

**Project Acronym:** **SemanticHEALTH**

**Project Title:** **Semantic Interoperability Deployment and Research Roadmap**

**Contract Number:** IST-27328-SSA

**Starting date:** January 01, 2006

**Ending date:** December 31, 2007

**Deliverable Number:** **D 2.1**

**Title of the Deliverable:** **Towards a semantic interoperability roadmap: Technological issues**

**WP/Task related to Deliv.:** **WP 2, Task 1.2**

**Type:** **Internal, Final**

**Authors:** ***Pierre Lewalle, and reviewed by all partners***

**Partners:** Radboud University Nijmegen Medical Centre  
World Health Organization  
Egészségügyi Stratégiai Kutatóintézet - National Institute for Strategic Health Research  
Uppsala Universitet  
Universite Jean Monnet Saint Etienne  
University College London  
University of Manchester (subcontractor)  
empirica Information & Communication Technology Research mbH

**Project Co-ordinator**

**Organisation name:** Radboud University Nijmegen Medical Centre; Department of Medical Informatics

**Name of representative:** Pieter E. Zanstra

**Address:** 152 Medical Informatics, POB 9101, 6500 HB Nijmegen

**Phone number:** +31-24-3 61 54 30

**Fax number:** +31-24-3 61 35 04

**E-mail:** p.zanstra@mi.umcn.nl

**Project WEB site:** <http://www.semantichealth.org>

## Document History

Name	Date	Version	Description
Pierre Lewalle	January 31, 2007	1	Draft structure
Pierre Lewalle	2007-02-15	1	1st draft
Dipak Kalra	2007-02 21	1	Critical review
György Súrjan	2007-02 22	1	Critical review
Pieter Zanstra	2007-03-07	1	Critical review
Jean Marie Rodrigues	2007-03-07	1	Critical review
Jean-Marie Rodrigues	2007-03-10		Annex 1
Pierre Lewalle	2007-03-12	2	Final version

<b>Contractual Date of Delivery to the EC:</b>	<b>December 2006</b>
<b>Actual Date of Delivery to the EC:</b>	<b>May 2007</b>

# Table of contents

<b>1</b>	<b>Executive summary .....</b>	<b>4</b>
<b>2</b>	<b>Introduction.....</b>	<b>4</b>
	2.1 Capturing complexity with simplicity: an analogy .....	5
	2.2 Mercator projections of knowledge.....	6
<b>3</b>	<b>The scope of the <i>semanticHEALTH</i> vision.....</b>	<b>7</b>
<b>4</b>	<b>Summary of experience to date .....</b>	<b>9</b>
	4.1 Deliverable 1.2 conclusion (for details, see Deliverable 1.2).....	10
	4.2 Technological issues and other analytical consequences.....	10
	4.2.1 From linguistics.....	10
	4.2.2 ... to computational linguistics .....	10
	4.2.3 Capturing domains of knowledge: the conceptual evidence.....	14
	4.2.4 An actor-centric perspective: conceptual framework relevance .....	15
	4.2.5 Use cases: the conceptual framework performance.....	17
	4.3 Operational parameters of semantic interoperability .....	18
	4.3.1 Across time and across countries: what about across domains and across languages .....	18
<b>5</b>	<b>Complex technical issues come to light .....</b>	<b>19</b>
	5.1 Declarative, procedural and inferential representations .....	19
	5.2 Context-dependence of meaning .....	20
	5.3 Finite state semantics versus semantic emergence.....	21
<b>6</b>	<b>Conclusions and recommendations.....</b>	<b>22</b>
<b>7</b>	<b>Bibliography.....</b>	<b>23</b>
<b>8</b>	<b>Annex 1 Report of the 2006-05-30 Copenhagen Workshop.....</b>	<b>25</b>

# 1 Executive summary

The challenges for semantic interoperability result from the need to combine the information collected and used in the clinical setting perspective of seamless care provision across all elements of the health care value system, in the public health perspective of integrating individual clinical information for evidence-based decision-making at population level, and in the research perspective in integrating clinical and biomedical information across the continuum of biosciences. Many data sets, whether minimum or expanded, data and metadata definitions, ranging from proprietary to standardized, controlled or otherwise organized vocabularies in the form of classifications and terminologies have been devised to meet a wide range of requirements, from ad hoc and subject-specific to more generalized and widely recognized. The variety of proposed solutions is increasing to meet the growing demand for solutions, but the rapid change and the sophistication of products (and the underlying research) makes them barely affordable even in better off countries.

The *semanticHEALTH* vision is based on the perception that market solutions proposed so far are missing a consistent reference to robust interoperability principles. Many efforts have been made in recent years to address the issue, and most developments have focused on structural interoperability, with insufficient attention being devoted to semantic interoperability anchored in a shared formal, generalized representation of knowledge.

*SemanticHEALTH* is developing a *roadmap* and *recommendations* that pave the way to true semantic interoperability, taking advantage of state-of-the-art research outcomes. This deliverable, in particular, will look at the technology aspects involved. Assuming that the search for semantic interoperability may also concern other scientific fields dealing with complex systems, it will venture to investigate promising developments in non-health fields. While doing so, it keeps in mind the reality of the health field, where dependence on machines alone will never be sufficient, sophisticated as it may be. Semantic Interoperability in the medical and health fields will surely take advantage of the processing power of machines, but it will continue to have to be understood and controlled by humans. The proposed perspectives will therefore seek a compromise where complexity can be accommodated and the search for creative, yet simple, practical solutions that are manageable and safe for human interaction.

## 2 Introduction

A first deliverable 1.1 has proposed a conceptual framework to inform in the project.

*SemanticHEALTH* applies the following **interoperability (IOp) definition** developed by the EU *i2-Health* project (Interoperability Initiative for a European e-Health Area):

*Health system interoperability is the ability, facilitated by ICT applications and systems,*

- *to exchange, understand and act on citizens/patient and other health-related information and knowledge*
- *among linguistically and culturally disparate clinicians, patients and other actors and organizations,*
- *within and across health system jurisdictions in a collaborative manner.*

The *SemanticHEALTH* consortium aims to address the *transmission and use of meaning within the framework of seamless health care services*, between providers, patients, citizens and authorities. In essence the *SemanticHEALTH* goal is to assure **co-operability and collaboration** rather than only **interoperability**.

The project has defined three levels of semantic IOp (SIOP), the highest one being defined as *Level 3 or full semantic interoperability, or co-operability*, which is reached if users of system *B* are able to use information acquired automatically from system *A* with equivalent meaning to

*its local data*: the information can be processed homogeneously with data captured natively within System B, as if they were entered by a user B directly into system B.

The **three dimensions** of the semantic interoperability framework, which will pervade the entire project, are the **analytical dimension**, covering both technological interoperability and socio-economic aspects of semantic interoperability, the **application dimension** focusing on specific application fields, as well as the **research dimension** which will identify key future research topics.

In Part B of Deliverable 1.2, the main needs for specific semantic interoperability were addressed from the real world application dimension perspective, proposing a connectivity infrastructure allowing to analyse key interoperability experiences in Member States and at international level.

