation) is a subclass of the other one (the generalisation). This hierarchical organisation allows modellers to create a taxonomy of classes, which is a first

step towards the realisation of an ontology. Figure 5 describes four taxonomies of classes; one for each type of SysMod element.

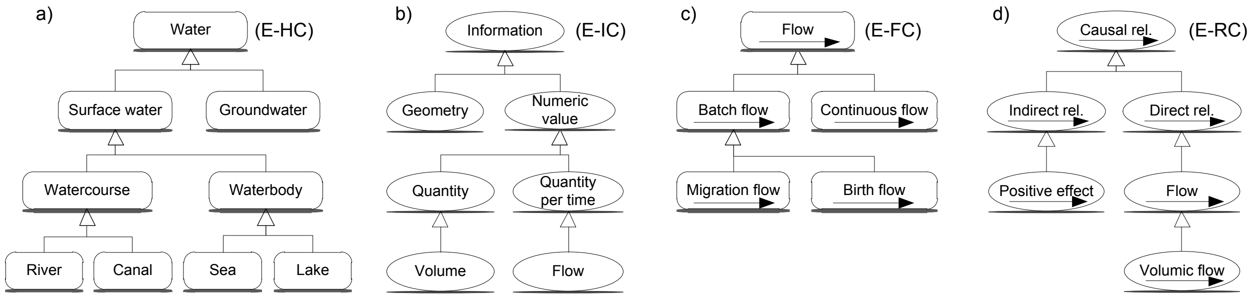


Figure 5 Taxonomies of classes (a. Holon, b. Information, c. Flow and d. Causal relationship), based on generalisation/specialisation relationships (R-G)

Mereotopology relationship (R-MT)

Mereotopology combines both set of relationships from mereology (relationships dealing with parts and their respective wholes) and topology (relationships about the spatial properties of elements, such as connectedness). Several mereotopological theories exist, such as GEMTC (Varzi 1996) or the Region Connection Calculi (RCC) (Cohn et al. 1997; Bennett et al. 2002).

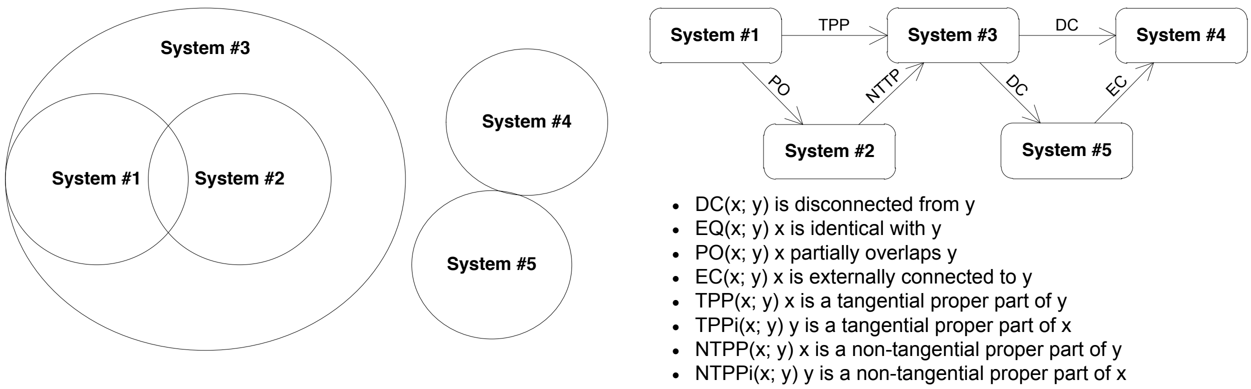


Figure 6 Example of mereotopological relationships between systems

The theory adopted in SysMod may be changed or extended according to the modelling requirements. Figure 6 shows for instance the use of the theory RCC8 (Cohn et al. 1997) to define the topological relationships between systems.

2.4 Examples of SysMod models

2.4.1 Ontologies

Ontologies aim to address the issue of semantic integration. An ontology can be a taxonomy, i.e. a hierarchical organisation of classes (Figure 5). At a more complex level, ontologies are obtained when classes, properties and possibly instances are combined. Figure 7 shows how a detailed ontology can be created from a set of classes, and then be used to produce an instance of system. The class River is for instance a subclass of Watercourse, it has the (holon) property Fish population which itself is documented by the (information) properties Birth rate and Abundance. These information properties may have default values such as Birth rate = 0.8 per year. Ontologies may also define relationships between properties, such as the Birth rate positively influences the Abundance of the Fish population.

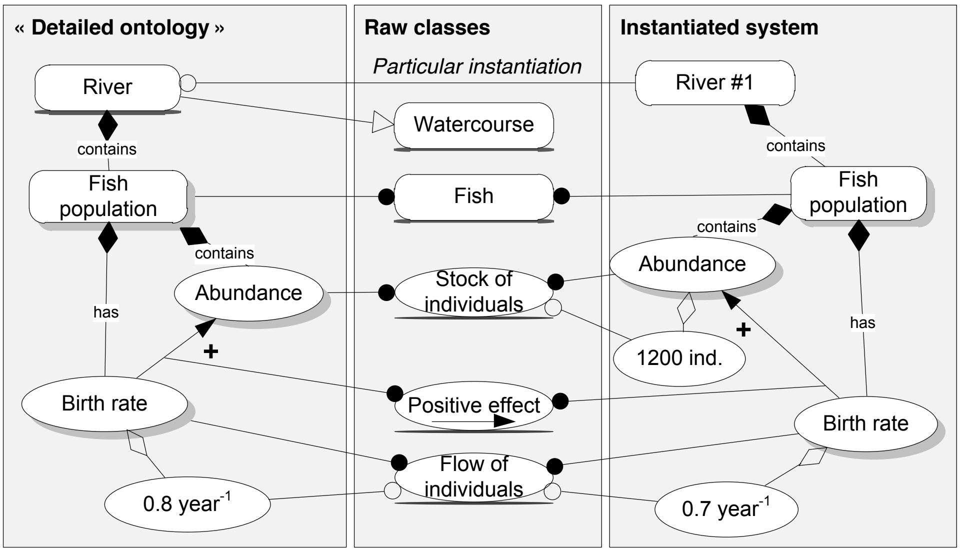


Figure 7 A detailed ontology and its instantiation into a particular system

An instantiated system can be created through the relationship “Particular instantiation”. By default, the resulting system (e.g. the River #1) inherits the properties of the class. Moreover, these properties and their members may be changed or completed to match the system to be described (see Section 2.4.2). For instance, the Birth rate of the Fish population is equal to 0.7 per year and the Abundance of fishes is 1200 individuals.

2.4.2 System inputs, outputs and relationships

Relationships between systems can be of different kinds; they either define how the systems are positioned relative to each other (the mereotopological relations) or how they interact (the cause-effects relationships and the flow of holons).

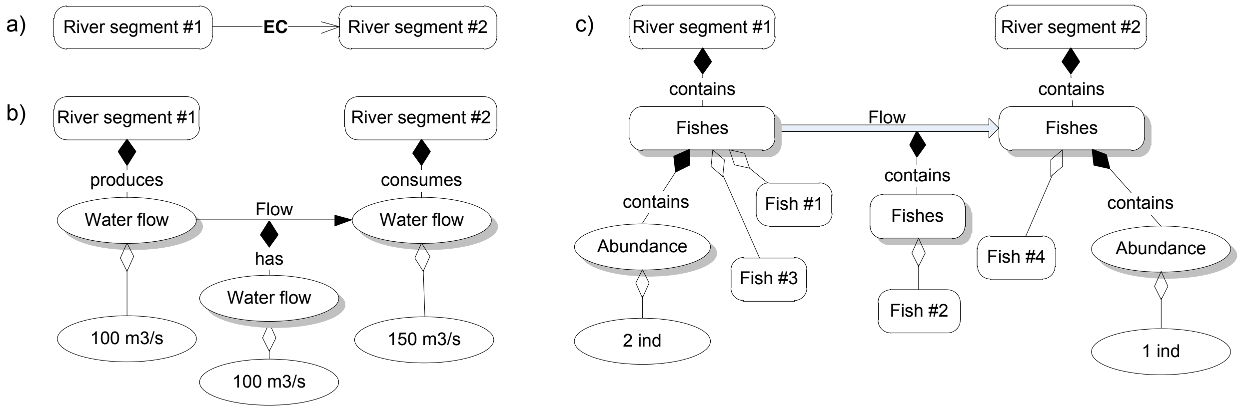


Figure 8 Relationships between systems: (a) mereotopological relationship, (b) causal relationship and (c) flow of holons

The types of possible relationships between systems are illustrated in Figure 8. (a) The mereotopological relationship EC means that the River segment #1 is externally connected to the River segment #2. (b) The flow from a river segment to another can be expressed as the direct effect (causal relationship) from the property “produces Water flow” of River segment #1 to the property “consumes Water flow” of River segment #2. The value of the outflow of the first segment is not necessarily equal to the value of the inflow of the second segment, as other (non-described) river segments may contribute to the inflow. (c) The flow between two stocks of fishes may be modelled as a flow of holons. Unlike the causal relationships, the flows of holons can model the movement of each fish between the river segments.

These relationships can also be used to model the dynamics of complicated systems, such as causal loop diagrams or stock and flow diagrams. Figure 9 illustrates the predator-prey model (widely used in the literature of system dynamics) as a stock and flow diagram based on the SysMod formalism.

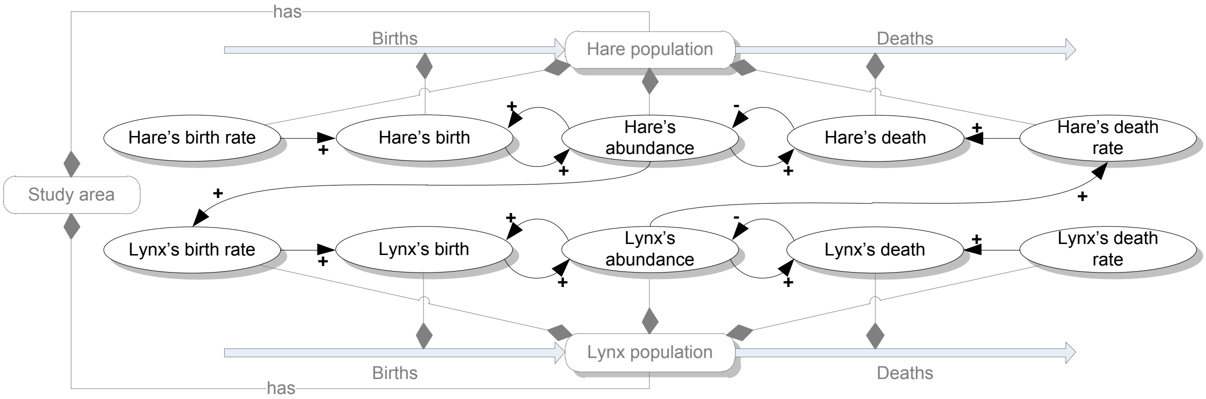


Figure 9 SysMod System dynamics: the case of the predator-prey model

The dynamic of the system is controlled by a collection of causal relationships that form some feedback loops. Reinforcing loops, which have no or an even number of negative relationships, tend to destabilise the system with some (exponential) increases or decreases, e.g. the Lynx's abundance has a positive effect on the Lynx's births and vice-versa. Balancing loops have an uneven number of negative relationships and tend to stabilise the system, e.g. the Lynx's deaths regulates the Lynx's abundance.

2.4.3 System control mechanisms

Detailed representation of control mechanisms in a system generally involves the representation of feedback loops, as shown in Figure 9 with the example of the predator-prey model. However, in some cases, such level of detail is not required or is too complex to be modelled. An alternative is to describe (using the property roles "interprets" and "acts on") what element controls the properties of the system and based on which information.

A basic representation of this control mechanism is shown in Figure 10. It shows that a WWTP "requires" a manager, which is a person that manages (acts on) the different subsystems of the WWTP, according to the measures (interpretations) of water quality, wastewater inflow and sludge production.

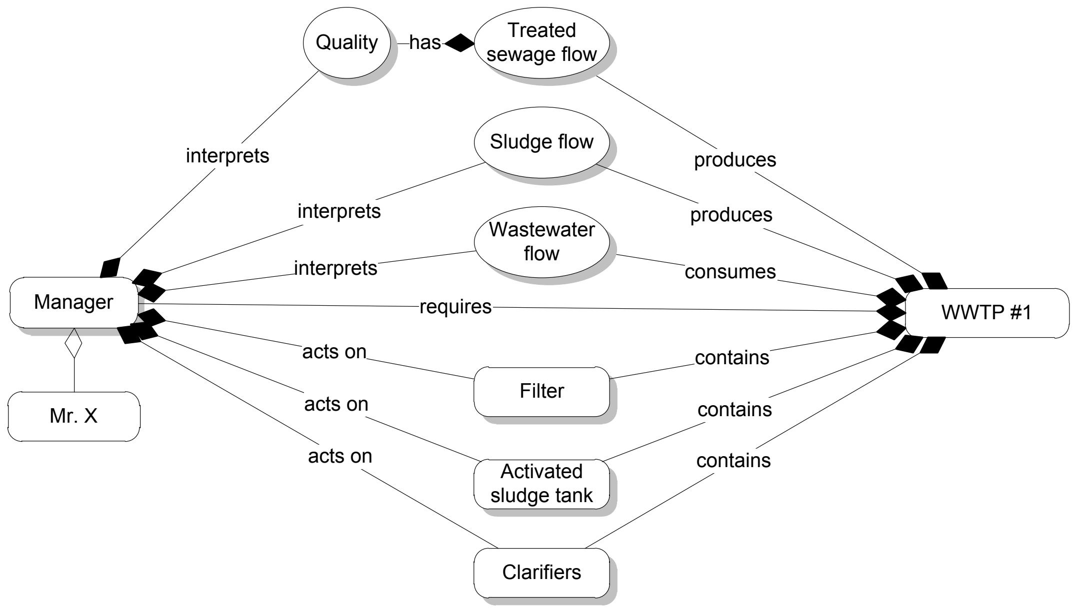


Figure 10 Basic representation of control mechanisms in SysMod: example of a wastewater treatment plant (WWTP)

2.4.4 Systems hierarchy and interdependence

As mentioned before, the systems (or holons) have the particularity to be organised hierarchically. Figure 11 shows, with the example of wastewater treatment plant (WWTP), how multiple levels of systems can be embedded. In order to facilitate the reading of the model, some rectangular frames have been added to symbolise the boundaries of the systems. In this case, the WWTP is a component of the watershed (the supra-system) and the clarifiers, activated sludge tank, and filter are some subsystems. The figure also shows how the flow of wastewater is conveyed through the systems, being produced by the households, carried by the sewer network and treated in the WWTP. At the level of the WWTP, the wastewater flow goes through the different subsystems and is released after treatment. The sludge issued from the second clarifier is partly reused in the activated sludge tank, while the surplus is evacuated from the WWTP.

It should be highlighted that properties can be assigned to other properties. This is particularly useful to aggregate the properties of a group of instances. As shown in Figure 11, a property “produces Wastewater flow” is connected to the property “Households” and gives the aggregated sum of the flows produced by the members of the property “Households”. Similarly the representation of all the segments of the sewer network is avoided by adding the property “contains Wastewater flow” to the property “Sewer network”.

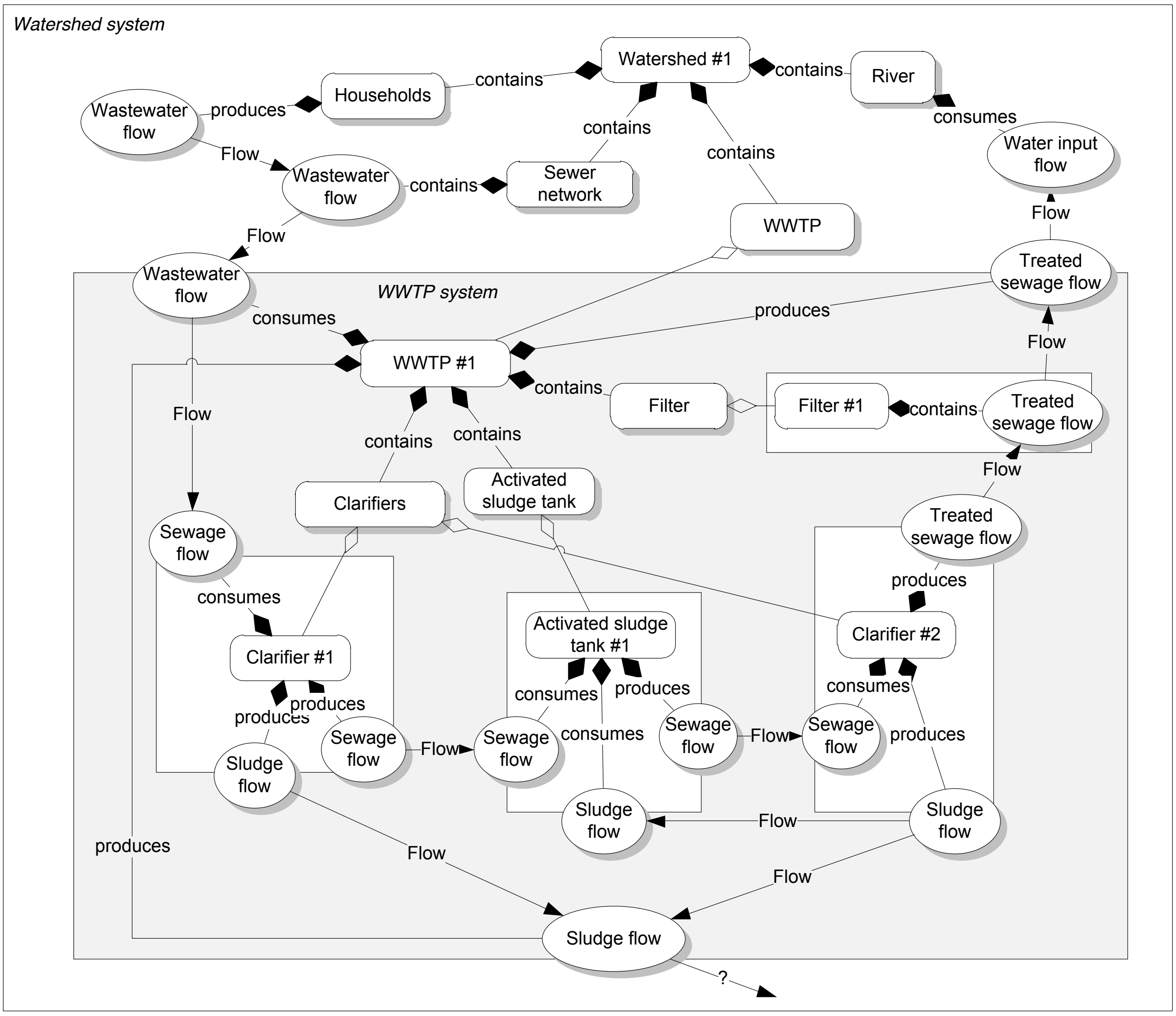


Figure 11 Hierarchical organisation of systems: example of a wastewater treatment plant

2.5 Synthesis

The present chapter has introduced a new generic modelling language – SysMod – whose main particularity is the ability to model systems and ontologies. The language is designed firstly for modelling environmental systems, but is prone to be applied to any kind of systems. While the ontology ensures the semantic integration and the domain-independence of the modelling language, the “holonic” perspective adopted guarantees that the systems modelled with SysMod are open systems.

In comparison, existing systems modelling languages are generally not generic and do not support the integration of ontologies to cope with the issue of semantic integration. On the other hand, existing ontology languages have more descriptive power, but are not particularly adapted for systems modelling. A specificity of the SysMod language, which is not available in other ontology language, is the categorisation of the properties based on different roles, which make a distinction between system inputs, outputs, components (or stocks), environmental requirements and other system features representing control and interpretation mechanisms.

SysMod is not usable as such and needs some compatible tools that implement the language constructs. This is the subject of the next chapter.

3. Implementation of SysMod: the Combined Water Information System (CWIS)

3.1 Introduction

Integrated resources management (IRM) is a relatively new domain. While the principles of IRM have been widely discussed and recognised, IRM and its related fields still need some new methods and tools to address the complexity of coupled human and natural systems (McDonnell, 2008). For this purpose, the first requirement was to find a formalism to represent this complexity. This led to the development of a new systems modelling language called SysMod (see Chapter 2). The objective now is to use this language to build an information system that facilitates the sharing of information between scientists, stakeholders, and decision makers involved in the IRM process[3].

This chapter introduces the Combined Water Information System (CWIS), which consists of a relational database and a modular software application that provides edition and visualisation tools (Schenk, 2010). For this purpose, the constructs of the SysMod language have been "translated" into a database schema, which provides a diagram representation of the data organisation in the database. A first prototype of CWIS was developed as a desktop program linked to a PostgreSQL[4] database. The application was composed of six modules: a system viewer, a geographic viewer, a reporting tool, a chart viewer, an indicator viewer and a data exchange and modelling tool. More recently, a second version of CWIS has been developed to resolve some of the software glitches and issues (which mainly resulted in unexpected crashes and slow loading/saving). Using the Microsoft Silverlight application framework[5] and the Microsoft SQL server 2008[6], the new version, now a Web-based

[3] The tools presented in this chapter result from developments performed in close collaboration with the SWITCH EPFL team and especially with my former colleague Colin Schenk (Schenk, C., 2010).

[4] PostgreSQL: http://www.postgresql.org/, Retrieved 29 December 2010

[5] Microsoft Silverlight: http://www.silverlight.net/, Retrieved 28 December 2010

[6] Microsoft SQL Server 2008: http://www.microsoft.com/sqlserver/2008, Retrieved 29 December 2010

application, is more reliable and powerful than the previous one. As a consequence of redesigning the application, only parts of the initial modules have currently been implemented in the new version.

This chapter first describes the structure of the database, and explains how the constructs of the SysMod language have been implemented. A short overview of the architecture and functionalities of CWIS is then provided. Finally, the advantages and limitations of the SysMod language are discussed, with respect to the three main viewing modules of CWIS (systemic, geographic and reporting modules).

3.2 Structure of the database

The design of CWIS's database is achieved with the tool Perceptory (Bedard, 2005; Bedard et al., 2004). Perceptory is a plug-in of Microsoft Visio[7] diagramming software that enriches the UML formalism and provides a code generator to convert a database schema into a "skeleton" code, which can be used afterwards in a relational database management system (RDBMS) to automatically generate the database.

The next subsections explain how the concepts introduced in Chapter 2 are integrated in the database schema (the complete diagram is available in Appendix A). The database schema is composed of entities (boxes) that represent the tables in the database and relationships that characterise the links between entities. Particular features of the data model are: (1) the *association class* (an entity connected to a relationship through a dotted line) that allows defining the attributes resulting from the association between two entities; (2) the *composition*, a particular type of association where an entity (connected to the end with a "black diamond") is composed by other entities; (3) the *generalisation relationship* (represented by a large arrow) that organises entities hierarchically, the specialised "children" entities inheriting the properties (attributes) from their "parent" entity.

To improve its readability the schema has been divided in two parts that describe respectively the SysMod constructs and the types of information.

3.2.1 SysMod database schema

Most ontology or systems modelling languages are based on specific file formats. In contrast the SysMod language is based on relational databases, and therefore benefits from the

[7] Perceptory is compatible with Microsoft Visio 2002, Office Visio 2003 or Office Visio 2007 (but not with Visio 2010) : http://visio.mvps.org/, Retrieved 28 December 2010

proven tools and techniques of the RDBMS to control the creation, maintenance and use of data.

The constructs of SysMod introduced in Chapter 2 have been translated into a database schema (Figure 12), which characterises the organisation of the “lexical elements” (the system element constructs) and the “grammatical rules” (the system relationship constructs) of the language. As a result, the SysMod data model is composed of two main entities (SYSTEM ELEMENT and SYSTEM RELATIONSHIP) that group the other constructs of the language.

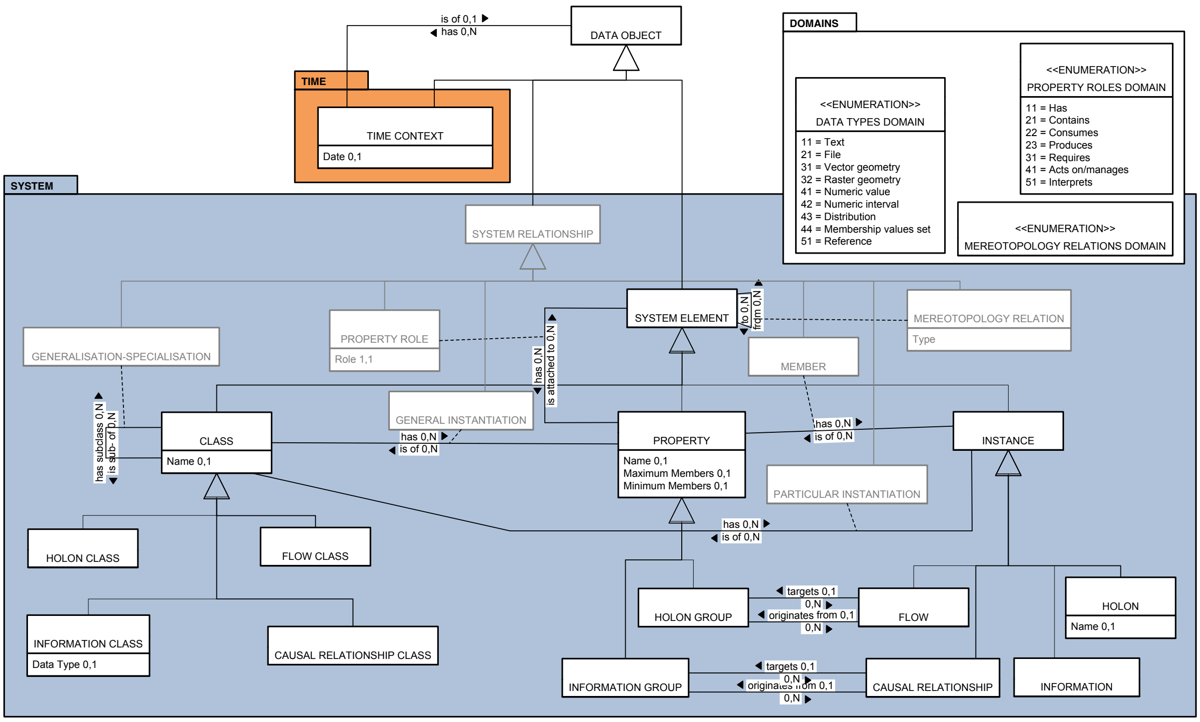


Figure 12 Part of the database schema that defines the SysMod constructs

In Figure 12, the large arrows are generalisation relationships that group the concepts sharing the same attributes. It means for example that the HOLON PROPERTY is a particular type of PROPERTY and the PROPERTY is itself a specific kind of SYSTEM ELEMENT. Through the generalisation relationship, the entities inherit the attributes of their “parent”; therefore the HOLON PROPERTY can have a Name, a Minimum and a Maximum number of Members.

The SYSTEM ELEMENT is divided into three sub-categories that represent the “instantiation levels” of the SysMod language:

- The CLASS groups the four types of SysMod classes, i.e. the HOLON CLASS, the INFORMATION CLASS, the FLOW CLASS and the CAUSAL RELATIONSHIP CLASS. The INFORMATION CLASS has an attribute Data Type, which can take the values proposed in the DATA TYPES DOMAIN.
- The PROPERTY groups the two types of SysMod properties. The specialised entities are the HOLON PROPERTY and the INFORMATION PROPERTY.
- The INSTANCE groups the four types of SysMod instances: the HOLON, the INFORMATION (detailed in Figure 13), the FLOW which is originated from/targets a HOLON PROPERTY and the CAUSAL RELATIONSHIP which is originated from/targets an INFORMATION PROPERTY.

On the other hand, the SYSTEM RELATIONSHIP groups the six entities (in grey) that correspond to the “system relationship constructs” of SysMod:

- The relationship linked (with a dotted line) to the PROPERTY ROLE defines the assignment of a property to a system element. The attribute Role characterises the role (component =”contains”, input =”consumes”, output =”produces”, etc.) of the property as regards the system element. The available roles are listed in the PROPERTY ROLES DOMAIN.
- The GENERAL INSTANTIATION associates a property to its corresponding class.
- The PARTICULAR INSTANTIATION associates an instance to its corresponding class.
- The GENERALISATION-SPECIALISATION is used to hierarchically organise the classes in SysMod, i.e. to define a class as a subclass of another class.
- The MEMBER entity allows defining an instance as a member of a property
- The mereological or topological relationship between two system elements is defined by the MEREOTOPOLOGICAL RELATION. The types of mereotopological relationships can be listed in the MEREOTOPOLOGY RELATIONS DOMAIN.

In order to support the modification of the modelled systems over time, a Date can be attributed to the SYSTEM ELEMENTS and SYSTEM RELATIONSHIPS through the relationship between the DATA OBJECT (which groups the both types of SysMod constructs) and the TIME CONTEXT. Based on the ISO 8601 standard, the attribute Date can be used to store dates, times, time intervals and durations.

3.2.2 Information database schema

The SysMod language is designed to support several types of information, which are stored in different ways in the database. Figure 13 provides an extension of the SysMod database schema (Figure 12) that defines the types of information and their organisation.

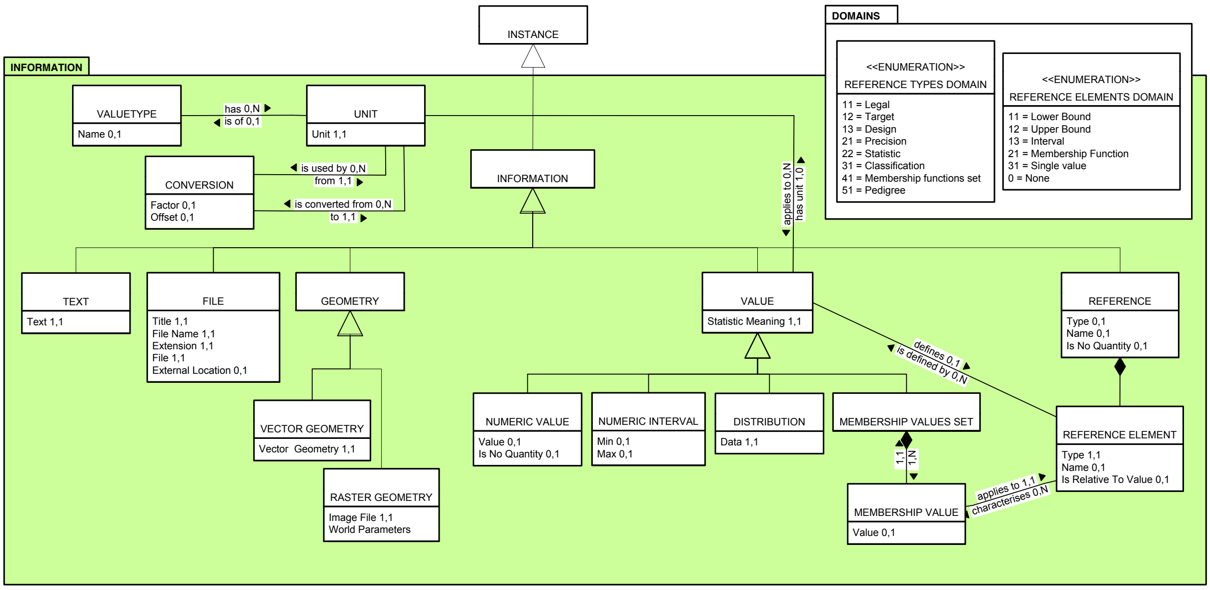


Figure 13 Part of the database schema that defines the types of information linked to SysMod

The entity "INFORMATION" is refined into several specialized entities, which cover the information types used by CWIS: Texts; Files (documents, images, videos…); Geometries; References; Numerical values; Membership values sets (to represent fuzzy values and classifications); Numeric intervals; and Numerical distributions (e.g. probability density functions or membership functions).

The TEXT information stored in the database are either some short chains of characters such as names, labels and web links or some longer texts such as descriptions, comments or definitions.

A FILE is saved as a binary large object (blob), along with its title, file name, extension and a possible external location (such as a reference in a library) if the file is not in an electronic format or if it is too big to be stored in the database.

Two kinds of GEOMETRY can be stored: the VECTOR GEOMETRY (vector shapes) and the RASTER GEOMETRY (digitized images mapped into a grid of pixels).

Any VALUE is connected to an UNIT and has an attribute which may host its “Statistic meaning” (e.g. average, maximum, standard deviation, quantile, etc.). A VALUE can take four different forms:

- The NUMERIC VALUE stores basic decimal number. The Boolean attribute “Is no quantity” serves to distinguish between physical (quantitative) values and qualitative ones. If the “Is no quantity” is true, it means the value is a “classification marks” such as 0 = low, 1 = medium, 2 = high, etc.
- A NUMERIC INTERVAL is defined by a minimum and a maximum value. If the maximum is omitted, the value is a “lower bound” and if there is no minimum, the value is an “upper bound”.
- The entity DISTRIBUTION stores probability density functions (PDF) and membership functions (MF). The probability density functions are expressed using the syntax PDF(‘*name of the distribution*’, parameters): for example the uniform distribution PDF(‘uniform’, min, max), the normal (or Gaussian) distribution PDF('normal', mean, variance), the Poisson distribution PDF(‘poisson’, lambda) where lambda is the expected number of occurrence during a given interval, or the log-normal distribution PDF(‘log-normal', mu, sigma) where mu is the mean and sigma the standard deviation.
- A “custom” discrete probability distribution can be defined to store distributions resulting uncertainty propagation methods such as the Monte-Carlo simulation. The distribution is expressed as PDF(‘custom’, min, max, [values]), where the min and max are the bounds of the function and the [values] is an array of ‘N’ values that contains the probabilities corresponding to the ‘N’ fixed intervals between min and max.
- Also handled by the entity DISTRIBUTION, the membership functions are used in fuzzy logic to characterise the “degree of membership” of a value to a fuzzy set (such as low, medium or high). They are defined as follow, with the syntax MF(a, b, c, d)(x) where the values a, b, c and d are the parameters of the function and x is the variable:

$$MF(a,b,c,d)(x) = \begin{cases} 0 & \text{if } x \leq a; \\ \dfrac{x-a}{b-a} & \text{if } a < x < b; \\ 1 & \text{if } b \leq x \leq c; \\ \dfrac{d-x}{d-c} & \text{if } c < x < d; \\ 0 & \text{if } x \geq d. \end{cases}$$

- The membership function has a trapezoid shape if a<b<c<d, a triangular shape if a<b=c<d or a rectangular shape if a=b<c=d.
- The entity MEMBERSHIP VALUES SET is composed of one or several MEMBERSHIP VALUE. The name of "Membership value" is drawn from the fuzzy logic where it represents a "truth value" that ranges between 0 and 1 and, based on a membership function, characterises the "degree of truth" of a statement.

Membership functions (stored in the table REFERENCE ELEMENT) allow quantifying information when numerical values are not available. They are defined by a linguistic label that gives textual information about the quantity. For example, instead of numerical values, a set of membership functions "low", "medium", "high" can indicate a quantity (as shown in Figure 14).

Reference

Id	Type	Name	Is no quantity
11	Target	*null*	*null*
12	Legal	*null*	*null*
13	Precision	*null*	*null*
14	Membership functions set	*null*	*null*
...	...	...	...

Reference Element

Id	Ref_Id	Type	Name	Is Relative To Value	Values	Examples of applications
1	11	Interval	*null*	false	(50,60)	e.g. The target value is between 50 and 60.
2	12	Upper bound	*null*	false	(- ,10 ppm)	e.g. The maximum legal concentration is 10 ppm.
3	13	Single value	*null*	true	1 mm	e.g. The precision of a value is 1 mm.
4	14	Membership function	Low	false	MF(0,1,1,3)	e.g. A set of membership functions
5	14	Membership function	Medium	false	MF(1,3,4,4)	
6	14	Membership function	High	false	MF(4,4,9,9)	
...	...	...	...	...	...	

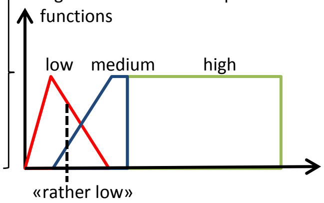


Figure 14 Examples of "Reference elements" and their applications

A fuzzy value "rather low" could be defined by a "membership values set" that contains two membership values (e.g. 0.3 "medium" and 0.7 "high"). While the entities MEMBERSHIP VALUE SET and MEMBERSHIP VALUE are used for defining the "truth values", the REFERENCE and REFERENCE ELEMENT define the membership functions. Besides fuzzy values, these four entities are designed to handle other types of composite data such as classifications or pedigree matrices (see Section 4.4.2)

Furthermore, the entities REFERENCE and REFERENCE ELEMENT can express different types of references (defined in the REFERENCE TYPES DOMAIN) such as legal values, target values or precision values. Figure 14 describes some of the possibilities offered by the

concept of “reference element” to define for instance: a target interval (Id=1), a maximum (upper bound) legal concentration (Id=2), the precision of a value (Id=3) and a sequence of membership functions (Id=4,5,6). The types of elements (lower bound, upper bound, interval, membership function or single value) are listed in the REFERENCE ELEMENTS DOMAIN.

