Collaborative and Distributed E-Research:

Innovations in Technologies, Strategies and Applications

Angel A. Juan
IN3 - Open University of Catalonia, Spain

Thanasis Daradoumis
University of the Aegean, Greece, & Open University of Catalonia, Spain

Meritxell Roca
IN3 - Open University of Catalonia, Spain

Scott E. Grasman
Rochester Institute of Technology, USA

Javier Faulin
Public University of Navarre, Spain

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Hannah Abelbeck
Acquisitions Editor: Erika Gallagher
Typesetter: Russell A. Spangler
Cover Design: Nick Newcomer, Lisandro Gonzalez

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Collaborative and distributed e-research: innovations in technologies, strategies, and applications / Angel A. Juan ... [et al.], editors.
p. cm.
Includes bibliographical references and index.
Summary: "This book offers insight into practical and methodological issues related to collaborative e-research and furthers readers understanding of current and future trends in online research and the types of technologies involved"--Provided by publisher.
ISBN 978-1-4666-0125-3 (hardcover) -- ISBN 978-1-4666-0127-7 (print & perpetual access) 1. Internet research. 2. Group work in research. I. Juan, Angel A., 1972-
ZA4228.C65 2012
001.4'202854678--dc23
2011039614

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 8
An Ontological Structure for Gathering and Sharing Knowledge among Scientists through Experiment Modeling

Luis Casillas
University of Guadalajara, Mexico

Thanasis Daradoumis
University of the Aegean, Greece & Open University of Catalonia, Spain

ABSTRACT

This chapter presents a proposal for modeling / simulating experiments conducted by scientists working in common scientific problems, based on gathering and exploiting knowledge elements produced among them. The authors' approach enables the adaptation of knowledge structures (bounded to scientific problems) and is based on recurrent refining processes that are fed by indicators, which come from collaboration among the scientists involved. This scheme captures a web-based infrastructure, which allows scientists to collaborate on synthesizing experiments online. The proposed model is approached as an ontology that contains scientific concepts and actions. This ontology is linked to the scientific problem and represents both the "common understanding" for such a problem and the way it could be managed by the group. This dynamic ontology will change its structure according to the collaboration acts among scientists. Frequent collaboration over certain elements of the experiment will make them prevail in time. Besides, this process has been defined in a way that provides a global understanding of the scientific treatment that could be applied on any scientific problem. Hence, the ontology represents a virtualization of the scientific experiment. This whole representation is aimed at providing the media for developing e-research among scientists that are working on common problems.

DOI: 10.4018/978-1-4666-0125-3.ch008

INTRODUCTION

Most of the activities developed by humans are based on sets of concepts and the relationships among such knowledge elements. The human's understanding of the environment is defined by the concepts and ideas acquired before. These elements are organized as nets of perceptions in which the relationships are modeled by links of proximity. The nets of concepts provide meaning by clustering related ideas. These meaningful structures could be understood as ontologies.

Our proposal aims at defining an innovative approach for the simulation of experiments through a formal process of gathering and sharing the knowledge from activities related to research, which results in defining an ontological structure. This ontology has dynamic capabilities in order to allow the addition of new slots for knowledge categories, concepts and relationships. The elements of the ontology can be shared among the researchers on common projects, with the purpose of establishing a *common understanding* of a scientific challenge and the topics to be observed and managed along the experimentation experience.

Scientists collaborating on the same scientific problem could modify the already defined structure, and such changes are allowed for all the collaborators. Researchers do not need to vote expressly for concepts (topics) or relationships, since the prolonged use of such elements will imply their confirmation, and the forgotten elements will eventually disappear.

Different tools could be used to build an artificial ontology with standardized format. The resulting construction could be shared among the scientists involved in the same research. We are proposing an abstract tool that may be fed by this model, where the concepts as well as their relationships are immediately shown.

This proposal is, by some means, framed by the regular understanding for e-science suggested by Atkins et al. (2003) and Jankowski (2007). We are trying to fulfill the remaining tasks in the simulation of experiments. In most of the cases, e-science is committed to offer information resources and is frequently focused on social sciences. The spirit of the e-science proposals is the capture of the web-based infrastructure. Thus, our proposal might be rather settled in the way of e-science experimentation, as explained by Walton and Barker (2004). The advance in information and communications technologies has enabled an innovative understanding for the web and its new capabilities allow an enhanced simulation of natural and/or technical phenomena.

KNOWLEDGE GRIDS AS ONTOLOGIES

It is a fact that knowledge has become an asset for most of the organizations, which are conscious about the resources of awareness and the ways to manage them. Knowledge, by itself, is hard to define and its handling is rather difficult. In such context, any knowledge managing technique acquires some attention from people related to the creation, storing and handling of knowledge. One of the most frequent mechanisms used to manage knowledge is grids (Li & Liu, 2007; Goble, et al., 2005; Zettsu, et al., 2008).

Every concept has its own charge of knowledge, which is a piece of information or even a piece of primitive knowledge. In order to represent a higher level of meaning, concepts can be assembled as nets. Along the process of assembling the nets of knowledge, the semantics bound to the nets becomes complex. Thus the action of assembling is important for meaning, although meaning does not depend on the assembling by itself, it depends on the concepts involved and the kind of relationships established among these concepts. Not every connection of concepts will imply meaningful structures. The connections should be rationally founded in the common understanding of reality.