To assess a connectivity infrastructure, *semanticHEALTH* decided at its Copenhagen workshop (30/09/06) to assess the relevant Member States and international experiences of SIOp on one hand on their ability to cope with the three dimensions of the framework and to use the level 3 definition of semantic interoperability, and on the other hand on the three main characteristics supporting the semantic interoperability of such experiences;

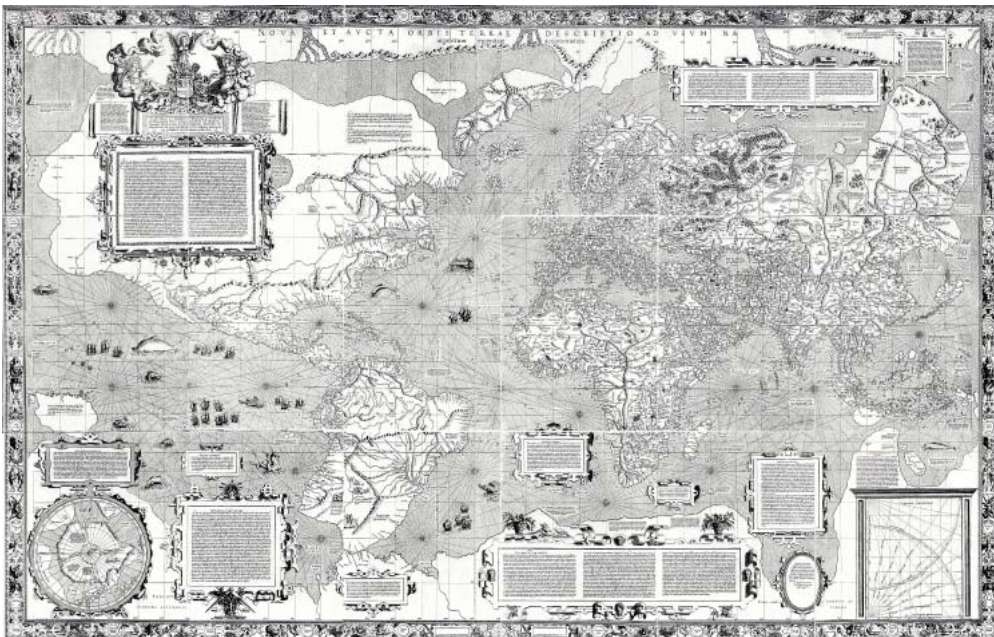
1. The first semantic specificity is to start such experiences from use cases or scenarios
2. The second one is to link the field of direct care to the patient to public health and secondary use
3. The third is to integrate terminologies with ontology and multilingualism.

These levels should be related to the **desiderata for the SIOp**.

- a) *Readable - with potential for reuse*
- b) *Searchable*
- c) *Systematic identification with meta data*
- d) *Systematic coding with hierarchy*
- e) *Systematic coding with context information*

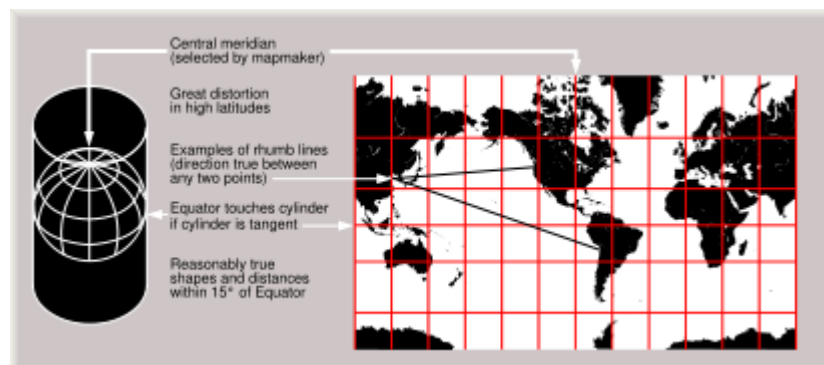
## 2.1 Capturing complexity with simplicity: an analogy

The best roadmap analogy that can be found to managing the world of health information is that of seafarers navigating efficiently and safely around the globe.



Historical maps show that diverse regional representations had been drawn in many parts of the world. They were all centered on the originator's territory. With the newly discovered world, it appeared that a unified cartography was needed, but the spherical shape of the planet was difficult to master, resulting in seafaring being confronted with many sources of danger and inefficiencies. To enable them to draw safe and operational routes for the nascent international trade, the complex reality of the world's geography needed to be captured in more manageable representations. Geographers imagined powerful solutions in the form of projections. Among the most famous ones are the Mercator projections, which are cylindrical projections of the world map.

The name and explanations given by Mercator to his world map (*Nova et Aucta Orbis Terrae Descriptio ad Usum Navigatum Emendate*: "new and augmented description of Earth corrected for the use of navigation") show that it was expressly conceived for the use of marine navigation.



While it is recognized that such projections are distorted representations of the world (for instance, with Greenland appearing as big as Africa), they have offered a practical and manageable representation on which distances and travel times can be calculated using adequate mathematical functions. They are used to the present day: [Google Maps](#) currently uses a Mercator projection for its map images. Despite its relative scale distortions, the Mercator is well-suited as an interactive world map that can be panned and zoomed seamlessly to local maps.

## 2.2 Mercator projections of knowledge

The cartographic paradigm [Lee F. et al. 2000] suggests a useful perspective for shaping research in the service of a knowledge modeling and health information system development that would be truly interoperable while remaining adaptable to local needs.

Describing parts of the knowledge world requires the constant reference to a complex system of areas of knowledge with inclusion, covered zones and intersection zones. Maps obtained from Internet are generally of unknown type: alignment and modeling of any transformation is difficult. In a similar way, semantic landmark mapping could provide reliable anchor points to assist in defining the semantic space. Working with texts, the landmark mapping technique could define the semantic space based on the registration of maps according to predefined criteria. Thus the collection of existing semantic maps with their underlying particular ontologies could provide pointers to the semantic landmarks. From there permitted semantic transforms could be derived. Building on the previous analogy, Global Atlas is a geographical search-engine. It indexes maps, satellite and aerial pictures, as well as HTML documents available on the World Wide Web. The Global Atlas leverages on the cartographic paradigm to provide a very natural support for indexing, searching and sharing information. It allows the design of intuitive user interfaces and the use of natural visual feedback. HTML documents are best indexed according to the geographical regions to which they are topically associated, and maps in the form of GIF and JPEG images are indexed to create a huge patchwork of maps. However, maps come in a variety of unspecified coordinate systems and projections. This entails calibrating different maps to a single reference coordinate system. [Bressan et al.

1999] discuss the design issues in building a Geographical Search Engine, and focus on the calibration of maps. This might constitute a bridge between the visual representation and text describing it.

### 3 The scope of the *semanticHEALTH* vision

Starting from the initial vision captured in the semantic interoperability framework described earlier, technological considerations of interoperability are reviewed in the light of (a) current practice and applications, and (b) published basic and applied research materials addressing aspects to be improved or suggesting development avenues.

Although interoperability is a declared a desirable feature of health-related information systems, only limited evidence exists that current systems are actually fully interoperable as defined for *semanticHEALTH* purposes. A number of issues come to mind:

1. Is the situation due to insufficient or inadequate scientific evidence to have an impact on practice?
2. Do development and operating costs better match the resources available to Member States individually or collectively?
3. What are the experiments conducted and the experience gained in better resourced areas? Would their results, if any, be transposable to other domains and under what conditions?

This Deliverable 2.1 discusses possible answers to those questions, in order to advance semantic interoperability of information systems in the health field.

The three dimensions of the semantic interoperability framework will be envisaged. In addition, technological issues will be anchored on the three main characteristics an ideally interoperable set of information sources should have to satisfy the requirements of the envisaged conceptual framework: (1) comparing use cases and other scenarios, what are the horizontal multidisciplinary commonalities across applications, and to what extent do they comply with *de facto* or *de jure* standards or research outcomes; (2) how do existing systems cater for the information interchange needs at different levels of specialization (e.g. from genomics to clinical care), and at the same time accommodate rules for information aggregation for epidemiological and statistical purposes at population level; and (3) drawing on a an appropriate terminology that carries the actual meaning of terms rather than a role in a pre-established system, in order to ensure in essence cross-cultural and cross-language interoperability.

The dimensions relevant to semantic interoperability were identified in Deliverable 1.2 as tabulated in Table 1.