3.3 Software architecture and functionalities

The Combined Water Information System (CWIS) is based on a multi-tier architecture, using the Microsoft Silverlight application framework, the Windows Presentation Foundation (WPF)[8] and ASP.net 4.0[9]. The client tier is installed on the computer of the users and mainly contains the user interface (i.e. the part of the software application that users see and interact with). The middle tier controls the functionalities of the application, processes the commands and orchestrates the transfer of data between the client tier and the database server (data tier). Each tier is independent and can be upgraded or replaced without affecting the others. The architecture is summarised in Figure 15.

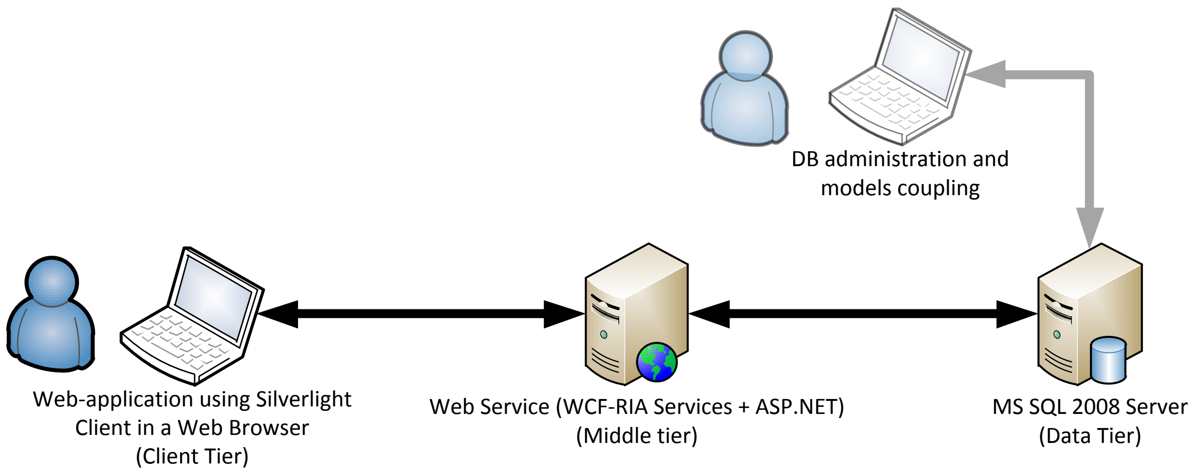


Figure 15 Software architecture of CWIS

The data tier consists of the database server and its database management system (DBMS), where data are stored and from which they are retrieved. In the DBMS (PostgreSQL for the 1st prototype and Microsoft SQL server 2008 for the second), the data are organised in the form of interconnected tables that follow a predefined database design, in this case the database schema described in Section 3.2.

[8] WPF: http://windowsclient.net/wpf/, Retrieved 28 December 2010

[9] ASP.net 4.0: http://www.asp.net/, Retrieved 28 December 2010

The middle tier is a web server that controls the application functionalities and the transfer of data between the database server and the client(s) of the application. The tier is divided in two sub-tiers: the Service Layer (SL) that contains part of the application logic and the Data Access Layer (DAL) which provides an access to data of the database. Within the DAL, the mapping between the relational data stored in a Microsoft SQL Server 2008 and the programmatic objects used by the application is made by the “ADO.NET Entity Framework”[10]. The Windows Communication Foundation for Rich Internet Applications (WCF-RIA) services[11] has been used to produce the code that coordinates the application logic between the client tier and the middle tier (web server).

The client tier contains the second part of the logic and the user interface, which consists of a modular interface that provides a series of visualisation tools (Figure 16).

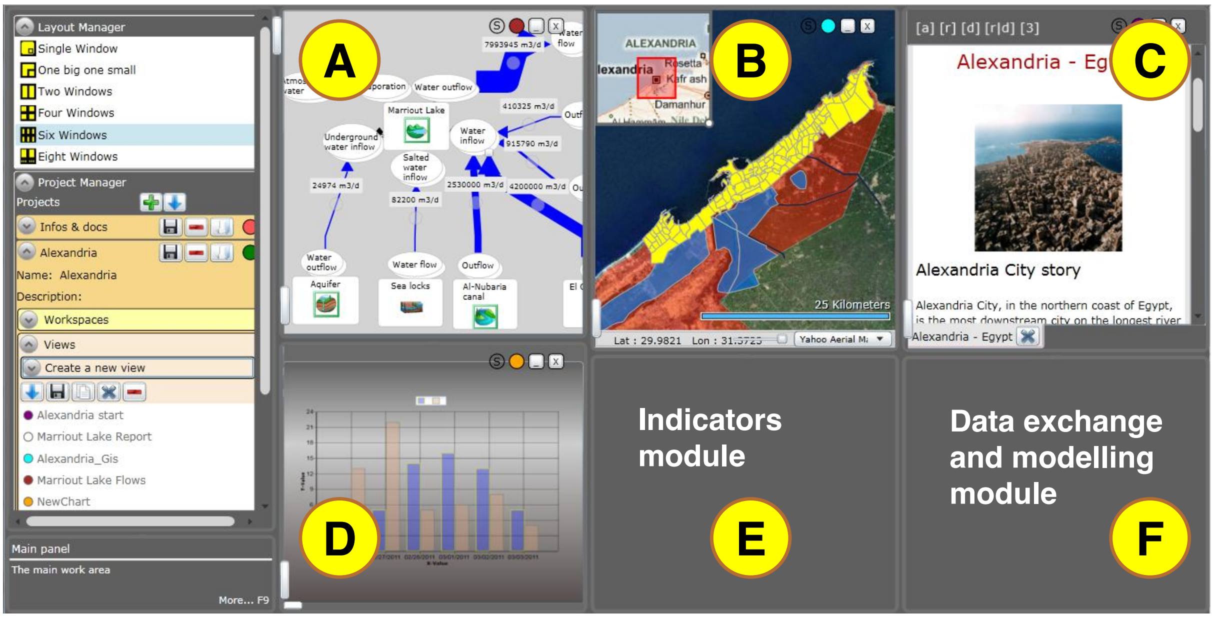


Figure 16 Overview of the CWIS user interface

The three main modules of the application (A) the system viewer, (B) the geographic viewer and (C) the active and reporting tool (ART) provide three different, but complementary, ways of managing heterogeneous and transdisciplinary data. The additional modules (D) the chart viewer and (E) the indicator viewer allow representing and comparing numeric values. The

[10] ADO.NET Entity Framework: http://msdn.microsoft.com/en-us/data/ef.aspx, Retrieved 30 November 2010

[11] WCF-RIA Services: http://www.silverlight.net/getstarted/riaservices/, Retrieved 30 November 2010

task of the last module, (F) the data exchange and modelling tool, is to import and export data, as well as to couple models to CWIS. All the modules were available in the first prototype of CWIS, however two of them (E and F) still have to be implemented in the second version.

A) System module

The system module enables the management of ontologies and the modelling of systems, based on the SysMod language. The "systemic view" allows editing systems and their properties (such as components, inputs, outputs, requirements or descriptions) and their interactions (flows or influences) Beyond the representation of systems, the module enables the creation of thematic views such as problem trees, views of flows, causal loops diagrams, etc. As of today, the mereotopological relationships introduced in Section 2.3.2 still need to be implemented in the module.

B) Geographic module

The geographic module allows visualising and managing the spatial properties of system elements, i.e. their locations and spatial shape. The module can display multiple layers of objects and edit their appearances, including their colour, transparency or line thickness. It also enables the creation of thematic maps, where the appearance of the spatial elements corresponds with some indicator values. The background cartography displayed by the geographic module, is based on maps from the major internet map providers (Google Maps, Bing Maps, Open Street Maps and Yahoo Maps).

C) ART (reporting) module

The "Active Reporting Tool" (ART) is primarily a tool for data management. ART is designed for the edition and visualisation of the non-spatial attributes linked to a system element, such as numeric values, texts, files (images, videos, pdfs, etc.). In addition, it allows creating reports that looks like web pages. These reports are based on Flow documents[12] (a feature of WPF) that are written in the Extensive Application Mark-up Language (XAML)[13]. Each report can "dynamically" incorporates data from the database to provide up-to-date information, for example by showing the last available value of a property, by updating data according to changes in the database, or by filtering the content of the report based on a time interval.

[12] Flow Document Overview: http://msdn.microsoft.com/en-us/library/aa970909.aspx, Retrieved 28 December 2010

[13] XAML Overview: http://msdn.microsoft.com/en-us/library/ms752059.aspx, Retrieved 28 December 2010

D) Chart module

In addition to the ART that displays numeric values as lists of data, the charts view shows charts of numeric values series. It supports any kind of traditional formats (e.g. scatter charts, pie charts, histograms, radar plots), as well as some more advanced formats (e.g. Gantt charts).

E) Indicators module

This module shows numerical values in tables of results, instead of lists like in the ART module. Its main purpose is to support the comparisons between similar system elements, scenarios (described in Chapter 4) or simulation results (described in Chapter 5).

F) Data exchange and modelling module

The “data exchange and modelling module” provides support to data import/export, as well as to perform simulations based on models externally connected to CWIS. Using some “export data structures” defined in system views, the module generates a list of the expected systems elements to export (or import). It works as a shopping basket or checklist of the input/output requirements as explained in Chapter 5.

3.4 Specifics related to the SysMod language

CWIS differs from traditional information systems by its data organisation based on the SysMod language. In order to highlight the advantages and limitations of SysMod, the next paragraphs examine the contributions of SysMod to the some of the main functionalities of CWIS: the ability to represent and navigate through the system being modelled, the integration of the concepts of classes, inheritance and polymorphism, the use of ontologies to assist system modelling, and the capacity to filter system’s data based on temporal conditions. For each of these functionalities, the assessment considers successively the three main modules of the application: the system module, the geographic module and “ART” the active reporting tool

3.4.1 System representation and navigation

The SysMod language has two main features: the creation of ontologies that describe the elements and relationships of a field of knowledge and the modelling of systems based on these ontologies. The structure of the SysMod constructs is reproduced in the database schema and is therefore also reflected in the CWIS modules.

System module

The system module allows editing both ontologies and systems in the same environment. It also provides the functionality of "dynamic exploration", which is used to navigate the system through the different types of relationships available in SysMod, as shown in Figure 17. In the exploration mode, all the elements connected to the initially selected item are retrieved and temporarily displayed. The latter can then be added to the view and used for further exploration.

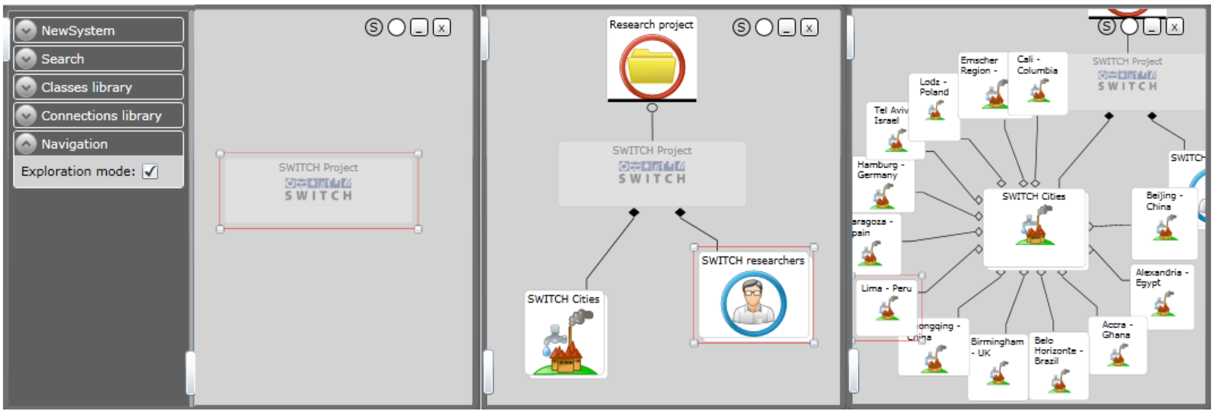


Figure 17 Dynamic exploration of data in the system view

Geographic module

Unlike conventional geographic information systems (GIS), the structure of the database is not based on the principle of "one table per layer of objects". The spatial data are organised according to the SysMod formalism, the geometries being specified in the "geometry" property of the objects. Although it does not alter the functionalities of the module, the increased data size imposed by the data structure of SysMod (currently) causes some slowdown of the application, due to the large quantities of objects required by the geographic layers.

Using the SysMod language, the objects represented in the geographic layers are considered as systems that can be represented with their properties and relationships in the system module. By synchronising the modules of CWIS, it is for instance possible to select an object in a system view in order to display its localisation and spatial shape in a geographic view and vice-versa.

Furthermore, the geographic module allows creating thematic maps based on indicator values (Figure 18). For this purpose, the indicators' layers are based on the collection of objects' properties having the same name and the same parent class.

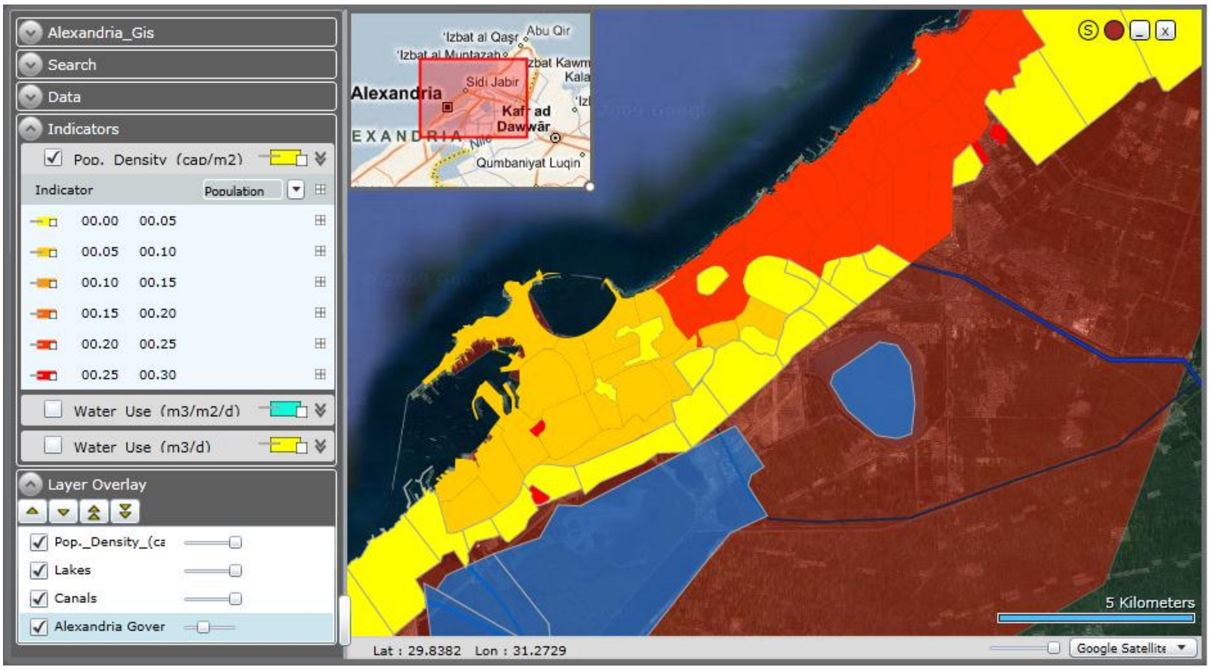


Figure 18 Thematic map with the geographic viewer

Reporting module

Systems diagrams are not suitable for managing data such as texts, files or values, while, on the contrary, the reporting module is specifically designed for this task. The coordinated use of the system module and the reporting module provides therefore a simple and flexible way to structure information; the "information properties" of a system element are edited with the system module, and for each property a tab is automatically generated to host the corresponding data in the reporting module.

Each page opened in the reporting module displays the information linked to a specific system element. Besides reports and data sets, a page provides also a series of "SysMod" hyperlinks, which characterise the different systems' relationship of the considered element (such as the relationships to the parent class, to the subclass or to the property members). Thus, the module allows the users to navigate the information simply by clicking on the hyperlinks.

3.4.2 Classes, inheritance and polymorphism

In the same way than programming languages, the SysMod language is characterised by the features of relationship class-instance, inheritance and polymorphism are applied to the SysMod language.

The classes correspond to the elements of the ontology, they can be instantiated in real objects ("holons" or pieces of information). When instantiated, a holon (or an information item) inherits the properties of the class, i.e. the properties of the instance are by default identical to those of the class. This is the concept of inheritance, which helps to reuse the information embedded in the ontology. However, the inheritance "rule" is not strict; some flexibility is provided through polymorphism. In the present case, the concept of polymorphism means that a subclass or an instance of a class may exhibit different properties than the class.

Moreover, SysMod permits multiple inheritance. This means that a class may have several super classes or that an instance can be derived from several classes. In both cases, the multiple ineritance results in a merging of the properties inherited form the parent classes.

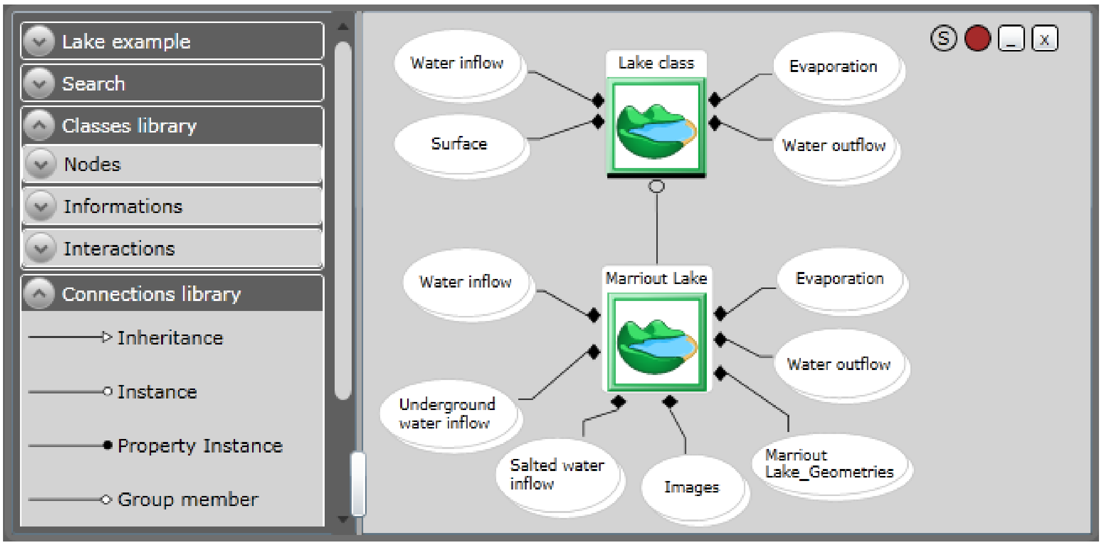


Figure 19 Example of instance with its specific properties and those inherited from the class

System module

In the system module, the concept of inheritance is characterised by the automated creation of properties when a subclass or instance is produced. The concept of polymorphism ensures that once the element created, the properties and their content can be modified or even removed, and new properties can be added without producing any conflict. Figure 19 illustrates these concepts by showing the example of a particular lake (the Lake Marriout) which is an instance of the class “Lake”. Through inheritance, the Lake Marriut received the same properties than the class “Lake” and based on the concept of polymorphism, new properties specific to the Lake Marriout have been added.

Geographic module

Through the concept of polymorphism, an instance of object can possess a “geometry” property even if its corresponding class does not have one. This is particularly convenient in this case, given that conceptual objects like classes generally do not have a geographic location.

Once the editing capabilities implemented in the geographic modules, the inheritance could be used to provide a predefined spatial shape for the instantiation of an object on a map. It however implies that the shape defined by the class of the object does not have a geographic location.

Reporting module

In addition to the visualisation and edition of inherited properties and property values, the reporting module also applies the concepts of inheritance and polymorphism to the reports. Therefore, the reports created for a class are also displayed (read only) when the page focuses on a subclass or an instance of the class. Therefore, even if an object (holon or information, as well as instance, property or class) does not have its own report, the report inherited from its parent class is displayed.

Figure 20 shows two report views: the first describes the class “Lake”, the second an instance of the class (the Marriout Lake). The report of the class named “Lake definition” is also available in the view of the instance.

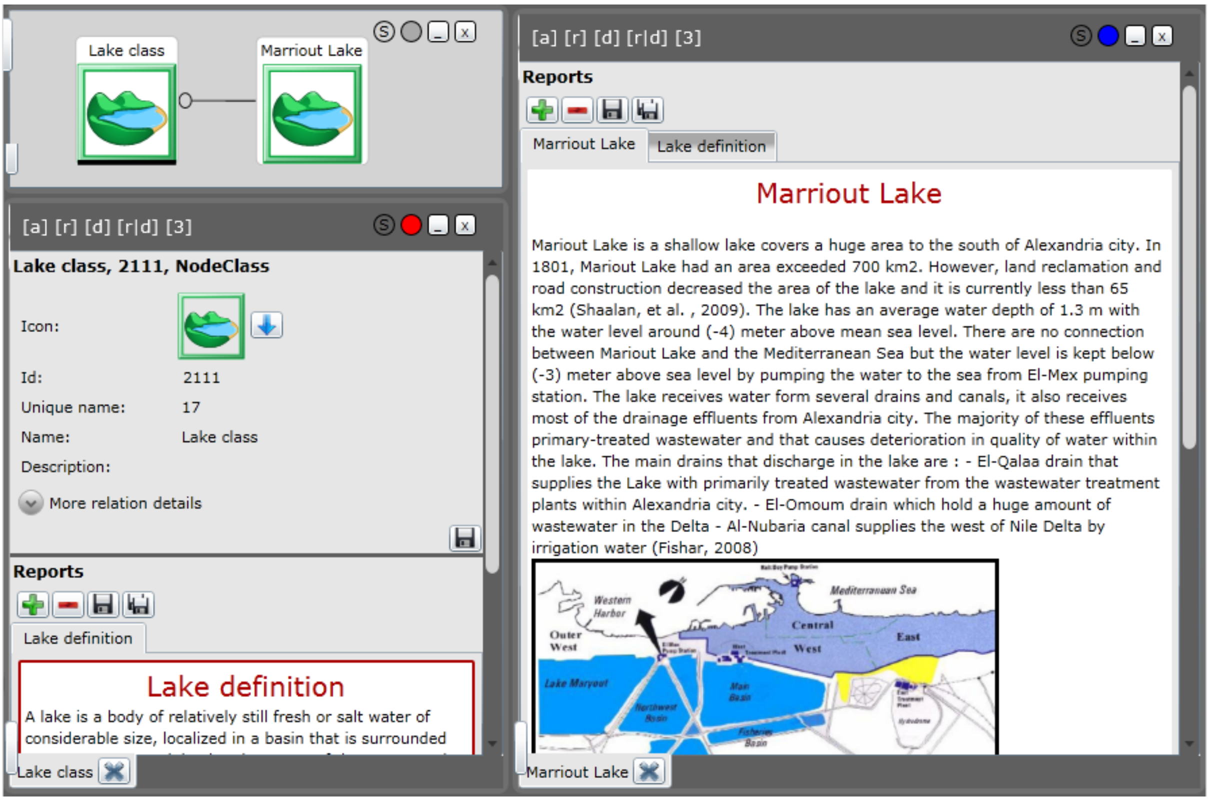


Figure 20 Example of specific report and report inherited from the class

3.4.3 Ontologies to assist system modelling

As explained in Chapter 2, SysMod enables the creation of ontologies, which consists in the definition of all the elements (the classes with their properties and relationships) required to describe a field of knowledge. Therefore, beyond the inheritance of properties, ontologies such as “the water management system model” (Schenk, 2010) provide an effective way to assist the modelling of systems.

System module

While inherited properties are automatically created or suggested to the instances of a class, other features such as the (SysMod) flows, causal relationships and mereotopological relationships may also be inferred from the ontology to help system modelling.

Therefore the ontology may help to identify the elements, properties and relationships to be added to a system model. For example, while modelling a sanitation system, the ontology may suggest that the wastewater treatment plant (WWTP) should be connected to a sewer network, and that a “flow” is required between the sewer network and the WWTP (Figure 21).

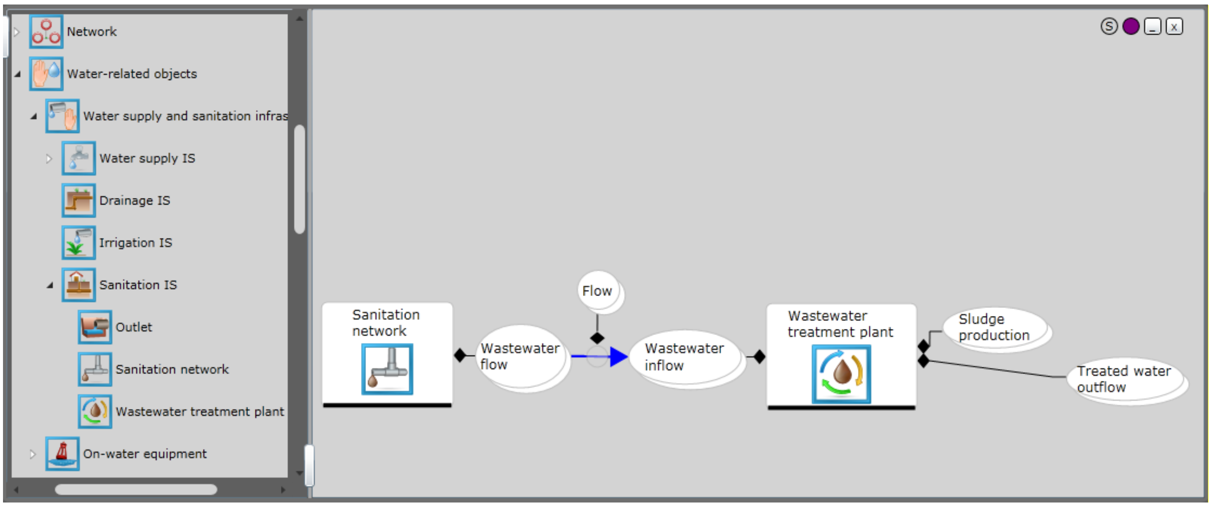


Figure 21 Definition of a waste water treatment plant (WWTP) in the ontology

Geographic module

Although the use of ontologies is less interesting in the case of the geographic module, the integration of mereotopological relationships in the ontologies can help positioning objects on the maps. For instance, the ontology may define through a mereotopological relationships that a WWTP should be located on the shore (the WWTP "is a non-tangential proper part (NTTP) of" the shore) and connected to a sewer network (the WWTP "is externally connected (EC) to" the sewer network).

Reporting module

The main benefit of ontologies in the reporting module is the automated creation of properties, property values and reports through inheritance. Ontologies also offer interesting perspectives to check for data inconsistencies and to assist the validation of properties and values, through comparison with the classes of the ontology.

3.4.4 Temporal filtering

The systems being represented with CWIS are natural or human systems whose properties and values evolve with time. For this reason, CWIS need to handle the modifications of systems data (system elements and system relationships) over time. While the temporal information is already included in the database, some programmatic developments are still needed to make the temporal filtering operational in the software application.

This feature is derived from the concept of "Time context", introduced in the database schema of SysMod. The "Time context" attributes some start and end times to the system

elements. As a result, to provide readable views, the system-, geographic- and reporting modules of CWIS need to filter the systems data based on chosen periods or moments in time.

System module

Based on the time context, the system module can display the systems evolution over time. The types of change highlighted by the temporal filter concern:

- The life time (time context) of the system, e.g. the WWTP is shut down in 2015 and replaced by a new one;
- The system composition and inputs/outputs (based on the holon properties), e.g. a new treatment step for electricity cogeneration is added in the WWTP;
- The information properties, e.g. a new property "amount of electricity produced";
- The connections (mereotopological relationships) with other elements, e.g. a second sewer network is connected to the WWTP;
- The flows and causal relationships, e.g. the WWTP receives wastewater from the second sewer network and the increase of the inflow affect negatively the treatment performance.

Geographic module

In the geographic module, the temporal filter allows the user to follow the spatial changes of systems. For instance, it can display the expansion of a city over the years or show the temporal changes of a parameter through various indicators' map obtained by filtering the values.

Reporting module

Like the other modules, the reporting module beneficiates from the temporal filtering capability. By applying the filter, the catalogue of data therefore only displays the properties and data that belong to the selected time context. The situation of the reporting section is slightly different as one may want to include in the reports some values or pieces of information referencing a time period that differ from the filter definition. The reporting tool is therefore designed to either display a specific data item (selected according to its database identifier value) or to apply the filter on the "cells" that references the data. If no filter is applied, the last available value is shown by default.

3.5 Synthesis

CWIS consists of a modular software application and a database built upon the SysMod language. Altogether, the three main modules of CWIS allow representing any kind of system, be it human or natural: the system module provides a schematic view of the system structure, the geographic module allows representing the spatial attributes of the system, and the reporting module covers the management of information such as values, texts, files, etc. The additional chart and indicator modules propose alternative way to display systems data, with charts and table of indicators.

Associated with the SysMod language, the concepts of class-instance, inheritance, polymorphism, ontology, and temporal and thematic filtering are taken up by the CWIS modules. The contributions of SysMod can be summarised as follows:

- The ontology allows representing a knowledge domain, it is composed of classes hierarchically organised and characterised by properties. Some instances can be derived from the classes to model objects/systems from the real world. The ontology can be adapted to address any kind of issues, as the SysMod language is not dependant of a particular knowledge domain.
- The concepts of inheritance and polymorphism enhance the reusability and flexibility of the information gathered in the ontology, by replicating the properties of a class to its instances (inheritance) while subsequently allowing their modification or the addition of new properties (polymorphism).
- The "time context" is a feature provided with the SysMod language to address the evolution of the structure, state and dynamic of systems. To be effective, the modules of CWIS need some temporal filters to only display the information relevant for a specific moment or time period.

The SysMod language, however, poses challenging questions regarding the management of large series of data. When the number of items to be considered exceeds one hundred, their graphical representation (with the system module) becomes confused and unreadable. Moreover, the management of large data series (more than a thousand items) tends to slow down the application, because of the data organisation imposed by SysMod which involves the unnecessary multiplication of identical data for each element of the series. As regards the IRM process, a more problematic limitation of the current database and software application is their incapacity to handle several scenarios, preventing the use of CWIS for scenario planning.

4. Extension of the SysMod concepts to address scenario data

4.1 Introduction

Visualising and understanding the complexity of coupled human and natural systems is a key challenge in the field of integrated resources management (IRM). To address this challenge, a new systems modelling language (SysMod) has been presented in Chapter 2 and used for the creation of CWIS, the Combined Water Information System introduced in Chapter 3. At this stage, the information storage and visualisation capabilities of CWIS are, however, not sufficient to handle the requirements of scenario planning. For that reason, this chapter extends SysMod and CWIS to handle multiple possible futures, and to enhance the reliability of scenario/strategy assessment through the concepts of data provenance and uncertainty.

This chapter first presents the additions required by the database schema to support the definition of scenarios and strategies. Second, it provides an overview of the concept of data provenance and describes its implementation into the database schema. The data provenance gives information about the source and history of data, its main purpose is thus to capture the chain of data processes at the origin of the data sets. The management of data provenance is particularly relevant when dealing with multiple interconnected mathematical models or processes, as for instance to track data uncertainties through coupled climate-hydrology models or to detect possible errors through the sequence of data treatments in large data sets. The list of processes traditionally considered in the field of data provenance has been extended to integrate non-computer based processes such as measurements or expert judgements, and thereby capture the provenance of any kind of data. Third, the chapter presents the definition of the uncertainty dimensions of mathematical model provided by Walker et al. (2003), shows how the definition can be generalised to any kind of data process and describes its implementation into the CWIS database. The main interest of this definition is its capacity to handle many different types of uncertainties such as statistical, qualitative and scenario uncertainty, in a single integrated framework.

4.2 Scenario planning and scenario data

According to the principle of “bounded rationality” (Simon, 1982), any decision making process is hampered by three inevitable constraints: (1) the available information about how the system works is limited and often unreliable, (2) human cognitive abilities to evaluate and process the information is also limited, and (3) there is a finite amount of time to make the decisions. Therefore, the rationality of the choices is always bounded and it is not possible to determine with certainty an optimal decision.

These limitations are particularly significant in the field IRM which may involve decisions affecting the environmental, social, political and economic systems all together. The holistic framework of IRM approaches implies that decision making inevitably faces situations where the knowledge of the system is incomplete (varying from statistical uncertainties to complete ignorance), the understanding of the situation and the motivations may greatly differ between stakeholders, and a trade-off is required between the need of prompt decisions and the time required to further assess their feasibility and repercussions.

In this context, finding a “satisfying solution” that cope with a wide range of possible future, is more relevant than finding the “optimal solution”; since finding an optimal solution requires optimisation methods which suppose a very low level of uncertainty (Pahl-Wostl, 2007; Simon, 1956). A list of management methods is given in Figure 22, based on the levels of uncertainty and controllability.

When the key factors influencing the IRM issues are very uncertain but remain controllable, the system needs to be managed according to the concept of adaptive management, which aims to adapt the management strategies and increase the knowledge of the issues, by monitoring the system properties (Medema et al., 2008; Pahl-Wostl, 2007; Pahl-Wostl et al., 2007). In contrast, scenario planning is intended for situations where the key factors are uncertain and uncontrollable (Peterson et al., 2003). Scenario planning extends the concepts of adaptive management through scenarios, which intend to represent the most important consequences of the uncertain and uncontrollable factors (see detailed definition in the Thesis Introduction).

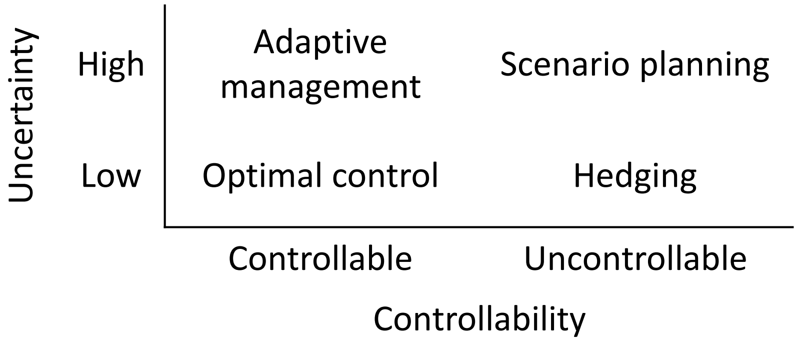


Figure 22 Classification of management approaches depending on the uncertainty and controllability of the factors (Peterson et al., 2003)

4.2.1 Definition of scenario data

In the field of scenario planning, the system under consideration is generally defined by a base case scenario, which describes the past evolution until today and a probable future based on the current trends. The creation of alternative scenarios implies that data can be affected to one scenario or the other, each alternative scenario being characterised by its differences with the base case scenario. These differences may concern various system features. They can affect the life time of the system, the values of the system's properties (e.g. the quantity of inputs, outputs and stocks), the system's properties themselves (e.g. a different systems composition), and the relationships with other systems (e.g. different flows, influences or topological relationships between the systems).

Scenario data can be defined as the selection of data that do not belong to the base case scenario, but only to an "alternative story" of the system. In that sense, it also includes the concept of strategies which are some "fragments of system's story" to be applied to a scenario. It should be noted that the combination of a strategy and a scenario also results in the creation of new scenario.

4.2.2 Implementation of scenario data

From a practical point of view, SysMod is not usable as such in the context of scenario planning, as it only provides data for a "single story" of the system. In order to allow several scenarios to coexist, the database schema of the CWIS database has been updated to integrate some "Themes", which are used to group the data per scenarios and or strategies (Figure 23).

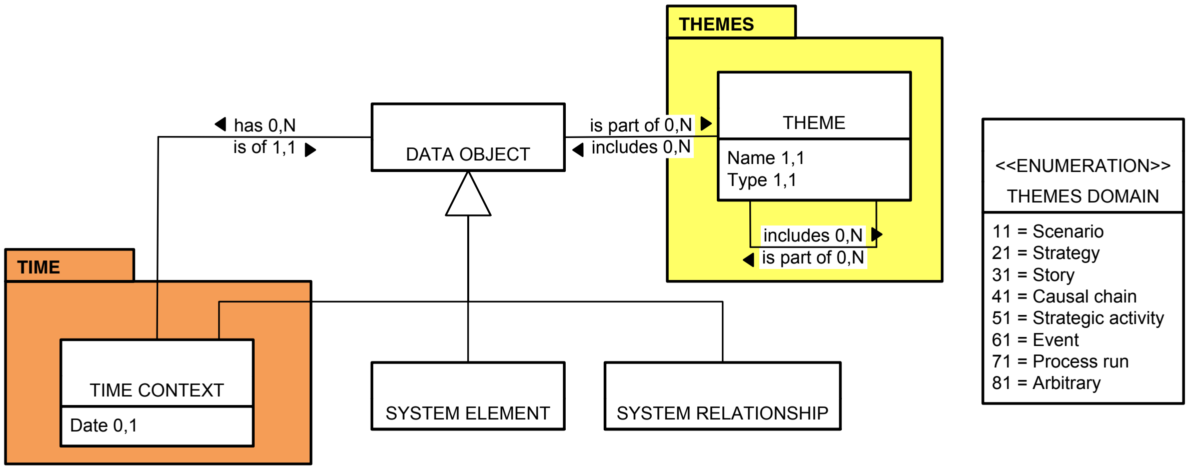


Figure 23 Extension of the database schema to scenarios and other "thematic" groups

As mentioned earlier, any SysMod construct (the SYSTEM ELEMENT and SYSTEM RELATIONSHIP entities) can be marked with a TIME CONTEXT. It includes an attribute Date, which can store dates, times, time interval and durations based on the ISO 8601 standard. As the time context applies to both SysMod elements and relationships, both the composition and the structure of the systems and ontologies can change over time. Given that multiple scenarios may involve different lifetime for a system element, the latter can therefore host several "time contexts", i.e. one per connected theme.