The human brain is always trying to relate the stimuli, arriving from the environment, to the elements already stored by the previous experiences. The brain's goal is to produce meaning to the current experience. The network of concepts, made during this linking process in the brain, is able to produce meaning. Artificial grids of knowledge could produce the very same support for dealing with synthetic concepts.

According to the Stanford Encyclopedia of Philosophy[1], *"ontology is the study of what there is."* The classical approach of ontology is bound with the mere existence of things. More recently, the concept of ontology has been associated to the capacity of machines to *synthetically* represent entities from the real world. Hence, concepts from reality could eventually be represented throughout ontological representations in machines.

Grids of knowledge can be understood as ontology. A synthetic ontology, from this perspective, can hold a family of concepts related to the understanding about some field of knowledge. These concepts are bound among themselves to form a grid.

RELATED WORK

The e-science seeks to provide order and support for the recent transformations in the science endeavors. In the current context of advanced technologies for information and communications, there is an infrastructure that undergirds modern societies. Science is undoubtedly seizing such infrastructure (Jankowski, 2007).

The report from Atkins et al. (2003) is highly focused in the exploitation of the available cyber-infrastructure. According to them, there is a revolution in science due to advances in information and communication technologies. This revolution is based in the innovative capabilities to successfully **emulate reality in the digital dimension**. Specifically, these authors argue *"...the classic two approaches to scientific research, theoretical / analytical and experimental / observational, have been extended to* ***in silico*** *simulation and modeling to explore new possibilities and to achieve new precision..."* (p. 4); with important achievements in Forestry, Ocean Science, Environmental Science, and Engineering, Space Weather, Computer Science and Engineering, Information Science and Digital Libraries, Biology / Bioinformatics, Medicine, Physics, Astronomy, Engineering, Materials Science and Engineering, and Social and Behavioral Sciences.

For Jankowski (2007) e-science is strongly supported by key aspects of the information and communication technologies, which are: *"1. International collaboration among researchers; 2. Increasing use of high-speed interconnected computers, applying Grid architecture; 3. Visualization of data; 4. Development of Internet-based tools and procedures; 5. Construction of virtual organizational structures for conducting research; 6. Electronic distribution and publication of findings."*

Besides, Jankowski (2007) has identified specific fields in the e-science activity: *"1. Managing collaboration and communication among researchers separated by distance; 2. Developing and using Internet-based tools for data collection, analysis, and visualization of findings; 3. Archiving and providing access to data; 4. Publishing and disseminating results."*

In this context, there is a clear need of mechanisms to successfully represent diverse phenomena in the digital dimension. Since 1960's, computers had been used formally in the scientific field. The tasks that used to be assigned to computers were mainly connected to calculations in iterative, recurrent and concurrent contexts. More recently, the advances in computers' graphics allowed improved simulations. Nowadays, computers are not the actor using "brute force" anymore. Of course, they will do the calculations needed; however, the main goals have changed. The goal for computers in this context is to support the collaboration among scientists, in the already mentioned fields for e-science.

According to Anandarajan and Anandarajan (2010), scientists have worked for the last decades under a quite regular fashion: isolated groups of researchers developing solutions (*ideas, projects, and inventions*), based on common understanding for a problem and its solution. For these editors, the advent of the Web technologies paraphernalia has established a new comprehension for collaboration among researchers: social networks and collaboration tools have enabled the knowledge sharing. In this context, they argue: *"…currently, tens of thousands of researchers are using research networks, ushering in a new paradigm for research. In this paradigm, collaboration is made much easier, and sharing of research knowledge is instant. Synergies from routine collaboration will yield huge advances in research productivity and innovation…"* (p. v).

One of the mechanisms that could support such forms of collaboration is the knowledge representation through formal structures. If it is admitted that there is knowledge underlying in the scientific problems, knowledge representation achieves forcefulness. The act of "sharing knowledge" would narrow under a standardized scheme, making available an improved capability to model advanced concepts of the analyzed problem.

During the last decade the ontology-based approach has become a goal in diverse fields related with knowledge handling. Human interaction through artificial channels requires meta-structures for representing and handling the knowledge involved. Unfortunately the practice of the ontology's design and its managing imply the resolution of complex challenges. Ontology mapping has a range of options to be modeled (Kalfoglou & Schorlemmer, 2003).

Newman et al. (2009) have proposed a complete framework for: *"1. Facilitate management and sharing of Research Objects (ROs); 2. Support a social model; 3. Provide an open extensible environment; 4. Provide a platform to action research."*

Their project (*myExperiment*), is a model that has been influenced by these four capabilities in order to provide the support to the scientists' interaction. *myExperiment* is based on a Resource Description Framework (RDF), this description is aimed at sharing definitions of experiments. These specifications include different components, such as contributions and experiments involved. This approach is an important reference regarding the backbone for the knowledge required in *e-Research*. Unfortunately the Newman's proposal does not include clear ideas regarding the interaction processes among scientists and it focuses on the complexity bound to the ontology mapping.