Dimension	Issue	Indicator
A) Analytical dimension	<p>1. Technological interoperability issues (technical, syntactical, semantic aspects)</p> <ul style="list-style-type: none"> <li>- Information model</li> <li>- Architecture and messages</li> <li>- Archetype-template-interface with clinical terminology</li> <li>- Clinical terminology, ontology</li> <li>- Multilingual approach (encoding, language generation)</li> <li>- Standards (incl. de facto standards; actors in standard development: ISO, WHO, CEN; W3C; user groups – e.g. HL7)</li> <li>- Imaging</li> <li>- Security-integrity-availability</li> <li>- Open source</li> </ul> <p>2. Socio-economic issues</p> <ul style="list-style-type: none"> <li>- Health policy</li> <li>- Legal, regulatory issues</li> <li>- Economic issues (costs, benefits)</li> <li>- Management : planning, organization (structures, processes), controlling, including change management</li> <li>- Stakeholder involvement</li> <li>- Professional culture</li> </ul>	<p>At1</p> <p>At2</p> <p>At3</p> <p>At4</p> <p>At5</p> <p>At6</p> <p>At7</p> <p>At8</p> <p>At9</p> <p>As1</p> <p>As2</p> <p>As3</p> <p>As4</p> <p>As5</p> <p>As6</p>
B) Application dimension	<p><b>1. Clinical settings:</b> EHR, patient summary, interface with DSS, CPOE</p> <p><b>2. Public health</b></p> <ul style="list-style-type: none"> <li>- Classifications</li> <li>- Demography (birth registration, mortality, morbidity, disability, risk factors,</li> <li>- Health indicators, etc.</li> </ul> <p><b>3. Secondary uses:</b> epidemiology, preparedness, patient safety, bioterrorism, healthcare management, financing, research, education</p> <p><b>4. Biomedical terminologies:</b> from genomics to population health (linkage and congruence)</p>	<p>Ap1</p> <p>Ap2</p> <p>Ap3</p> <p>Ap4</p> <p>Ap5</p> <p>Ap6</p>
C) Research dimension	To identify in each item of dimensions A) and B) where are the needs for research as opposed to the opportunity to apply immediately the existing knowledge	R1

**Table 1 Dimensions identified in semanticHealth as relevant to semantic interoperability**

## 4 Summary of experience to date

For the key to indicator codes, see Table 1 on page 8

Possible indicator codes (At1, At2, At3, At4, At5, At6, At7, At8, At9); (As1, As2, As3, As4, As5, As6); (Ap1, Ap2, Ap3, Ap4, Ap5, Ap6); (R1)	Conceptual Framework Dimensions Conformity Score			Connectivity Infrastructure Assessment Score			
	Analytical	Application	Research	Total	Use cases	PH and Sec	Ontologies
Germany	7 At1, At2, At6, At8, At9; As1, As5	1 Ap5	0	8	1	0	0
Denmark	10 At1, At2, At4, At6, At7, At8; As1, As2, As5, As6	1 Ap1	0	11	1	0	0
Finland	8 At1, At2, At5, At6, At7; As1, As2, As5	1 Ap1	0	9	1	0	0
Norway	7 At1, At2, At6, At7, At8; As2, As5	2 Ap1, Ap5	0	9	0	0	0
Sweden	11 At1, At2, At6, At7, At8; As1, As2, As3, As4, As5, As6	2 Ap1, Ap5	0	13	1	0	0
United Kingdom / England	10 At1, At2, At4, At6, At7, At8; As1, As2, As4, As5	1 Ap1	0	11	0	0	0
France	5 At6, At7; As1, As2, As5	0	0	5	0	0	0
The Netherlands	8 At1, At2, At6, At7, At8; As2, As4, As5	1 Ap1	0	9	0	0	0
Hungary	7 At1, At2, At6, At7, At8; As2, As5	0	0	7	0	0	0
IHE (XDS-MS)	4 At2, At6, At7; As5	0	0	4	1	0	0
Australia	12 At1, At2, At3, At4, At6, At7, At8; As1, As2, As3, As4, As5	2 Ap1, Ap2		14	1	1	0
Canada	13 At1, At2, At4, At6, At7, At8, At9; As1, As2, As3, As4, As5, As6	2 Ap1, Ap2	1	16	1	1	0
USA	7 At2, At4, At6, At8; As1, As2, As5	1 As1	0	8	0	1	0

**Table 1 Summary of SemanticHealth Conceptual Framework Conformity Scores**

## 4.1 Deliverable 1.2 conclusion (for details, see Deliverable 1.2)

1. The inventory of these different so-called semantic interoperability initiatives within EU and the world shows limited progress as far as full (level 3) semantic interoperability, or co-operability, is concerned.
2. Achievements across the initiatives are numerous. There is therefore a rich and varied mix of experiences for cross-national cooperation and for unlocking synergies.
3. As a short term or midterm consequence the present situation provides immediate opportunities to learn from other experiences and namely from the commonality of Canadian and Australian initiatives. This opportunity shall be illustrated by showcase implementations.
4. For a longer-term approach to semantically interoperable e-health solutions, sizeable resources from EU , WHO and public institutions in Member States are required to link the field of direct patient care to public health and secondary uses. This can only be achieved by unfolding the full potential of the new generation of clinical and other specialized terminologies based on robust ontology representations, and an adequate realization of the knowledge interchange in a multilingual knowledge space.

The review of the systems currently put in place in the reference countries reveals that, so far, steps taken are not commensurate with the rationale put forward to initiate the project elements that have started.

## 4.2 Technological issues and other analytical consequences

### 4.2.1 From linguistics...

Systems are designed to meet the information needs for actors in particular domains. It is natural to consider that any domain of knowledge features a mosaics of **terms** [Hébert 1993], understood as a string of letters or signs (*morpheme*) with which some meaning is associated (*semems*). Semems themselves are more or less complex constructs of primitive meaning-carriers, or *sems* , i.e. fundamental meaning units, the sum of which best describes the underlying concept as the mental representation of a notion.

The relationship between the constituent sems of a semem has traditionally been described in human language as the **informal (i.e. natural language) description** of the term, a statement capturing the *intensions* of the terms. These linguistic elements are typically underlying, albeit oftentimes intuitively, the companion glossaries that are attached for easy reference to most knowledge domain descriptions. With complex concepts, a definition by *extension* is often conveniently used, leaving it up to the intuitive intelligence of the reader to determine the commonalities that constitute the head notion.

### 4.2.2 ... to computational linguistics

For automatic processing, the definition must be converted into a formal representation, i.e. a format that is semantically processable by machines. This requires rigorous and unambiguous analysis of data elements. Data elements (with attributes), have been largely standardized, and markup frameworks have been recommended in various standards. They include:

#### Systems to manage terminology, knowledge and content

- ✔ [ISO 1087-2:2000](#) Terminology work -- Vocabulary -- Part 2: Computer applications
- ✔ [ISO 12200:1999](#) Computer applications in terminology -- Machine-readable terminology interchange format (MARTIF) -- Negotiated interchange
- ✔ [ISO 12620:1999](#) Computer applications in terminology -- Data categories
- ✔ [ISO 16642:2003](#) Computer applications in terminology -- Terminological markup framework

Furthermore, additional information is required to interchange terms across systems. A number of standards have been or are being jointly developed at international level under the aegis of ISO and IEC:

Parts	Description
<a href="#">19763-1</a>	<b>Part 1: Reference Model</b> -- This part of ISO/IEC 19763 describes the concepts and an overall architecture of the metamodel framework standard to be applied in the development and the registration of the following individual metamodel frameworks.
<a href="#">19763-2</a>	<b>Part 2: Core Model</b> -- This part of ISO/IEC 19763 specifies the core model which is required to describe metamodel items, and which may be used in situations where a complete metadata/metamodel registry is appropriate.
<a href="#">19763-3</a>	<b>Part 3: Metamodel for Ontology Registration</b> -- This part of ISO/IEC 19763 specifies the metamodel that provides a facility to register administrative information about ontologies.
<a href="#">19763-4</a>	<b>Part 4: Metamodel for Model Mapping</b> -- This part of ISO/IEC 19763 provides a normative metamodel for describing differences regarding formats and types of objects to be exchanged or shared. This metamodel framework also provides a capability for describing transformation rules between different objects in term of a metamodel instance.