The THEME provides a selection of system elements and relationships. It allows the modeller to create scenarios or to symbolise other systems features such as events or strategies by just displaying the elements and relationships that belong to the thematic selection. Systems elements and relationships can belong to several themes.

Both time contexts and themes do not have a visual representation in SysMod diagrams. They just allow filtering the data to depict the systems at a given time and for a given scenario. For this reason among others, the SysMod language needs a software interface to provide a way to create and consult SysMod models. This task is achieved by the interface of CWIS introduced in Section 3.3.

4.3 Provenance of scenario data

The concept of data provenance has been applied to various computer-related domains such as database management systems and scientific data processing, but to our knowledge, this concept has never been used in the fields of scenario planning and environmental management.

4.3.1 Definition of data provenance

The term provenance (also known as lineage) pertains to the identification and description of the sources of data. When more than one operation is involved in the history of a data generation, data provenance can be tracked through workflows (some chains of interconnected processes), which allow identifying step by step the processes and the initial and intermediate data involved in data creation.

As stated by Gobel (2002) and Stimmhan et al. (2005), provenance techniques can be applied in several situations, which may all be relevant in the context of integrated resources management:

- To check the **reliability** and **quality** of data, by providing information about its source(s) or the process(es) that produced it. Knowing the source of data is a prerequisite before any further quality assessment, since it is impossible to consider the reliability of the sources and processes if they remain unknown.
- To provide an "**audit trail**" of data, for example to possibly detect errors in data generation or to determine the resources spent in the process.
- To support the **reusability, reproducibility and repeatability** of a workflow process. Reusability means that the data sets, the workflow and the processes themselves can be reused in the same or in other contexts. A typical case of reusability is the update of some results to take into account some data or process modifications. Reproducibility implies that the workflow of processes can be reproduced by other operators, possibly using other but similar processes, in order to validate or invalidate the results. It differs from the repeatability, which implies that the provenance keeps track of all the information required to reproduce the identical conditions for a specific workflow invocation (same data, same software versions, same processes…).
- To establish the **copyright and ownership** of data. Provenance information may be used by the data providers to charge the data acquisition. Alternatively, it may help determine the liability in case of erroneous data. It may be helpful when trying to distribute the responsibility for a decision.
- To support **data discovery** by providing an alternative way to consult data and processes based on the provenance workflow. A typical illustration of this situation is the exploration of articles based on the consultation of other article's references.

Provenance techniques have been developed principally for two kinds of tools. The first ones are the database management systems (Benjelloun et al., 2008; Buneman et al., 2001), where the provenance is characterised by "annotation" of the derivation history of the data or

through the “inversion” of the executed queries. An example of such database management system is TRIO (Agrawal et al., 2006) which addresses the issue of provenance through query inversion. The second types of tools are the scientific workflows systems (SWSs), such as Taverna (Oinn et al., 2004) or Kepler (Altintas et al., 2006), which are designed to compose and execute some chains of data processes for scientific applications. A more detailed description of the provenance techniques and tools is given in Appendix B.1.

The types of processes considered in SWSs are limited to the executable processes that can be controlled by the software system. The processes generally handled by the SWSs are:

- **Data transformations** covering a wide range of applications such as string manipulation, treatment of numerical data series, file parsing and writing, image processing and treatment of geospatial information.
- **Data analysis** such as error detection, data cleaning, statistical analysis or data classification.
- **Mathematical operations and models** such as the predefined operations proposed in the mathematical library of Ptolemy II (Eker et al., 2003) or some mathematical expressions using modelling environments such as Matlab (www.mathworks.com/products/matlab) and R (www.r-project.org).
- **Queries** such as SQL queries for relational databases, SPARQL for RDF graphs or XQuery for XML documents.
- **Display processes** which can display the outcomes of the workflow in various ways, such as graphical plots, data series, etc.
- **Composite processes (or nested workflows)**, as workflows can be composed of sub-workflows. This hierarchical organisation promotes a better reusability of the processes and is effective to perform particularly complex tasks.

However, when tracking provenance and the sources of uncertainty in a more general perspective, additional processes are required, such as:

- **Expert judgments**, which are used in the present case as a broad definition for all the processes that involves the subjective judgment of one or several individuals to produce an informed estimation of a data value (or state).
- **Measurements and observations**, which aim to quantify (or sometimes qualify) a phenomenon, possibly using a measuring instrument: e.g. the quantification of rainfall with a rain gauge, the measure of pollutant concentration based on a lab experiment or the definition of geographic coordinates with the help of a GPS.

- **Quotations and references**, which identify the source of data when no particular procedure or transformation is applied.

One can observe that the provenance recording is automated within tools such as SWSs, assuming that “humans are unlikely to record all the necessary information manually” (Frew and Bose, 2002). This is undoubtedly a helpful feature to support the collection of provenance metadata, but it also one of the reasons that non-computable processes such as expert judgments, observations or measurements are not be taken into account in such system.

It should be noted that CWIS (the software in which these concepts will be implemented) does not handle the execution of processes. Therefore both computer-based and non-computer-based processes must be recorded manually to establish the data provenance.

4.3.2 Implementation of data provenance

The provenance of data is defined as the specification of the processes and inputs involved in the data production. For this purpose, the concept of “hierarchical workflow” (some chains of interconnected processes, where each process can itself be a workflow) is implemented in the database. The part of the database schema dealing with provenance information is shown in Figure 24.

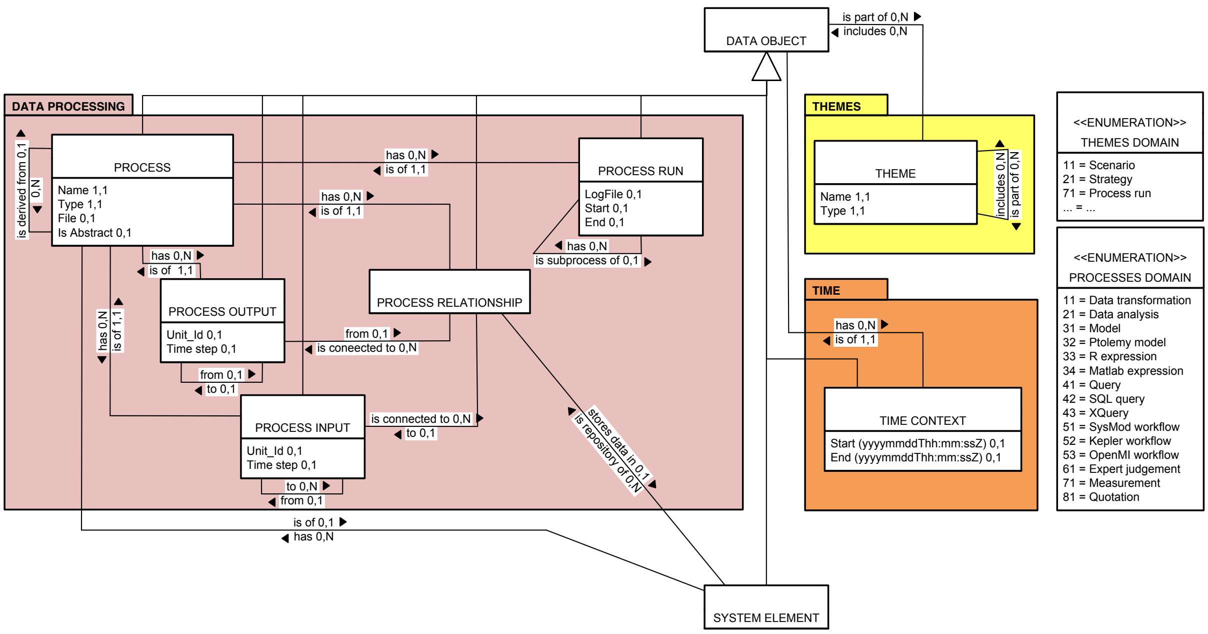


Figure 24 Part of the database schema that defines the processes used for tracking data provenance

Abstract vs. concrete process

The entity PROCESS is characterised by a Name, a Type (the list of available types is in the PROCESSES DOMAIN), and possibly a File which contains an executable or a workflow configuration (e.g. a workflow file for a scientific workflow system such as Kepler). The relationship between the entities PROCESS and SYSTEM ELEMENT indicates that a process can be carried out by a system or by the part of a system. A process defined as "Is Abstract" means the process is a kind of blueprint which can be used to derive other processes. An abstract process aims to describe all the components of the process, including the type of expected inputs and outputs, but without specifying real input, intermediary and output data sets.

Process inputs, outputs and relationships

Each process owns zero-to-many inputs (PROCESS INPUT) and may produce zero-to-many outputs (PROCESS OUTPUT). If required, both entities can reference the Unit and Time step of the input/output data. The connection between processes is defined through the PROCESS RELATIONSHIP, which connects the output of a process to the input of another one. If the relationship corresponds to an output of the main workflow, it is not connected to a PROCESS INPUT as no process follows the workflow. Similarly, relationships that define the inputs of the main workflow are not connected to a PROCESS OUTPUT. Each relationship can be linked to a SYSTEM ELEMENT that acts as repository for the data. For abstract processes, this system element is normally a "SysMod class" that can define a value type (flow, length, etc.) and a default unit. For concrete processes, this system element is a "SysMod property" which is derived from the previous class and can contains series of values.

Process run

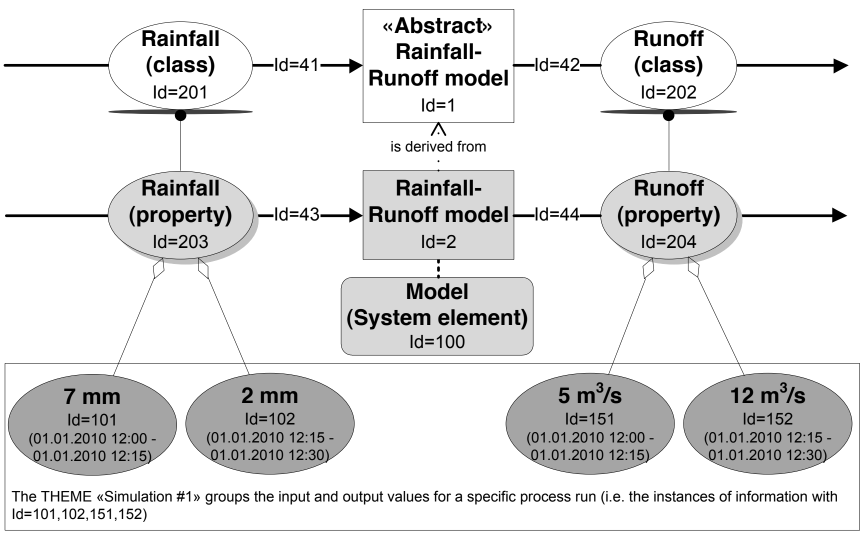
The PROCESS RUN is defined as the operation of a process during a continuous period. A model simulation, an expert judgment or an unbroken series of measures are specific cases of process run. Moreover, each run is characterised by a set of attributes: a Start timestamp, an End timestamp and possibly a Log-file which records the chronological events of the run.

Hierarchical workflows

The PROCESS RUN has a self-loop "is subprocess of" that hierarchically defines the sub-processes of the workflow. In the same way, the self-loops of the PROCESS INPUT and PROCESS OUTPUT allow defining the transfer "from" an input of the workflow "to" one or several sub-processes and respectively conveying the output "from" a sub-process "to" the output of the workflow.

Data provenance

Given that all system elements and process-related entities (Figure 24) are specific kinds of DATA OBJECT, they can be linked to a THEME or a TIME CONTEXT. In the framework of data provenance, the THEME is used, in this case, to group all the information linked to a specific PROCESS RUN. Thus, the provenance (and utilisation) of values can be retrieved by identifying the workflow linked to the property that contains the values and by selecting the workflow input and intermediate data (and output data) based on the values' THEME (as shown in Figure 25).



DataObject_Theme

DataObject _Id	Theme_Id
11	301
101	301
102	301
151	301
152	301

Theme

Id	Name	Type
301	"Simulation #1"	Process run

Process Relationship

Id	From_Id	To_Id	SystElem_Id
41	*null*	21	201
42	31	*null*	202
43	*null*	22	203
44	32	*null*	204

Process

Id	Name	Type	SystElem_Id	Is_Abstract	Derived_from
1	"Abstract" Rainfall-runoff model	Model	*null*	true	*null*
2	Rainfall-runoff model	Model	100	false	1

Process Input

Id	Process _Id	Unit _Id	Time step	Parent _Id
21	1	90	15mn	*null*
22	2	90	15mn	*null*

Process Output

Id	Process _Id	Unit _Id	Time step	Parent _Id
31	1	91	15mn	*null*
32	2	91	15mn	*null*

Process Run

Id	Process _Id	Parent _Id	LogFile	Start	End
11	2	*null*	*null*	01.01.2010 12:00	24.11.2010 12:00

Figure 25 Illustration of the use of process-related information to display a "provenance workflow"

In the database, the entities of the database schema are transformed in tables that host the data. To illustrate how provenance-related data are stored in the database, Figure 25 shows an example of "provenance workflow" and the corresponding content in the tables. The

entities in white represent an “abstract” workflow, which is a template for the implementation of a “concrete” workflow (in light grey) for simulations or data processing. The data required for the creation of these workflows are stored in the tables *Process, Process Relationship, Process Input* and *Process Output*. The colours of the workflows correspond to the colours of the rows. The “concrete” workflow can be used to perform several runs. The data relative to a specific run are shown in dark grey. The location of the inputs and outputs are the same for each run of the workflow, the distinction between the data of different runs is made with the *Theme*, which groups all the data that belong to the same simulation.

So, now if someone wants to know the source of the runoff value 5m3/s, one can perform some database queries to find which process and which input data have been used to generate the value. Moreover, the “concrete” workflow can serve as a basis to reproduce the process, while the “abstract” workflow allows to set up (i.e. to reuse) the process in another context.

4.4 Uncertainty of scenario data

Addressing uncertainty in the field of IRM is a challenging task, because of the variety of data sources and the different forms of uncertainty. Some frameworks, such as the Harmoni-CA Guidance for Uncertainty Assessment (Refsgaard et al., 2007; Refsgaard et al., 2005b), may help choosing adequate methods for uncertainty assessment. A brief description of how this guidance can support the selection of methods is given Appendix B.2.

In order to facilitate the identification and analysis of uncertainties, this section proposes a generalisation of the typology of uncertainties provided by Walker et al. (2003). Originally describing the uncertainty in model-based decision support, the typology has been generalised to consider other types of data processes, such as measurements, expert judgements and or data transformations (see list of data processes in Section 4.3.1). The section then describes the implementation of the typology in the CWIS database.

4.4.1 Definition of uncertainty data

The typology of uncertainties is based on the three dimensions defined in the next sections: the location, level and nature of the uncertainty associated with data and processes. Together, these dimensions can form an uncertainty matrix (Walker et al., 2003), which is a suitable tool to identify, characterise and monitor uncertainty. An example of uncertainty matrix is given in Section 5.3.2.

Location of uncertainty

Initially defined for characterising model-related uncertainties, the list of uncertainty locations provided by Walker et al. (2003) can be generalised to take into account other types of data processes, such as measurements, expert elicitations or data treatments. The locations are sorted into five categories:

A. *Context and framing:* The relation between the process application and its context is an important uncertainty location. It mainly depends on the application domain of a mathematical model, the expertise domain of an expert or the environmental constraints to be respected when installing a measuring instrument. As an illustration, a non-heated rain gauge may undervalue the precipitation value in case of snowfall, meaning that the domain of applicability of the instrument is not respected. More generally, context uncertainties may depend on various environmental, economic, social, political and technological criteria.

B. *Structure of the data process:* In the case of a mathematical model, the structure of the data process corresponds to the model formulation. As models are simplifications of the real world, incomplete understanding or over-simplified description of the reality may be significant sources of uncertainty. In the case of measurements, the structural uncertainty may depend on a conceptual misconception of the measuring instrument or result from the intrinsic variability of the technical process. Finally, from the perspective of the complete workflow itself, structure uncertainty depends on the adequacy of the relationships connecting processes. For instance: Does the output of a first process match the expected input of a next one? Are the scale and the time intervals of the models compatible?

C. *Technical implementation:* Implementation uncertainties are uncertainties linked to the misapplication of the process. For instance when implementing a mathematical model, uncertainties may emerge from numerical approximations, from resolution in space and time, or from bugs in the software. For expert judgements, implementation uncertainty comes from the way the process is applied. It can arise from ambiguities in the question formulation or from conclusions based on a non-representative sample of experts. In the case of measuring instruments, some breakdowns, bad calibration or wrong manipulations are potential sources of this kind of uncertainty.

D. *Parameters:* Like any other data, parameters are affected by uncertainties. When the parameters are defined through calibration, it should be highlighted there is a close relationship between structure and parameters uncertainties. As stated by Refsgaard et al. (2006), “an inadequate model structure will, however, be compensated by biased

parameter values to optimise the model fit with field data during calibration". It results in an increase of parameters uncertainty and an underestimate of the uncertainty related to the structure of the process.

E. *Data in general*, such as numerical values, time series, geographic data or temporal data. The input and output of the process are included in this definition. The challenge is to determine how output's uncertainties are affected by the inputs', as well as the other sources mentioned just above.

The total uncertainty of the workflow (i.e. the uncertainty of the final output) can be assessed by uncertainty propagation techniques (see Appendix B.2), but prior to that the level and nature of uncertainties need to be identified for each location.

Level of uncertainty

The level of uncertainty, the second dimension suggested by Walker et al. (2003), aims to characterise the state of the knowledge for a specific issue. It is defined as a level between an ideal (and unachievable) complete certainty and a total ignorance.

Defining the uncertainty level is based on the "state of confidence" (Brown, 2004a) in the accuracy of the information. Thus, positive or negative surprises may happen, whether the result matches the reality while the uncertainty was high, or conversely the result is incorrect because of ignorance despite a previous high confidence in the process and data.

Figure 26 Levels of uncertainty (based on Refsgaard et al., 2007; and Walker et al., 2003)

As shown Figure 26, the levels of uncertainty that can be addressed are bounded between the "certainty" which is the ideal case of complete knowledge and the "indeterminacy" which refers to situation where the uncertainty cannot be determined. The different levels are:

- *Statistical uncertainty* applies to any value whose deviation from reality can be characterised statistically. This type of uncertainty may have various sources, such as the precision of the measuring instrument, the representativeness of the sample used to derive the data or the adequacy of a model to describe real-world phenomena.

- *Qualitative uncertainty* characterises the uncertainty with qualitative information. Its main purpose is to provide an insight into data quality when there is no statistical uncertainty available. A method frequently used for the assessment of qualitative uncertainties is the pedigree matrix (Funtowicz and Ravetz, 1990; Weidema and Wesnæs, 1996). Qualitative uncertainty can be used jointly with statistical uncertainty, especially when the confidence in the latter is limited.
- *Scenario uncertainty* refers to incomplete or missing information needed to execute a process. It generally concerns external forces that are not controlled by the system to be modelled, but for which plausible values are known. In the field of IRM, classical examples of scenario uncertainties are the predictions of the demography and climate change. To cope with this type of uncertainty, methods such as the scenario analysis (or scenario planning) enable the creation of plausible future situations (scenarios), in order to explore their possible consequences and help defining strategic measures.
- *Recognised ignorance***,** when no information is available to characterise the data uncertainty.

Nature of uncertainty

The nature of uncertainty is the third dimension of the concept of uncertainty proposed by Walker et al (2003). This dimension makes the distinction between stochastic (irreducible) uncertainty and epistemic (reducible) uncertainty. Although these two “natures” are theoretically sufficient to describe any kind of uncertainty, an extension can be useful to consider practical implementation aspects.

- *Stochastic uncertainty* (also named irreducible, objective, random or variability uncertainty) depends on the inherent variability of the phenomena being described. The uncertainty cannot be reduced but it can be assessed through statistical analysis.
- *Epistemic uncertainty* (also named reducible, subjective or ontological uncertainty) is due to the imperfection of our knowledge. It means that additional research may improve the quality of our knowledge and thereby reduce the uncertainty.
- *Undefined nature*: This additional point aims to take into account the lack of knowledge about the nature of the uncertainty.
- *“Postponed” (epistemic) uncertainty* describes uncertainties which are theoretically reducible but irreducible in practice because of data scarcity, or due to some lack of operational and economic resources. As stated by de Rocquigny et al. (2008): “For practitioners, the reducibility issue may therefore be more of a context-dependant feature or even of a modelling choice.”

Data are often affected by stochastic and epistemic uncertainties at the same time, i.e. a part of the uncertainty is reducible while the other part remains irreducible. In this case, and from a practical point of view, the uncertainty nature should be set as epistemic.

4.4.2 Implementation of uncertainty data

The database of CWIS is designed to handle many types of uncertainty and uncertainty values. The characterisation of the uncertainties is based on the three dimensions proposed by Walker et al. (2003): the location, level and nature of uncertainty. These dimensions are hold by the entity UNCERTAINTY SOURCE, which assigns the uncertainties to the data. The utilisation of the uncertainty sources and uncertainty values is summarised in the fragment of the database schema shown in Figure 27 (see Appendix A for the complete database schema).

The attributes of the UNCERTAINTY SOURCE are:

- The Name that textually identifies the uncertainty source;
- The Location, which is the “Process context”, “Process structure, “Process implementation”, “Process parameter” or “Data”, as defined by the UNCERTAINTY LOCATIONS DOMAIN;
- The Statistical level, Qualitative level, Scenario level and Recognised ignorance, which document the uncertainty level of the source using the THREE SIZES DOMAIN (Small, Medium or Large);
- The Nature of uncertainty whose possible values are defined by the UNCERTAINTY NATURES DOMAIN (Undefined, Epistemic, Stochastic or “Postponed” epistemic);
- The Significance that gives qualitative information about the impact of the uncertainty, using the THREE SIZES DOMAIN.

The UNCERTAINTY SOURCE can affect all the entities grouped under the term DATA OBJECT: the SysMod constructs (not detailed in the schema), some of the process-related entities (the PROCESS, PROCESS OUTPUT and PROCESS RUN).

The attribution of the uncertainty value is based on the relationship “has zero to many” between the UNCERTAINTY SOURCE and the VALUE connected to the uncertainty source. The entity VALUE can be used to save both statistical and qualitative uncertainties. The NUMERIC VALUE enables the management of single statistical values, such as standard deviations, variances, means or probabilities of occurrence. The NUMERIC INTERVAL deals with composite values such as confidence intervals for quantiles and the DISTRIBUTION handles probability density functions.

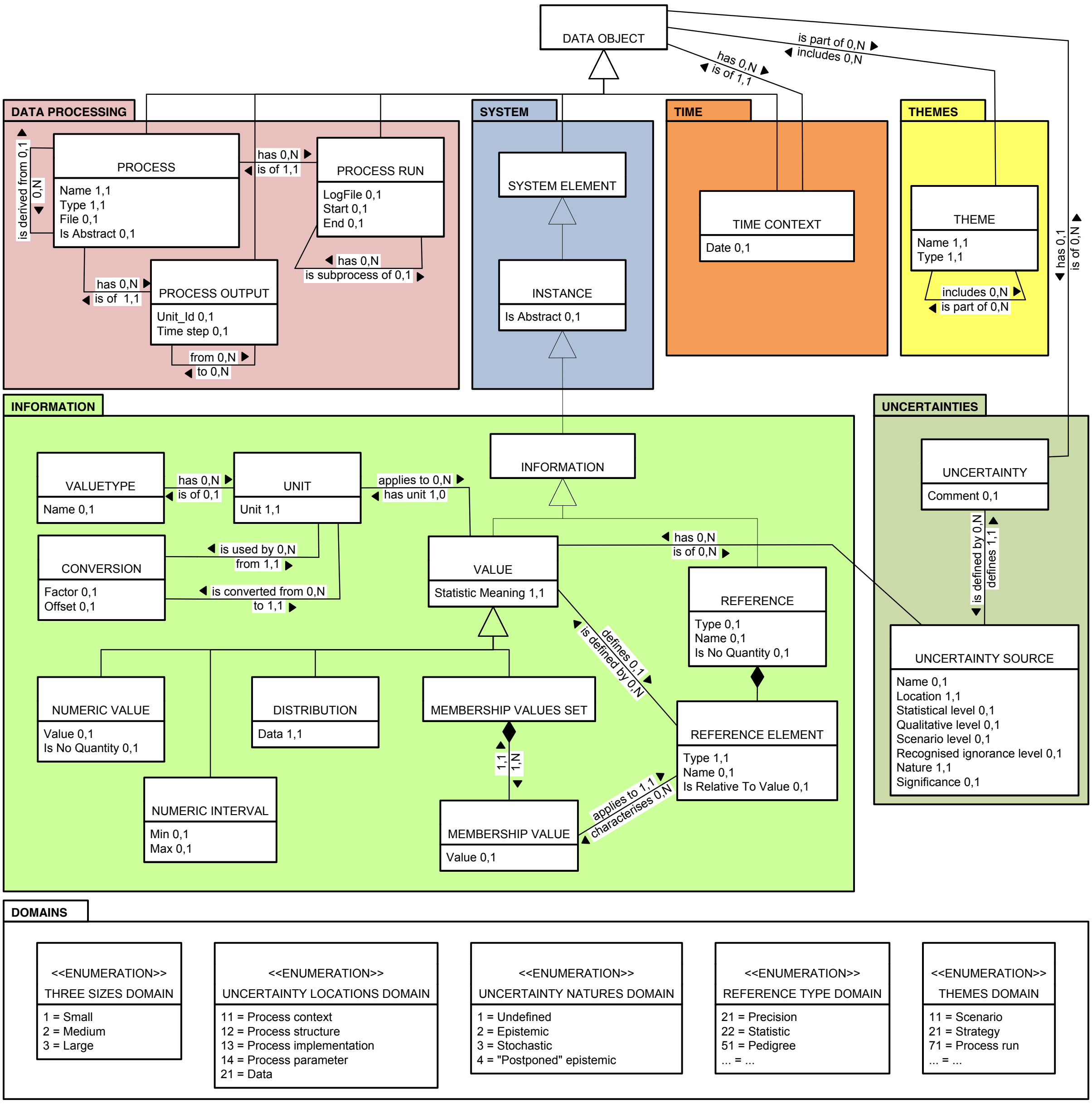


Figure 27 Part of the database schema that enables the identification of uncertainty sources and the storage of uncertainty values

Besides providing all the required data (i.e. uncertainty locations, levels and natures) to create uncertainty matrices (see Section 5.3.2), the main originality of the uncertainty part of the database schema is the ability to manage qualitative uncertainties. For this purpose, the schema is designed to support the creation of pedigree matrices (Refsgaard et al., 2006; van der Sluijs et al., 2005), which provide a standard structure to incorporate qualitative expert

judgements about processes' and data uncertainties. Table 3 shows an example of pedigree matrix to characterise the uncertainties related to the structure, parameters and overall validity of a mathematical model (such as the rainfall-runoff model shown in Figure 25). The pedigree matrix is composed of criteria, which are evaluated using a discrete numerical scale from 0 (weak) to 4 (strong). A description is given for each criterion's score to facilitate and reduce the subjectivity of the assessment. In the present case, the pedigree matrix shows that the model structure is a "transfer function", and that the model is not calibrated and validated, since the parameters are some "expert guesses" and the results have not been analysed at all.

Table 3 Example of pedigree matrix for evaluating model uncertainties, adapted from Funtowicz and Ravetz (1990)

Score	Model structure		Parameter(s)		Testing	
0		Definitions	1	Expert guess	1	None
1		Statistical processing		Calculated		Sensitivity analysis
2	1	Transfer function		Experimental		Uncertainty analysis
3		Finite-element approximation		Historic/field		Comparison
4		Comprehensive		Review		Corroboration

The construction of the pedigree matrix is based on the same concept than the membership functions described in Section 3.2.2. Figure 28 shows how the data of the pedigree matrix are stored in the database. The highlighted rows in the figure summarise the data required for generating the score "1 – Transfer function" (highlighted in Table 3).

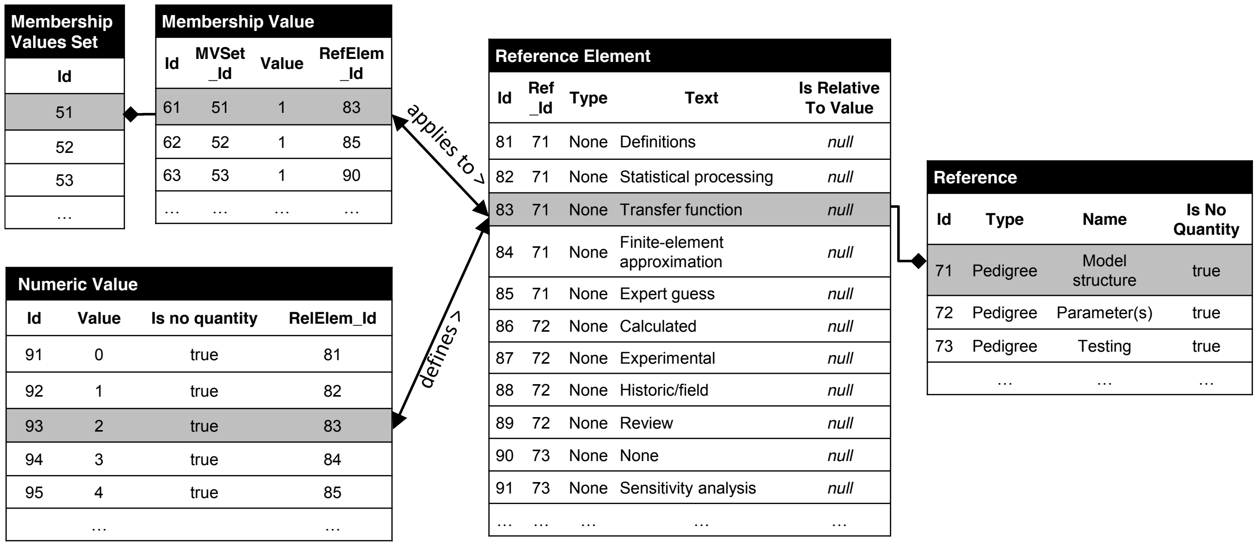


Figure 28 Organisation of the pedigree data in the tables of the database.

The table “Reference” contains the different criteria of evaluation (Model structure, Parameter(s) and Testing). The values (the scores from 0 to 4) of the discrete numeral scale are stored in the table “Numeric Value” and their corresponding descriptions per criterion (e.g. 2 = “Transfer function” for the criterion “Model structure”) are saved in the table “Reference Element”. The score of a criterion is attributed by specifying that the Value is equal to 1 in the table “Membership Value” and by linking the Value to the appropriate “Reference Element”. If the experts do not agree on a specific score, several “truth values” (membership values) can be defined, for instance 0.3 for the score 1 – “Statistical processing” and 0.7 for the score 2 – “Transfer function”.

4.5 Summary of CWIS features

Taking into account the additional features developed in this chapter, the data handled by CWIS can be grouped into three categories: (1) the SysMod constructs, which define the system elements and relationships, completed by the “time context”, which model the systems’ evolution introduced in Section 3.2.1, (2) the information types detailed in Section 3.2.2, which covers the various data types being supported by the CWIS database and (3) the scenario data, which includes the concepts of theme (used for the definition of scenarios and strategies), data provenance and uncertainty. Although all these types of data have been considered in the database schema of CWIS, only a portion is currently implemented in the software modules. Table 4 gives an overview of the implementation of these features in the CWIS prototypes and defines which modules are (or will be) used for editing and viewing the data.

Table 4 Synthesis of the database features and their implementation in CWIS

Features	Database	CWIS (Prototype 1)	CWIS (Prototype 2)	System module	Geographic module	ART module	Comments
	ok = available, (ok) = partially available and (-) = not implemented			*E = edition, V = visualisation and ( ) = to be implemented in the 2nd prototype*			
SysMod							
System elements	ok	ok	ok	E		V	
System relationships	ok	(ok) 4.5/6	(ok) 4.5/6	E		V	Property roles and mereotopology relationships are not yet implemented
Time context	ok	(ok)	(ok)	(V)	(V)	E	Feature implemented, but filtering capabilities not yet available
Information							
Text	ok	ok	ok			E	
File	ok	ok	ok			E	
Vector geometry	ok	ok	ok		V, (E)	E	Editing capabilities to be transferred in the Geographic module
Raster geometry	ok	(-)	(-)		(V), (E)		
Numeric value	ok	ok	ok			E	
Numeric interval	ok	(-)	(-)			(E)	
Distribution	ok	(-)	(-)			(E)	
Membership values set	ok	(-)	(-)			(E)	
Reference	ok	(-)	(-)			(E)	
Scenario data							
Theme	ok	(ok)	(-)	(V)	(V)	(E)	Theme definition and filter implemented in the ART of the first prototype
Provenance/ Processes	ok	(-)	(-)	(E)		(E)	Schematic representation in the System module and data edition + file upload/download in the ART
Uncertainty source	ok	(-)	(-)	(V?)		(E)	Possible visual representation in the System module to be explored

4.6 Synthesis

This chapter proposes an adaptation of the database schema given in Chapter 3 to incorporate the concepts of scenario, data provenance and uncertainty. These changes extend the functionalities of CWIS to support the creation of scenarios and strategies, and to facilitate the communication about data sources and data quality between scientists, decision-makers and stakeholders.

Once fully implemented in the CWIS software application, the new features will allow users to:

- Create and manage themes representing scenarios and strategies. This functionality involves the use of filters to only display the data that belong to the selected themes.
- Track data provenance, using some workflows of data processes made in the CWIS system module. The types of processes taken into account are the mathematical models, data transformations, queries, expert judgements, measures and quotations.
- Store almost any kind of uncertainty data and metadata. An overview of all the identified uncertainties is accessible through the generalisation of the concepts of uncertainty location, level and nature provided Walker et al. (2003).

Some new approaches are required to take advantages of these features (scenario, data provenance and uncertainty) in the framework of scenario planning. Next chapter will focus on developing a modelling approach that integrates the concepts of data provenance and uncertainty, to better assess scenarios and strategies.

5. Model integration to assess scenarios and strategies

5.1 Introduction

Scenario planning is a strategic planning method that helps decision makers to anticipate unforeseen issues and prepare a strategic plan accordingly. It involves testing some strategies against a series of scenarios that represent plausible (including unexpected) future situations. Scenarios are based on the combination of the possible states of the most important and most uncertain factors that cannot be controlled by the decision makers.

The outline of the scenarios generally consists of narrative descriptions and numerical values for the “key” uncertain factors (such as the population growth, the economic growth or the increase/decrease of rainfall due to climate change). A range of processes such as expert judgements or mathematical models can be applied to model the impacts of these key factors. These processes can produce a coherent set of variables to enrich the scenarios and allow testing strategies. As scenario planning deal with a large number of uncertainties, it is particularly important to have a consistent and reliable modelling approach.

The challenges of using mathematical models (and other data processes) for scenario planning, and decision making in general, are threefold:

1. **Integrating data processes**, given that decisions dealing with complex situations often involve data that cannot be produced by a single process, such as a mathematical model, an expert judgement, a data transformation or a measurement.
2. **Assessing data uncertainties**, in order to understand the limitations and risks involved by the decisions.
3. **Keeping track of data provenance**, to provide information about the sources of data, as well as to allow checking for errors and to facilitate the processes reproduction.

Taken individually or in pairs, these three aspects have been the subjects of multiple methodological developments, but to our knowledge none has attempted to unite them into a single method.

This chapter aims to provide an approach that combines these three aspects. The chapter first gives an overview of the techniques for mathematical models/processes integration and explains the choices adopted for coupling data processes with the Combined Water Information System (CWIS). Designed to leverage the features of the CWIS database and application, the modelling approach is based on the concepts of provenance and uncertainties described in the previous chapter. The different modules of CWIS contribute to set up the process integration, to manage the data provenance and uncertainty, to prepare the input data and to visualise the results. The adopted techniques for process integration are then described in detail and the pros and cons of each approach are discussed.

As for other features of CWIS, the developments presented in this chapter have only been partly implemented in the latest version of the software tool; some of them haven't even been implemented in the first version either. Some of the figures in this chapter are montages combining snapshots of the application and schematic representation of the missing elements.

5.2 Methodological background

In the field of natural resources management, integrated modelling frameworks (IMFs) aim to promote the reuse and integration of data and mathematical models, in order to produce some decision support systems (DSSs) or integrated assessment tools (IATs) of high quality, in short period of time and at reasonable cost (Rizzoli et al., 2008). There are basically three types of IMFs:

- **Software frameworks:** The first type of IMFs consists of particular software frameworks that provide reusable components (code libraries and programmatic classes) for integrating models and building simulation tools. Examples of such IMFs are The Invisible Modelling Environment (TIME) (Rahman et al., 2004; Rahman et al., 2003), Tornado (Claeys et al., 2006), the Object Modelling System (OMS) (David et al., 2002), JAMS (Kralisch and Krause., 2006), ModCom (Hillyer et al., 2003) or OpenMI (Gregersen et al., 2007).
- **Numerical computing environments:** Using dedicated programming languages and providing rich libraries of components, numerical computing environments such as MATLAB (The MathWorks Inc., 2009) or Mathematica (Wolfram Research Inc., 2009), can also be considered as IMFs. They may facilitate the development of DSSs using their existing mathematical libraries and visualisation tools.