There is another group interested in sharing high level knowledge among scientists. The project FEARLUS-G of Pignotti et al. (2004), which allows large scale experiments distributed among participants throughout a semantic grid. Pignotti's group has understood that scientific research always involves collaboration between individual scientists. They are clear about the Internet's capability to share and distribute data, but they noticed that such channel does not offer a natural way to manage and coordinate computational resources. There is a set of models under the FEARLUS project. They are oriented to provide semantics to atomic elements bound to the experiment. The aim is to provide an infrastructure for simulating the collaboration that would happen among scientists cooperating under the same ceiling. This infrastructure could handle experiments involving participants from remote locations. The main constraints from FEARLUS project is that is focused in specific domains related to land use research, although their discoveries are useful in general approaching for social/scientific interaction.

The project STIN (Walker & Creanor, 2009) is an effort to go beyond the eLearning experience to explore the relationships between people and technology. The most interesting aspect from this project is the fact that its authors recognize networks as a way / tool to capture the complexity of relationships among people through technology artifacts. The STIN's framework is capable to handle different levels of resolution. Another interesting aspect from their proposal is the elements used to build the network, which is composed by objects and properties involved in the learning process instead of people as the center of the network. Unfortunately STIN project does not reach the following step in the semantics stair, the ontology approach.

David (2004) makes an astonishing utterance regarding the capability of scientists to perform e-research in the scopes provided by modern ICT's. This author recognizes the significant advance in information and communication technologies, as well as the internet computing and grid technologies. Nevertheless, the author also argues that *"engineering breakthroughs alone will not be enough to achieve the outcomes envisaged for these undertakings. Success in realizing the potential of e-Science—and other global collaborative activities supported by the 'cyber-infrastructure'—if it is to be achieved, will more likely be the resultant of a nexus of interrelated social, legal and technical transformations"* (p. 3). Hence, a technical effort, as the presented in this chapter, will provide a framework or, at least, a tool for gathering, handling and sharing the knowledge related to certain scientific activities. Our proposal tries to enhance the effects of data and information related to scientific research. According to David (2004), *"scientific research collaboration is more and more coming to be seen as critically dependent upon effective access to, and sharing of digital research data. Equally critical are the information tools that facilitate data being structured for efficient storage, search, retrieval, display and higher level analysis, and the codified and archived information resources that may readily be located and reused in new combinations to generate further additions to the corpus of reliable scientific knowledge"* (p. 5-6); our proposal could be framed in such understanding for e-science.

MODEL DEFINITION

The idea is to allow the collaboration among scientists through a common framework. This structure is composed by a set of concepts and the relationships among them. Under the assumption that: *it is possible to deal with a specific problematic through the treatment for a set of variables, ranging along a group of values*, the problem's decomposition will be oriented to gathering the main variables of the problem.The collaborative stages that have been considered for this experiment simulation model are:

- Discovering variables (direct and indirect). Variables are filtered to achieve an optimal list.
- Defining ranges for variables. Ranges are filtered to achieve an optimal definition of values.
- Building a Cartesian product to merge the variables through their ranges.
- Applying constrains to filter the set of *tuplets* resulting from the Cartesian product.
- Binding specific resulting *tuplets* to specific states in the contextual reality of the studied problem. Hence, defining a list of problem states based on *tuplets* of values.
- Defining or establishing possible actions over a studied problematic situation.
- Applying constrains to filter up to an optimal list of actions.
- Merging the discovered states of the problem with the resulting actions that are allowed in the context of the problem.

Discovering Variables Bound to the Problem / Solution

We understand every concept bound to the problem / solution as a specific assembly of variables. In order to compose this ensemble of variables, scientists should coordinate to discover and define the most significant aspects of the scientific problem. The set of variables could be classified as direct and indirect variables. Direct variables have an immediate influence in the outcome of the studied problem; meanwhile the indirect ones have an effect on the situation through the affection over direct variables or even other indirect variables. Understanding that every variable will imply specific dealing efforts, the list of variables must be filtered in order to achieve an optimal list. This list should have only the most representative indicator for the problem, but it should not miss any influent aspect of the problem / solution. Coordination among collaborating scientists is required in order manage this optimal list of variables.

Defining Ranges for Variables

Once the whole set of variables has been defined and refined, scientists should work over the definition for the ranges to every variable. This definition of ranges includes the data type specification for every variable. Some variables could have a declarative or qualitative specification for their range, e.g. a temperature could be expressed as "very cold," "cold," "warm," "hot," and "very hot." The use of these declarative elements will allow a handling with higher semantics for the aspects related to the problem.

Discovering and defining the variables and their ranges is a *non computable task*. These processes are clearly related to mental performance in humans, which remain away from current computers capabilities. It clearly demands some assistance from humans. Nevertheless, some automatic elements can be integrated in order to assist the scientists while they are collaborating in defining these main aspects, concepts and relationships of the problem. Distributed applications can support the sharing of knowledge among the researchers involved. Hence, a tool for cooperative work could be adapted to allow the variables' definitions and their ranges.

Merging Variables to Define the States of the Problem

The ranges of variables, resulting from the previous stage, are merged through a Cartesian product. This task is clearly assignable to machines. The wide list of ensembles of variables must be analyzed by the group of collaborating researchers. At this point, the task is oriented to discriminate those ensembles that do not have compatibility with problem understanding. This is an important activity. If redundant or incompatible ensembles are kept, the problem could become not amenable or even illogical for the approach. Two ideas are central in order to achieve this goal: on the one hand, the scientists involved are able to vote for the ensembles they consider as crucial. On the other hand, a set of constraints must be defined. If every ensemble resulting from the Cartesian product is understood as a configuration of the analyzed scope, we believe that these configurations are the states of the scientific problem. Thus, the constraints are defined as a set of states which represent the unreal or unpractical configurations of the problem.