More work has also been done to ensure availability of common platforms at international level, in the form of Metadata Registries (MDR).<sup>1</sup>

Parts	Description
<a href="#">11179-1</a>	<b>Part 1: Framework</b> , introduces and discusses fundamental ideas of data elements, value domains, data element concepts, conceptual domains, and classification schemes essential to the understanding of this set of standards and provides the context for associating the individual parts of ISO/IEC 11179.
<a href="#">11179-2</a>	<b>Part 2: Classification</b> , provides a conceptual model for managing classification schemes. There are many structures used to organize classification schemes and there are many subject matter areas that classification schemes describe. So, this Part also provides a two-faceted classification for classification schemes themselves.
<a href="#">11179-3</a>	<b>Part 3: Registry Metamodel and Basic Attributes</b> , specifies a conceptual model for a metadata registry. It is limited to a set of basic attributes for data elements, data element concepts, value domains, conceptual domains, classification schemes, and other related classes, called administered items. The basic attributes specified for data elements in ISO/IEC 11179-3:1994 are provided in this revision.
<a href="#">11179-4</a>	<b>Part 4: Formulation of Data Definitions</b> , provides guidance on how to develop unambiguous data definitions. A number of specific rules and guidelines are presented in ISO/IEC 11179-4 that specify exactly how a data definition should be formed. A precise, well-formed definition is one of the most critical requirements for shared understanding of an administered item; well-formed definitions are imperative for the exchange of information. Only if every user has a common and exact understanding of the data item can it be exchanged trouble-free.
<a href="#">11179-5</a>	<b>Part 5: Naming and Identification Principles</b> , provides guidance for the identification of administered items. Identification is a broad term for designating, or identifying, a particular data item. Identification can be accomplished in various ways, depending upon the use of the identifier. Identification includes the assignment of numerical identifiers that have no inherent meanings to humans; icons (graphic symbols to which meaning has been assigned); and names with embedded meaning, usually for human understanding, that are associated with the data item's definition and value domain.
<a href="#">11179-6</a>	<b>Part 6: Registration</b> , provides instruction on how a registration applicant may register a data item with a central Registration Authority and the allocation of unique identifiers for each data item. Maintenance of administered items already registered is also specified in this document.

<sup>1</sup> [ISO/IEC 11179, Information Technology -- Metadata Registries \(MDR\)](#) For a more complete description, <http://metadata-standards.org/11179/>, including expected delivery dates of various components.

The European contribution to developments in the field is quite substantial:

ENV	<a href="#">1064</a>	Medical informatics - Standard communication protocol - Computer-assisted electrocardiography
ENV	<a href="#">1068</a>	Medical Informatics - Healthcare information interchange - Registration of coding schemes
CR	<a href="#">1350</a>	CEN Report: Investigation of syntaxes for existing interchange formats to be used in healthcare
ENV	<a href="#">1613</a>	Medical informatics - Messages for exchange of laboratory information
ENV	<a href="#">1614</a>	Healthcare informatics - Structure for nomenclature, classification and coding of properties in clinical laboratory sciences
ENV	<a href="#">1828</a>	Medical informatics - Structure for classification and coding of surgical procedures
ENV	<a href="#">12017</a>	Medical Informatics - Medical Informatics - Vocabulary
ENV	<a href="#">12018</a>	Identification, administrative, and common clinical data structure for Intermittently Connected Devices used in healthcare (including machine readable cards)
ENV	<a href="#">12052</a>	Medical Informatics - Medical Imaging Communication
ENV	<a href="#">12251</a>	Health Informatics - Secure User Identification for Healthcare - Identification and Authentication by Passwords - Management and Security
ENV	<a href="#">12264</a>	Medical informatics - Categorical structures of systems of concepts - Model for representation of semantics
ENV	<a href="#">12265</a>	Medical informatics - Electronic healthcare record architecture
ENV	<a href="#">12381</a>	Health care informatics - Time standards for health care specific problems
ENV	<a href="#">12388</a>	Medical Informatics - Algorithm for Digital Signature Services in Health Care
ENV	<a href="#">12435</a>	Medical informatics - Expression of the results of measurements in health sciences
ENV	<a href="#">12443</a>	Medical informatics - Medical informatics healthcare information framework
ENV	<a href="#">12537-1</a>	Medical informatics - Registration of information objects used for EDI in healthcare - Part 1: The Register
ENV	<a href="#">12537-2</a>	Medical informatics - Registration of information objects used for EDI in healthcare - Part 2: Procedures for the registration of information objects used for electronic data interchange (EDI) in healthcare
ENV	<a href="#">12538</a>	Medical informatics - Messages for patient referral and discharge
ENV	<a href="#">12539</a>	Medical Informatics - Request and report messages for diagnostic service departments
CR	<a href="#">12587</a>	CEN Report: Medical Informatics - Methodology for the development of healthcare messages
ENV	<a href="#">12610</a>	Medical informatics - Medicinal product identification
ENV	<a href="#">12611</a>	Medical informatics - Categorical structure of systems of concepts - Medical Devices
ENV	<a href="#">12612</a>	Medical Informatics - Messages for the exchange of healthcare administrative information
ENV	<a href="#">12623</a>	Medical Informatics - Media Interchange in Medical Imaging Communications
CR	<a href="#">12700</a>	CEN Report: Supporting document to ENV 1613:1994 - Messages for Exchange of Laboratory Information

ENV	<a href="#">12922-1</a>	Medical Image Management - Part 1: Storage Commitment Service Class
ENV	<a href="#">12924</a>	Medical Informatics - Security Categorization and Protection for Healthcare Information Systems
ENV	<a href="#">12967-1</a>	Medical Informatics - Healthcare Information Systems Architecture - Part 1:Healthcare Middleware Layer
CR	<a href="#">13058</a>	Medical data interchange - Mapping between the models specified in ENV 12539:1997 and NEMA PS3 Supplement 10
ENV	<a href="#">13606-1</a>	Health informatics - Electronic healthcare record communication - Part 1: Extended architecture
ENV	<a href="#">13606-2</a>	Health informatics - Electronic healthcare record communication - Part 2: Domain term list
ENV	<a href="#">13606-3</a>	Health informatics - Electronic healthcare record communication - Part 3: Distribution rules
ENV	<a href="#">13606-4</a>	Health informatics - Electronic healthcare record communication - Part 4: Messages for the exchange of information
ENV	<a href="#">13607</a>	Health informatics - Messages for the exchange of information on medicine prescriptions
ENV	<a href="#">13608-1</a>	Health informatics - Security for healthcare communication - Part 1: Concepts and terminology
ENV	<a href="#">13608-2</a>	Health informatics - Security for healthcare communication - Part 2: Secure data objects
ENV	<a href="#">13608-3</a>	Health informatics - Security for healthcare communication - Part 3: Secure data channels
ENV	<a href="#">13609-2</a>	Health informatics - Messages for maintenance of supporting information in healthcare systems - Part 2: Updating of medical laboratory-specific information
CR	<a href="#">13694</a>	Health informatics - Safety and security related software quality standards for healthcare
ENV	<a href="#">13728</a>	Health informatics - Instrument interfaces to laboratory information systems
ENV	<a href="#">13729</a>	Health informatics - Secure user identification for healthcare strong authentication using microprocessor cards
ENV	<a href="#">13734</a>	Health Informatics - Vital signs information representation
ENV	<a href="#">13735</a>	Health Informatics - Interoperability of patient connected medical devices
ENV	<a href="#">13939</a>	Health informatics - Medical data interchange: HIS/RIS-PACS and HIS/RIS - Modality Interface
ENV	<a href="#">13940</a>	Health Informatics - Systems of Concepts to Support Continuity of Care
CR	<a href="#">14300</a>	Health Informatics - Interoperability of healthcare multimedia report systems
CR	<a href="#">14301</a>	Health Informatics - Framework for security protection of healthcare communication
CR	<a href="#">14302</a>	Health Informatics - Framework for security requirements for intermittently connected devices
CEN/ TS	<a href="#">14463</a>	Health informatics - A syntax to represent the content of medical classification systems (CiaML)