- **Software modelling environments:** The third group of IMFs consists of a particular kind of software applications that support the creation and integration of models, such as the Modular Modelling System (MMS) (Leavesley et al., 1996), the Dynamic Integration Architecture System (DIAS) (Sydelko et al., 2001), the Interactive Component Modelling System (ICMS) (Reed et al., 1999; Rizzoli et al., 1998), Tarsier (Watson and Rahman, 2004), or the Spatial Modelling Environment (SME) and its associated module specifications (Maxwell, 1999; Voinov et al., 1999; Voinov et al., 2004). This list can be extended with the scientific workflow systems, such as Kepler (Ludascher et al., 2004) and Taverna (Oinn et al., 2004), which allow executing some chains of data transformations.

In itself, the Combined Water Information System (CWIS) introduced in Chapter 3 does not belong to any of these IMF categories. Even so, two procedures have been developed to enable the linkage of data processes with CWIS. The first procedure is based on the development of a software module called the "*Data exchange and modelling module*", which allows selecting and checking the input data, exporting them in a file usable by the process and importing the results back in the application (Figure 29.a). The second procedure aims to directly link the processes to the CWIS database using some IMFs such as OpenMI or Kepler (Figure 29.b). In this case, the CWIS application is only required for preparing the input data and displaying the results.

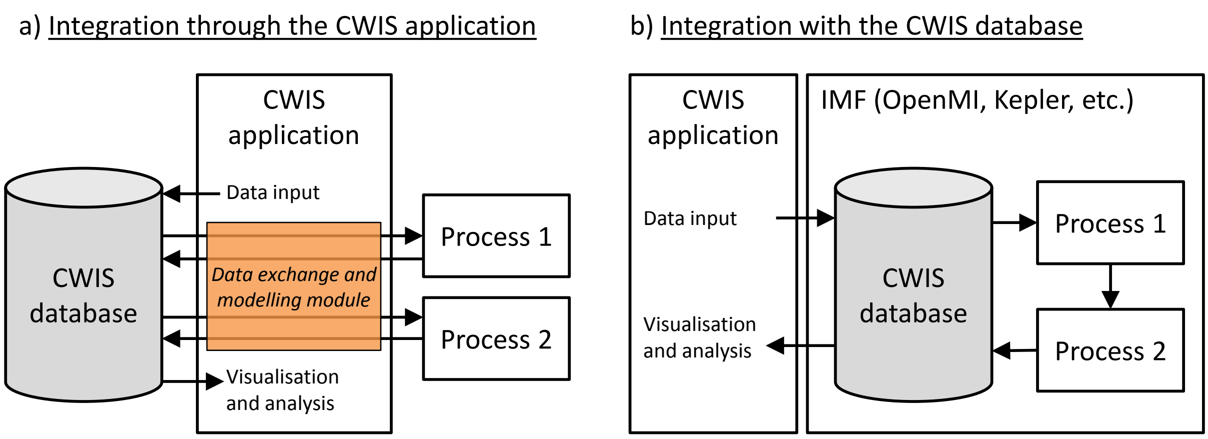


Figure 29 Techniques for process integration in CWIS

As regards other integration techniques, the main originality of CWIS is to support the integration of data, mathematical models and processes through ontologies. As expressed by Rizzoli et al. (2005), "ontologies can be used to define the extended semantics of a model

interface, in order to abstract from a specific modelling framework and to support effective and sound model component linking." In other words, the ontology can be used to formalise the inputs, outputs and parameters of the models and ensure the semantic compatibility of the connected inputs/outputs.

5.3 A process-based modelling approach

The uncertainty management framework proposed by Refsgaard et al. (2007), aims to support the management of uncertainties in the field of model-based decision support. This framework provides a methodology for uncertainty assessment, which is applied at each step of the modelling process. The main limitation regarding a general approach for uncertainty management is that the method focuses only on mathematical models. Processes such as data transformations, measurements or expert judgements are not included.

The new modelling approach generalises the assessment method provided by Refsgaard et al. (2007) to integrate any kind of data process. The resulting extended modelling approach is a stepwise methodology for integrating data processes (including mathematical models), while simultaneously assessing the data uncertainty and recording the data provenance.

5.3.1 Methodological steps

The extended modelling approach is based on five successive steps, each of them followed by a short uncertainty assessment, using the concept of uncertainty matrix that provides an overview of the uncertainties and their characteristics (see Section 5.3.2). The flowchart of Figure 30 details the different actions to be taken by the "modeller" to apply the method.

Step 1 Initialisation: According to Refsgaard et al. (2007), the first step of the process is the creation of a "Model Study Plan", which aims to define the objectives as well as the technical and human resources available for the modelling process. It also specifies that "*a very important (but often overlooked) task is then to analyse and determine what are the various requirements of the modelling study in terms of the expected accuracy of modelling results*".

This first step dedicated to the planning of the modelling process aims to (1) frame the modelling problem and define the objectives, (2) review the data and models/processes already available in the CWIS database, (3) define the data and models/processes to be included and (4) specify the accuracy requirements of the outputs.

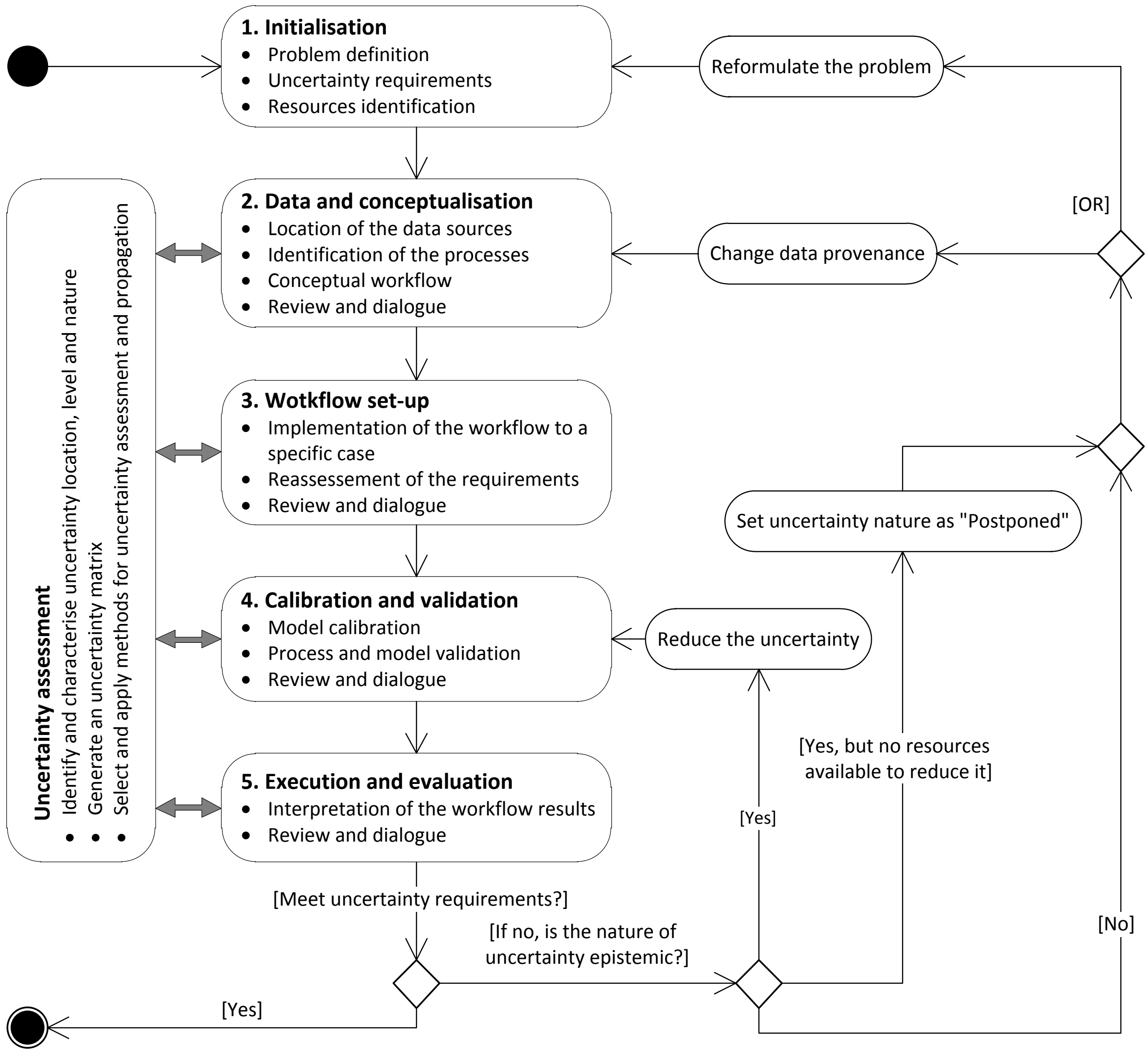


Figure 30 Methodological steps of the extended modelling approach

To ensure an effective integration, an outline workflow can be defined to check the compatibility between data and models (and processes). CWIS provides two procedures for model integration (Section 5.4: *Model integration through the CWIS application* and Section 5.5: *Model integration with other integrated modelling frameworks*); the choice of using one or the other can be set at this stage. If the integration requires the models to exchange data at runtime, an IMF such as OpenMI (Gregersen et al., 2007) is necessary. Otherwise, the models can directly be "controlled" from the CWIS application, provided they are made compatible by adding a routine that converts the data exported from the application in the input format of the model and conversely for the model outputs.

Step 2 Data and conceptualisation: This step involves the identification and acquisition of the data, the definition of the processes and their organisation as a workflow. In order to meet the requirements specified in the initial stage, data and processes need to be chosen depending on their compatibility and complementarity at the conceptual (or semantic), data precision and temporal levels.

Using the system module of CWIS, the approach introduced in Section 4.3 to track data provenance allows the modeller to specify the inputs, outputs and linkages of models (or any data processes) and create some workflows. The inputs and outputs of each model are characterised by a unit, a time step and a SysMod class. This class is either an "information class" that specifies the data type (text, numeric value, file…) and a possible value type (length, volume, flow…), or a "holon class" that defines an expected set of properties of the inputs/outputs (Figure 31).

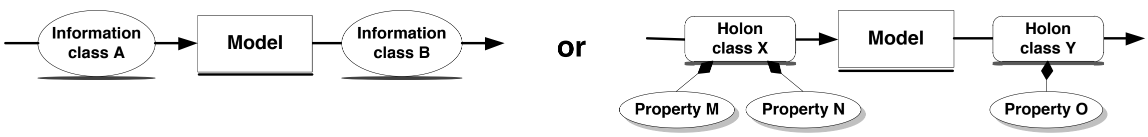


Figure 31 Characterisation of the model inputs and outputs by the SysMod classes

An example of rainfall-runoff model is given in Figure 32, which illustrates how classes can be used to define the inputs and outputs of the model.

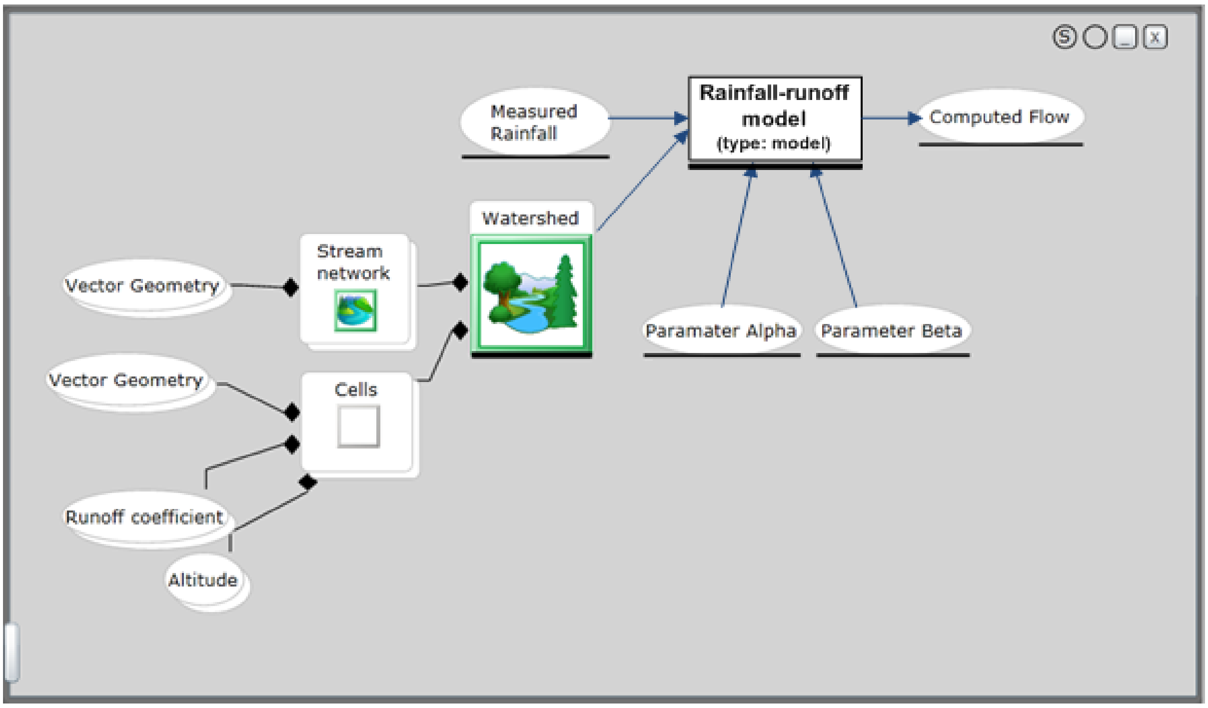


Figure 32 Example of rainfall-runoff model in CWIS system view

Workflows are created by combining models, the output of one model being reused by the input of another. In order to check the consistency of the connections between models, the linked output and input need to have the same time step and to be defined by the same class (or possibly by an equivalent classes, e.g. a subclass or a class that share the same data type, value type and properties) (Figure 33). When output and input have the same value type but different units, the conversion of the values can be made on the fly.

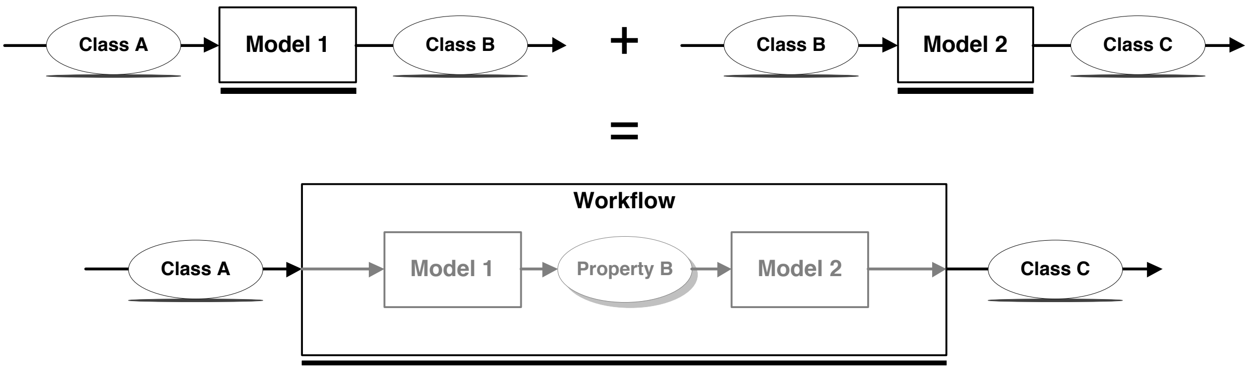


Figure 33 Combination of models to create a workflow

By definition, the (SysMod) classes are the elements that form the backbone of the ontology. They can be derived into properties and instances that represent the elements and values of the system to be modelled. It means the classes that define the inputs and outputs of the models can be reused to help preparing the input data and checking the data validity when loading back the results.

Based on the elements of the workflow, a first screening assessment of the uncertainties can be performed using the uncertainty matrix and some preliminary investigations can be carried out to evaluate the achievability of the accuracy requirements. The main principle of the approach is to allocate the uncertainty metadata to their corresponding data or data process, like some tags attached to the different elements of the workflows.

Step 3 Workflow set-up: This step involves setting up the various processes described in the conceptual workflow. Besides mathematical models, it may involve a range of other processes such as measurements, expert judgements, data transformations (possibly using some scientific workflow systems) or the coupling of mathematical models.

In the CWIS system view, the workflows can be combined with the elements of the SysMod language. The information properties of SysMod (ellipses) can be used as repository for storing the input, intermediate and output data of the workflow.

Gradually, as new information is collected, the uncertainty matrix needs to be updated and used to possibly suggest some researches for enhancing the reliability of the processes and data.

Step 4 Calibration and validation: To be effective, models and measuring instruments need to be calibrated and validated against some data of reference. In a more general way, any process can be validated through internal reviews (by the modeller) and external reviews (by peers or experts).

CWIS does not provide any specific tool to help calibrating models, but the data used for the calibration can be stored in the database and supplied through the application.

During this step, most of the uncertainties related to the structure, parameters and implementation of the processes can be assessed. The implementation of the workflow requires the assessment of the uncertainty that were not previously defined in the workflow definition (Figure 34).

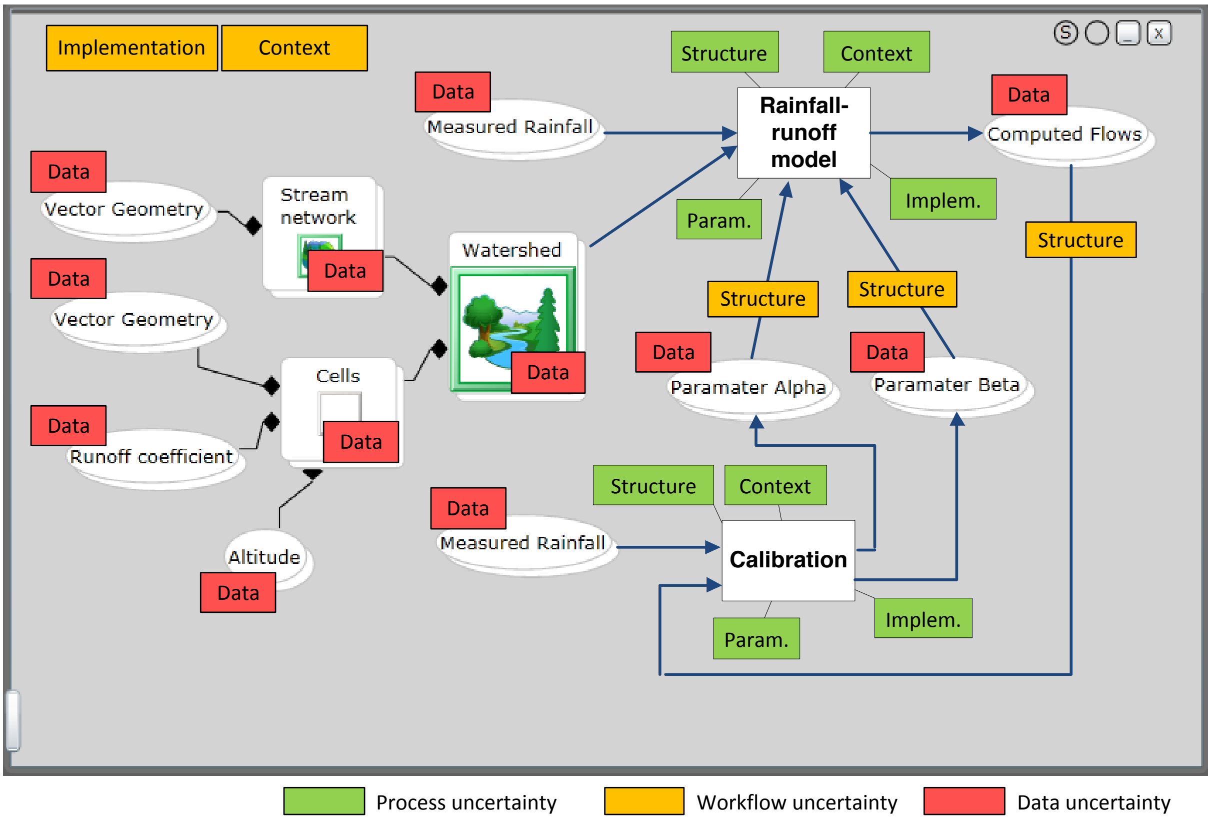


Figure 34 Locations of uncertainties for an implemented workflow

For each location, the assessment of the uncertainty involves the definition of its level (statistical, qualitative, scenario or recognised ignorance) and nature (epistemic or stochastic). The results are then collected and organised in the uncertainty matrix to provide an overview of all the sources of uncertainty. For instance, taking the example of Figure 34, the level of structural uncertainty of the rainfall-runoff model may be considered as “recognised ignorance” and its nature as “epistemic”, if no information is available to quantify or qualify it. The context uncertainty may be considered as not significant if the data inputs correspond with the model requirements and the objectives of the simulation match the domain of applicability of the model. The parameters’ uncertainty could have a “statistical” level and a “stochastic” nature, provided the parameters are properly calibrated with some representative data sets. The input data may have various levels varying between statistical uncertainty and recognised ignorance depending on whether uncertainties metadata are provided by the source.

Step 5 Execution and evaluation: The models, as well as any processes of a workflow, can interact asynchronously (the processes are loosely-coupled) or synchronously (the processes are tightly-coupled). Both types of interactions may exist in the same workflow. In the example of Figure 34, the calibration is made manually and therefore processes are asynchronous. Conversely, synchronous integration requires the processes to exchange data at runtime; it requires an IMF like Kepler (Ludascher et al., 2004) for performing data transformations or OpenMI (Gregersen et al., 2007) for the integration of complex models.

The execution of the workflow may generate a series of metadata, such as: a date and time of the process execution; metadata describing the inputs of the process (their start times and end and their belonging to a specific simulation or scenario; the name of the author of the simulation; some observations; a log files describing the events occurred during the execution. In addition to the data and process uncertainties, the features mentioned above can be used to evaluate the output data from the workflow. The uncertainty of the results can be quantified (or qualified) using the propagation techniques mentioned in Appendix B.2 and compare with the objectives of data accuracy defined at the beginning of the procedure. If the expected data accuracy is not achieved, one may try to reduce the epistemic uncertainty through further research or measurements, the input data and the structure of the workflow may be reconsidered or the purpose and accuracy requirements of the whole activity can be redefined.

The models are executed either through the “data exchange and modelling module” of the CWIS application or using the CWIS database directly with another IMF. (Both techniques are describes in the next sections 5.4 and 5.5.)

Once loaded in the database, the results can be displayed through the modules of CWIS: the geographic viewer allows representing the results through thematic maps; the system viewer can display them as flow diagrams or other systemic representation; values and series of values are browsed with the ART; and the graphs of the results are obtained through the “chart module”.

Through propagation methods (see Appendix B.2), the uncertainty related to the results can be quantified (or qualified). Such methods, like the Monte Carlo Analysis or the Sensitivity Analysis, may involve reusing the workflow to carry out multiple simulations.

5.3.2 Uncertainty assessment

The uncertainty assessment is carried out after each step of the modelling process. The methodology is mainly based on the concept of uncertainty matrix (Walker et al., 2003),

which can be defined as "*a heuristic tool to classify and report the various dimensions of uncertainty, thereby providing a conceptual framework for better communication among analysts as well as between them and policymakers and stakeholders*". The rows of the uncertainty matrix characterise the various uncertainty sources, while the processes and the uncertainty locations, levels and natures are detailed in the columns (Table 5).

Table 5 Columns of the uncertainty matrix, adapted from Walker et al. (2003)

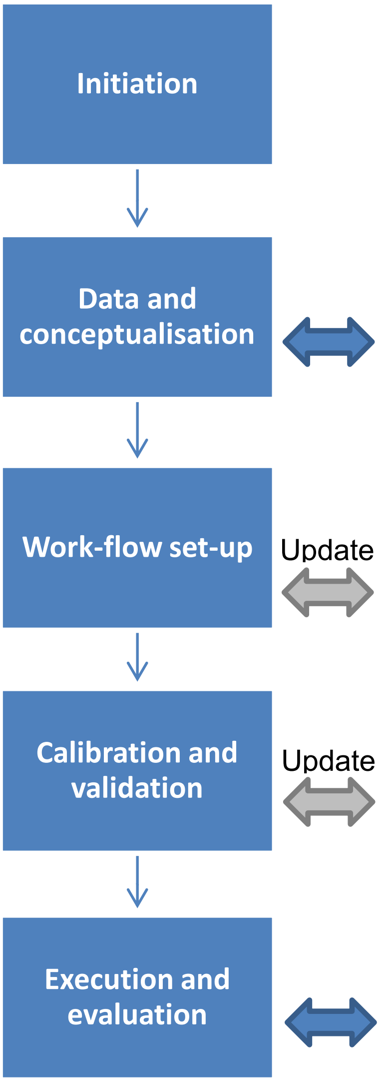
Source of uncertainty	**Process**	**Location**	**Level**	**Nature**	**Significance**
(description /name of the source)	(name of the process)	Context, Structure, Implementation, Parameter or Data	Statistical, Qualitative, Scenario or Recognised ignorance	Stochastic, Epistemic, Undefined or "Postponed"	Rank between 1 and 5 (1 = not significant and 5 = predominant)

An extra column "Significance" has been added to the original matrix, in order to rank the significance of the source of uncertainty on a scale of 1 (weak) to 5 (strong). The "significance score" characterises the importance of the uncertainty as regards its impacts on the final result (the output of the workflow).

An example of database schema that allows storing these uncertainty metadata (location, level, nature and significance) is given in Section 4.4. From this database structure, it is then easy to create an uncertainty matrix, simply by performing database queries on the uncertainty locations attached to the workflow and its sub-processes.

Figure 35 illustrates how the uncertainty matrix can be applied to a workflow (WF) that contains two processes (a rainfall-runoff model and a calibration process of type "expert judgement"). As and when the project evolves, the content of the uncertainty matrix is updated to represent the state of knowledge about the uncertainties of the workflow elements. The assessment starts from the identification of the uncertainty sources with no information about their level and nature and is gradually completed through multiple updates.

At each step of the modelling process, the matrix can be used as a scanning tool to identify areas where specific uncertainty assessments are required. For instance, further assessments should focus on uncertainty sources of high significance and epistemic nature. The choice of the appropriate uncertainty method may depend on several criteria, including those mentioned in the uncertainty matrix. A guide to select uncertainty assessment methods is given in Appendix B.2.



Source	**Process**	**Location**	**Level**	**Nature**	**Significance**
Input	-	Data	Ignorance	Undefined	
...	WF	Context	Ignorance	Undefined	
...	WF	Structure	Ignorance	Undefined	
...	WF	Implem.	Ignorance	Undefined	
...	WF	Parameter	Ignorance	Undefined	
...	Model	Context	Ignorance	Undefined	
...	Model	Structure	Ignorance	Undefined	
...	Model	Implem.	Ignorance	Undefined	
...	Model	Parameter	Ignorance	Undefined	
...	Calibration	Context	Ignorance	Undefined	
...	Calibration	Structure	Ignorance	Undefined	
...	Calibration	Implem.	Ignorance	Undefined	
...	Calibration	Parameter	Ignorance	Undefined	
Output	-	*Data*	*Ignorance*	*Undefined*	-

Source	***Process***	**Location**	**Level**	**Nature**	**Significance**
Input	-	Data	Statistical	Stochastic	3
...	WF	Context	Qualitative	Epistemic	1
...	WF	Structure	Ignorance	Undefined	1
...	WF	Implem.	Qualitative	Postponed	1
...	WF	Parameter	Ignorance	Undefined	1
...	Model	Context	Qualitative	Stochastic	2
...	Model	Structure	Qualitative	Epistemic	1
...	Model	Implem.	Ignorance	Epistemic	1
...	Model	Parameter	Ignorance	Undefined	1
...	Calibration	Context	Qualitative	Stochastic	3
...	Calibration	Structure	Qualitative	Epistemic	2
...	Calibration	Implem.	Qualitative	Epistemic	4
~~...~~	~~Calibration~~	~~Parameter~~	~~Statistical~~	~~Stochastic~~	~~1~~
Output	-	*Data*	*Qualitative*	*Epistemic*	-

Figure 35 Update of the uncertainty matrix for each step of the modelling process

5.4 Model integration through the CWIS application

The first approach to model integration with CWIS is based on the use of a software module dedicated to data import and export. This module, called the "Data exchange and modelling module", is also designed to perform simulations by linking models to the CWIS application. Using this module, only one model can be invocated at a time, the integration of multiple models at run time being currently not possible. The steps below describe briefly the general principles for exporting/importing data and running models (Figure 36).

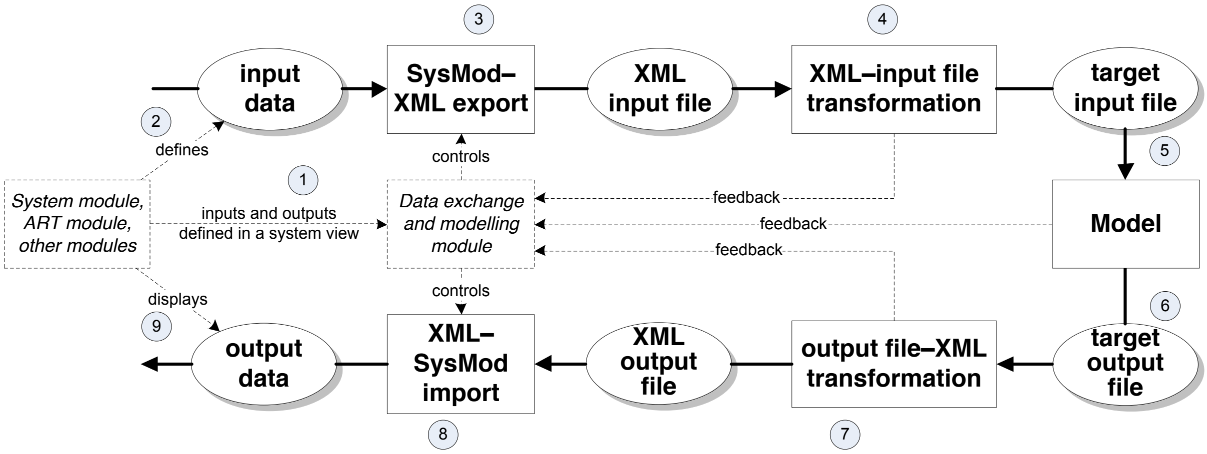


Figure 36 Workflow representing the processes and data involved in model integration with the CWIS application

1. The inputs and outputs of the models are schematically defined in a “Model’s system view”, which is used by the “Data exchange and modelling module” to create a list of the expected inputs/outputs. This view corresponds to the conceptual workflow created during the Step 1 (see Section 5.3.1) of the modelling process.
2. The input data can be edited in the system, geographic and ART modules of CWIS. Then they can be dragged from these different modules and dropped on the list of expected inputs generated by the “Data exchange and modelling module”.
3. Upon completion of the selection process, the data can then be exported in a file based on the SysMod XML exchange format.
4. If required, the xml file can be transformed (through an intermediate data translation routine) into specific formats, such as CSV, Excel or proprietary model input files. Before this stage, the file is first checked and possible problems are notified back to the Data Exchange and Modelling view.
5. The target (translated) file is used as input for the model.
6. The model is executed, resulting in the creation of an output file.
7. The output file is then transformed back into XML format of SysMod-format.
8. The results contained in the XML file can be loaded and checked by the “Data exchange and modelling module”. In order to distinguish data belonging to different runs of the workflow, a new “Theme” of type “Process run” is generated for each run. (Details about the use of Themes to group data belonging to a same “thematic cluster” are given in Chapter 4).
9. Finally, the results may be displayed using the different modules of CWIS.

In addition to the model results, the intermediary data (XML input/output files and target input/output files) and the various feedbacks generated by the processes (models or data transformations) can be automatically recovered by CWIS and saved as "files" or "log files" in the database.

5.4.1 Model's system views

The system module of CWIS allows modellers to define some model's system views, which structure the input/output requirements of models using the SysMod languages. Figure 37 illustrates one of these views for a specific model, City Water Balance (CWB), a scoping tool to assess alternative strategies for urban water management (Mackay and Last, 2010).

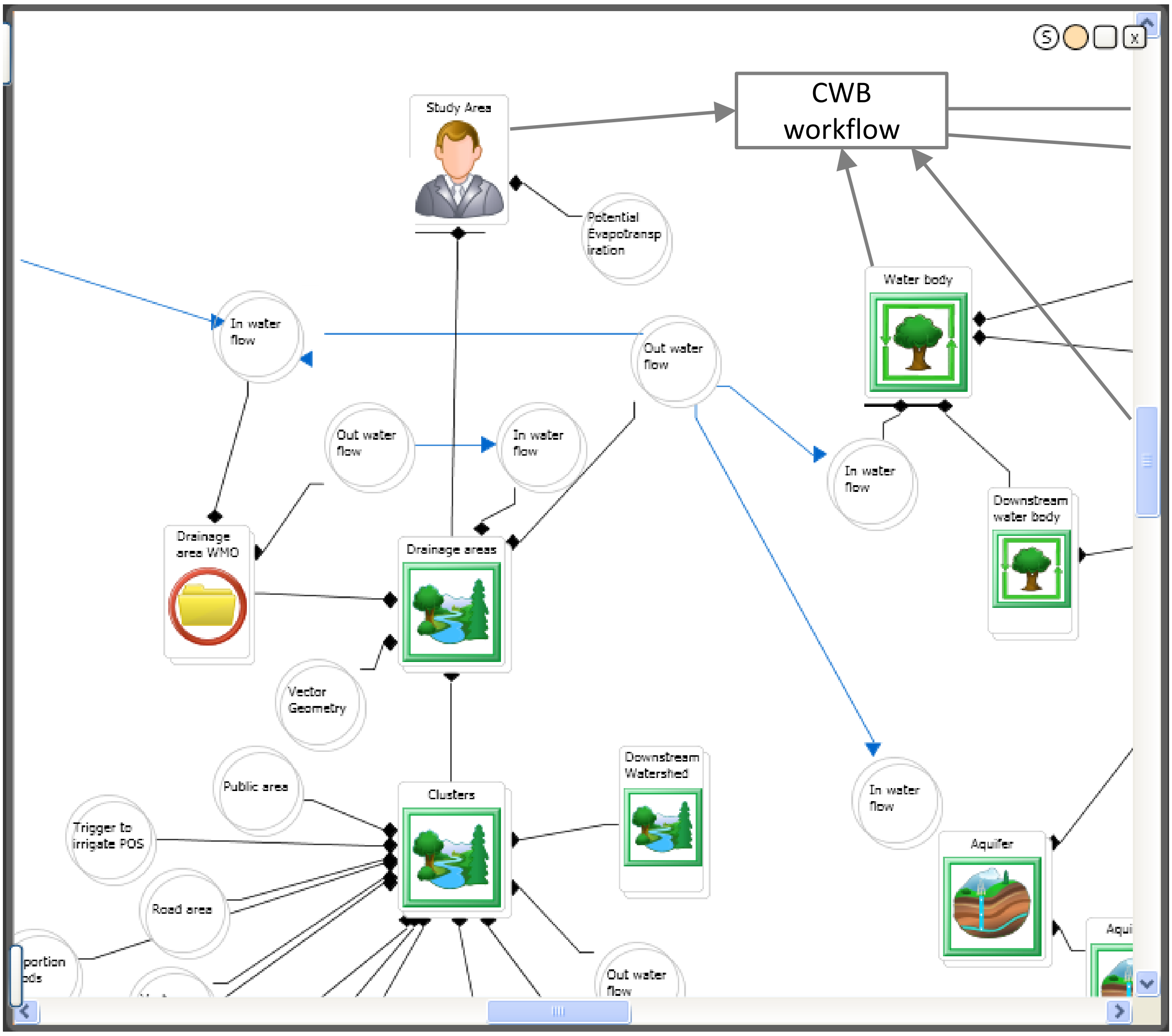


Figure 37 An example of model's system view, the case of the model City Water Balance

The model's system view is composed of several elements: (1) A workflow that combines the model with data transformation processes to export and import SysMod data. (2) The inputs and outputs of the workflow symbolised by the arrows. (3) The input/output classes (underlined elements) and their linked properties, which semantically define the model requirements and help validate the input/output. A set of input/output is valid if the data properties, data types (numeric value, text…) and value types (flow, volume…) match the conditions given by the input/output classes. Besides, the input/output classes are members of the ontology and can be utilised to create some new (instances of) inputs.

5.4.2 The Data exchange and modelling module

The "Data exchange and modelling module" works as a shopping basket, providing a list of input/output requirements based on the content of the "model's system views" The export view of the module allows preparing the set of input data by manually drag-and-dropping elements into the view (Figure 38). If the inputs are already organised according to the required data structure, a selection algorithm can be applied to automatically populate the export list. The selected inputs are then exported in a XML file and possibly (if the model is not compatible with the SysMod format) transformed by a programmatic routine in another format which is compatible with the model.

Once the simulation completed, the results (translated in the SysMod XML format) are checked and loaded through the import view. During this process, the results can be visualised in the import view and then rejected or approved. In the case of approval, the pre-existing elements/values are updated and a new element/value is created for each imported data item not yet referenced in CWIS.