Eventually, the list of states of the problem will be strictly defined by the cooperative work of the scientists involved in the treatment of the studied problem. At the same time, these states are the elements for constructing the ontology.

Defining the Applicable Actions for Problem / Solution

A set of scientific activities must be defined by the group of researchers. In this case, scientists must propose sets of activities. These activities are analyzed by the whole group. Once again, they will vote and confront them with the constraints (at this point, the constraints to perform certain action). Activities could be defined in different phases. At first, every activity is defined only by the name. Later on, the conceptualization of the activities should be specialized through the definition of inner steps. This exercise will allow the detection of redundant or missing activities. Inner steps could be established as a narrative inside a script.

It is possible to use a CSCW tool to undertake this task. In the first phase, a forum would be useful for providing the actions' listings. In the following phase, which concerns the activities' bodies, common spaces could be created to allow placing in the proposals; these spaces could be assessable by the collaborators.

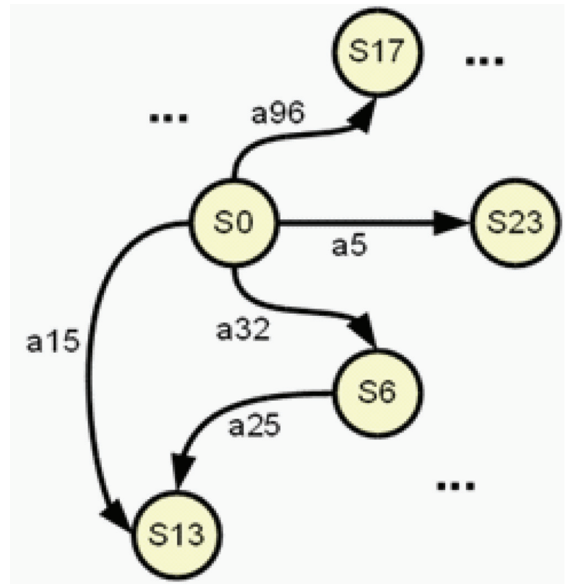
Building the Network of States and Actions

The refined set of activities is merged with the states through a graph. The resulting network combines the different states of the problem, as nodes, and the actions which are responsible of the state changes, as edges. This network is the kernel for the ontology's construction. A sketch of this graph is shown in Figure 1. This figure shows a segment from a diagram for states' transition. It is built with the possible states and actions, in the scientific scope, bound to the problem. This diagram models the common awareness among scientists collaborating in the same problem.

The network of states and actions could be understood as a synthetic representation of the experiment. The main advantage of having this representation is the possibility to share and replicate the experiment among scientists throughout the cyber-infrastructure. It is also possible for scientists to operate the experiment and simulate the performance, without the necessity of being *in situ*. Hence, scientists can collaborate in the same experiment even if they are in different geographical places. A distributed approach is allowed. The number of variables and the fine tuning in the actions elected for the problem implies an improved granularity available for the scientific management.

This network is the ontology linked to the specific scientific problem. Its elements are bound

Figure 1. A segment from a diagram for states' transition



to the variables through the states (nodes) and the scientific actions through the edges.

Use of the Ontology

This ontology could be understood as the plastic representation of the hypotheses that are supporting the research effort. This is an important achievement for this model. This plastic representation could be always matched with the actual results during the experiment's development, as well as the capacity to correct the hypotheses or the tasks to perform. Hence, the ontology becomes a map to guide the research efforts made by the different participants, which constitutes a common reference built democratically by the participants.

Regarding the control of the interaction among the scientists, the model will be measured from the perspective of affiliation networks proposed by Faccioni and Panzarasa (2006). This approach is aimed at discovering the amount of knowledge transferred among the social actors collaborating in common spaces. These quantitative elements will provide additional semantics on the interaction acts shown by participants in the common research effort. Given the complexity of this framework, this approach presents a significant achievement to grasp "meaning" from the collaboration among humans.

SPECIFICATIONS FOR THE ONTOLOGY-HANDLING TOOL

A standardized way to provide semantics over the Internet channels is the use of RDF "Resource Description Framework" model (Tauberer, 2006). Such approach allows the definition and /or naming for pieces of knowledge involved in the transferences among nodes in the Internet. This perspective is useful for the proposed model in this chapter, but is not enough to cover all the necessary aspects to represent the knowledge involved, which has different semantic-levels.

According to the arguments in the previous sections, scientists must agree over the variables and scientific actions that are bound to the problem or experiment to be handled by the group. As it would be understood, the nature of these elements will differ among the experiments or problem regarded. Besides, the scientists and even scientific groups will have different understanding for the very same problem. Thus, the understanding and therefore the RDF for a scientific problem are strongly bound to such problem and the approach established by the people involved. This situation will imply the production of an RDF that can be used only in the analyzed problem. Although this treatment is valid and useful for dealing with a specific problem, it is not useful for dealing with any scientific problem. An additional structure for definitions must be considered. This additional arrangement contains definitions for global elements that are useful for any scientific treatment.