More are at various phases of development:

METAVOC	Health informatics -Vocabulary for terminological systems
T&C-base	Health informatics - Vocabulary - Maintenance Procedure
CLAML	Health informatics - A syntax to represent the content of medical classification systems
METKNOW	Health informatics - Clinical knowledge resources - Metadata
CATANAT	Health informatics - Categorial structure for anatomy
CATFIND	Health informatics - Categorial structure for documentation of patient findings and problems
CATERMCON	Health informatics - Categorial structure of representation of conditions, coding systems and clinical terminologies

The languages used to express relationships between data elements is usually based on *description logic (DL)*, i.e. a formal logic that represents a compromise between expressivity and decidability. Several languages have been developed over the years, increasing in sophistication, expressivity and powerfulness, or a balanced mixture of these. The range of description logic languages is growing. Standards organizations have identified the need to have a common framework for future developments. This resulted in more standards being developed such as [ISO/IEC 24707 Information technology -- Common Logic \(CL\) -- A framework for a family of logic-based languages<sup>2</sup>](#), currently at development stage 40.99<sup>3</sup>.

### 4.2.3 Capturing domains of knowledge: the conceptual evidence

Corpus-based linguistics has harnessed computer power to analyze vast amounts of text materials that suddenly became more accessible. Applied to the literature of a specific scientific domain, descriptive investigations and other statistical processes generated an inventory of domain terms (or *terminological tokens*), increasing dramatically the breadth and depth of the domain coverage of existing collections [Nazarenko et al. 2001], simultaneously improving collections through disambiguation [Liu 2002]. The end result, however is hardly manageable for practical purposes<sup>4</sup>.

Alternatively, reductionist solutions are proposed, e.g. in the form of *controlled vocabularies*. These are easy to enforce with the use of pick lists. They are largely unsatisfactory, as argued by [Rogers 2006]. Another, tiered approach, based on a utility principle, pervades the literature. Communities of practice in the domain in point reportedly draw on subsets of the domain

<sup>2</sup>

<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39175&ICS1=35&ICS2=60&ICS3=&scopelist=PROGRAMME>

<sup>3</sup> For an overview of the international harmonized stage codes, see <http://www.iso.ch/iso/en/widepages/stagetable.html>

<sup>4</sup> Several aspects of such terminologies may explain the lack of manageability: size is dissuasive for human processing; relevance of terms and currency of definitions cannot easily be ascertained. Machine-processing allows for massive data processing, but the structure of information behind the words is either unknown, or inadequate. Conformity with standards is often limited, for unclear or unconvincing reasons.

terminology, or use terms at various levels of analytical granularity for practical purposes<sup>5</sup>. Clinical laboratory terminology, for example, will be different for the research community, the specialist in a tertiary hospital, the general practitioner, the nursing profession or the general public. Classifications of medical procedures may include a variable number of entries, ranging from some 20 000 (like in OPCS4) to be used in resourced health systems, to some 2000 in a proposed Condensed Classification of Health Interventions, for use in resource-limited settings. Special applications may be limited to an even smaller number of categories, for instance in hospital data monitoring. It is therefore important to ascertain that actual products are tailored to the needs of corresponding use cases, regardless of the necessity (all too often ignored) to ensure that they properly capture behind the scene the complexity of their study field.

The actual meaning assigned to the domain terms hinges on the view point of domain experts describing the reality they deal with. Short of entering a philosophical debate, reality is interpreted subjectively by domain experts, who organize information about it in the most suitable and efficient manner. This information is anchored on terms that are willfully selected and put in a meaningful relationship to one another in what is referred to as an **ontology**. While some researchers may claim that an ontology should describe the objective reality, others have declared it impossible. Pragmatically, ontologies are declarative (data-driven) *expressions of an agent's "world"*: the objects, operations, facts and rules that constitute the logical space within which an agent performs [BAILIN et al. 2002b]. This view implies that several agents may wish to describe their "world" differently. Within one domain, there may therefore be several interpretative representations or explanatory attempts that would become varied ontological representations of that universe portion. Even on one single study field, it is therefore essential to ensure that the representation makes sense to the human users confronted with it.

Aiming for a standardized ontology now seems inadequate, as it will not do justice of the variety of view points, known or still to come, that are congenial to scientific work. The focus of research should rather be on the interoperability of multiple ontology representations. Further investigations are required to delineate the most appropriate boundaries of concept descriptions, including the intrinsic features that will be carried along with the labeling term wherever it is used and the extrinsic or relational features, that will be crucial for the functioning of the term in specific environments.

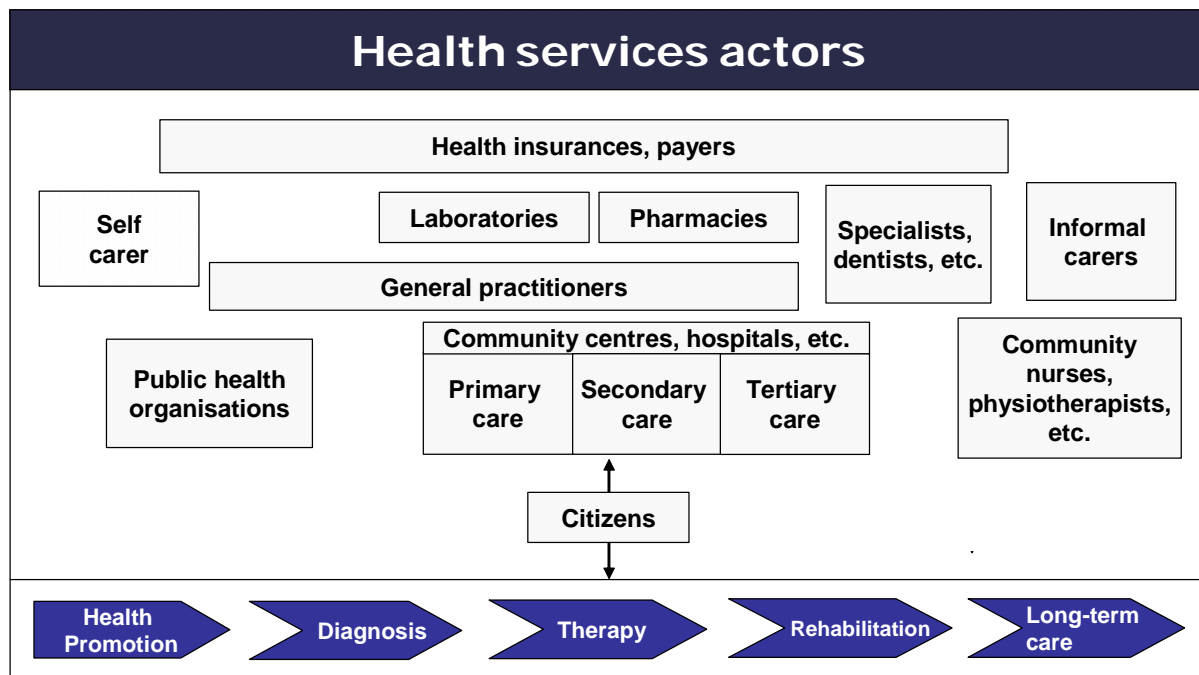
#### 4.2.4 An actor-centric perspective: conceptual framework relevance

If the ontological representations of health-related domains is to be true to any agent's world vision, an inventory of the relevant actor categories become indispensable, together with a clear description of the functions they perform and the role they play individually performing those functions in the health systems. A large number of actors have been identified, including the citizen patient, who is often confronted with many other actors interacting with a series of processes addressing health condition at various stages. For all and every one of those actors, it is essential that the conceptual proposed framework in this field **make a lot of sense**.

The Deliverable 1.2 review of existing information systems at country level visually represented a number of recognized actor categories, along with the processes they are involved in.