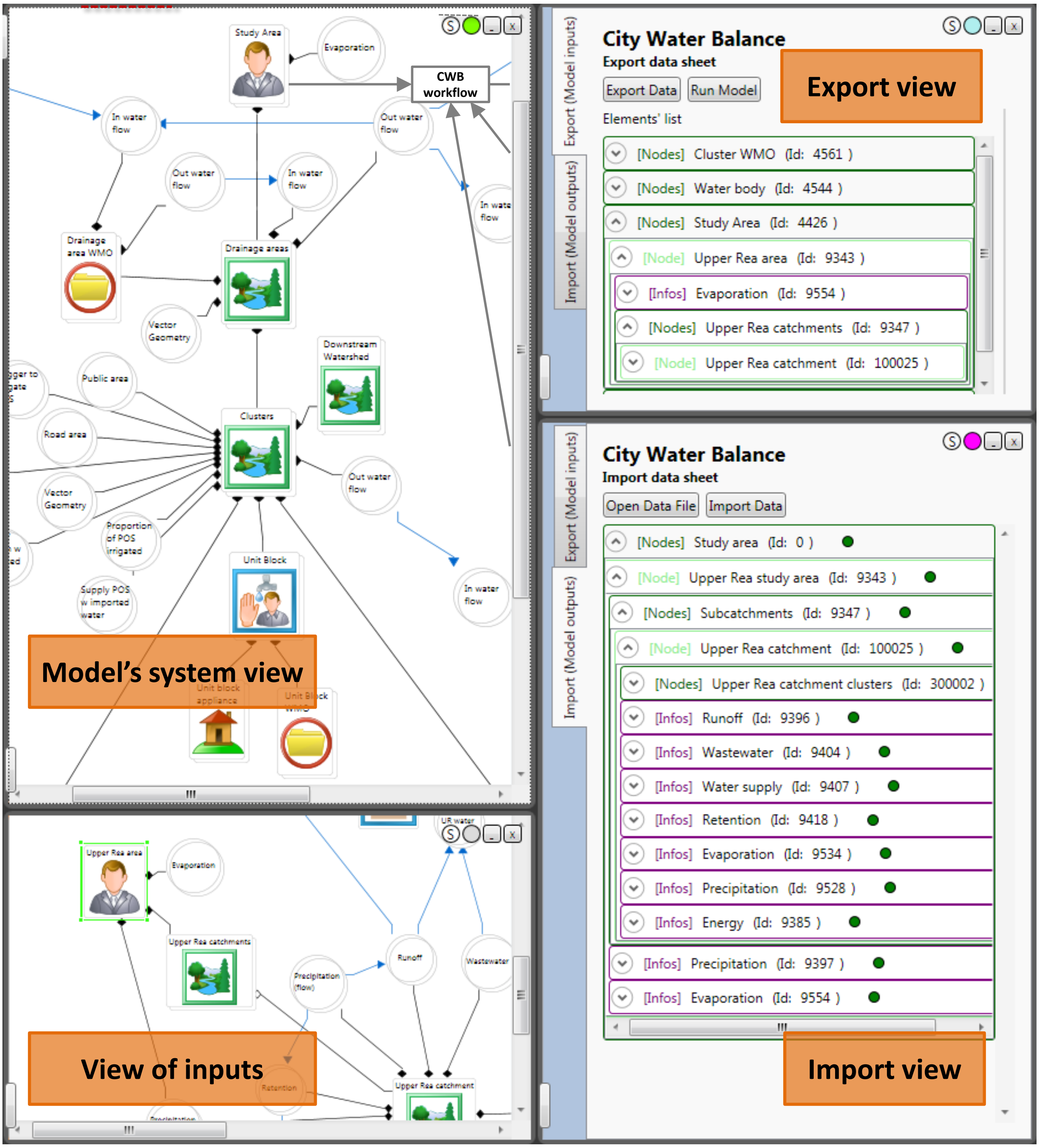


Figure 38 Export and import views of the Data exchange and modelling module

5.4.3 The SysMod XML exchange format

The Extensible Markup Language (XML) is a simple and flexible text format, which is used for documents containing structured information. Like HTML, XML uses tags to define the purpose of each information item contained in the file, with the main distinction that some "customised" tags can be created to fit specific requirements.

The XML files exported or imported with CWIS are based on a custom set of tags that reproduces the data structure set by the SysMod language (Figure 39). In the CWIS application, these files are generated by XML serialisation (the process of transforming programmatic objects into an XML file) of some "replicas" of the (SysMod) programmatic objects. Each element or relationship modelled with the SysMod language is enclosed in a programmatic object and for each data item to be exported, a replica (i.e. a "simplified" object that contains only the information related to the SysMod language) of the programmatic object is created. The collection of object replicas is then serialised in an XML file.

```
<?xml version="1.0" encoding="utf-8"?>
<ArrayOfNodeGroupReplica xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<NodeGroupReplica>
  <Id>4426</Id>
  <Properties />
  <Name>Study Area</Name>
  <Instances>
    <NodeReplica>
      <Id>100003</Id>
      <Name>Birmingham StArea</Name>
      <Properties>
        <PropertyReplica xsi:type="NodeGroupReplica">
          <Id>4428</Id>
          <Properties />
          <Name>Drainage areas</Name>
          <Instances>

… list of NodeReplica (Drainage areas) …

          </Instances>
        </PropertyReplica>
        <PropertyReplica xsi:type="InformationGroupReplica">
          <Id>4478</Id>
          <Properties />
          <Name>Potential Evapotranspiration</Name>
          <Instances>

… list of InformationReplica (Values of potential evapotranspiration) …

          </Instances>
        </PropertyReplica>
      </Properties>
    </NodeReplica>
  </Instances>
 </NodeGroupReplica>

…(etc.)…

</ArrayOfNodeGroupReplica>
```

Figure 39 Example of (SysMod) XML file that contains some model inputs

When importing the results back into CWIS, the deserialization of the XML file generates new object replicas, which can be used to update the original programmatic objects or create some new ones.

5.5 Model integration with other integrated modelling frameworks

Without making use of the “Data exchange and modelling module”, the integration can be achieved by linking the models directly to the CWIS database using an integrated modelling framework (IMF). If the model is a software application, the integration can be performed with a modelling software framework such as OpenMI (Gregersen et al., 2007). Otherwise, in the case of data transformations, data analyses or models expressed in mathematical languages such as Matlab (http://www.mathworks.com/products/matlab/) or R (http://www.r-project.org/), the integration can be achieved by a scientific workflow system (SWS) such as Kepler (Ludascher et al., 2004).

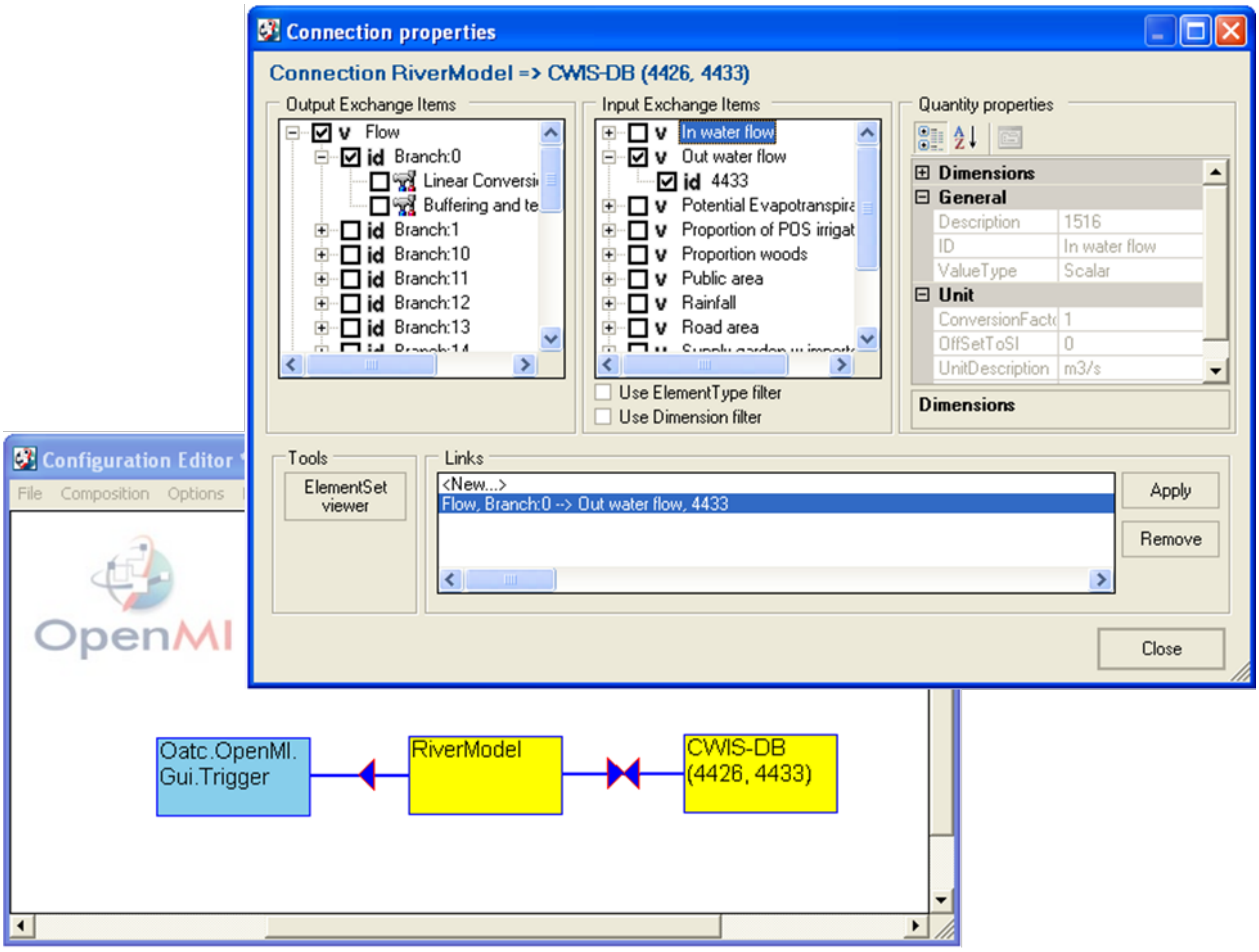


Figure 40 Integrating the CWIS database with other models using OpenMI

It should be noted that, both types of IMFs (software frameworks and SWSs) need the development of an additional programmatic component, to allow the querying the database. This task has been completed for OpenMI, with the development of a component that “wraps” the database and makes it compliant to the OpenMI standard and thus directly usable in the OpenMI Configuration Editor (Figure 40).

Given that models cannot be integrated at run-time using the CWIS application, the adoption of such IMFs is particularly interesting. However, the drawback of this approach is that the process of data validation carried out by the “Data exchange and modelling module” is not applicable here. Therefore, there is (currently) no way to check the input data and to approve or reject the database changes caused by the model outputs.

5.6 Synthesis

In the field of model integration, there are basically three categories of integrated modelling frameworks (IMFs). The first IMF category consists of software frameworks that provide reusable components (code libraries and programmatic classes) to integrate models and build simulation tools. The second category represents the numerical computing environments that possess dedicated programming languages and provide rich libraries of components for model integration. The third category characterises the software modelling environments, which are some particular software applications to create and integrate models. This last category includes the scientific workflow systems, used for the execution of chains of data transformations.

CWIS is not in itself an IMF, since it does not directly support the execution of mathematical models and other data processes. However, CWIS can store model inputs and outputs. On this principle, two procedures for model integration have been developed: the first is based on a dedicated module of the CWIS application and the second proposes the use of an IMF, such as OpenMi (Gregersen et al., 2007), to directly access the data of the CWIS database. The advantage of using an IMF is the ability to integrate multiple processes at run-time. In comparison, the CWIS application supports only one model invocation at a time, but offers interesting features to support the validation of the model inputs and outputs and to automatically record the provenance of the results.

A general modelling approach has been developed to set up the integration of computable data processes (e.g. mathematical models, data transformations and queries) and non-computable data processes (e.g. experts judgments, measurements and quotations). The novelty of this modelling approach is the combination of the concepts of data provenance

and data uncertainty to provide more reliable modelling results. Data provenance provides valuable information to keep track of the data history, and thus allow for error detection and the reproduction of the modelling process. On the other hand, the concept of data uncertainty is essential for sound decision-making, in order to handle the risks associated to strategic decisions.

An added value of CWIS for model integration is the ability to handle ontologies to ensure the semantic interoperability of models. Ontologies are composed of classes that allow storing some descriptive variable, such as a data type (numeric value, text…), a value type (flow, volume…), some units and cardinalities, as well as some series of properties. These classes can semantically characterise the model inputs and outputs, and therefore to ensure the compatibility of the models.

6. Application to scenario planning

6.1 Introduction

In the previous chapters, the SysMod language has been introduced, which addresses through ontologies the problem of semantic in the field of systems modelling. Later, the concepts of SysMod have been extended to manage scenario data, and handle the concept of data provenance and uncertainty. In view of using these concepts for scenario planning, SysMod and its extensions have been implemented into an information system called CWIS, and methods to couple mathematical models with CWIS have been developed.

In order to illustrate how CWIS can support the process of scenario planning for integrated urban water management (IUWM), the present chapter first describes the configuration options of CWIS to tune the system for a specific purpose and make it usable by non-specialists. The chapter then gives some details about the ontology that has been developed to address the requirements of IUWM and shows how to use CWIS for scenario planning, based on a case study of the city of Alexandria (Egypt).

6.2 Configuration of CWIS

CWIS is a generic tool that can be applied to model and structure the information of any kind of systems, be they environmental, economic, technical, social or political. Its ability to handle interdisciplinary information comes from the adoption of the SysMod language, which enables the creation of ontologies and the modelling of systems in a unique framework. CWIS requires a specific database, whose core structure reflects the constructs of the SysMod language. Thus, it does not need to (and cannot) integrate heterogeneous databases and data structures.

Three types of actions can be carried out to configure the application, in order to suit the needs of the users. These actions address three different layers of CWIS (the Knowledge base, the Modules and the Interface elements) shown in Figure 41.

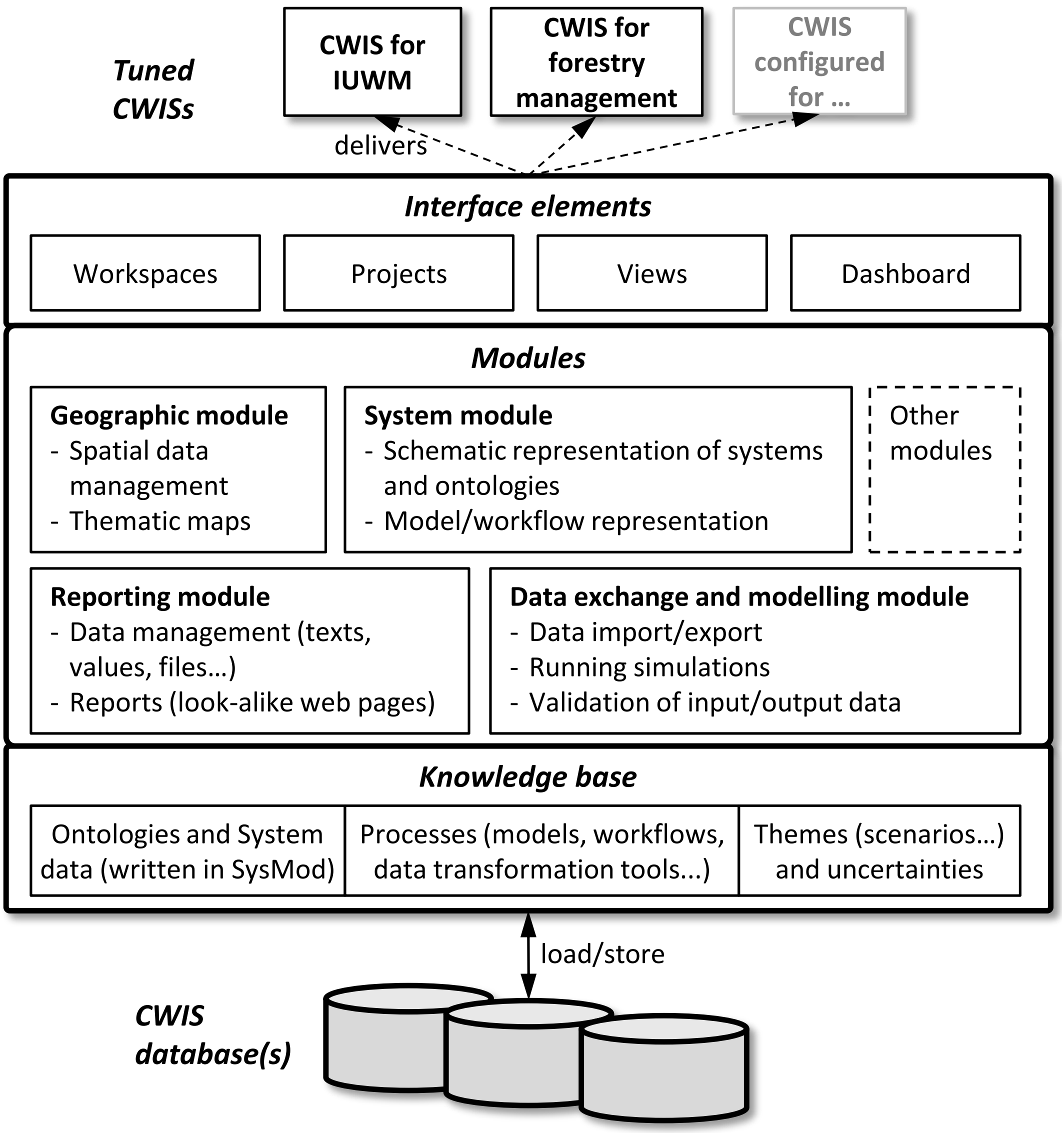


Figure 41 Overview of the CWIS features, adapted from Rizzoli et al. (2008)

Defining new ontologies

The knowledge base is the fundamental layer that contains all the information needed to use CWIS. The structure and semantic of the data is ensured by the ontology, a collection of documented classes representing the types of objects and information available for system modelling. The ontology is the backbone of CWIS, and therefore any new set-up of the application requires the prior development of an appropriate ontology. At the beginning, the ontology can remain quite simple, it just needs to cover the basic requirements of the application (i.e. the initial objects and types of information one wants to use, e.g. images, files, numeric values…). Then, gradually, the ontology will be expanded to handle new objects and data.

Creating a new module

Each module of CWIS is an independent piece of software that can be updated without altering the functionalities of the application and other modules. As a result, some additional modules may be created to provide new functionalities to the application (Schenk, 2010).

Configuring the interface

CWIS offers several configuration options to personalise the application: (1) The **workspaces**, which provide a way to save the state of the application interface, by defining the layout of the windows' regions (which can host the views) and assigning a view to each region. (2) The **projects**, which bring together some sets of views and workspaces. (3) The **views**, which allow organising and presenting the information. (4) The **dashboard**, which is a "welcome" panel configured by the CWIS administrator to facilitate the access to some workspaces or views without being forced to browse the different projects (Figure 42).

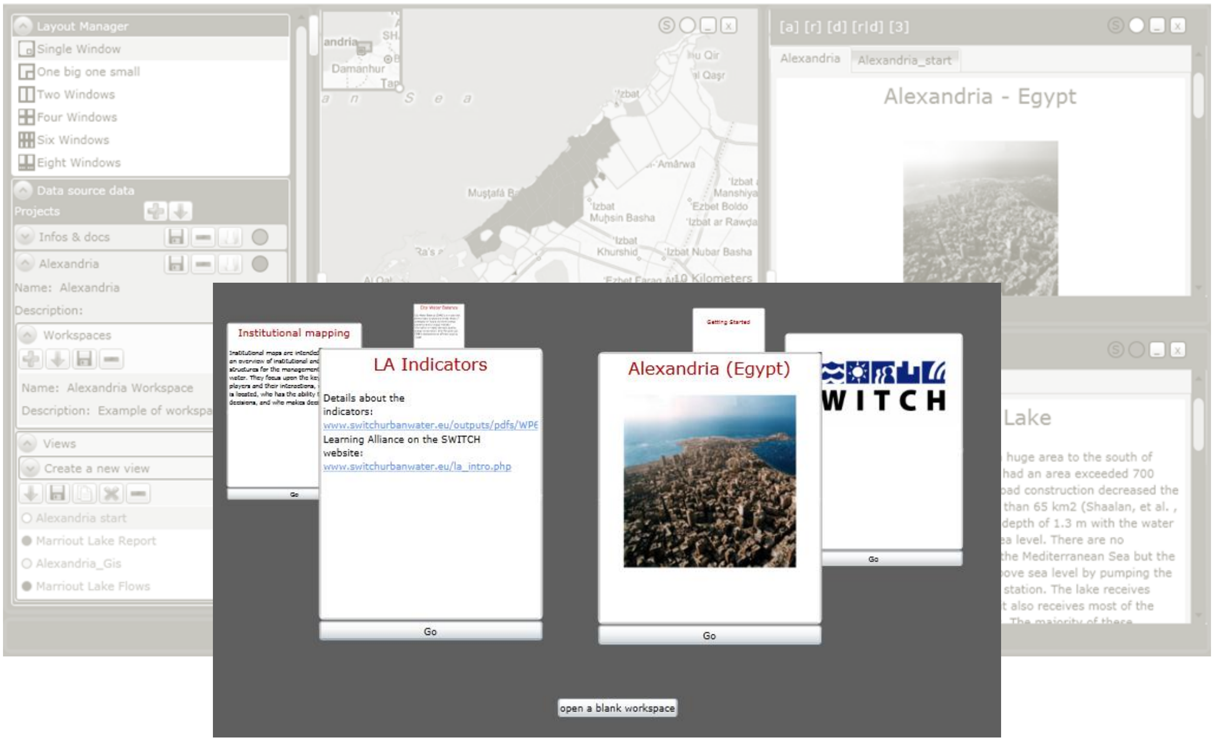


Figure 42 CWIS dashboard, an access to project workspaces and views

6.3 Ontology of the water system and its related components

As mentioned above, the development of an ontology is a prior requirement for the use of CWIS. To support the implementation of CWIS in the cities of the SWITCH project, a “set up” database has been created that contains the basic data needed to start working with CWIS. This database contains three libraries of classes that hierarchically organise the elements of the ontology, and some other basic data such as lists of data types (text, file, numeric values…), value types (volume, flow, length, area…), units, types of processes (quotation, model, expert judgement…) and several others.

The first library of classes (left part of Figure 43) contains the objects that can be used to model systems. The objects are drawn from the article of Schenk et al. (2009), which enumerates the structural elements (ecosystems, fauna, flora, water bodies, urban elements such as water supply and sanitation infrastructures, etc.) and non-structural elements (the water policies, the stakeholders, etc.) linked to the urban water system.

The second library of classes (upper right of Figure 43) covers the types of information available to document the systems being modelled. A particularly wide range of type of numeric values has been defined to address as many value types and units as possible. A basic list of the main information classes is available in Appendix C.

The third library (bottom right of Figure 43) lists the different classes of systems relationships (interactions), such as flows and causal relationships.

6.4 Case Study

The case study illustrates, with a series of screen captures, how CWIS can support the process of scenario planning for IUWM. The examples are made on the basis of documentation produced by the learning alliance of Alexandria, one of the demonstration cities involved in the SWITCH project.

A learning alliance is a group of representative stakeholders from different sectors that work together to address common issues, by sharing their experience and understanding, and implementing innovative approaches (Butterworth et al., 2009). “This group of interconnected players typically includes public sector (e.g. line ministries, utilities, regulators, educators, research institutes), private sector (e.g. industry, financial services), and civil society players (e.g. NGOs, media, professional bodies and unions, advocacy organizations).” (Morris, 2006)

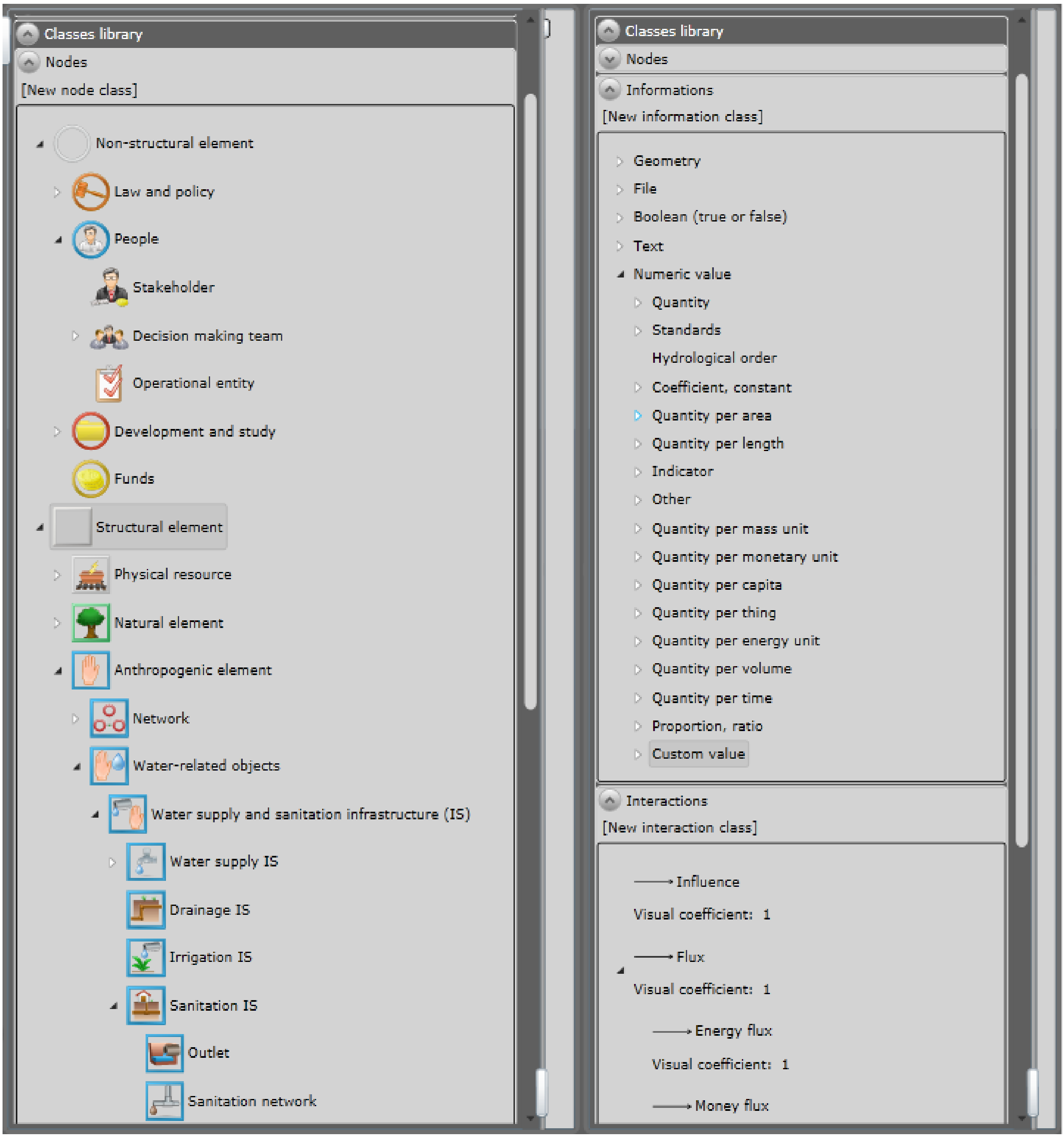


Figure 43 Libraries of classes (objects, types of information and interactions)

In the context of the SWITCH project, the learning alliances are involved in the process of visioning and scenario planning. The overall process aims to identify by consensus some shared long-term objectives, examine some strategic activities to achieve these objectives, elaborate a plan for the implementation of an agreed strategy that judiciously combines a selection of strategic activities, and the monitoring of the overall process. The project management cycle that describes this approach is given in Figure 44.

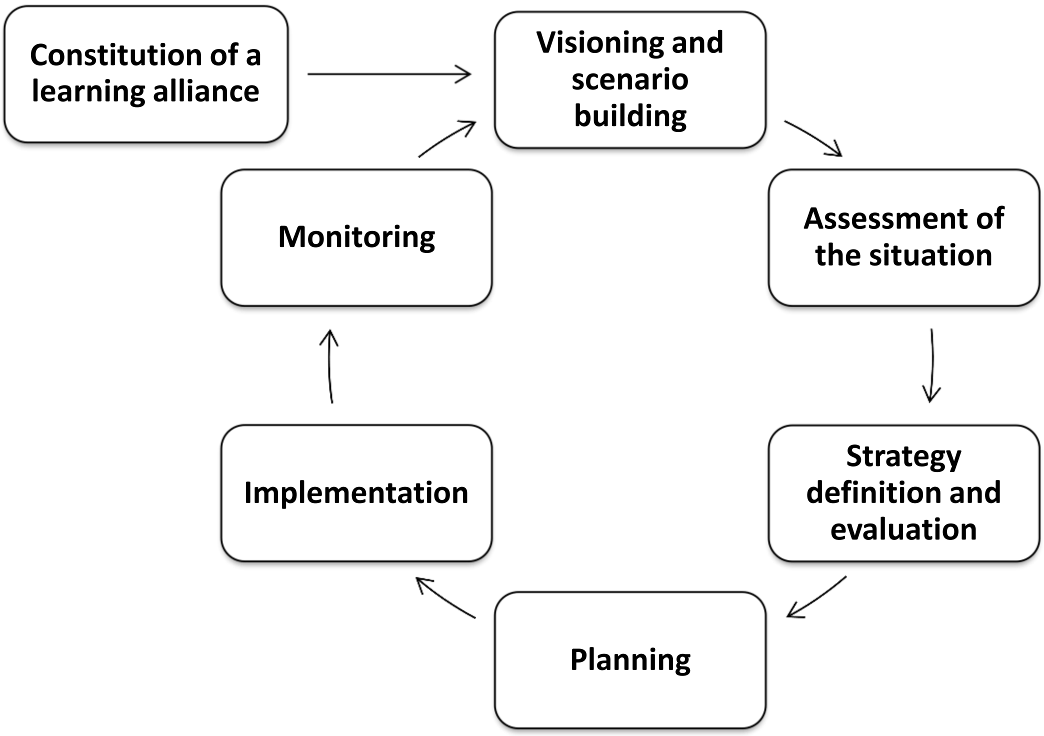


Figure 44 Project management cycle adopted by the SWITCH learning alliances, adapted from Batchelor and Butterworth (2008c)

Some short definitions of the concepts of vision, scenario and strategy are shown in Figure 45, using CWIS as a support to structure and access the relevant information.

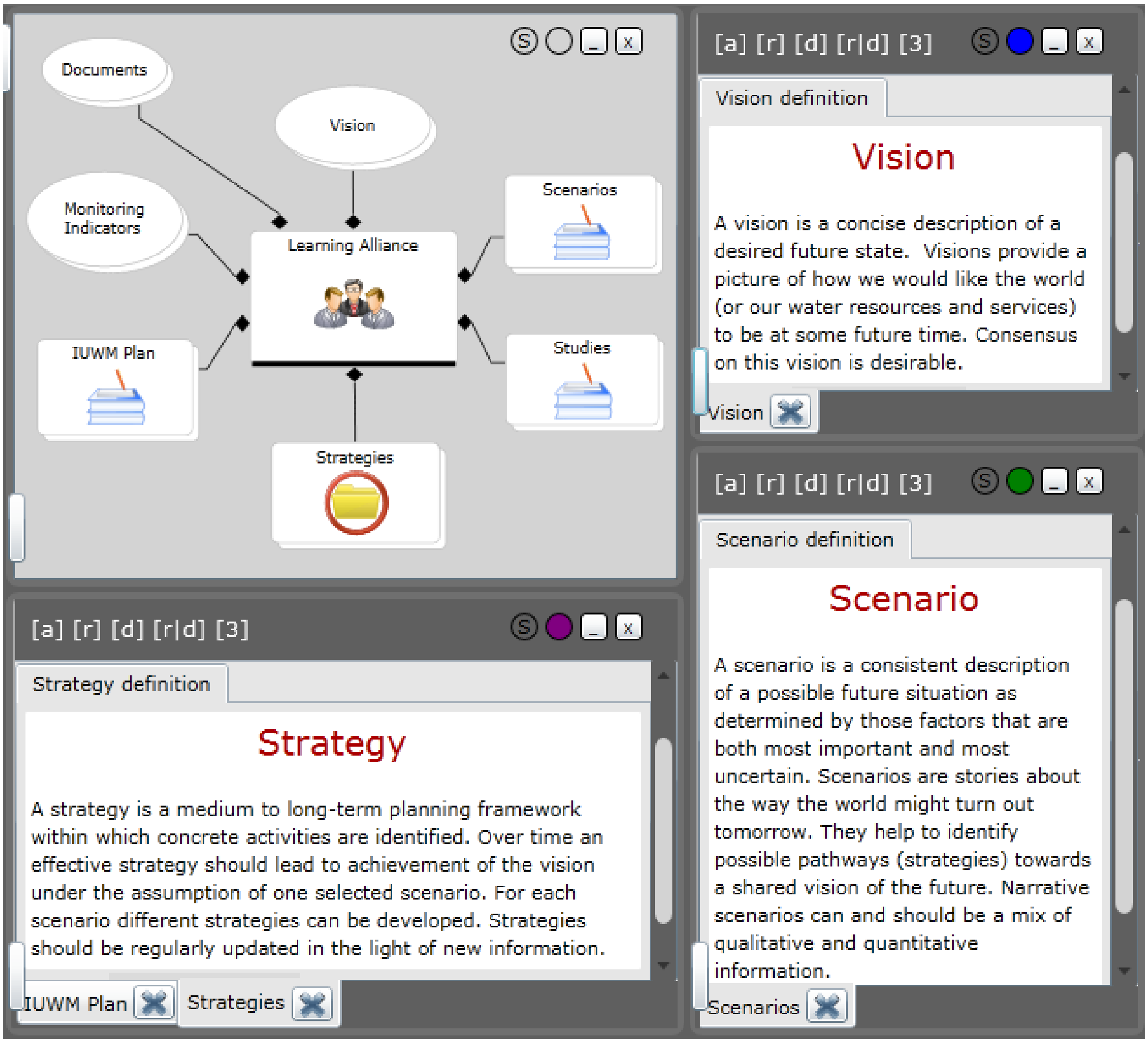


Figure 45 Information about the visioning and scenario planning in CWIS

With the help of the learning alliances and the SWITCH scientific partners, CWIS has been tested and some database were set up for three demonstration cities of the SWITCH project: Birmingham (Dessimoz, 2008; Schenk, 2010), Belo Horizonte (Brandenberg, 2008) and Alexandria whose case is discussed below.

The case study is structured in six sections that describe the steps of the project management cycle (Figure 44) of the learning alliances.

6.4.1 Constitution of the Learning Alliance in Alexandria City

Created in March-April 2007, the Alexandria learning alliance groups stakeholders from the public sector (several national ministries and agencies, the local research institutes and the inter-governmental organization CEDARE), the semi-private sector (the national Holding Company for water and Wastewater and its sub-branches in Alexandria)), and the civil society (some environmental NGOs, media and representatives of the local community…). (CEDARE, 2007)

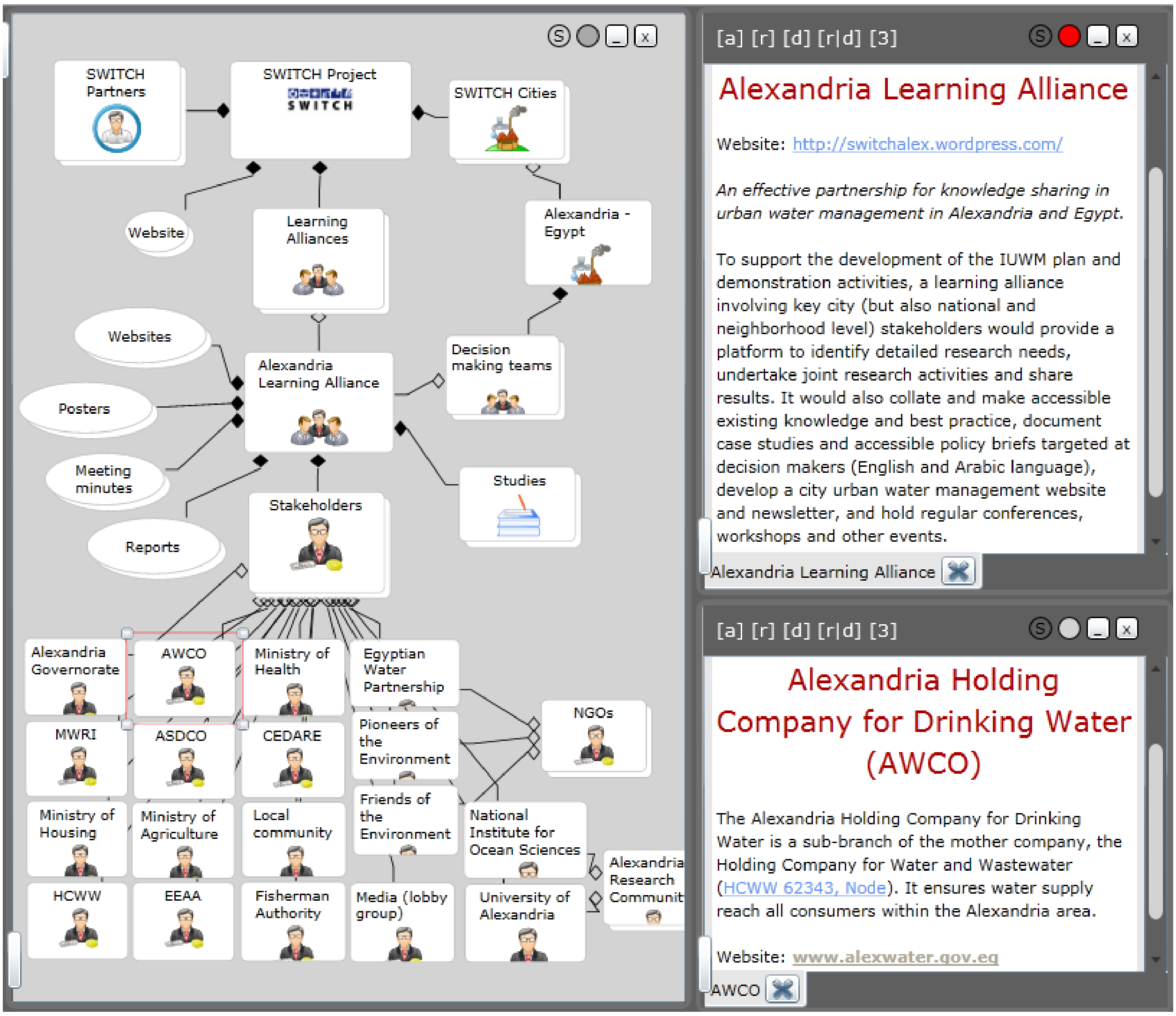


Figure 46 Organisational chart of the Alexandria learning alliance and the involved stakeholders

An organisational chart of the learning alliance of Alexandria is shown in Figure 46. In this case, the SysMod diagramm allows sharing information about who is who in the learning alliance and the activities of each stakeholder.