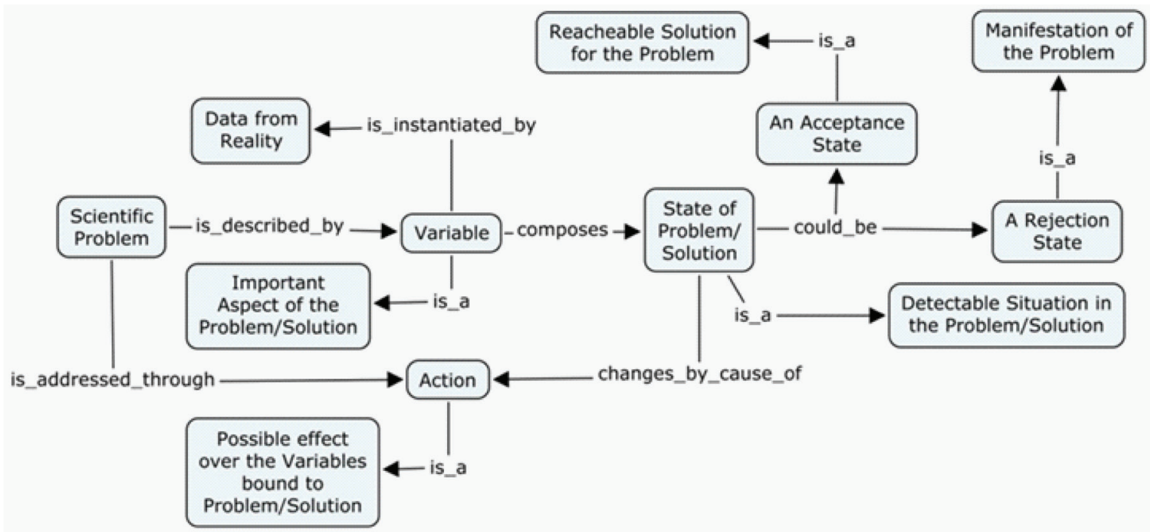
The higher level definitions in the upper RDF are aimed at gathering the main concepts and actions from any scientific challenge. As mentioned in the previous section, such elements are the direct and indirect variables as well as the scientific actions that can be performed over the variables. This meta-structure is the backbone for supporting a common handling for scientific problems. The lower level is bound to specific aspects of the problem, although is compatible to the higher understanding provided by the upper RDF. This lower RDF has definitions for concrete variables, states, actions and events from the scientific problem.

Meta-definitions can be defined as a framework for guiding the scientific efforts. The standard RDF provides mechanisms to crumble the knowledge into small pieces (Tauberer, 2006). The primitive elements of knowledge are organized as triplets and some rules are included to provide meaning. According to Tauberer (2006), an RDF document is built on the following bases: *"1. A fact is expressed as a Subject-Predicate-Object triple, also known as a statement. It's like a little English sentence. 2. Subjects, predicates, and objects are given as names for entities, also called resources (dating back to RDF's application to metadata for web resources) or nodes (from graph terminology). Entities represent something, a person, website, or something more abstract like states and relations. 3. Names are Uniform Resource Identifiers (URIs), which are global in scope, always referring to the same entity in any RDF document in which they appear. 4. Objects can also be given as text values, called literal values, which may or may not be typed using XML Schema datatypes."*

Hence, the graphical understanding for an ontology can be modeled as metadata that explains the treatment for a scientific problem. Figure 2 sketches the ontology both for the upper level of the scientific treatment for a problem and the breakdown of the associated complexity.

The RDF that can be built from the upper ontology for scientific problems, shown in Figure 2, is roughly approached through the tabular notation that is shown in Table 1. This approach is applied during the treatment of scientific problems and the effort to reach a solution. As such,

Figure 2. Graphical representation for the general ontology approach



it can be understood as the ontology that supports an automated breakdown of the complexity in scientific problems.

Now, we have enough elements to build the RDF of the upper ontology meant for scientific problems. In order to achieve this knowledge representation, the first stage is to use a notation that mimics the tabular representation shown in Table 1. This representation style is called "Notation 3" or simply N3 (Tauberer, 2006). An advantage of this specification is that the notion of the underlying graph prevails and it is possible to grasp the elements organization from the reading. The next step is to build the RDF/XML standard representation, which is more compatible to the computer's approach but it is more obscure regarding the graph view.

Once the upper RDF has been defined, the following task is to build the lower RDF. This new effort is oriented to discover the specific aspects of the problem, based on the approach proposed through the higher ontology (represented in the upper RDF). During this task, it is precise to gather all the knowledge elements bound to the problem. These elements are gathered under the higher ontology's basis. Hence, the scientific

Table 1. Tabular representation for the general ontology approach

Start Node	Edge Label	End Node
Scientific Problem	is_described_by	Variable
Scientific Problem	is_addressed_through	Action
Variable	is_a	Important Aspect of the Problem / Solution
Variable	is_instantiatied_by	Data from Reality
Variable	composes	State of Problem / Solution
Action	is_a	Possible effect over the Variables bound to Problem / Solution
State of Problem / Solution	is_a	Detectable Situation in the Problem / Solution
State of Problem / Solution	could_be	An Acceptance State
State of Problem / Solution	could_be	A Rejection State
State of Problem / Solution	changes_by_cause_of	Action
An Acceptance State	is_a	Reachable Solution for the Problem
A Rejection State	is_a	Manifestation of the Problem

problem must be observed in order to unveil its passive and active descriptors. The passive descriptors could be bound to problem's variables; meanwhile the active descriptors will be bound to the possible actions in the problem's scope. This effort requires an abstraction exercise, divided in different inductive stages.