---

<sup>5</sup> Here too, manageability by humans appears to be a determining factor; the step to machine processing is often discouraged by poorly ergonomic interfaces, additional overhead for the users, and unconvincing rationalization.



**Figure 1 Illustrative map of actors and processes in the health value system**

Traditionally, actors in the health sector have been grouped in a conveniently small number or categories for easy reference in operational terms. In linguistic terms, however, neither the *intensions* nor the *extensions* of such terms are formally described, let alone the relations between them. If we assume the three broad categories enumerated in Deliverable 1.2,

- Healthcare authorities
- Healthcare professionals
- Patients/Consumers/Citizens Associations

it is impossible to develop a commonly agreed narrative of the functions, roles and tasks that are expected of health professionals. One would instead have a list of tasks that may (or may not) be assigned to them. WHO has long been confronted with the definition, for instance, of the health professions (*intensions*). An comprehensive list of all health professions that may be subsumed under healthcare professionals will probably continue to be elusive, the correspondence of roles across countries is less than perfect and will be changing with time. Yet how do we capture the reality of those key actors if we want to have a reasonably stable health information systems.

Conversely, similar tasks may be entrusted to different professionals in different countries and in different environments. The nomenclature may expand indefinitely, the same difficulty is encountered: there are as many lenses to see reality as there are actors interacting with it.

Citizens, patients and informal carers; physicians and other healthcare professionals; hospitals and other healthcare provider entities or organizations; pharmacies, pharmaceutical industry; central, regional or local government representatives, including health policy makers, sometimes third party payers or owners of health care facilities, public and private health insurance organizations and systems, (home) care management service providers, case managers, all these actors have roles to play. Exchanging information about them, from simple counting to recruitment, from performance assessment to impact evaluation, from cost-effectiveness considerations to patient safety concerns, from an employment equity to a gender equality perspective, required clear definitions against which any analysis can be tested, and not only with respect to the intrinsic nature of those actors, but pragmatically with reference to the contextual system in which they operate.

It is therefore necessary to have a machine-processable **formal representation of actors perspectives**, formally equipped with adequate connector identifiers, associated with permissible values for *ad lib* plug-and-play interoperability in a manner similar to bios in computers. In this manner, ontology-matching issues might be restricted to properly defining **low-level bios characteristics** rather than pondering on the entire structure of complex systems.

#### 4.2.5 Use cases: the conceptual framework performance

Use cases are sets of constraints that must be met for operational reasons. They represent the true measure of the kind of information that is needed for particular purposes. It is therefore absolutely crucial to define clearly, explicitly and formally the characteristics of the use case, including its boundaries or limits. They would act in the information world as the technical specifications prescribed in engineering work long before the knowledge engineers start their work. With respect to health information, there is no guarantee that data elements collected or worked out prior to the actual definition of a use case will be adequate as long as the match between them has not been irrevocably and positively assessed and confirmed.

Superimposing an ontology on a collection of concepts that doesn't have one and put together for unrelated or undefined purposes is in violation of conceptual framework integrity. Similarly, applying a set of terms in an unspecified or unknown relation, based on a superimposed ontology, onto an unforeseen use case is likely to result in misfits. The mere existence of a terminology or of an ontology doesn't necessarily mean that either can be used for a particular purpose. This has to be evaluated and ascertained with adequate processes supported by quality assurance methodology.

**Operations research** as an interdisciplinary science which deploys scientific methods like mathematical modeling, statistics, and algorithms to decision making in complex real-world problems<sup>6</sup> could provide insight into the terms and conditions to be met in the health information area for optimum benefit. The nature of organization is essentially immaterial. The eventual intention behind using this science is to elicit scientifically a best possible solution to a problem, which improves or optimizes the performance of the organization. This could provide an opportunity to adapt true solutions to actual problems, rather than the other way round.

It would be presumptuous to attempt to imagine a conceptual framework that would solve all problems at once. During the Copenhagen workshop (see Annex 1 Report of the 2006-05-30 Copenhagen Workshop) the practicality of use cases was recognized. To that end, it was found preferable to address actual use cases rather than theoretical ones, as is often done for modeling purposes. The proposed approach would be to identify those co-operable use cases that seem to offer the best opportunities in the short to medium term, be it with regard to potential impact on problem solving or to optimization of resources to be allocated. Building on the strong drive to define adequate electronic health records, for instance, additional consideration should be given to facilitating aggregation of individual health record data to inform the emerging need to derive accurate, evidence-based public health information that would guide decision making at the highest policy level. Similarly, proper recording of hospital data on procedures would permit to infer useful information for optimum allocation of resources in the health system, including in the perspective of their evolution.

---

<sup>6</sup> [http://en.wikipedia.org/wiki/Operations\\_research](http://en.wikipedia.org/wiki/Operations_research)

## 4.3 Operational parameters of semantic interoperability

### 4.3.1 Across time and across countries: what about across domains and across languages

As described in Deliverable 1.2, various generations of solutions to improve interoperability have been hypothesized.

1. Everyone adopts *a single model*
2. Everyone has its own model but interchange standards need bilateral/multilateral agreements.
3. Everyone agrees about *common data elements* with systematic unambiguous format e.g. data descriptions (data types, terminologies, coding), metadata, information model.
4. Everyone uses *knowledge representation framework* (classes, attributes, definitions, identification principles) and *inference mechanisms* (inclusions, exceptions, constraints, reasoning, etc.)
5. Everyone adopts the higher layer of a tiered system of ontologies that would guarantee the overall consistency of subsequent developments.

The adoption of a **single model** at international level (even if implemented in various applications) has been elusive and short-lived. Consensus-based products endorsed by the highest authorities at international level have failed to be adopted uniformly by all countries. Only where a unique model can be imposed by regulatory authorities has it come to life, in spite of serious doubts and resistance in the scientific community of the countries concerned.

A number of countries have adopted ICT-for-health policies. The immediate advantages for national health care management are obvious in terms of harmonization of procedures across the country, which in turn creates the conditions for efficiency of the resource allocation and reimbursement processes. Even when a national institution has been designated to develop a framework for the national system, reusability of information thus collected or generated is seldom put forward. The guiding principle is essentially one of sectoral efficiency, cost effectiveness or optimization of resource allocation, rather than one of message content integrity across systems. Reasons usually put forward given include lack of specificity in different environments, inadequacy for locally intended purposes, incompatibility with other elements of the systems, non localization of the language or terminology, opacity of the underlying structure or information model.

The second option based on **interchange layers** has been used extensively to make up for the disharmony between particular applications. Even in the central area of its activities, morbidity and mortality statistics, WHO has been confronted with the emergence of national so-called clinical modifications. In order to preserve the compatibility between the versions and the comparability of data collected under them, the preparation of a reference interchange layer is now envisaged. These solutions, however, rely mostly on negotiated declarative agreements of compatibility between originally independent items. The probability of a true congruence is therefore fairly limited: it may be sufficient in particularly stable areas, such as the message-driven processing of reimbursement claims, but less satisfactory in user-driven exchanges, e.g. for clinical purposes.

The third option around the adoption and clear definition of **common data and metadata elements** seems to have been the most developed so far, particularly at the initiative of standards development organizations and scholarly societies. Much progress has been achieved, but on the machine-processing aspects and intimate knowledge of the many parameters at play.

**Knowledge representation** standards, **ontology** metadata and **inference engines** have opened up new avenues for an information processing that is closer to reality. Under the impulse of the W3C community, semantic web opportunities seem to have emerged at the

global level. The Consortium is jointly administered by the [MIT Computer Science and Artificial Intelligence Laboratory](#) (CSAIL) in the USA, the [European Research Consortium for Informatics and Mathematics](#) (ERCIM) (in [Sophia Antipolis, France](#)), and [Keio University](#) (in [Japan](#)), which gives a truly global resonance to the collective effort.