CWIS can support the activities of the learning alliance in different ways. The CWIS database can host some key documents, such as the Learning Alliance Briefing Notes[14], which are some guidelines that describe, among others, the processes of visioning and scenario planning. CWIS is also a platform to share information and innovations with other stakeholders, researchers or learning alliances. CWIS is also an appropriate tool to conduct some process documentation (Schouten, 2007), i.e. any kind of information produced during the activities of the learning alliance (for instance some comments, anecdotes, reports, presentations, pictures and videos) can be saved, organised and disseminated using CWIS.

At a different level, "advanced users" such as engineers, modellers or scientists, can apply CWIS to design models of the urban water system, store some series of data and possibly perform simulations, as explained in Chapter 5.

6.4.2 Visioning and scenario building

The process of visioning aims to develop a consensus amongst the stakeholders of the learning alliance and lay the basis for further collaborations. The process can be split up into several steps (Batchelor and Butterworth, 2008b):

- creation of the learning alliance,
- definition of the scope of the vision (geographic boundaries of the area of interest and temporal horizons),
- review of existing visions, such as those of the National Water Resources Plan for Egypt - 2017 (NWRP, 2005) and the Alexandria Master Plans for water supply and sanitation for the year 2037,
- identification of the main issues that need to be integrated in the vision,
- development of an outline vision combining a descriptive narrative and some quantifiable criteria,
- verification of the consistency of the vision with existing projects, plans and policies, at the different geographic and administrative scales, and
- finalisation of the vision after some feasibility assessment and wider consultation.

[14] The Learning Alliance Briefing Notes: http://www.switchurbanwater.eu/la_guidance.php (accessed the 18 February 2011)

CWIS can be used in several ways to assist these activities. For instance, the geographic viewer is an adequate tool to discuss the boundaries of the area of interest, in this case the governorate of Alexandria (Figure 47) with a specific focus on the city and its neighbourhood.

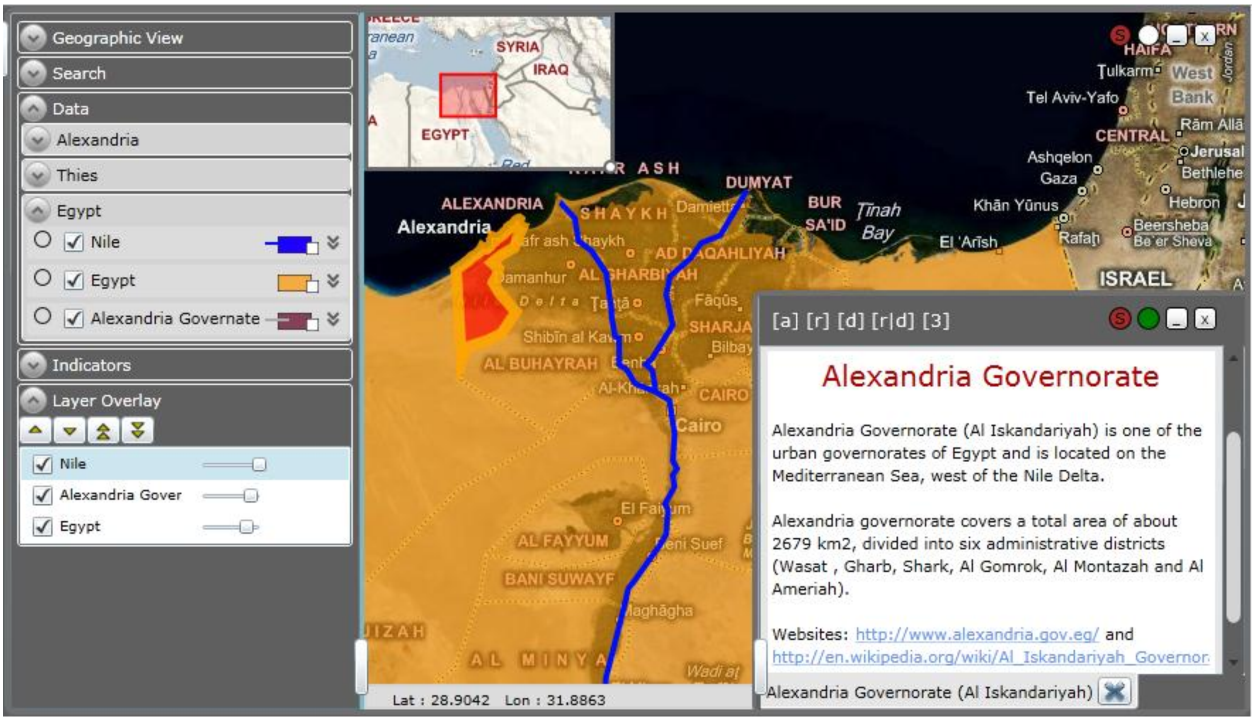


Figure 47 Geographic view of the Alexandria Governorate

Besides defining the organization of data, the CWIS “system views” allow developing some (conceptual) models of the water system that take into account the flows and causal relationships between the properties of the system elements. These diagrams are particularly helpful for representing the complexity of the interactions between the components of the water system, be they structural (e.g. rivers, supply and sanitation infrastructures, building and ecosystems) or non-structural (e.g. laws, environmental policies, fiscal policies and studies). This type of diagram is particularly helpful to identify the main issues of the urban water management, for instance by showing the results of a problem-tree analysis (a useful tool to identify the causes and consequences of a key problem). These diagrams can be reused as a basis for computing simulations, using Bayesian belief networks or system dynamics tools, such as causal loop diagrams or stock and flow diagrams.

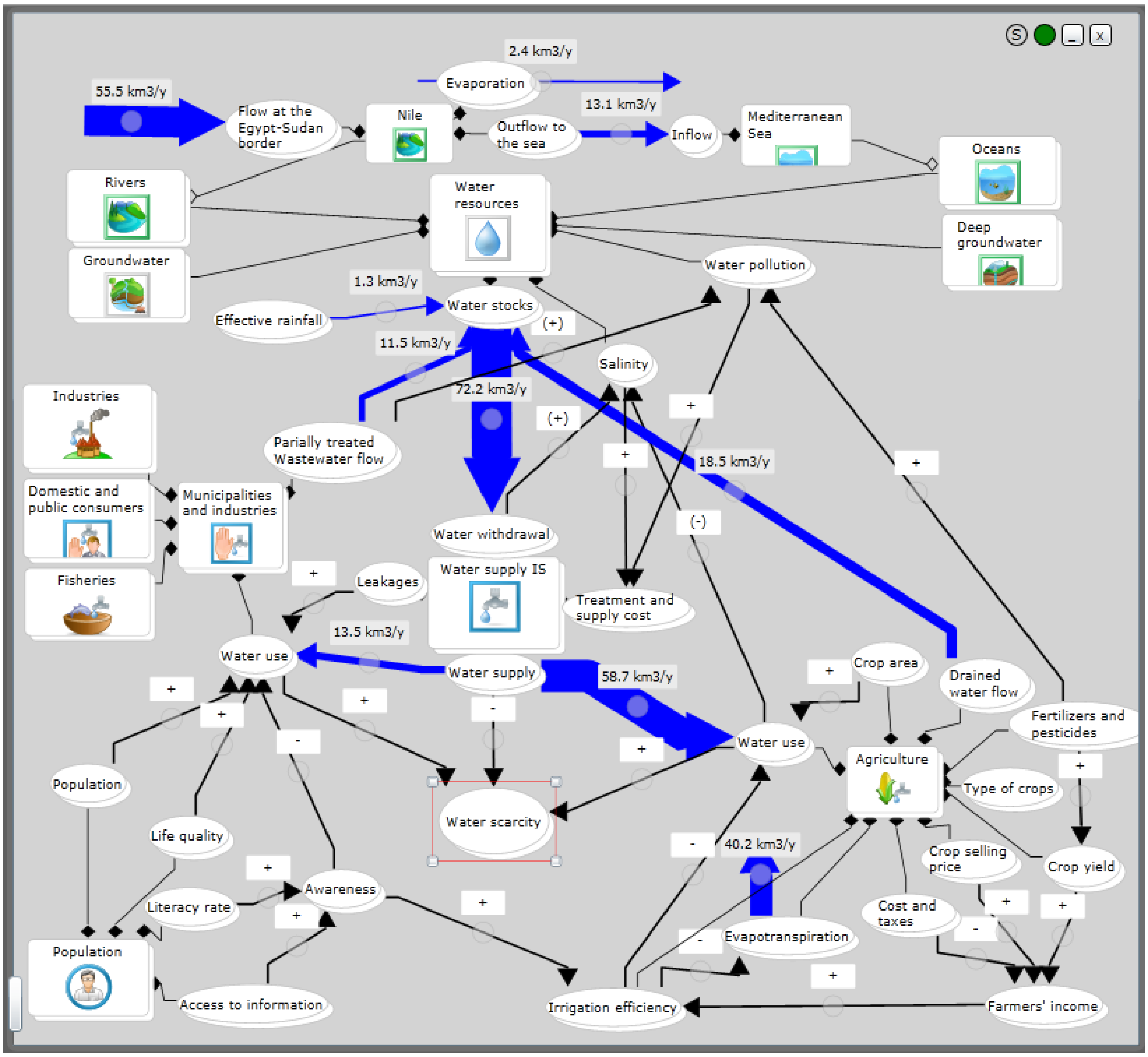


Figure 48 Diagram of flows and influences between the elements of the Egyptian water cycle (1997)

As an illustration of such “complex” diagram, Figure 48 summarises the components and flows of the water cycle for the year 1997, along with some key properties and their influence (positive or negative) on the issue of water scarcity in Egypt. The values of flows are quoted from the National Water Resources Plan (NWRP, 2005) and the causal relationships have been defined based on the article of Abdin and Gaafar (2009) about the issue of water use in Egypt.

In Egypt, economic and demographic developments lead to an increase in the water demand of all sectors (industry, domestic and agriculture) associated with a deterioration of the water quality; agriculture is the main water consumer accounting for almost 96% of the total consumption in 1997. This situation is particularly delicate, as 96% of the freshwater is provided by the Nile River, while the current allocation of 55.5 BCM per year allocated to Egypt could be revised downwards depending on the agreements with the other countries of the Nile basin.

The water supply in Alexandria (located near the Nile Delta) is endangered by the upstream consumption and pollution of the Nile River. In addition to these external factors, the city of Alexandria suffers from local problems such as a high level of physical and commercial losses in the water supply system, some malfunctioning sanitation infrastructures, a pollution of the water bodies such as the lake Marriout, potential risks in case of rising sea level and inequities in access to water supply and sanitation infrastructures.

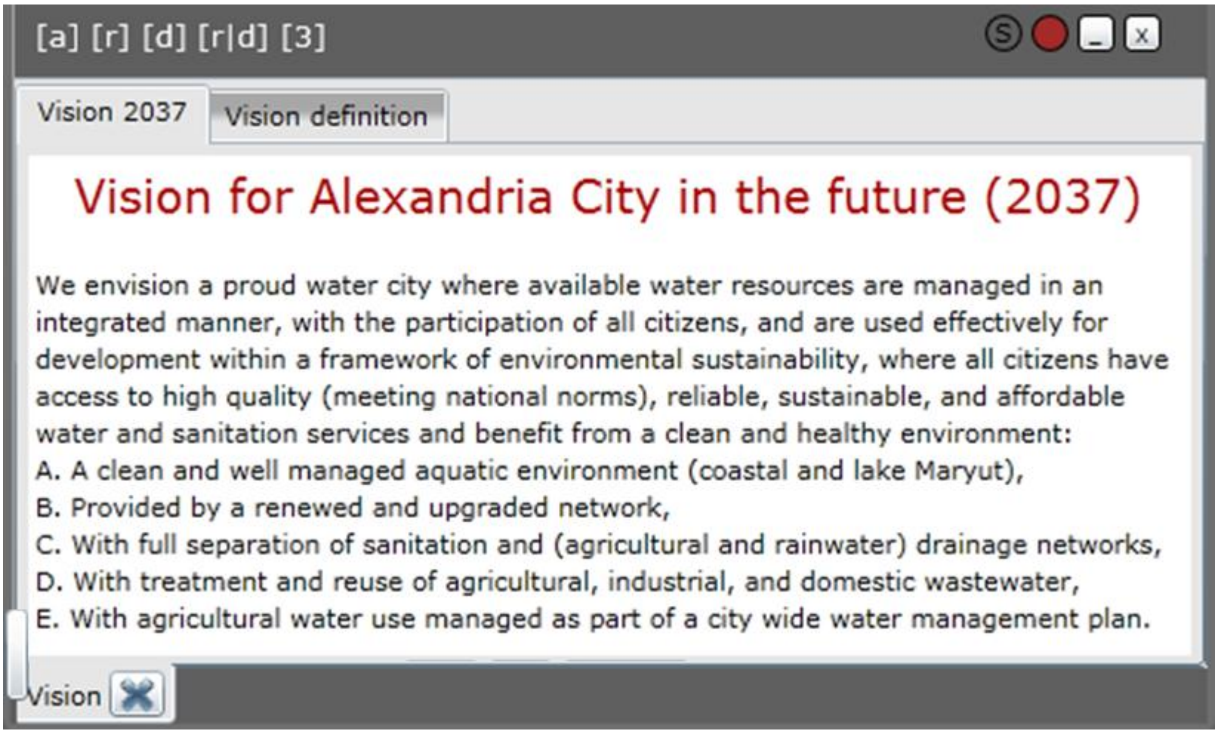


Figure 49 Vision for Alexandria City (time-horizon 2037)

After agreement between the various stakeholders of the learning alliance, a shared vision has been defined in July 2007. In addition to a concise description of the desired future state for the year 2037 (corresponding to the time horizon adopted for the Alexandria Master Plans

for water supply and sanitation), the vision contains five criteria that could be turned into quantifiable indicators for monitoring the progress of the plan over the years (Figure 49).

Following the visioning exercise, the learning alliance went through the process of scenario building. The main objectives of scenario building are to identify and discuss uncertainties and risks that may prevent the achievement of the vision, and to make use of scenarios to create flexible and adaptive strategies that cope with a range of possible future changes. Details about the methodology applied by the learning alliance can be found in one of the SWITCH briefing note (Batchelor and Butterworth, 2008a).

Scenarios are meant to take into consideration possible extreme futures whose causes cannot be controlled at the current decision level. Therefore, scenarios need to consider the uncontrollable factors that are the most important and the most uncertain as regards the potential impacts on the urban water system. In Alexandria, the factors selected to build the scenarios are the population growth, the allocation of the Nile water between the various regions, the rising sea level, the economic growth and the energy cost.

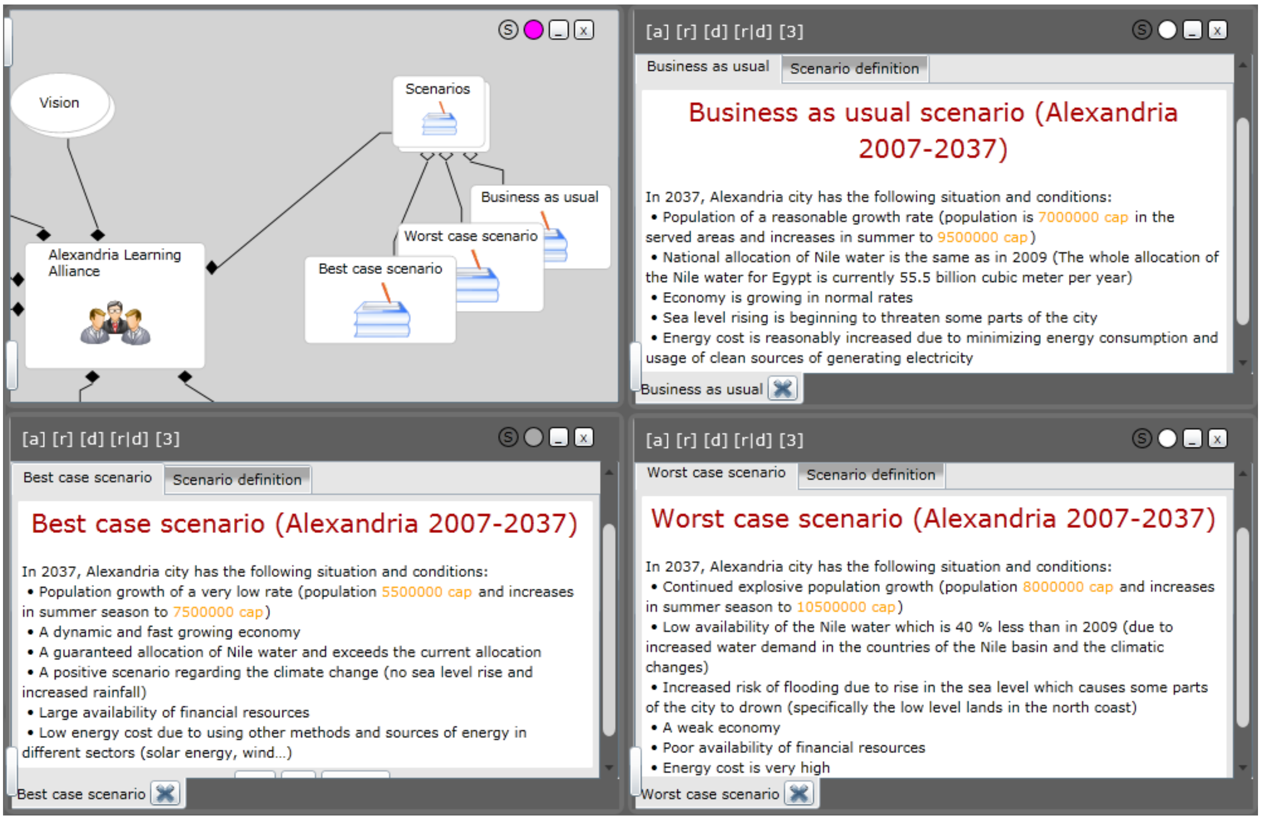


Figure 50 Scenarios developed by the Alexandria learning alliance

The process of scenario building resulted in the creation of three scenarios (Business as usual, Best case scenario and Worst case scenario) shown in Figure 50. The scenarios mix narrative descriptions (to be understandable by everyone) and numerical information that allows performing quantitative analysis and simulations. The highlighted values shown in the scenario reports are examples of such numerical information, the colour indicates that the values are taken from the database and that they belong to a property of the urban water system (e.g. in this case the property “Population” of the object “Alexandria City”).

6.4.3 Assessment of the situation

The objectives of the assessment stage are to collect and analyse data to form a shared information accessible to all stakeholders and to further develop the vision and scenarios, in order to have reliable and consistent information for decision making (Moriarty et al., 2005).

This process consists in bringing together existing data, especially those owned by the stakeholders, and conducting additional surveys to fill the missing information. The information base needs to cover (1) the available water resources, (2) the infrastructures for water supply, sanitation and protection against water-related hazards, and (3) the demand for water by the different water consumers, as well as (4) the accessibility of each consumer to the infrastructures.

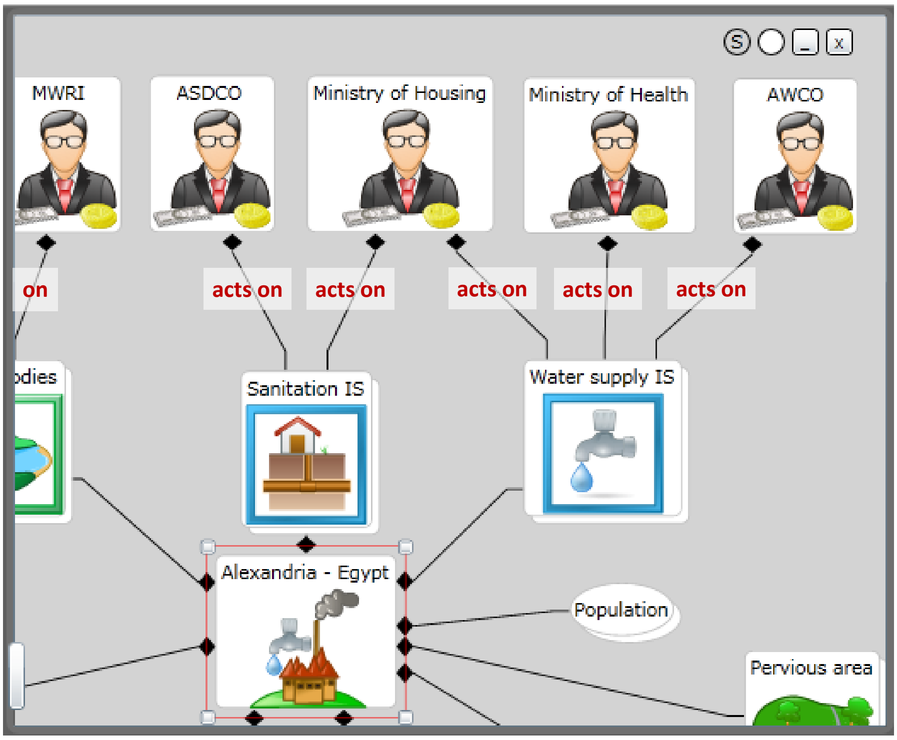


Figure 51 Schematic representation of the domain of responsibilities of the stakeholders

As shown in Figure 51, CWIS allows representing (through system views) the responsibilities of each stakeholder as regards the different elements of the urban water system. The complete information about the role of the institutional stakeholders and the legal framework are available in the report about institutional mapping and water governance analysis in the city of Alexandria (CEDARE, 2010a)

Depending on the type of collected data, the CWIS user has to choose the appropriate module (geographic viewer, system viewer or reporting tool) to organise and display the information. It is interesting to note that the three modules offer three complementary views to describe an object and can be synchronised. Figure 52 illustrates this point by showing three different views that depict the Lake Marriout located south of Alexandria City. The geographic view displays its location and geometry. The system view displays the properties of the lake and its interactions (the input and output flows) with the other elements of the water system. Finally, the reporting tool provides a report (similar to a web-page) and the list of data that documents the considered object (the lake).

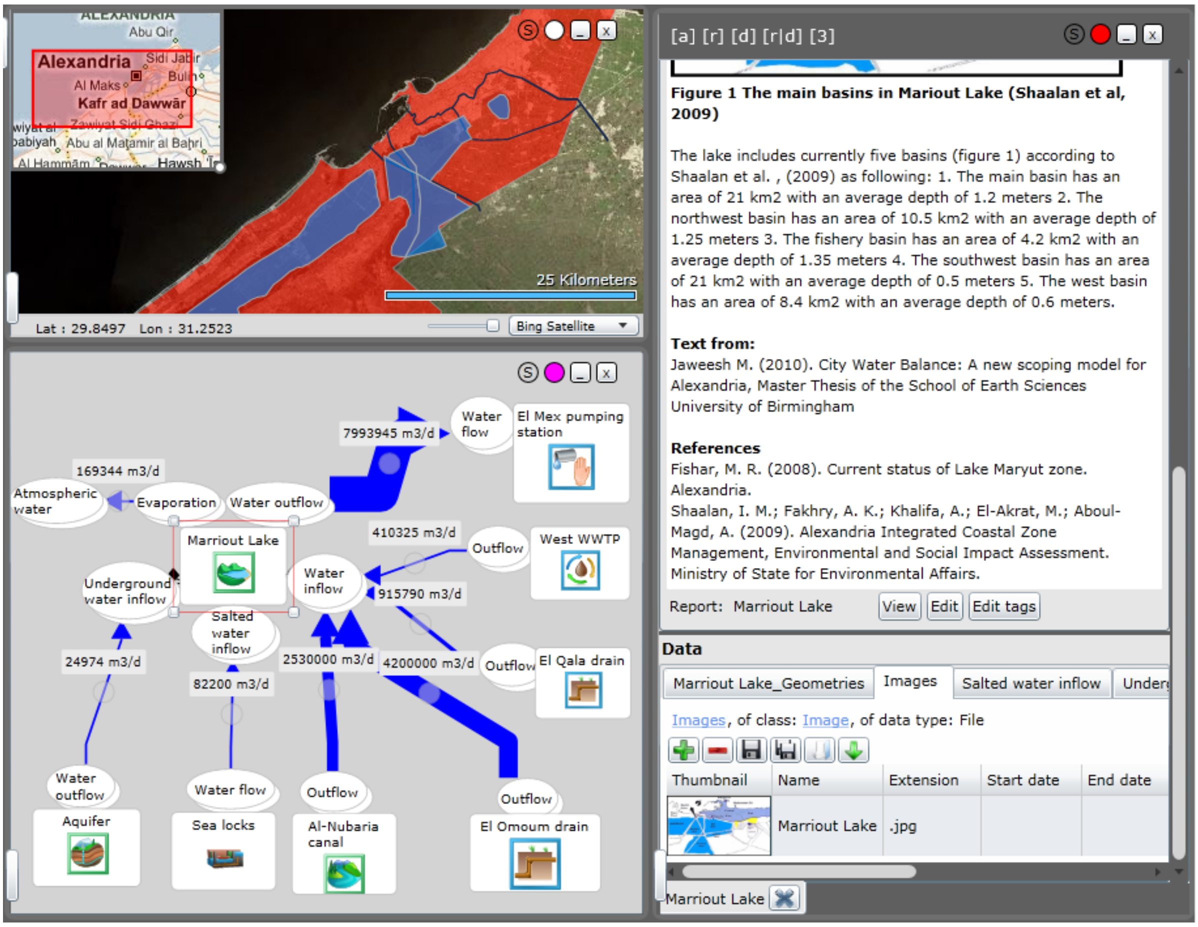


Figure 52 Three views (geographic view, flow diagram and report) of the Lake Marriout

Thematic maps are particularly helpful to analyse spatial variations and facilitate the identification and communication about water-related problems (provided they can be represented geographically). Figure 53 shows two thematic maps that describe the population density (in capita/m^2) and the water consumption (in m^3/m^2/day) for some clusters of the Alexandria City. Each cluster is a group of buildings sharing the same affectation (industrial area, dense populated area, suburban houses, hotels, informal settlement…). The very dense area surrounded by a red circle is an informal settlement called Ma'awa Sayadeen. Situated on the shore of the lake Marriout, Ma'awa Sayadeen is a "fishermen's village" which currently doesn't have an adequate sewage system. It has more than 10000 residents for an area of 273'000 m^2. This area has been chosen by the learning alliance to perform some "action researches", an original approach that aims to bring together scientists and stakeholders to find innovative solutions to address the problems affecting this neighbourhood (Moriarty, 2007).

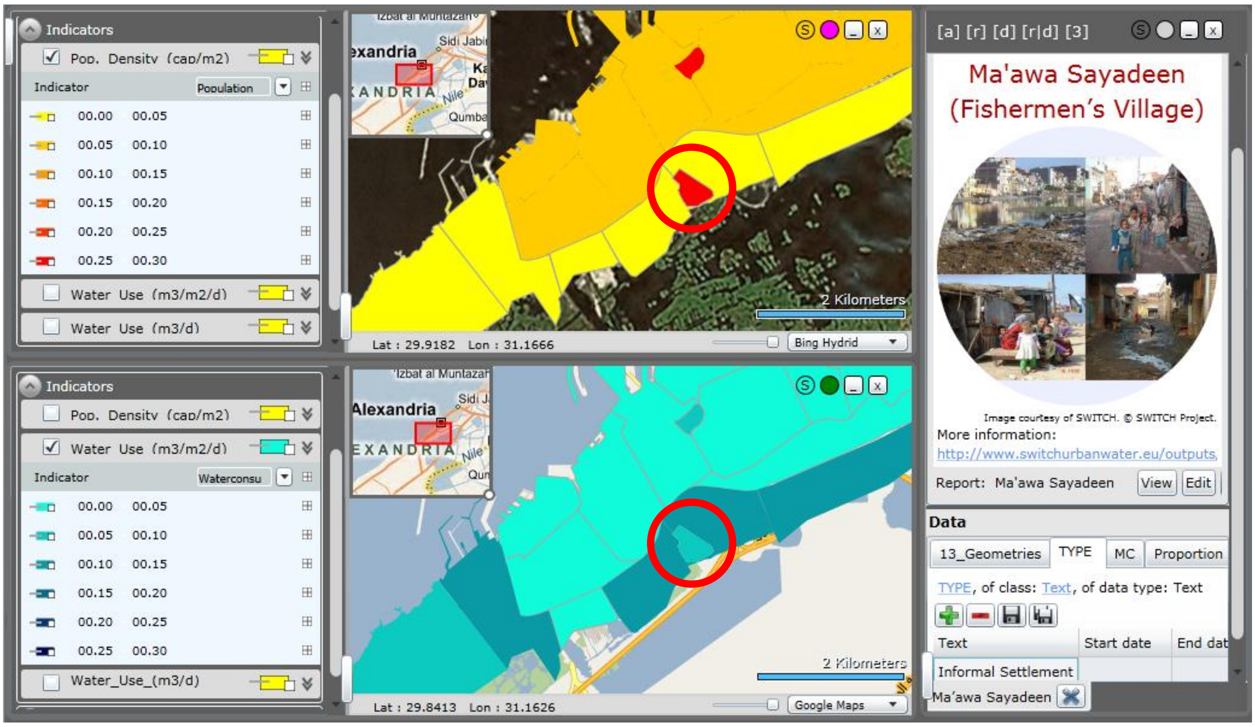


Figure 53 Thematic maps of Alexandria and location of the Ma'awa Sayadeen settlement

Finally, the assessment stage may require the use of mathematical modelling, in order to determine future characteristics of each scenario and to verify their validity. The model City Water Balance (CWB) (Jaweesh, 2010) has been applied to evaluate the future water cycle of Alexandria and the need to import water.

CWB is a scoping model used to simulate the water cycle in urban water system and to provide some output indicators about the water demand, quality, energy consumption and life-cycle cost (Mackay and Last, 2010). It is particularly adequate for evaluating future scenarios and testing the effect of strategies.

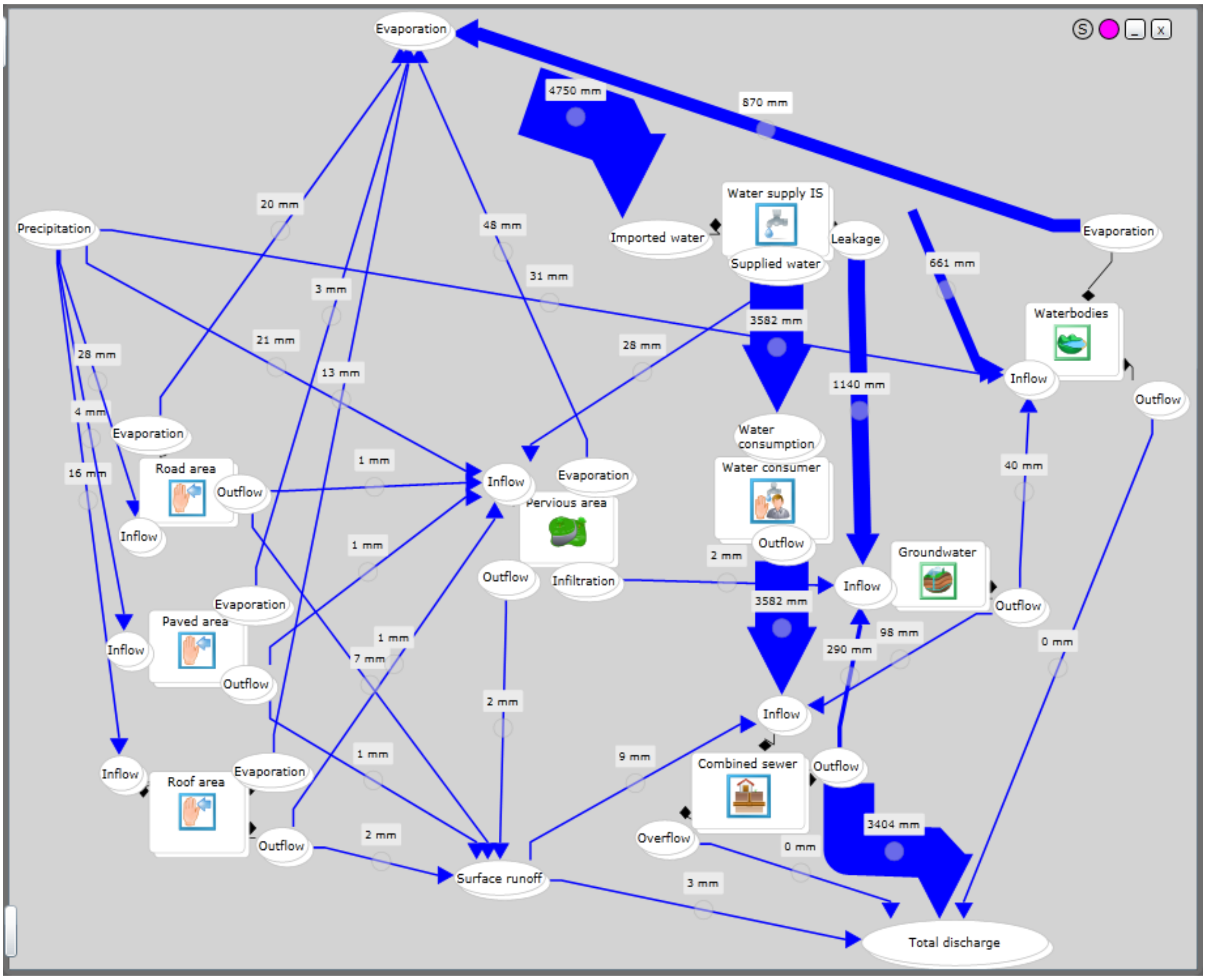


Figure 54 Results of the CWB model, scenario "business as usual" for the year 2037, values drawn from Jaweesh (2010)

Results of the simulation for the scenario "business as usual" are shown in Figure 54. The values are in mm/year. The "future" surface of the study area being approximately 238 km^2, the water demand will be 1.13 MCM in 2037, compared to 0.80 MCM in 2006 (Jaweesh, 2010).

6.4.4 Strategy definition and evaluation

The main objective of the strategy development is to propose “a robust adaptable strategy that has the potential to achieve a shared vision under a whole range of different scenarios” and to “encourage stakeholders to take the leading role in an IUWM strategy development process” (Batchelor and Butterworth, 2008c).

For this purpose, the learning alliance has conducted additional surveys to explore a range of possible strategies: to decrease the water consumption through measures such as leakage reduction, consumption monitoring, increased tariff or awareness campaign (CEDARE, 2010c) and to develop alternative water sources, such as wastewater treatment and reuse, use of storm water (CEDARE, 2009a, b), groundwater (CEDARE, 2009c) and desalination (CEDARE, 2010b). A synthesis of the different studies and strategies is given in Figure 55.