The first inductive stage is oriented to detect the main aspects and manifestations of the scientific problem. During this stage, the scientists involved must define a set of direct variables and a set of indirect variables. Direct variables are those that produce an immediate effect on problem's situation when they change; meanwhile indirect variables produce collateral effects in the problem. All scientists involved in the problem could provide sets of variables. A Problem / Solution administrator could collect those proposals and build unified lists of variables. At this point all the proposed variables are added and only those similar variables are combined into one common-identification. The discrimination of variables will be done automatically when a variable is left behind because some other variables are more effective to model the treatment of the problem.

These lists of variables, once defined, are treated through the assignment of ranges for every variable. The ranges are established, as it was mentioned before, through discrete elements. The use of discrete elements settles any exuberance of the problem. The variables are combined using a Cartesian product, which generates a set of *tuplets* that model the understanding of the problem, for the scientists involved, within the different dimensions that have been regarded as valid for the problem /solution. Every *tuplet* represents a possible state of the problem.

This inner approach to the scientific problem can be shown throughout an example: *the design of an intelligent suspension of a car*, which will have the goal of providing automatic-response to changes in the road and driving conditions. This is shown in Figure 3. Some interesting aspects in the car suspension are: *oil pressure inside the shock absorbers, oil pressure in the steering pump, compression in coil springs, speed of the car and lateral inertia;* besides, the following aspects are interesting too: *air pressure in tires, oil temperature in the shock absorbers, oil temperature in the steering pump and temperature of tires*.

These aspects could be different according to the experience and abstraction approach of the observer(s). The first group of aspects is clearly influencing over the intelligent suspension; meanwhile the second group of aspects has not direct influence over the performance of the suspension. Hence, the first group could be handled as the direct variables and the second one refers to the indirect variables. The following stage consists in defining discrete ranges for the variables. Table 2 shows the discrete ranges for those variables that have been considered a direct. These variables, as well as the acronyms that shorten the variables' names, play an influencing role in an intelligent suspension. Once again, the discrete ranges could be different according to the approach of every person.

In addition, Table 3 shows the discrete ranges for those variables considered as indirect influence for the problem. It is understood that indirect

Figure 3. Graphical model for the hydraulic steering system analyzed

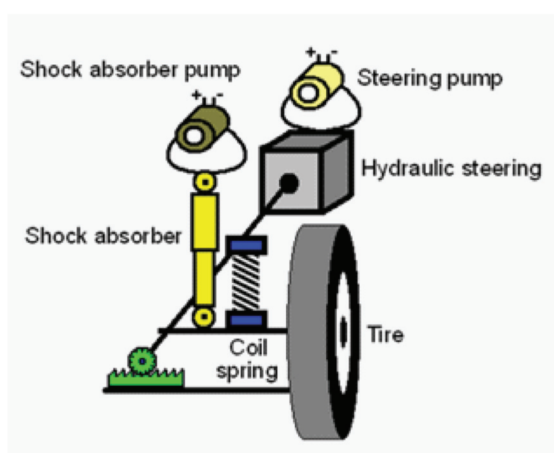


Table 2. Discrete ranges for the variables considered as direct

Oil pressure inside the shock absorbers (OPISA)	High
	Medium
	Low
Oil pressure in the steering pump (OPISP)	High
	Medium
	Low
Compression in coil springs (CICS)	Very shaky
	Frequent and heavy
	Frequent but brief
	Regular but intense
	Regular and brief
	Sporadic but intense
	Sporadic and brief
	Soft
Speed of the car (SOC)	First quintile
	Second quintile
	Third quintile
	Fourth quintile
	Fifth quintile
Lateral inertia (LI)	Left high
	Left medium
	Left low
	Equilibrium
	Right low
	Right medium
	Righ high

variables will influence the scientific problem through the changes they produce over the direct variables.

It is the time now for the variables to get mixed. In order to finish this task, the ranges of the variables are combined using the Cartesian product. Due to the simplification achieved for representing the problem's milestones, the effort to combine

Table 3. Discrete ranges for the variables considered as indirect influence in an intelligent suspension, as well as the acronyms that shorten the variables' names

Air pressure in tires (APIT)	High
	Medium
	Low
Oil temperature in the shock absorbers (OTISA)	Very hot
	Hot
	Temperate
	Cold
	Very Cold
Oil temperature in the steering pump (OTISP)	Very hot
	Hot
	Temperate
	Cold
	Very Cold
Temperature of tires (TIT)	Very hot
	Hot
	Temperate
	Cold
	Very Cold

the ranges of the variables can be performed by an automatic algorithm.

A program could automatically produce the set of *tuplets* from merging the variables. Every *tuplet* is one of the possible states for the scientific problem according to the abstraction approach that supports the group of variables elected. Box 1 contains a subset of 945000 states found to the problem.

The following stage consists in defining the set of possible actions over the variables detected. Sometimes it is not possible to modify a variable's manifestation, such as modifying the time of continuing to passing, stopping it or returning it. Nevertheless, a set of actions must be made for all the variables that can be changed. All those variables that cannot be changed must be treated as implicit changes, bound to time or other unhandled events.