OWL has attracted much interest in the last decade, due to its clearly defined syntax, its ability to process semantically relevant information. Many ontologies have been modeled using OWL and ontology editors are available on the market, that are producing outputs in a variety of formats, including for OWL. This is conducive the development of **inference engines** and other **reasoners** that will form the basis of decision support, warning schemes and further knowledge developments. Based on **description logics** a range of commercial or non commercial inference engines are available today, reasoning at class level or at instance level.

Opinions are now being voiced to refine the original monolithic view of ontology-based models. For all practical intents and purposes, the feasibility of using a more stratified model is being discussed in the scientific community. The view is that of a very generic, **top-level ontology** that would cement the commonality of vision and perception. It would address such elements as the occurments and continuants, objects, processes and functions. The underlying principle would then serve as a basis for domain-specific **foundational ontologies**. At a more operational level, **application ontologies** would be tailored to specific tasks or communication scenarios, yet relate to the higher level ontologies for consistency.

## 5 Complex technical issues come to light

### 5.1 Declarative, procedural and inferential representations

Scientific literature is the most familiar repository of science in progress. Books are the more stable portion of it, journals the more volatile and rapidly changing. Supplemented with grey literature and other materials, they constitute a vast **corpus** from which knowledge can be extracted.

Terminology and lexicography research has been conducted for decades to elicit snapshot views of the knowledge expressed in a corpus at a given point in time. From there, elements of knowledge have been identified and subsequently arranged in two main kinds of particular representations: declarative representation and procedural representations<sup>7</sup>.

Declarative representations are static, capturing knowledge about objects, events *etc.* and their instantiated relationships as the surface or **upper layer** of the terms. An extraneous program is then required to decide what to do with the encapsulated knowledge and how to do it.

Procedural representations on the other hand are such that control information necessary to use the knowledge is embedded in the knowledge itself. *e.g.* how to find relevant facts, make inferences *etc.* Rather than a program, they require an interpreter to follow the instructions embedded in the knowledge construct. They refer to the **underlying nature** of the terms, addressing the rules that seem to be embedded under the surface.

Given the visible characteristics of knowledge constructs and the underlying properties associated with them, which describe how they operate in real world settings, those **semantic objects** include the necessary **features and “methods”** of basic building blocks of knowledge.

Adding the **inference rules** to operationalize them further, new entities emerged as **bottom up** sophisticated constructs contributing to more complex knowledge engineering schemes.

---

<sup>7</sup> <http://www.cs.cf.ac.uk/Dave/AI2/node51.html>

In spite of all these promising developments, knowledge representation soon reaches its limits. How is it that systems around the world continue to show so little semantic interoperability ?

## 5.2 Context-dependence of meaning

The modern health care enterprise lies at the nexus of several different industries and institutions. Within a single hospital, different departments (e.g. internal medicine, medical records, pharmacy, admitting, billing) maintain separate information systems yet must share data in order to ensure high levels of care <and cost-effective management of records>. Medical centers and local clinics not only collaborate with one another but with State and Federal regulators, insurance companies, and other payer institutions. This sharing requires reconciling differences such as those of procedure codes, medical supplies, classification schemes, and patient records. [Lee P. W. 2003b]

The claim that existing terminologies and existing knowledge representation systems are not suitable for use outside their original context stresses an underestimated factor with considerable impact on applications, namely that of context-dependence of meaning. One would have to decide whether domain ontologies must be inferred from the properties and methods of the entities populating their space, or whether their universe includes relational, functional features in which objects could play certain roles if they have the appropriate docking knowledge configurations, regardless of other docking configurations that would allow them to be used in different contexts, therefore in different ontologies.

As a result, vast territories of applied knowledge representation might well revolve around contextual ontology research. This pre-conditional approach has been used in ensuring interoperability of financial and industrial information sources from around the world, i.e. developed in different contexts. [BAILIN et al. 2002a] presents an overview of different methods for resolving ontology mismatches and motivates the Ontology Negotiation Protocol (ONP).

[Lee P. W. 2003a] has reviewed the requirements to be met for the representation of knowledge with appropriate characteristics and metadata for context mediations. Assuming that context are by nature different, the ontology-based visions of the respective universes are *per force* different. Interoperability can only be ascertained if the knowledge represented by the ontologies of the two systems can be interchanged. The COntext INterchange Mediator had been suggested in the late 1990s, and [Bressan, S. 1997] has proposed a practical application in Prolog. The Context Interchange strategy addresses those concerns with an emphasis on resolving problems arising from semantic heterogeneity, i.e. inconsistencies arising from differences in the representation and interpretation of data. This is accomplished using three elements: a shared vocabulary for the underlying application domain (in the form of a domain model), a formal object-relational data model (COIN), and an object-deductive language (COINL). Semantic interoperation is accomplished by making use of declarative definitions corresponding to source and receiver contexts; i.e. constraints, choices, and preferences for representing and interpreting data. The identification and resolution of potential semantic conflicts involving multiple sources are performed automatically by the context mediator. Users and application developers can express queries in their own terms and rely on the context mediator to rewrite the query in a disambiguated form. [Bressan, S. et al. 1998] uses a context-oriented linguistic descriptor equipped with corresponding language (COINL) This has found applications in other fields such as financial information management in heterogeneous context systems.[Firat et al. 2002]

Given the large number of possible actors and corresponding information sources that have to be made truly interoperable in the health sector, there is a need to look for distributed solutions development. The ARGUGRID<sup>8</sup> model should be examined with a view to serve

---

<sup>8</sup> Partners in that project include the following:

- [Department of Computing, Imperial College London, UK](#)

possibly as a model for programming the Grid at a semantic, knowledge-based level of abstraction through the use of **argumentative agent technology**. Agents are associated with service/resource requestors and service/resource providers on the Grid. Argumentation technology is used to support rational decision making, internal to agents, as well as negotiation, among agents, all required facilitating the dynamic composition of Grid services/resources into executable workflows. Argumentation is also used to support the creation, management and dynamic evolution of virtual organizations, understood as societies of agents, to compose individual services into more complex ones.<sup>9</sup>

### 5.3 Finite state semantics versus semantic emergence

One common feature of present-day knowledge representation systems here is that they capture snapshot views of information situations, i.e. they fail to convey the continuity of intelligent information sharing.<sup>10</sup> It is an anomaly that it attempts to depict an evolutionary development with a series of finite state semantic snapshots. Considering that any given complex system has more value than the total sum of its individual components, the information universe with which the health information system is confronted can only be meaningfully represented if the non-finite meaning relationships based on contextual semantic emergence are duly captured.<sup>11</sup>

By extrapolation, concepts can be viewed as the atomic elements of information. Not unlike the periodic table of chemical elements, concepts constitute the elementary building blocks which can be put together in various combinations resulting in a series of constructs that are equipped with new potentialities depending on their use context. At the same time the way those elementary building blocks can be put together is determined by the sub-atomic semantic primitives which include both semantic particles and the organizing semantic forces [Fabre F. 2005]

There is therefore an entire field of research that should be devoted to defining the semantic primitives of individual concepts in a kind of **bottom down** approach. The evidence provided by the contextual use of concepts would support the construction of a **functionality map repository**. A variety of theoretical frameworks exist, for instance those based on Explanatory and Combinatorial Lexicology (ECL) and cognitive semantics (Mel'cuk<sup>12</sup> and the biomedical applications<sup>13</sup>).