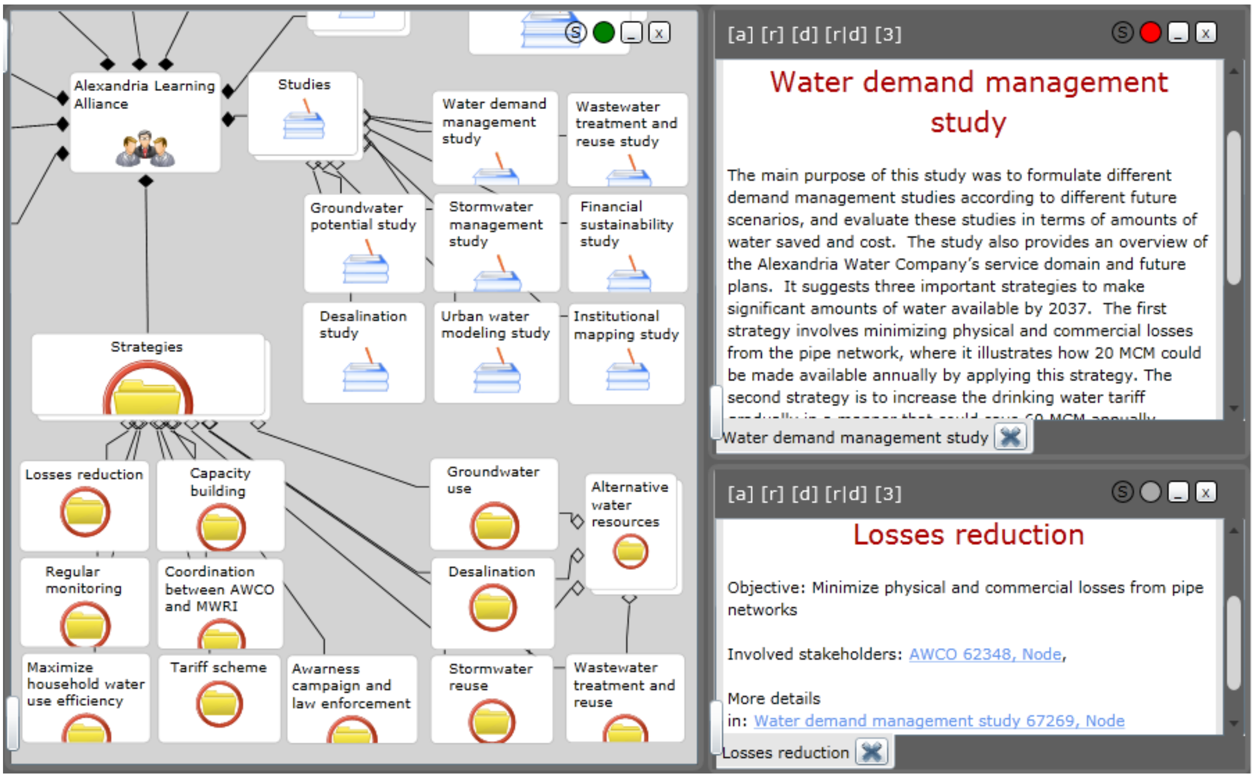


Figure 55 Studies conducted by the learning alliance and resulting strategies

Simulation tools may be particularly useful to test and evaluate the effects of the strategies. For instance, the CWB model has been applied to test the strategies of leakage reduction, maximisation of the water use efficiency and treatment and reuse of the wastewater (Jaweesh, 2010). A life cycle assessment of several approaches such as losses reduction in

the water supply, extension of the current sanitation network and decentralised wastewater treatment has been carried out by El-Sayed Mohamed Mahgoub et al. (2010).

The integration of mathematical models with CWIS can be achieved in two different ways, as described in Chapter 4:

1. By running the model directly from the CWIS interface; this is possible if a programmatic routine has been developed to transform the CWIS data in the model format and vice-versa, or if the model is compliant with the CWIS format.
2. By using the CWIS database with an integrated modelling framework (IMF) such as OpenMI (Gregersen et al., 2007). In this case the model needs to be compatible with the IMF, for instance by adopting the OpenMI standards.

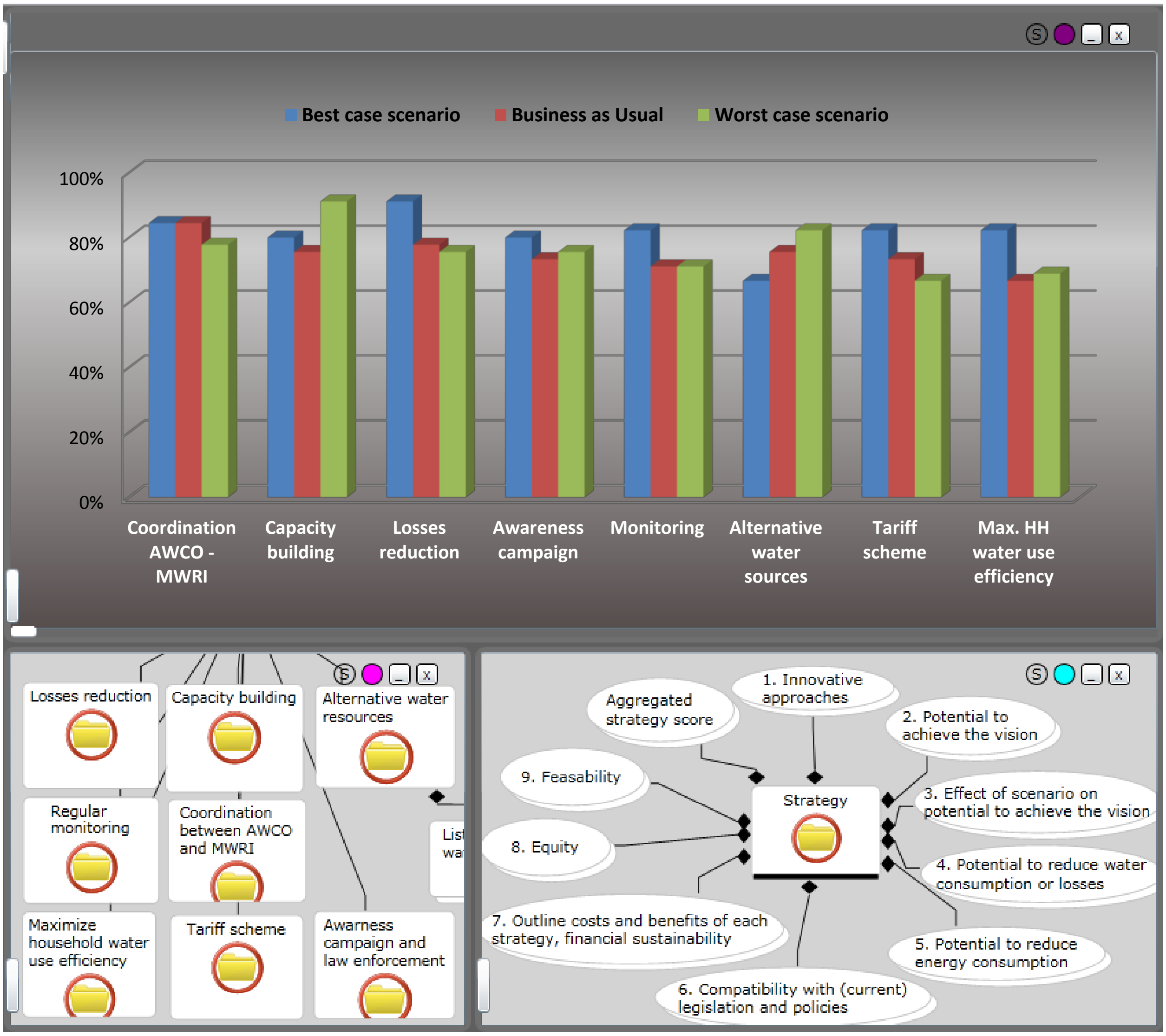


Figure 56 Scenario-based assessment of the strategies (CEDARE, 2010c)

The results of the simulations and strategy assessments can be browsed with the different viewers of CWIS. Figure 56 presents the preliminary assessment of eight strategies (CEDARE, 2010c). The evaluation is based on a series of qualitative indicators (see bottom right of Figure 56).The chart describes the aggregate results of the assessment in descending order of performance, 100% being the best possible score. Each indicator has the same weight in the evaluation.

6.4.5 Planning and implementation

The planning phase aims to create a detailed plan that defines which parts of the agreed strategy need to be implemented in priority and depending on what schedule. The use of CWIS is similar to the previous examples. CWIS allows defining the roles and responsibilities of each stakeholders, and the temporal changes of the urban water system according to the planned implementation of the strategies can be modelled using the system views of the application.

During the implementation phase, CWIS can be used for documenting and monitoring the implementation process.

6.4.6 Monitoring and continuous assessment

As a general rule, the monitoring process can be applied to any type of activities linked to the process of scenario planning carried out by the learning alliance. Most of all, the monitoring process is relevant for managing indicators to:

- Assess the progresses towards the achievement of the learning alliance's vision,
- Follow the implementation of the IUWM plan, and
- Monitor the work of the learning alliance itself.

As shown by the example of strategy evaluation in Figure 56, CWIS is particularly adapted for handling indicators. An additional illustration of the use of CWIS in this field is given in Figure 57, which describes a set of indicators developed for monitoring the progress carried out by the SWITCH learning alliances (Butterworth and Dasilva, 2008).

Afterwards, the indicator scores can be used to plot some charts, such as radar charts, bar charts or line charts in order to graphically display the evolution of the indicators.

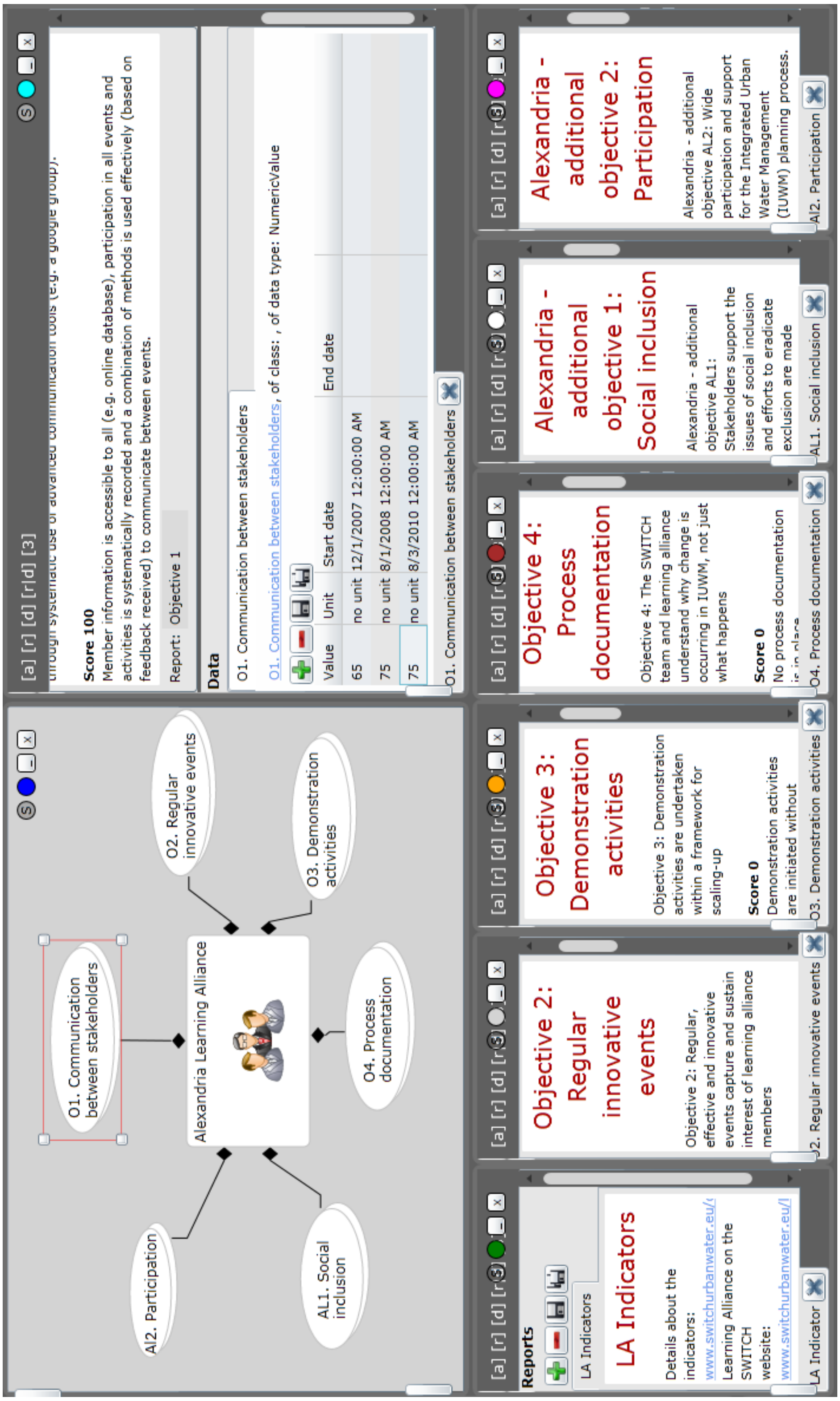


Figure 57 Use of indicators to monitor the progress of the learning alliance

6.5 Discussion and synthesis

The objective of the current case study was to illustrate one of the possible applications of the CWIS tool and its underlying concepts. In the context of scenario planning for IUWM, the main added-value of CWIS resides in its capacity to integrate and structure data from various sources and display the information in different complementary ways like geographic views, system views and textual reports.

The holistic, systems thinking perspective adopted by the SysMod language makes CWIS a particularly flexible and scalable tool for IUWM, even though the formalism imposed by SysMod may be confusing at a first sight. While reading the SysMod diagrams is quite straightforward, their edition requires a specific training.

To match the needs of the users, CWIS can be configured at different levels. Firstly, at a knowledge management level, the implementation of some ontologies allows defining some lists of concepts and terminologies to be used for the modelling of the systems under consideration (e.g. the urban water system). Then, from a programming perspective, CWIS is based on a modular architecture, which can eventually be extended with other modules providing extra functionalities. CWIS offers several tools (projects, views, workspaces, and dashboard) to personalise the application and facilitate the access to information.

The current prototype of CWIS still has missing features such as the ability to handle processes for describing data provenance and capacity to handle uncertainty metadata. As a consequence, these features were not presented in the case study, although they may have provided some meaningful inputs. The concept of data provenance would have been particularly helpful to characterise the sources of information and to enrich the documentation of the processes carried out by the learning alliance. On the other hand, the framework for uncertainty assessment would have provided valuable tools, such as the uncertainty matrix, to assess the quality of the data. Moreover, the use of pedigree matrices would have been particularly adequate to evaluate the quality and reliability of the vision, scenarios and strategies.

As illustrated by the case study, CWIS has the ability to store, structure and display information for a wide range of different situations. The application could be extended with some management and planning features (such as projects organisation with timeline, tasks and responsibilities), to provide users with a complete and operational suite of tools for planning and monitoring the implementation of the strategies for IUWM.

7. Conclusion

In this thesis, a systems modelling framework has been developed to provide a set of theoretical principles and tools to model the knowledge of human and natural systems and support the process of integrated resources management (IRM). The system modelling framework rests on four cornerstones:

- Systems thinking, which offers a perspective to address the complexity of IRM, by representing the human and natural systems as interacting systems hierarchically organised.
- Semantic integration, to tackle the issues of integrating data from various sources. This integration is based on ontologies, which are (in computer science) some structured representations of the elements that compose a field of knowledge.
- Scenario planning, a participative, strategic planning method particularly adapted for the sustainable management of natural resources. It goes through the processes of visioning, scenario building and strategy development to elaborate adaptive plans that address the most significant and most uncertain factors influencing the future systems' evolution.
- Simulations, to provide forecast of the systems' evolution and assess the impacts of various strategies.

The development of a modelling language (SysMod) has addressed the fields of systems thinking and semantic integration. SysMod is a hybrid language between system modelling languages and ontology languages. SysMod allows "modellers" to semantically define the types of elements that compose the considered human/natural systems. These elements are subsequently used for modelling the issues related to natural resources management.

The SysMod language has then been extended to include important features for scenario planning. The first feature is the capacity to define multiple scenarios representing possible systems' evolutions; this is achieved by the definition of "thematic groups" of data to gather systems elements and relationships that belong to the same scenarios or strategies. The second feature is the "data provenance", which is a method to keep track of the data sources and processes involved in data generation. The third feature is the "data uncertainty", which provides metadata about the quality and validity of information. Combined together, the

concepts of data provenance and uncertainty support more informed decision making, by providing a way to characterise the data reliability.

The constructs of the SysMod language and its extensions were used to create an information system called CWIS (Combined Water Information System). Developed by the EPFL Switch team, CWIS consists of a relational database and a web-application. CWIS possesses a modular interface that contains several visualisation tools. The three main modules of the application are: (1) the system viewer, which allows structuring the data according to the formalism of SysMod, (2) the geographic viewer, which is used for displaying the spatial properties of system elements, and (3) the reporting tool, which is designed to enable data edition (text, files, values…) and creation of reports (some kind of interactive fact sheets) that document the systems elements.

An additional module has been integrated into the CWIS interface to allow exporting and importing data, as well as linking simulation tools (mathematical models). This tool called the “data and exchange modelling module” is a part of a general modelling approach that has been developed to provide a methodology for model integration, taking advantage of the features of SysMod and the concepts of themes (scenarios), data provenance and uncertainty.

Finally, the research developments have been tested with a case study on scenario planning for integrated urban water management (IUWM) in the city of Alexandria (Egypt). The case study demonstrated the ability of CWIS to organise and share the information related to the activities of a stakeholders’ platform.

CWIS is an innovative and flexible information system, which can be used at each steps of the scenario planning process to model the knowledge of the urban water system, to support the documentation process, to perform simulations and to monitor indicators. While numerous information systems exist in the field of integrated water resources management, none have reached this level of integration and succeeded in providing an evolutionary environment to manage complex and interdisciplinary problems. Beyond water management, CWIS can be used to manage and share any kind of environmental information. This ability comes from the SysMod ontologies, which can cover and structure data from any field of knowledge. Ontologies are also particularly suited for inventorying some best management practises and other data of references; such information being very helpful to create scenarios and strategies. Compared to other information systems in the field of IRM, CWIS is able to handle quantitative and qualitative uncertainties, to track data provenance, to manage

scenarios, to share information in various forms (geographic views, system views and textual pages) and to integrate mathematical models.

The second version of CWIS is still in the prototype phase. Further improvements are required to finalise the implementation of features, such as the theme (scenario/strategy) and temporal filters, the edition of provenance and uncertainty data and the models' linkage. Taking into account the users feedbacks on the functionality of CWIS will also offer opportunities to improve the application and to question the concepts developed in this thesis.

Different directions for future research have been identified, to enhance and expand the functionalities of the systems modelling framework and the CWIS information system:

- The construction of an ontology can be costly and time consuming. Hence semi-automated tool, such as text2onto (Cimiano and Volker, 2005), could facilitate the creation of ontologies from documents using text mining and natural language techniques. Moreover, additional developments may enable the translation and reuse of ontologies defined in other ontology languages.
- Some inference algorithms could be implemented to better exploit the content of the ontologies. Using these algorithms, CWIS could detect data inconsistencies and assist the modelling of human and natural systems.
- The concept of declarative modelling (Villa, 2001), could offer some solutions to transform the current conceptual representation provided by CWIS into computable models. Declarative modelling involves the ability to characterise the causal relationships between the system input and output through equations or logical rules, as well as to develop algorithms that use these equations to perform simulations. In a similar vein, CWIS could be extended with reasonable efforts to handle Bayesian belief networks, which represent probabilistic relationships between system variables. This approach aims to define the various states of the system variables, the dependencies between them, and the conditional probabilities linked to those dependencies.
- Finally, further researches may investigate possible coupling with existing environmental assessment tools that rely on a systems representation, such as (1) material flow analysis, which is used to analyse the flows of a material in a system, (2) life cycle assessment, a technique to assess the environmental impacts of a product (or service) by quantifying and allocating the impacts of all the systems that play a role in the product's life cycle, and (3) life cycle cost, a tool to measure the economic costs of a product (or a service) over its entire life cycle, i.e. from the production of the raw material to the disposal of the product.

Appendix A

A.1 Database schema of the CWIS database

DOMAINS

<<ENUMERATION>>
THEMES DOMAIN
11 = Scenario
21 = Strategy
31 = Story
41 = Causal chain
51 = Strategic activity
61 = Event
71 = Process run
81 = Arbitrary

<<ENUMERATION>>
UNCERTAINTY LOCATIONS DOMAIN
11 = Process context
12 = Process structure
13 = Process implementation
14 = Process parameter
21 = Data

<<ENUMERATION>>
UNCERTAINTY NATURES DOMAIN
1 = Undefined
2 = Epistemic
3 = Stochastic
4 = "Postponed" epistemic

<<ENUMERATION>>
THREE SIZES DOMAIN
1 = Small
2 = Medium
3 = Large

<<ENUMERATION>>
PROCESSES DOMAIN
11 = Data transformation
21 = Data analysis
31 = Model
32 = Ptolemy model
33 = R expression
34 = Matlab expression
41 = Query
42 = SQL query
43 = XQuery
51 = SysMod workflow
52 = Kepler workflow
53 = OpenMI workflow
61 = Expert judgement
71 = Measurement
81 = Quotation

<<ENUMERATION>>
REFERENCE ELEMENTS DOMAIN
11 = Lower Bound
12 = Upper Bound
13 = Interval
21 = Membership Function
31 = Single Value
0 = None

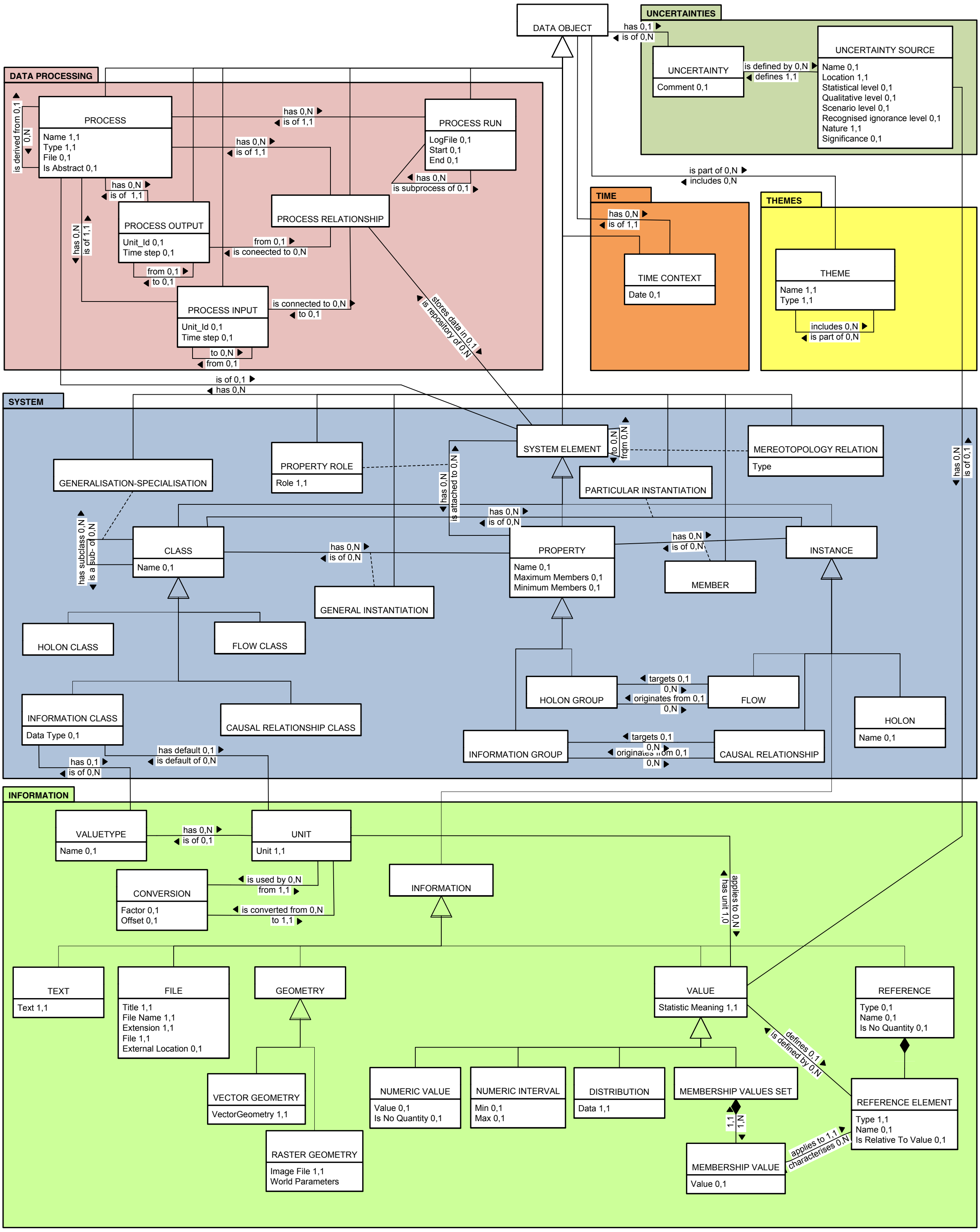
<<ENUMERATION>>
REFERENCE TYPES DOMAIN
11 = Legal
12 = Target
13 = Design
21 = Precision
22 = Statistic
31 = Classification
41 = Membership function set
51 = Pedigree

<<ENUMERATION>>
PROPERTY ROLES DOMAIN
11 = Has
21 = Contains
22 = Consumes
23 = Produces
31 = Requires
41 = Acts on/manages
51 = Interprets

<<ENUMERATION>>
DATA TYPES DOMAIN
11 = Text
21 = File
31 = Vector Geometry
32 = Raster Geometry
41 = Numeric Value
42 = Numeric Interval
43 = Distribution
44 = Membership Values Set
51 = Reference

Schema Information			
Project Name	SWITCH		
Schema Name	CWIS database		
Organization	ECHO/ EPFL		
Individual	Bastien Roquier		
File Name	E:\Thesis\CWIS_database_schema.vsd		
Version	3.1	Version Date	13.12.2010

Perceptory 2003


DATA OBJECT
UNCERTAINTIES
UNCERTAINTY
Comment 0,1
UNCERTAINTY SOURCE
Name 0,1
Location 1,1
Statistical level 0,1
Qualitative level 0,1
Scenario level 0,1
Recognised ignorance level 0,1
Nature 1,1
Significance 0,1
DATA PROCESSING
PROCESS
Name 1,1
Type 1,1
File 0,1
Is Abstract 0,1
PROCESS RUN
LogFile 0,1
Start 0,1
End 0,1
PROCESS RELATIONSHIP
PROCESS OUTPUT
Unit_Id 0,1
Time step 0,1
PROCESS INPUT
Unit_Id 0,1
Time step 0,1
TIME
TIME CONTEXT
Date 0,1
THEMES
THEME
Name 1,1
Type 1,1
SYSTEM
SYSTEM ELEMENT
MEREOTOPOLOGY RELATION
Type
PROPERTY ROLE
Role 1,1
GENERALISATION-SPECIALISATION
PARTICULAR INSTANTIATION
CLASS
Name 0,1
PROPERTY
Name 0,1
Maximum Members 0,1
Minimum Members 0,1
INSTANCE
MEMBER
GENERAL INSTANTIATION
HOLON CLASS
FLOW CLASS
HOLON GROUP
FLOW
HOLON
Name 0,1
INFORMATION CLASS
Data Type 0,1
CAUSAL RELATIONSHIP CLASS
INFORMATION GROUP
CAUSAL RELATIONSHIP
INFORMATION
VALUETYPE
Name 0,1
UNIT
Unit 1,1
CONVERSION
Factor 0,1
Offset 0,1
TEXT
Text 1,1
FILE
Title 1,1
File Name 1,1
Extension 1,1
File 1,1
External Location 0,1
GEOMETRY
VECTOR GEOMETRY
VectorGeometry 1,1
RASTER GEOMETRY
Image File 1,1
World Parameters
VALUE
Statistic Meaning 1,1
REFERENCE
Type 0,1
Name 0,1
Is No Quantity 0,1
NUMERIC VALUE
Value 0,1
Is No Quantity 0,1
NUMERIC INTERVAL
Min 0,1
Max 0,1
DISTRIBUTION
Data 1,1
MEMBERSHIP VALUES SET
MEMBERSHIP VALUE
Value 0,1
REFERENCE ELEMENT
Type 1,1
Name 0,1
Is Relative To Value 0,1

Appendix B

B.1 Data provenance techniques

At a common level, bibliography is a widespread form of provenance technique, which provides an effective way to reference the sources of data or information. Its reliability can be enhanced with identifiers, such as the Digital Object Identifiers (DOIs) that uniquely identify electronic documents or scientific data (Witten et al., 2010).

Provenance techniques may differ depending on the domain where they are applied. Some of them are specifically dedicated to database systems (Benjelloun et al., 2008; Buneman et al., 2001), tracking the provenance based on "annotation" of the derivation history of the data or through the "inversion" of the executed queries. An example of such database management system is TRIO (Agrawal et al., 2006) which addresses the issue of provenance through query inversion.

Scientific workflows systems (SWSs) are another domain where the concept of provenance has been extensively explored. SWSs are tools designed to compose and execute some chains of data processes for scientific applications. To get a comprehensive overview of provenance techniques in the field of SWSs, one can read the articles of Bose and Frew (2005), Simmhan et al. (2005) and Moreau et al. (2008). There are dozens SWSs integrating the concept of provenance that can be found in the literature, it includes Chimera (Foster et al., 2002) applied in physics and astronomy, myGrid (Zhao et al., 2004) used in biology and based on the SWS Taverna (Oinn et al., 2004), the Earth System Science Workbench (ESSW) (Frew and Bose, 2001) known for its capacity to manage large data set from environmental models and global satellite imagery, the Collaboratory for the Multi-scale Chemical Sciences (CMCS) (Pancerella et al., 2003) applied in chemical sciences, and Pegasus (Kim et al., 2008) and Kepler (Altintas et al., 2006) which both have been applied in a wide range of domains such as biology, earth sciences and physics.

Tracking provenance implies to consider the different types of elements involved in the data generation. For instance, Simmhan et al. (2005) make the distinction between two types of provenance: the data-oriented provenance and the process-oriented provenance, depending on a focus either on data or processes. In order to enlarge the provenance framework, Klasky et al. (2008) suggest to consider four types of provenance:

- **Process provenance** which records the operations performed within the scientific workflow;
- **Data provenance** which focuses on the input, output and intermediary data, within the workflow;
- **Workflow provenance** which aims to capture the composition and structure of the workflow and, in the case of executable scientific workflows, to record the log of workflow events for a specific run (or workflow invocation), e.g. a log file including events' execution time, invocation time steps and some identifiers;
- **System provenance** which is about recording the context information to help understand a workflow run or to reproduce it. System provenance aims to consider all the external factors which could affect the outcome of the workflow, such as the operating system of the computer, a software version or some environment variables at the time of compilation.

The types of provenance addressed vary between the tools: the TRIO database management system and the CMCS scientific workflow system are data-oriented, MyGRID and Chimera are process-oriented and ESWW and Kepler make use of both types (Altintas et al., 2006; Simmhan et al., 2005). As the size of provenance meta-information can become larger than the data it describes, the choice of the type of provenance may depend on a trade-off between the scientific needs and the computing and storage requirements.

B.2 Guide of methods for uncertainty assessment

Refsgaard et al. (2007) provide a "*Guide to select an appropriate methodology for uncertainty assessment*" consisting in a list of methods for uncertainty assessment classified according three different perspectives: *(1) the modelling process and level of ambition*, *(2) the sources* (location) *and type* (level) *of uncertainty* and *(3) the purpose of use.*

The first two classifications are summarised below (see Refsgaard et al. (2007) and van der Sluijs et al. (2004) for more details).

Selection based on the modelling [and data] processes

Among the modelling process, and more generally for any kind of survey implying data processes, methods dealing with uncertainties are required for three main types of situation:

Identify and characterise the sources of uncertainty:

The uncertainty matrix (Walker et al., 2003) is a suitable tool to identify and characterise the uncertainties by defining their location, level and nature. In case of doubt or ambiguity regarding the evaluation, extended peer review (EPR) and stakeholder involvement (SI) (Grimble and Wellard, 1997; Kloprogge and Van Der Sluijs, 2006; Refsgaard et al., 2007) may help completing this task by providing respectively insight and extra knowledge from non-scientific sources, and improving the involvement and accountability of the stakeholders in the operation.

Review-dialogue-decisions:

Quality assurance (QA) methods (Refsgaard et al., 2005a; Scholten et al., 2007) consist in the application of some guidelines or protocols to review the proper application of the processes, be they models, measurements, data transformations or expert consultations.

Moreover, the IRM framework (as well as IWRM), requires a permanent and structured dialogue between the modellers (or scientist), decision-makers and stakeholders. It can be achieved through stakeholder involvement or extended peer review.

Uncertainty assessment and propagation:

There is wide range of techniques to assess and propagate model-related statistical uncertainties. For instance, the Monte-Carlo analysis (MCA) (Beven, 2006; EPA, 1997) and the sensitivity analysis (SA) (Saltelli et al., 2000; Saltelli et al., 1999) are probably the most well-known methods for assessing the propagation of uncertainties through models, but there are many other methods such as the error propagation equations (EPE) (Bevington and Robinson, 2003; Mandel, 1984), the inverse modelling for predictive uncertainty (IN-UN) (Refsgaard et al., 2007), the response surface methodology or the Fourier amplitude sensitivity (Helton and Davis, 2003). Other types of quantitative uncertainties have their own approaches, intervals can be handled with interval arithmetic (Jaulin et al., 2001), and membership functions deals with fuzzy logic (Zadeh, 1978).

The assessment of parameter uncertainty can be performed through inverse modelling for the parameter estimation (IN-PA) (Banks and Bihari, 2001; Madsen, 2003). On the other hand, multiple model simulation (MMS) (Asselt, 2000; Refsgaard et al., 2006) is used to estimate of the uncertainty related to model structure.

Some tools such as the data uncertainty engine (DUE) (Brown and Heuvelink, 2007) integrate multiple methods into a single software environment; among others functionalities, DUE supports the assessment of uncertainty for numerical, spatial and temporal data, integrates some forms of expert judgement and allows propagating uncertainties through Monte-Carlo simulation.

When quantitative uncertainties are not available: Expert elicitation (EE) (Keeney and Vonwinterfeldt, 1991; Spetzler and Staelvonholstein, 1975) provides a solution to translate the subjective judgement from experts in "*a 'subjective' probability distribution function (PDF) reflecting the expert's degree of belief*" (Refsgaard et al., 2007). Scenario analysis (SC) (Alcamo, 2001; Van der Heijden, 2005) aims to tackle the lack of knowledge by creating plausible future scenarios whose implications are assessed quantitatively.

To incorporate qualitative information about the data validity, the NUSAP system (Refsgaard et al., 2006; van der Sluijs et al., 2005) suggests the use of pedigree matrices (PM) (Funtowicz and Ravetz, 1990). In this case, the term pedigree does not reflect the origin of data, but corresponds to an evaluative description of the mode of data production. The columns of the pedigree matrix represent the quality criteria of the "data production" and the rows characterise some grades to assess the criteria. Within the framework of the NUSAP system[15], several pedigree matrices have been created to assess the model structure (Refsgaard et al., 2006), the model parameters (van der Sluijs et al., 2001) and the data from measuring instrument (van der Sluijs et al., 2005). Chapter 4.4.3 highlights how pedigree matrices are handled in the CWIS database.

Selection based on the source and type of uncertainty

The uncertainty matrix provides a way to identify and characterise the uncertainty sources associated to data provenance. These uncertainties then require to be assessed and if possible propagate to the final data (which is the outcome of the provenance workflow). Table 6 shows an adaptation of the correspondences between methods and the uncertainty dimensions provided by Refsgaard et al. (2007). For each source and type of uncertainty, the table suggests some methods to be applied. The last row specifies which methods can be applied to propagate the uncertainty to the outcome of the process. The signification of the acronyms is exposed below the table. The choice of the appropriate method is up to the analyst and may depend on other considerations such as the uncertainty nature or the type of process.

[15] Website: http://www.nusap.net/

Table 6 Correspondence of the methods with the sources and types of uncertainty, adapted from (Refsgaard et al., 2007; van der Sluijs et al., 2004)

Source of uncertainty (locations)		Types of uncertainties (levels)			
		Statistical uncertainty	**Qualitative uncertainty**	**Scenario uncertainty**	**Recognised ignorance**
Context and framing	Natural, technological, economic, social, political	EE	EE, EPR, NUSAP, SI, UM, PM	EE, SC, SI	EE, EPR, PM, SI
Data	System data and driving forces	DUE, EPE, EE, QA	DUE, EE, EPR, PM	DUE, EE, SC, QA	DUE, EE. EPR
Process structure	Model structure	EE, MMS, QA	EE, PM, QA	EE, MMS, SC, QA	EA, PM, QA
Process implementation	Software and hardware implementation	QA, SA	QA, SA, PM	QA, SA	QA
Process parameters		IN-PA, QA	IN-PA, QA, PM	IN-PA, QA	QA
Process output uncertainty (via propagation)		EPE, EE, IN-UN, MCA, MMS, SA	EE, PM	EE, IN-UN, MMS, SA	EE, PM

Abbreviations and references of methodologies: **DUE**, data uncertainty engine (Brown and Heuvelink, 2007); **EPE**, error propagation equations (Bevington and Robinson, 2003; Mandel, 1984); **EE**, expert elicitation (Keeney and Vonwinterfeldt, 1991; Spetzler and Staelvonholstein, 1975); **EPR**, extended peer review (review by stakeholders) (Grimble and Wellard, 1997; Refsgaard et al., 2007); **IN-PA**, inverse modelling (parameter estimation) (Banks and Bihari, 2001; Madsen, 2003); **IN-UN**, inverse modelling (predictive uncertainty) (Refsgaard et al., 2007); **MCA**, Monte Carlo analysis (EPA, 1997); MMS, multiple model simulation (Asselt, 2000; Refsgaard et al., 2006); **PM**, Pedigree matrix (Funtowicz and Ravetz, 1990; van der Sluijs et al., 2005); **QA**, quality assurance (Refsgaard et al., 2005a; Scholten et al., 2007); **SC**, scenario analysis (Alcamo, 2001; Van der Heijden, 2005); **SA**, sensitivity analysis (Saltelli et al., 2000; Saltelli et al., 1999); **SI**, stakeholder involvement (Kloprogge and Van Der Sluijs, 2006).