Box 1. A subset of the states bound to the scientific problem of the intelligent suspension

```
State: 0
Directs[opisa(High), opisp(High), cics(Very shaky), soc(First Quintile),
li(Left High)]
Indirects[apit(High), otisa(Very Hot), otisp(Very Hot), tit(Very Hot)]

State: 1
Directs[opisa(High), opisp(High), cics(Very shaky), soc(First Quintile),
li(Left High)]
Indirects[apit(High), otisa(Very Hot), otisp(Very Hot), tit(Hot)]
…

State: 427341
Directs[opisa(Medium), opisp(Medium), cics(Very shaky), soc(Third Quintile),
li(Right Medium)]
Indirects[apit(Medium), otisa(Cold), otisp(Cold), tit(Hot)]

State: 427342
Directs[opisa(Medium), opisp(Medium), cics(Very shaky), soc(Third Quintile),
li(Right Medium)]
Indirects[apit(Medium), otisa(Cold), otisp(Cold), tit(Temperate)]

State: 427343
Directs[opisa(Medium), opisp(Medium), cics(Very shaky), soc(Third Quintile),
li(Right Medium)]
Indirects[apit(Medium), otisa(Cold), otisp(Cold), tit(Cold)]
…

State: 944997
Directs[opisa(Low), opisp(Low), cics(Soft), soc(Fifth Quintile), li(Right
High)]
Indirects[apit(Low), otisa(Very Cold), otisp(Very Cold), tit(Temperate)]

State: 944998
Directs[opisa(Low), opisp(Low), cics(Soft), soc(Fifth Quintile), li(Right
High)]
Indirects[apit(Low), otisa(Very Cold), otisp(Very Cold), tit(Cold)]

State: 944999
Directs[opisa(Low), opisp(Low), cics(Soft), soc(Fifth Quintile), li(Right
High)]
Indirects[apit(Low), otisa(Very Cold), otisp(Very Cold), tit(Very Cold)]
```

It now possible to build a network in which the nodes are the states generated by merging the variables' ranges and the edges of the network are the actions and the implicit changes detected. This structure can be made through automatic media. Every state is faced to every action and implicit change detected, and every edge is modeled by considering the current state and the following one (as a leap of state). The resulting graph can be understood as a state transition diagram, which mimics the experiment according to certain approach. Hence the experiment can be modeled as a set of triplets, in which every triplet contains the code in Box 2.

These edges are not arbitrarily established; they are settled on the basis of the discrete ranges defined for the variables. If a certain state has the oil pressure in the steering pump as medium and the action of pumping more oil is performed, the oil pressure in the steering pump will increase and might reach the state of *high*. Therefore, the algorithm that creates the edges has to follow up a set of rules in which the behavior of variables is described under the effect of actions and implicit changes.

The first approach of the graph will be somehow *overcrowded* with transitions, as a natural effect of the computer's action which mixes blindly all of the states with all of the change factors. The rules will control the uncontrolled growing, but some redundant growing will happen in some branches of the definition. This collateral congestion could be handled by the natural interaction among scientists. All those underused transitions will receive less voting and eventually will be disregarded of the synthetic experiment.

Hence, the lower RDF must be constructed under the same perspective used to build the upper one, but now considering the concrete variables of the scientific problem and their discrete ranges, the possible actions and the implicit changes are detected. A distributed tool, based on web, could handle these RDF definitions in order to manage their contents and enable the collaboration through the supervision of the use of each element. All those overused elements will imply a rhetorical invitation to maintain such element in the approach and all those underused elements will be a tacit invitation for removing them from the approach. Besides, a *system manager* or *collaboration supervisor* could always receive queries from scientists for adding or removing elements according to the own expertise or the own practical experiment definition.

DISCUSSION

The model presented in this work is a new proposal for modeling the knowledge undergirding scientific problems. Such a representation of knowledge can be used for supporting a series of tasks. The tasks considered during the construction of this proposal are the experiment modeling / simulation and the sharing of knowledge among scientists. Due to the nature of the representation proposed, RDF/XML, these tasks can be undertaken automatically by systems designed for such purpose.There are a series of feasible advantages from the perspective of this model:

- Scientists collaborating have the chance to narrow the exuberance that is bound to certain scientific problems, by synthesizing its treatment as a list of tasks. Though the act of following lists could kill creativity, our proposal provides some looseness regarding *collaboration mechanisms* and *human intervention* to **fix** or **complete** the approach taken for the problem/solution.

Box 2.

```
<tuplet_current_state, tuplet_following_state, action_or_event>
```

- The synthetic representation of experiments can travel along the Internet channels, enhancing the chances to capture the cyber-infrastructure. This representation can be easily replicated and shared among collaborators, while it provides ways for gathering feedback from specialists in the group, independently where they are.
- The RDF/XML representation for the experiment allows standardized transference and exploitation for the model. Different tools can grasp the meaning stored in the structure. The resulted ontology can feed adapted tools for virtual reality. Artificial agents could mimic the actors of the problem by performing as indicated in the diagram of the states and transitions, as explained by Walton and Barker (2004).
- Automatic testing tools could be implemented to run the tests over different aspects of the problem. Or, even, to verify automatically the presence of constraints disregarded by collaborating scientists. These tools can be fed with the proposed ontology.