Appendix C

C.1 Taxonomy of information types

List of basic information classes contained in the ontology

- **Boolean**
- **File**
  - ASCII
  - Audio
  - Compressed
  - Executable
  - Image
  - Map
  - PDF
  - Video
  - XLS
- **Geometry**
  - Raster Geometry
  - Vector Geometry
- **Text**
  - Address
  - Assessment
  - Comment
  - E-mail
  - Formula
  - Information
  - Problem
  - Request
  - Scenario
  - Strategy
  - Vision
  - Website
- **Numeric Value**
  - Coefficient, constant
    - Fluid conductance
    - Manning coefficient roughness
    - Infiltration rate
    - Hydraulic conductivity
    - Weir coefficient
    - Permeability rate
    - Aquifer transmitivity
  - Custom indicators
    - CWB Indicators (model city water balance)
    - CWE Indicators (economic model)
    - LA Indicators (learning alliance)
  - Quantity
    - Amount of substance
    - Area
    - Electric current
    - Energy
    - Force
    - Length
    - Luminous flux
    - Mass
    - Monetary value
    - People
    - Plane angle
    - Power, radiant flux
    - Pressure, stress
    - Temperature
    - Time
    - Voltage
    - Volume
  - Quantity per area
    - Area ratio
    - Energy per area
    - Mass per area
    - Monetary value per area
    - People per area
    - Power per area
    - Things per area
    - Volume per area
  - Quantity per capita
    - Area per capita
    - Energy per capita
    - Length per capita
    - Mass per capita
    - Monetary value per capita
    - People ratio
    - Things per capita
    - Time per capita
    - Volume per capita
  - Quantity per energy unit

- Area per energy unit
- Distance per energy unit
- Energy ratio
- Mass per energy unit
- Monetary value per energy unit
- People per energy unit
- Things per energy unit
- Volume per energy unit

Quantity per length
- Energy per length
- Length ratio
- Mass per length
- Monetary value per length
- People per length
- Things per length

Quantity per mass unit
- Energy per mass unit
- Mass ratio
- Monetary value per mass unit
- People per mass unit
- Things per mass unit
- Volume per mass unit

Quantity per monetary unit
- Area per monetary unit
- Energy per monetary unit
- Length per monetary unit
- Mass per monetary unit
- Monetary ratio
- People per monetary unit
- Things per monetary unit
- Volume per monetary unit

Quantity per thing
- Area per thing
- Distance per thing
- Energy per thing
- Mass per thing
- Monetary value per thing
- Number per thing
- Power per thing
- Proportion of things
- Time per thing
- Volume per thing

Quantity per time
- Area per time
- Energy per time
- Frequency
- Length per time
- Mass per time
- Monetary value per time
- People per time
- Rate of change per time period
- Things per time
- Time ratio
- Volume flow rate
- Volume per area per time

Quantity per volume
- Energy per volume
- Mass per volume
- Monetary value per volume
- People per volume
- Things per volume
- Volume ratio

Standards
- MDG Indicators
- WFD Indicators

Bibliography

Agrawal, P., Benjelloun, O., Sarma, A.D., Hayworth, C., et al., 2006. Trio: a system for data, uncertainty, and lineage, Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment, Seoul, Korea, pp. 1151-1154.

Alcamo, J., 2001. Scenarios as tools for international environmental assessments. Copenhagen.

Altintas, I., Barney, O., Jaeger-Frank, E., 2006. Provenance collection support in the Kepler Scientific Workflow System. Lect Notes Comput Sc 4145, 118-132.

Asselt, M.B.A.v., 2000. Perspectives on uncertainty and risk : the PRIMA approach to decision support. Kluwer Academic Publishers, Boston xvi, 434 p. pp.

Banks, H.T., Bihari, K.L., 2001. Modelling and estimating uncertainty in parameter estimation. Inverse Probl 17, 95-111.

Batchelor, C., Butterworth, J.A., 2008a. Learning Alliance Briefing Note 9: Visioning. SWITCH Project briefing note series. http://switchurbanwater.lboro.ac.uk/outputs/pdfs/WP6-2_BRN_9_Visioning_draft.pdf (accessed 6 February 2011)

Batchelor, C., Butterworth, J.A., 2008b. Learning Alliance Briefing Note 11: Scenario Building. SWITCH Project briefing note series. http://switchurbanwater.lboro.ac.uk/outputs/pdfs/WP6-2_BRN_11_Scenario_building.pdf (accessed 6 February 2011)

Batchelor, C., Butterworth, J.A., 2008c. Learning Alliance Briefing Note 12: Strategy Development. SWITCH Project briefing note series. http://switchurbanwater.lboro.ac.uk/outputs/pdfs/WP6-2_BRN_12_Strategy_Development.pdf (accessed 6 February 2011)

Bedard, Y., 2005. Perceptory 2006, a conceptual modeling tool for geospatial databases. http://sirs.scg.ulaval.ca/perceptory/english/enewindex2.asp (accessed 30 January 2008)

Bedard, Y., Larrivee, S., Proulx, M.J., Nadeau, M., 2004. Modeling geospatial databases with plug-ins for visual languages: A pragmatic approach and the impacts of 16 years of research and experimentations on perceptory, Conceptual Modeling for Advanced Application Domains, vol. 3289, pp. 17-30.

Bellamy, J.A., Walker, D.H., McDonald, G.T., Syme, G.J., 2001. A systems approach to the evaluation of natural resource management initiatives. Journal of Environmental Management 63, 407--423.

Benjelloun, O., Sarma, A.D., Hallevy, A., Theobald, M., et al., 2008. Databases with uncertainty and lineage. Vldb J. 17, 243-264.

Bertalanffy, L.v., 1973. General system theory : foundations, development, applications. Penguin, Harmondsworth xxii, 311 p. pp.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320, 18-36.

Bevington, P.R., Robinson, D.K., 2003. Data reduction and error analysis for the physical sciences, 3rd ed. McGraw-Hill, Boston xi, 320 p. pp.

Bosch, O.J.H., King, C.A., Herbohn, J.L., Russell, I.W., et al., 2007. Getting the big picture in natural resource management - Systems thinking as 'method' for scientists, policy makers and other stakeholders. Systems Research and Behavioral Science 24, 217-232.

Bose, R., Frew, J., 2005. Lineage retrieval for scientific data processing: A survey. ACM computing surveys 37, 1-28.

Brandenberg, P., 2008. Analyse de la gestion de l’eau par l’application d’un système d’information du système : le cas de Belo Horizonte Brésil, Master Thesis, Laboratory of Ecohydrology (ECHO). EPFL, Lausanne.

Brown, J.D., 2004a. Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. T I Brit Geogr 29, 367-381.

Brown, J.D., Heuvelink, G.B.M., 2007. The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. Computers & Geosciences 33, 172-190.

Brown, M.T., 2004b. A picture is worth a thousand words: energy systems language and simulation. Ecological Modelling 178, 83-100.

Buneman, P., Khanna, S., Tan, W.C., 2001. Why and where: A characterization of data provenance. Database Theory - Icdt 2001, Proceedings 1973, 316-330.

Bunge, M., 1977. Treatise On Basic Philosophy: Volume 3: Ontology I: The Furniture of the World.

Bunge, M., 1979. Treatise on Basic Philosophy, Volume 4. Ontology II A World of Systems.

Butterworth, J.A., Batchelor, C., Moriarty, P., Schouten, T., et al., 2009. Building more effective partnerships for innovation in urban water management. Crc Press-Taylor & Francis Group, Boca Raton 557-565 pp.

Butterworth, J.A., Dasilva, C., 2008. Learning Alliance Briefing Note 7: A framework for monitoring and evaluating project outcomes at city level. SWITCH Project briefing note series. http://switchurbanwater.lboro.ac.uk/outputs/pdfs/WP6-2_BRN_7_M_and_E.pdf (accessed 6 February 2011)

Cabrera, D., Colosi, L., Lobdell, C., 2008. Systems thinking. Eval. Program Plan. 31, 299-310.

Ceccaroni, L., Cortés, U., Sànchez-Marrè, M., 2004. OntoWEDSS: augmenting environmental decision-support systems with ontologies. Environmental Modelling & Software 19, 785-797.

CEDARE, 2007. SWITCH Stakeholders Analysis Report City of Alexandria, Egypt. http://switch.cedare.int/cedare.int/files28%5CFile2826.pdf (accessed 10 February 2011)

CEDARE, 2009a. Alexandria storm water management study 1. Alexandria, Egypt. Strategic studies. http://switch.cedare.int/cedare.int/files28/File2991.pdf (accessed 10 February 2011)

CEDARE, 2009b. Alexandria storm water management study 2. Alexandria, Egypt. Strategic studies. http://switch.cedare.int/cedare.int/files28/File2992.pdf (accessed 10 February 2011)

CEDARE, 2009c. Assessment of groundwater potential in Alexandria Governorate. Alexandria, Egypt. Strategy studies. http://switch.cedare.int/cedare.int/files28/File2974.pdf (accessed 10 February 2011)

CEDARE, 2010a. Institutional mapping and water governance analysis in the city of Alexandria. Alexandria, Egypt. Strategy studies.

CEDARE, 2010b. Introducing and implementing desalination process in Alexandria City 2037. Alexandria, Egypt. Strategy studies. http://switch.cedare.int/cedare.int/files28/File2977.pdf (accessed 10 February 2011)

CEDARE, 2010c. Water demand management study. Alexandria, Egypt. Strategy studies. http://switch.cedare.int/cedare.int/files28/File2975.pdf (accessed 10 February 2011)

Checkland, P., 1981. Systems Thinking, Systems Practice. Wiley, Chichester.

Chisholm, R.M., 1996. A Realistic Theory of Categories: An Essay on Ontology Cambridge University Press, New York 146 pp.

Cimiano, P., Volker, J., 2005. Text2Onto - A framework for ontology learning and data-driven change discovery. Natural Language Processing and Information Systems, Proceedings 3513, 227-238.

Claeys, F., Pauw, D.J.W.D., Benedetti, L., Nopens, I., et al., 2006. Tornado: A versatile and efficient modelling & virtual experimentation kernel for water quality systems, in: Voinov, A., Jakeman, A.J., Rizzoli, A.E. (Eds.), iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modelling and Software Society, Burlington, USA.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., 2003. Methodologies, tools and languages for building ontologies. Where is their meeting point? Data & Knowledge Engineering 46, 41-64.

Costanza, R., Gottlieb, S., 1998. Modelling ecological and economic systems with STELLA: Part II. Ecological Modelling 112, 81-84.

David, O., 1997. A kernel approach for interactive-oriented model construction in Java. Concurrency: Practice and Experience 9, 1319-1326.

David, O., Markstrom, S.L., Rojas, K.W., Ahuja, L.R., et al., 2002. The object modeling system, Agricultural System Models in Field Research and Technology Transfer. Lewis Publishers Inc, Boca Raton, pp. 317-330.

de Rocquigny, E., Devictor, N., Tarantola, S., 2008. Uncertainty Settings and Natures of Uncertainty. John Wiley & Sons, Ltd 199-211 pp.

Dessimoz, J.-J., 2008. Integrated water resources management with a non-geographic information system in the city of Birmingham UK, Master Thesis, Laboratory of Ecohydrology (ECHO). EPFL, Lausanne.

Dewhurst, S.M., Kessler, W.B., 1999. Scenario planning - Wading into the real world. J Forest 97, 43-47.

Edwards, L.B., Jaros, G.G., 1994. Process-based systems thinking--Challenging the boundaries of structure. Journal of Social and Evolutionary Systems 17, 339-353.

Eker, J., Janneck, J.W., Lee, E.A., Liu, J., et al., 2003. Taming heterogeneity the Ptolemy approach. P Ieee 91, 127-144.

El-Sayed Mohamed Mahgoub, M., van der Steen, N.P., Abu-Zeid, K., Vairavamoorthy, K., 2010. Towards sustainability in urban water: a life cycle analysis of the urban water system of Alexandria City, Egypt. J. Clean Prod. 18, 1100-1106.

EPA, 1997. Guiding Principles for Monte Carlo Analysis. Washington, DC. Risk Assessment Forum. http://www.epa.gov/ncea/pdfs/montcarl.pdf

Foster, I., Vockler, J., Wilde, M., Zhao, Y., 2002. Chimera: A virtual data system for representing, querying, and automating data derivation. Ieee Computer Soc, Los Alamitos 37-46 pp.

Frew, J., Bose, R., 2001. Earth system science workbench: A data management infrastructure for earth science products, in: Kerschberg, L., Kafatos, M. (eds.), Thirteenth International Conference on Scientific and Statistical Database Management, Proceedings. Ieee Computer Soc, Los Alamitos, pp. 180-189.

Frew, J., Bose, R., 2002. Lineage issues for scientific data and information,, Workshop on Data Derivation and Provenance, Chicago.

Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Dordrecht: Kluwer., in: Publishers, K.A. (ed.), Dordrecht , NL.

Giampietro, M., Allen, T.F.H., Mayumi, K., 2006. The epistemological predicament associated with purposive quantitative analysis. Ecological Complexity 3, 307-327.

Gobel, C., 2002. Position statement: Musings on provenance, workflow and (semantic web) annotations for bioinformatics, Workshop on Data Derivation and Provenance, Chicago.

Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P., 2007. OpenMI: Open modelling interface. Journal of Hydroinformatics 9, 175-191.

Grimble, R., Wellard, K., 1997. Stakeholder methodologies in natural resource management: A review of principles, contexts, experiences and opportunities. Agricultural Systems 55, 173-193.

Groves, D.G., Lempert, R.J., 2007. A new analytic method for finding policy-relevant scenarios. Global Environmental Change 17, 73-85.

GWP, 2000. Integrated Water Resources Management. Global Water Partnership -Technical Advisory Committee, Stockholm, Sweden 67 pp.

Hajkowicz, S., Collins, K., 2007. A Review of Multiple Criteria Analysis for Water Resource Planning and Management. Water Resour. Manag. 21, 1553-1566.

Hajkowicz, S., Higgins, A., 2008. A comparison of multiple criteria analysis techniques for water resource management. European Journal of Operational Research 184, 255-265.

Hamilton, M.H., Hackler, W.R., 2008. Universal Systems Language: Lessons Learned from Apollo. Computer 41, 34-+.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety 81, 23-69.

Hillyer, C., Bolte, J., van Evert, F., Lamaker, A., 2003. The ModCom modular simulation system. European Journal of Agronomy 18, 333-343.

Huang, D.B., Bader, H.P., Scheidegger, R., Schertenleib, R., et al., 2007. Confronting limitations: New solutions required for urban water management in Kunming City. Journal of Environmental Management 84, 49-61.

ICWE, 1992. The Dublin Statement on Water and Sustainable Development. http://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html (accessed 9 January 2010)

Jaulin, L., Kieffer, M., Didrit, O., Walter, E. (eds.) 2001. Applied interval analysis, 1rst Edition ed. Springer 379 pp.

Jaweesh, M., 2010. City Water Balance: A new scoping model for Alexandria, MSc, School of earth science. University of Birmingham, Birmingham, p. 61.

Kay, J.J., Regier, H.A., Boyle, M., Francis, G., 1999. An ecosystem approach for sustainability: addressing the challenge of complexity. Futures 31, 721-742.

KBSI, 2010. IDEF5: Ontology Description Capture Method. http://www.idef.com/IDEF5.htm (accessed December 16)

Keeney, R.L., Vonwinterfeldt, D., 1991. Eliciting Probabilities from Experts in Complex Technical Problems. Ieee T Eng Manage 38, 191-201.

Kim, J.H., Deelman, E., Gil, Y., Mehta, G., et al., 2008. Provenance trails in the Wings/Pegasus systems. Concurr Comp-Pract E 20, 587-597.

Kira, M., van Eijnatten, F., 2008. Socially Sustainable Work Organizations: A Chaordic Systems Approach. Systems Research and Behavioral Science 25, 743-756.

Klasky, S., Barreto, R., Kahn, A., Parashar, M., et al., 2008. Collaborative visualization spaces for petascale simulations. Int S Collab Technol, 203-211.

Kloprogge, P., Van Der Sluijs, J.P., 2006. The inclusion of stakeholder knowledge and perspectives in integrated assessment of climate change. Climatic Change 75, 359-389.

Koestler, A., 1967. The ghost in the machine. Hutchinson, London.

Koestler, A., 1969. Beyond atomism and holism — the concept of the holon, in: Koestler, A., Smythies, J.R. (eds.), Beyond reductionism; new perspectives in the life sciences. Macmillan, New York, pp. 192–216.

Kogut, P., Cranefield, S., Hart, L., Dutra, M., et al., 2002. UML for ontology development. The knowledge engineering review 17, 61-64.

Kolkman, M.J., Kok, M., van der Veen, A., 2005. Mental model mapping as a new tool to analyse the use of information in decision-making in integrated water management. Phys. Chem. Earth 30, 317-332.

Kralisch, S., Krause., P., 2006. JAMS – A Framework for Natural Resource Model Development and Application, in: Voinov, A., Jakeman, A.J., Rizzoli, A.E. (Eds.), iEMSs

Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modelling and Software Society, Burlington, USA.

Lal, P., Lim-Applegate, H., Scoccimarro, M.C., 2001. The Adaptive Decision-Making Process as a Tool for Integrated Natural Resource Management: Focus, Attitudes, and Approach.

Leavesley, G.H., Markstrom, S.L., Brewer, M.S., Viger, R.J., 1996. The modular modeling system (MMS) — The physical process modeling component of a database-centered decision support system for water and power management. Water, Air, & Soil Pollution 90, 303-311.

Liberatore, S., Sechi, G.M., Zuddas, P., 2006. Non Linear Optimization Models in Water Resource Systems, Global Optimization, pp. 227-242.

Liu, Y.Q., Gupta, H., Springer, E., Wagener, T., 2008. Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. Environmental Modelling & Software 23, 846-858.

Ludascher, B., Altintas, I., Berkley, C., Higgins, D., et al., 2004. Scientific workflow management and the Kepler system, GGF Workshop on Workflow in Grid Systems. John Wiley & Sons Ltd, Berlin, GERMANY, pp. 1039-1065.

Lundie, S., Peters, G.M., Beavis, P.C., 2004. Life Cycle Assessment for sustainable metropolitan water systems planning. Environ. Sci. Technol. 38, 3465-3473.

Mackay, R., Last, E., 2010. SWITCH city water balance: a scoping model for integrated urban water management. Reviews in Environmental Science and Biotechnology 9, 291-296.

Madin, J., 2007. An ontology for describing and synthesizing ecological observation data. Ecological informatics 2, 279-296.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources 26, 205-216.

Mahmoud, M., Liu, Y.Q., Hartmann, H., Stewart, S., et al., 2009. A formal framework for scenario development in support of environmental decision-making. Environmental Modelling & Software 24, 798-808.

Maia, R., Schumann, A., 2007. DSS Application to the Development of Water Management Strategies in Ribeiras do Algarve River Basin. Water Resour. Manag. 21, 897-907.

Makropoulos, C.K., Memon, F.A., Shirley-Smith, C., Butler, D., 2008. Futures: an exploration of scenarios for sustainable urban water management. Water Policy 10, 345-373.

Mandel, J., 1984. The statistical analysis of experimental data. Dover, New York xi, 410 p. pp.

Matos, M.A., 2007. Decision under risk as a multicriteria problem. European Journal of Operational Research 181, 1516-1529.

Mattsson, S.E., Elmqvist, H., Otter, M., 1998. Physical system modeling with Modelica. Control Engineering Practice 6, 501-510.

Max-Neef, M.A., 2005. Foundations of transdisciplinarity. Ecological Economics 53, 5-16.

Maxwell, T., 1999. A paris-model approach to modular simulation. Environmental Modelling and Software 14, 511-517.

McDonnell, R.A., 2008. Challenges for integrated water resources management: How do we provide the knowledge to support truly integrated thinking?, Workshop on Integrated Water Resources Management in Latin America. Routledge Journals, Taylor & Francis Ltd, Rio de Janeiro, BRAZIL, pp. 131-143.

Medema, W., McIntosh, B.S., Jeffrey, P.J., 2008. From Premise to Practice: a Critical Assessment of Integrated Water Resources Management and Adaptive Management Approaches in the Water Sector. Ecol. Soc. 13, 18.

Mendoza, G.A., Prabhu, R., 2006. Participatory modeling and analysis for sustainable forest management: Overview of soft system dynamics models and applications. Forest Policy Econ. 9, 179-196.

Midgley, G., 1992. Pluralism and the legitimation of systems science. Syst. Pract. Action Res. 5, 147-172.

Mitchell, B., 2005. Integrated water resource management, institutional arrangements, and land-use planning. Environment and Planning A 37, 1335--1352.

Moreau, L., Lud\, B., \#228, scher, et al., 2008. Special Issue: The First Provenance Challenge. Concurr. Comput. : Pract. Exper. 20, 409-418.

Moriarty, P., 2007. Learning Alliance Briefing Note 4: A brief introduction to action research concepts and practice. SWITCH Project briefing note series. http://www.switchurbanwater.eu/outputs/pdfs/WP6-2_BRN_4_Action_research.pdf (accessed 14 February 2011)

Moriarty, P., Batchelor, C., Laban, P., 2005. The EMPOWERS Participatory Planning Cycle for Integrated Water Resource Management. Amman, Jordan. EMPOWERS Working Paper http://www.project.empowers.info/page/1070 (accessed 14 February 2011)

Morin, E., 1990. Introduction à la pensée complexe, Seuil ed. Editions du Seuil, Lonrai, France.

Morris, M., 2006. Learning Alliance Briefing Note 1: An introduction to learning alliances. SWITCH Project briefing note series. http://www.switchurbanwater.eu/outputs/pdfs/WP6-2_BRN_1_Intro_to_LAs.pdf (accessed 6 February 2011)

Muetzelfeldt, R., Massheder, J., 2003. The Simile visual modelling environment. European Journal of Agronomy 18, 345-358.

Mulej, M., Potocan, V., Zenko, Z., Kajzer, S., et al., 2004. How to restore Bertalanffian systems thinking. Kybernetes 33, 48-61.

Naveh, Z., 2000. What is holistic landscape ecology? A conceptual introduction. Landscape and Urban Planning 50, 7-26.

NWRP, 2005. National Water Resources Plan for Egypt - 2017, in: Irrigation, M.o.W.R.a. (Ed.), Cairo.

Odum, H.T., 1960. Ecological potential and analogue circuits for the ecosystem. American Scientist 48, 1-8.

Odum, H.T., Odum, E.C., 2000. Modeling for all scales: an introduction to system simulation, San Diego, California, USA 80 pp.

Oinn, T., Addis, M., Ferris, J., Marvin, D., et al., 2004. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20, 3045-3054.

OMG, (Object Management Group), 2009. Ontology Definition Metamodel. OMG Document formal/2009-05-01. http://www.omg.org/spec/ODM/1.0/ (accessed 13 December 2010)

Opdahl, A.L., Henderson-Sellers, B., 2001. Grounding the OML metamodel in ontology. Journal of Systems and Software 57, 119-143.

Opdahl, A.L., Henderson-Sellers, B., 2004. A template for defining enterprise modelling constructs. Journal of Database Management 15, 39-73.

Pahl-Wostl, C., 2007. The implications of complexity for integrated resources management. Environmental Modelling & Software 22, 561-569.

Pahl-Wostl, C., Sendzimir, J., Jeffrey, P., Aerts, J., et al., 2007. Managing change toward adaptive water management through social learning. Ecol. Soc. 12, 18.

Pancerella, C., Hewson, J., Koegler, W., Leahy, D., et al., 2003. Metadata in the collaboratory for multi-scale chemical science, Proceedings of the 2003 international conference on Dublin Core and metadata applications: supporting communities of discourse and practice---metadata research \& applications. Dublin Core Metadata Initiative, Seattle, Washington, pp. 1-9.

Paucar-Caceres, A., Rodriguez-Ulloa, R., 2007. An application of soft systems dynamics methodology (SSDM). J. Oper. Res. Soc. 58, 701-713.

Peterson, G.D., Cumming, G.S., Carpenter, S.R., 2003. Scenario planning: a tool for conservation in an uncertain world. Conserv Biol 17, 358-366.

Rahman, J.M., Seaton, S.P., Cuddy, S.M., 2004. Making frameworks more useable: using model introspection and metadata to develop model processing tools. Environmental Modelling & Software 19, 275-284.

Rahman, J.M., Seaton, S.P., Perraud, J.M., Hotham, H., et al., 2003. It's TIME for a new environmental modelling framework, International Congress on Modelling and Simulation. Univ Western Australia, Townsville, AUSTRALIA, pp. 1727-1732.

Ralston, B., Wilson, I., 2006. The scenario-planning handbook : a practitioner's guide to developing and using scenarios to direct strategy in today's uncertain times. Thomson/South-Western, Australia ; Mason, Ohio viii, 264 p. pp.

Rammel, C., Stagl, S., Wilfing, H., 2007. Managing complex adaptive systems - A co-evolutionary perspective on natural resource management. Ecological Economics 63, 9-21.

Reed, M., Cuddy, S.M., Rizzoli, A.E., 1999. A framework for modelling multiple resource management issues--an open modelling approach. Environmental Modelling and Software 14, 503-509.

Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholten, H., et al., 2005a. Quality assurance in model based water management - review of existing practice and outline of new approaches. Environmental Modelling & Software 20, 1201-1215.

Refsgaard, J.C., van der Sluijs, J.P., Brown, J., van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. Advances in Water Resources 29, 1586-1597.

Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process - A framework and guidance. Environmental Modelling & Software 22, 1543-1556.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2005b. Uncertainty Analysis. Harmoni-CA Guidance No. 1.

Rivera, E.C., de Queiroz, J.F., Ferraz, J.M., Ortega, E., 2007. Systems models to evaluate eutrophication in the Broa Reservoir, São Carlos, Brazil. Ecological Modelling 202, 518-526.

Rizzoli, A.E., Davis, J.R., Abel, D.J., 1998. Model and data integration and re-use in environmental decision support systems. Decision Support Systems 24, 127-144.

Rizzoli, A.E., Donatelli, M., Athanasiadis, I.N., Villa, F., et al., 2005. Semantic links in integrated modelling frameworks, International Congress on Modelling and Simulation (MODSIM05), Melbourne, AUSTRALIA, pp. 412-423.

Rizzoli, A.E., Leavesley, G., Ascough, J.C., Argent, R.M., et al., 2008. Integrated modelling frameworks for environmental assessment and decision support., in: Jakeman, A.J., Voinov, A.A., Rizzoli, A.E., Chen, S.H. (eds.), Environmental Modelling, Software and Decision Support. Elsevier, pp. 101-118.

Rumbaugh, J., Jacobson, I., Booch, G., 2005. The unified modeling language reference manual, 2nd ed. Addison-Wesley, Boston xx, 721 p. pp.

Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity analysis as an ingredient of modeling. Stat Sci 15, 377-395.

Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41, 39-56.

Saravanan, V.S., 2008. A systems approach to unravel complex water management institutions. Ecological Complexity 5, 202-215.

Schenck, D., Wilson, P.R., 1994. Information modeling : the EXPRESS way. Oxford University Press, New York xxviii, 388 p. pp.

Schenk, C., 2010. A Systems-Based Generic Decision Support System Application to Urban Water Management, Doctoral Thesis, School of Architecture, Civil and Environmental Engineering. EPFL, Lausanne, p. 179

Schenk, C., Roquier, B., Soutter, M., Mermoud, A., 2009. A system model for water management. Environmental Management 43, 458-469.

Scholten, H., Kassahun, A., Refsgaard, J.C., Kargas, T., et al., 2007. A methodology to support multidisciplinary model-based water management. Environmental Modelling & Software 22, 743-759.

Schouten, T., 2007. Learning Alliance Briefing Note 6: Process Documentation SWITCH Project briefing note series. http://switchurbanwater.lboro.ac.uk/outputs/pdfs/WP6-2_BRN_6_Process_documentation.pdf (accessed 6 February 2011)

Scott, M., 2007. Quantifying uncertainty in multicriteria concept selection methods. Research in Engineering Design 17, 175-187.

Simmhan, Y.L., Plale, B., Gannon, D., 2005. A survey of data provenance in e-science. Sigmod Record 34, 31-36.

Simon, H.A., 1956. Rational Choice and the Structure of the Environment. Psychol Rev 63, 129-138.

Simon, H.A., 1982. Models of bounded rationality. MIT Press, Cambridge, Mass. v. <1-3 > pp.

Skyttner, L., 2005. General systems theory : problems, perspectives, practice, 2nd ed. World Scientific, Hackensack, NJ x, 524 p. pp.

Smith, C., Felderhof, L., Bosch, O.J.H., 2007. Adaptive management: Making it happen through participatory systems analysis. Systems Research and Behavioral Science 24, 567-587.

Smith, M.K., Welty, C., McGuinness, D.L., 2004. OWL Web Ontology Language Guide. http://www.w3.org/TR/owl-guide/ (accessed January 22, 2010)

Spetzler, C.S., Staelvonholstein, C.A.S., 1975. Probability Encoding in Decision Analysis. Management Science 22, 340-358.

Srdjevic, B., Medeiros, Y.D.P., Faria, A.S., 2004. An Objective Multi-Criteria Evaluation of Water Management Scenarios. Water Resour. Manag. 18, 35-54.

Sydelko, P.J., Hlohowskyj, I., Majerus, K., Christiansen, J., et al., 2001. An object-oriented framework for dynamic ecosystem modeling: application for integrated risk assessment. Science of the Total Environment 274, 271-281.

Tan, K.C., Lee, T.H., Khor, E.F., 2002. Evolutionary Algorithms for Multi-Objective Optimization: Performance Assessments and Comparisons. Artificial Intelligence Review 17, 251-290.

The MathWorks Inc., 2009. MATLAB - The Language Of Technical Computing. http://www.mathworks.com/products/matlab/ (accessed 11/17/2009)

Tudorache, T., 2006. Employing Ontologies for an Improved Development Process in Collaborative Engineering, Dipl.–Ing, Institut für Telekommunikationssysteme Technische Universität Berlin, p. 171.

Van der Heijden, K., 2005. Scenarios : the art of strategic conversation, 2nd ed. John Wiley & Sons, Chichester, West Sussex ; Hoboken, N.J. xxiv, 356 p. pp.

van der Sluijs, J., Janssen, P.H.M., Petersen, A.C., Kloprogge, P., et al., 2004. Tool Catalogue for Uncertainty Assessment. 90-393-3797-7, Bilthoven (Utrecht, Netherlands). RIVM/MNP Guidance for Uncertainty Assessment and Communication. http://www.nusap.net/downloads/toolcatalogue.pdf

van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., et al., 2005. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. Risk Analysis 25, 481-492.

van der Sluijs, J.P., Potting, J., Risbey, J., van Vuuren, D., et al., 2001. Uncertainty assessment of the IMAGE/TIMER B1 CO2 emissions scenario, using the NUSAP method. Bilthoven.

Villa, F., 2001. Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. Ecological Modelling 137, 23-42.

Villa, F., Athanasiadis, I.N., Rizzoli, A.E., 2009. Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. Environmental Modelling & Software 24, 577-587.

Voinov, A., Costanza, R., Wainger, L., Boumans, R., et al., 1999. Patuxent landscape model: integrated ecological economic modeling of a watershed. Environmental Modelling and Software 14, 473-491.

Voinov, A., Fitz, C., Boumans, R., Costanza, R., 2004. Modular ecosystem modeling. Environmental Modelling & Software 19, 285-304.

W3C, 2009. OWL 2 Web Ontology Language: Document Overview. http://www.w3.org/TR/owl2-overview/ (accessed 16 December 2010)

Walker, W.E., Harremoes, P., Rotmans, J., van der Sluijs, J.P., et al., 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment 4, 5 - 17.

Wand, Y., Weber, R., 1988. An ontological analysis of some fundamental information system concepts, in Proceedings of the International Conference on Information Systems, pp. 213--225.

Wand, Y., Weber, R., 1990. An Ontological Model of an Information System. IEEE Trans. Softw. Eng. 16, 1282-1292.

Watson, F.G.R., Rahman, J.M., 2004. Tarsier: a practical software framework for model development, testing and deployment. Environmental Modelling & Software 19, 245-260.

Weidema, B.P., Wesnæs, M.S., 1996. Data quality management for life cycle inventories--an example of using data quality indicators. J. Clean Prod. 4, 167-174.

Willard, B., 2007. UML for systems engineering. Computer Standards & Interfaces 29, 69-81.

Winz, I., Brierley, G., Trowsdale, S., 2009. The Use of System Dynamics Simulation in Water Resources Management. Water Resour. Manag. 23, 1301-1323.

Witten, I.H., Bainbridge, D., Nichols, D.M., 2010. How to build a digital library, 2nd ed. Morgan Kaufmann Publishers, Amsterdam ; Boston xxiii, 629 p. pp.

Wolfram Research Inc., 2009. Mathematica. http://www.wolfram.com/products/mathematica/ (accessed 17 November 2009)

Xevi, E., Khan, S., 2005. A multi-objective optimisation approach to water management. Journal of Environmental Management 77, 269-277.

Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3-28.

Zhao, J., Wroe, C., Goble, C., Stevens, R., et al., 2004. Using semantic web technologies for representing e-Science provenance, in: McIlraith, S.A., Plexousakis, D., VanHarmelen, F. (eds.), Semantic Web - Iswc 2004, Proceedings, vol. 3298. Springer-Verlag Berlin, Berlin, pp. 92-106.

Zitzler, E., Künzli, S., 2004. Indicator-Based Selection in Multiobjective Search, Parallel Problem Solving from Nature - PPSN VIII, pp. 832-842.

Bastien Roquier

Date de naissance: 18 août 1982

Profession: Ingénieur en environnement

Nationalité: Suisse

Email: broquier@gmail.com

Etat civil: marié

Profil:

Ingénieur spécialisé dans le domaine des écobilans et de la gestion intégrée des ressources naturelles.

Intérêt pour les approches interdisciplinaire et participatives, avec un accent particulier pour le domaine de la gestion de l'eau.

Autonomie, esprit d'analyse, excellente capacité d'adaptation et de travail en équipe.

Cherche à s'engager dans des projets de « recherche–action » visant à améliorer les collaborations entre scientifiques, praticiens et décideurs.

Langues:

Français (langue maternelle), Anglais (courant), Allemand (intermédiaire)

Formation:

2007-2011 Doctorat en science de l'environnement, Laboratoire ECHO, EPFL
- Sujet de thèse: Gestion intégrée des ressources naturelles (développement d'outils pour le partage de l'information et l'aide à la décision)

2000-2005 Diplôme d'ingénieur en environnement, EPFL
- Sujet de diplôme: Analyse du cycle de vie (ACV) environnementale et sociale appliquée à l'industrie du verre en Inde.
- Prix EPFL pour le Développement Durable (prix BG Ingénieurs-conseils)

2000 Maturité fédérale scientifique (type C), Gymnase cantonal de La Chaux-de-Fonds

Expérience d'enseignement:

2007-2010 **Laboratoire d'Ecohydrologie (ECHO), EPFL**

2008-2010 *Chargé de cours* – Quartiers urbains durables, partie gestion de l'eau

2008-2010 *Assistant* – Cours de gestion intégrée des eaux

2007-2010 *Assistant* – Travaux pratiques d'hydrologie

Expérience professionnelle:

01.06-03.07 **Unité de Développement Durable, Etat de Vaud, Lausanne**

01.07 - 03.07 Responsable de projets

Mise en place du système de gestion EcoEntreprise® au Département des Infrastructures (Etat de Vaud) et accompagnement à la certification EcoEntreprise®.
Intégration de critères environnementaux dans les appels d'offre pour les achats de véhicules.

01.06 - 07.06 Service civil

Conception de fiches pour la sensibilisation au « Développement durable au travail ».
Participation au développement d'outils d'évaluation de projets selon des critères de durabilité.

07.06-11.06 **Resource Optimization Initiative (ROI), Bangalore, Inde**

Spécialiste ACV (Ecobilan)

Etude d'impact environnemental et social de la production de biocombustibles à partir des résidus agricoles et évaluation des risques pour la santé humaine associés à l'usage des combustibles domestiques.

Compétences informatiques:

Logiciels (Matlab, MicroStation, AutoCAD, MapInfo, ArcGIS, Simapro, Kepler, OpenMI), base de données (PostgreSQL, Oracle, Microsoft SQL Server) et programmation (.Net, Visual Basic, C/C++, C#)

Publications:

Roquier, B., C. Schenk, M. Soutter and A. Mermoud (2011) SYSMOD: A systems modelling language for environmental information. (In revision, submitted to Ecological Modelling).

Schenk, C., B. Roquier, M. Soutter and A. Mermoud. (2009) A system model for water management. Environmental Management 43 (3): 458-469.

Conférences:

Roquier, B. and M. Soutter. A Systems-Based Generic Suite of Tools to Support Information and Knowledge Sharing. Joint SWITCH/UNESCO-IHP Conference, Paris, 24-26 January 2011.

Roquier, B., P. Brandenberg, C. Schenk and M. Soutter. City Water: an information sharing platform to support LAs in exploring new strategies. 3rd SWITCH Scientific Meeting, Belo Horizonte, Brazil, 2008.

Schenk, C., M. Soutter, B. Roquier and A. Mermoud. Towards an information system on the water system. 2nd SWITCH Scientific Meeting, Tel-Aviv, Israel, 2007.

Haye, S. and B. Roquier. Assessing the environmental and socio-economic impacts of biocombustibles: the use of agricultural residues as household fuel in rural India. Scientific Workshop on Industrial Ecology, University of Lausanne, Switzerland 2006