CONCLUSION

Our proposal is aimed at creating a model that can support the scientific interaction among researchers involved in common problems. This very first step, described in this chapter, aims at creating procedures to approach effectively the scientific problem to be managed by the group. Such knowledge is used to build an artificial representation of the experiment. We believe that most of scientific problems could be approached by this perspective and, therefore, could be handled by the proposed model. Our method has the capacity to capture the main aspects linked to scientific problems: variables (direct and indirect), ranges for variables, states of the problem/solution, actions over the variables, implicit changes and state transitions. These elements, presented in synthetic fashion in machines, enable the idea that scientific problems can successfully be emulated in synthetic scopes, capturing the cyber-infrastructure. This is an important achievement, since a synthetic representation demands specific knowledge from the concrete object / situation represented; this knowledge, once isolated and represented in machines, promotes the enforcement of the indicators linked to complexity's control. By enabling this knowledge in machines, the synthetic representation can be easily transferred through regular channels that actually are used to transfer data and information in the Internet, as suggested by David (2004). The pending tasks in this project are the design of the interaction/collaboration model among participants and the implementation of the model in specific platforms. Nevertheless, the model is already defined; the whole design and implementation could also imply the development of a framework that supports the interaction among scientists cooperating in common problems.

ACKNOWLEDGMENT

This work has been partially supported by the FP7 European project ALICE, under grant FP7-ICT-2009-5-257639

REFERENCES

Anandarajan, M., & Anandarajan, A. (2010). *E-research collaboration: Theory, techniques and challenges*. Berlin, Germany: Springer-Verlag.

Atkins, D., Droegemeier, K., Feldman, S., Garcia-Molina, H., Klein, M. L., & Messina, P. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure. *National Science Foundation.* Retrieved June 16th 2011, from www.nsf.gov/ cise/ sci/ reports/ atkins.pdf.

David, P. A. (2004). Towards a cyberinfrastructure for enhanced scientific collaboration: Providing its 'soft' foundations may be the hardest part. *Oxford Internet Institute, Research Report No.4, August 2004*. Retrieved June 30th 2011, from http:// www. oii.ox.ac.uk/ resources/ publications/ RR4.pdf.

Faccioni, M., & Panzarasa, P. (2006). Knowledge transfer within affiliation networks. In *Proceedings of the IEEE International Engineering Management Conference*, (pp. 226-230). IEEE Press.

Goble, C., Stevens, R., & Bechhofer, S. (2005). The semantic web and knowledge grids. *Drug Discovery Today*, *2*(3), 225–233. doi:10.1016/j. ddtec.2005.08.005

Jankowski, N. (2007). Exploring e-science: An introduction. *Journal of Computer-Mediated Communication, 12*(2). Retrieved June 15th 2011, from http:// jcmc.indiana.edu/ vol12/ issue2/ jankowski.html.

Kalfoglou, Y., & Schorlemmer, M. (2003). Ontology mapping: The state of the art. *The Knowledge Engineering Review*, *18*(1), 1–31. doi:10.1017/ S0269888903000651

Li, H., & Liu, L. (2007). A decentralized resource discovery based on keywords combinations and node clusters in knowledge grid. In *Proceedings of the Intelligent Computing 3rd International Conference on Advanced Intelligent Computing Theories and Applications,* (pp. 738-747). Qingdao, China: Springer.

Newman, D., Bechhofer, S., & De Roure, D. (2009). *myExperiment: An ontology for eResearch*. Paper presented at the Workshop on Semantic Web Applications in Scientific Discourse in Conjunction with the International Semantic Web Conference. Washington, DC.

Pignotti, E., Edwards, P., Preece, A., Gotts, N., & Polhill, G. (2004). FEARLUS-G: A semantic grid service for land-use modelling. In *Proceedings of the ECAI-04 Workshop on Semantic Intelligent Middleware for the Web & the Grid*, (vol 111). IEEE Press.

Tauberer, J. (2006). What is RDF? *O'Reilly XML from Inside Out*. Retrieved February 4th, 2011 from http:// www.xml.com/ pub/ a/ 2001/ 01/ 24/rdf.html.

Walker, S., & Creanor, L. (2009). The STIN in the tale: A socio-technical interaction perspective on networked learning. *International Forum of Educational Technology & Society*, *12*(4), 305–316.

Walton, C., & Barker, A. (2004). An agent-based e-science experiment builder. In *Proceedings of the 1st International Workshop on Semantic Intelligent Middleware for the Web and the Grid*, (pp. 247-264). Valencia, Spain.

Zettsu, K., Nakanishi, T., Iwazume, M., Kidawara, Y., & Kiyoki, Y. (2008). Knowledge cluster systems for knowledge sharing, analysis and delivery among remote sites. In *Proceeding of the 2008 Conference on Information Modelling and Knowledge Bases XIX*, (pp. 282-289). Amsterdam, The Netherlands: IOS Press.

KEY TERMS AND DEFINITIONS

Cyberinfrastructure: The set of cybernetic implements which constitute the support to different services and functionality.

E-Science: The set of resources offered to support science practice throughout Internet channel.

Experiment Modeling: Achieving a synthetic representation of an experiment.

Knowledge Grid: A set of concepts linked through different abstract or concrete bounds.

Knowledge Transfer: The transference of concepts: ideas, awareness, etc. regarding a matter.

Online Collaboration: People collaborating throughout Internet channels: mainly the Web.

Ontology Mapping: Gathering pieces from reality and insert them into the slots bound to main concepts.

ENDNOTES

1 Stanford. (2011). *Website.* Retrieved from http://plato.stanford.edu/.