Trying to include the functionality potential of concepts as one of the semantic primitives very much limits the capacity to be creative in the use of various concepts. It may reasonably be suspected that what is required of systems is how concepts are put together and operationalized, the actual notions are to play being defined in other contexts. Why do medical doctors probably agree as to what they need to know in recording or retrieving

- 
- [Department of Computer Science, Royal Holloway, University of London, UK](#)
  - [Dipartimento di Informatica, Pisa University, Italy](#)
  - [Institute of Communication and Computer Systems, National Technical University of Athens, Greece](#)
  - [School of Engineering and Technology, Asian Institute of Technology, Thailand](#)
  - [InforSense LTD UK](#)
  - [GMV S.A., Spain](#)
  - [cosmoONE Hellas Market-site S.A. Greece](#)

<sup>9</sup> <http://www.argugrid.org/index.php>

<sup>10</sup> For theoretical considerations and applications to physics, biological or sociological systems, the seminal book by Nobel Prize Laureate [Laughlin, R. M. 2006]

<sup>11</sup> <http://www.automatesintelligents.com/echanges/2007/fev/emergence.html>

<sup>12</sup> <http://www.olst.umontreal.ca/melcuk/index.html>

<sup>13</sup> <http://www.olst.umontreal.ca/biomedeng.html>

information on a patient, or a health situation, they hate to be constrained in selecting the terms the system would use (preferred terms, controlled vocabulary, pick list) rather than the phraseology they are familiar with.

In terms of representation, one could imagine differential sets of semantic primitives around the semantic core common to all application fields. That could constitute the essence of a top-level, or core, ontology. Additional information could then be supplied to form more customized systems. The emergence of the Internet as the de facto Global Information Infrastructure enables the construction of decision support systems that leverage the panoply of on-line information sources. This highly dynamic environment underscores a critical need for a flexible and scalable strategy for integrating the disparate information sources while respecting their autonomy.

## 6 Conclusions and recommendations

*Whence, then, do my errors arise? Only from the fact that the will is much more ample and far-reaching than the understanding, so that I do not restrain it within the same limits but extend it even to those things which I do not understand. Being by its nature indifferent about such matters, it very easily is turned aside from the true and the good and chooses the false and the evil. And thus it happens that I make mistakes and that I sin.<sup>14</sup>*

*René Descartes, Meditations on First Philosophy, Meditation Four, "Of the True and the False" (cited in [Howard 2003])*

The search for semantic interoperability has so far been elusive, the demeanor to achieve it oscillating between acknowledgment of complexity and the urge to implement practical solutions. Many have been successfully implemented, either as versatile intelligent algorithms that can be adapted in a variety of contexts, or applied to broader schemes, or exploring deeper layers of knowledge.

Theoretical considerations have been developed: they help make sense of what we see, in an increasingly sophisticated manner: they have names theory of chaos, fractals, complex systems theory, etc. At the same time, ad hoc solution are being developed in very focused contexts: short lists, sentinel lists, light systems, relational systems, distributed solutions. They are all fine and useful but largely not interoperable, for they do not take into account the complexity of the substrate.

[Rogers J. 2004]has recognized the dilemma in practical terms. *Medicine needs useful formal ontologies, but formal ontologies that are simple to use are not useful, while useful ontologies appear to be too complex to be directly useable.* He went on to propose a methodology whereby the complexity of ontologies may be harnessed in order to construct a simpler interface with the users. Whether this is referred to as context interchange mediators, argumentation grid, of user interface doesn't make much of a difference.

In order to approach semantic interoperability it is therefore recommended to:

1. Revisit aspects of knowledge representation with a view to taking advantage of progress made and **solutions adopted in other domains** also confronted with managing dynamically information emanating from **complex systems**;

<sup>14</sup> D'où est-ce donc que naissent mes erreurs? C'est à savoir, de cela seul que, la volonté étant beaucoup plus ample et plus étendue que l'entendement, je ne la contiens pas dans les mêmes limites, mais que je l'étends aussi aux choses que je n'entends pas; auxquelles étant de soi indifférente, elle s'égare fort aisément, et choisit le mal pour le bien, ou le faux pour le vrai. Ce qui fait que je me trompe et que je pêche.

2. To that end, explore the merits of new approaches, combining a **top down** conceptual framework approach that would be conducive to defining rules generally applicable, and a **bottom up approach** identifying the more promising use cases, in terms of potential impact and best return on investment.
3. In that perspective and in addressing those domain-specific issues, **identify gaps in knowledge engineering**, such as those dealing with context dependence, semantic emergence, interchange protocols, including in cross-cultural and multilingual environments.
4. Work out the conditions to capitalize on the promising technological developments, such as the grid technology, that would unfold the processing power of machines and networks to **new frontiers**.
5. Finally, invest in **simplified** (rather than simplistic) **graphical representations of complex systems**, in order to serve all actors in the health field with means **to navigate through oceans of knowledge and to face emerging situations**, in a manner that **makes sense at their finger tips**, without compromising the quality of the information thus produced.

## 7 Bibliography

BAILIN S. C., Truskowski W 2002a. Ontology negotiation between intelligent information agents. The Knowledge Engineering Review 2002a;17(01):7-19.

BAILIN S. C., TRUSZKOWSKI W 2002b. Ontology negotiation between intelligent information agents. The Knowledge Engineering Review 2002b;17(01):7-19.

Bressan S. Overview of a Prolog Implementation of COntext INterchange Mediator. Sloan School of Management, Massachusetts Institute of Technology; 1997.

Bressan S, Cheng HG. Answering Queries in Context. 1998.

Bressan S., Ibrahim IK 1999. Semantic Query Transformation for the Integration of Autonomous Information Sources. Proceedings of the 12th Int'l Conference on Application of Prolog (INAP' 99) 1999.

Fabre F. Emergence et représentation 2005.

Firat A., Madnick S, Grosz B 2002. Financial Information Integration in the Presence of Equational Ontological Conflicts. Proceedings of the Workshop on Information Technology and Systems (WITS), Barcelona, Spain 2002;211-6.

Hébert L. Petit glossaire du sémanticien. 1993.

Howard D. 2003. Reduction and Emergence in the Physical Sciences: Some Lessons from the Particle Physics/Condensed Matter Physics Debate. 2003.

Laughlin RM. A Different Universe: Reinventing Physics from the Bottom Down. Basic Books; 2006.

Lee F, Bressan S, Ooi BC. Global Atlas: Calibrating and Indexing Documents from the Internet in the Cartographic Paradigm.: IEEE Computer Society Washington, DC, USA; 2000.

Lee PW. Metadata Representation and Management for Context Mediation 2003a.

Lee PW. Metadata Representation and Management for Context Mediation 2003b.

Liu H. 2002. Corpus-based Ambiguity Resolution of Biomedical Terms Using Knowledge Bases and Machine Learning. 2002.

Nazarenko A., Zweigenbaum P, Habert B, Bouaud J 2001. Corpus-based Extension of a Terminological Semantic Lexicon. Recent Advances in Computational Terminology 2001;327-51.

Rogers J. Development of a methodology and an ontological schema for medical terminology 2004.

Rogers J. E. 2006. Quality assurance of medical ontologies. *Methods Inf Med* 2006;45(3):267-74.

## **8 Annex 1 Report of the 2006-05-30 Copenhagen Workshop**





Semantic Interoperability  
RTD roadmap project  
EU IST 6th fw 27328



Document title Report for 2006-09-30 Copenhagen workshop

Original author Jean marie Rodrigues

Document type

Version

2

Changed by

Change date

2006-10-15

## Report for 2006-09-30 Copenhagen Workshop



## **Plan**

### **Executive summary**

### **1 Organisation**

### **2 Requirements for the workshop**

### **3 Minutes**

### **4 Conclusions**

### **5 Annexes**

## 8.1 Executive Summary

SemanticHEALTH consortium held its first workshop in Copenhagen Hotel Clarion on 2006/09/30. All the partners attended as well as invited experts from Australia, Germany, Sweden, USA, UK (see chapter 1 Organisation).

The goal of the workshop was to identify the main issues and problems (and not the solutions) challenging the semantic interoperability (SIOP) of the e-health initiatives across the world

Each work-package leader introduces the issues in a requirements paper distributed before the workshop (see chapter 2 Requirements for the workshop and chapter 5 Annexes) and each work-package was discussed and the issues summarised (see chapter 3 Minutes)

The workshop was very successful and a commonality of thinking was set not only across the different work-packages but as well between members of the consortium and outside experts some of them coming from non EU countries.

As a conclusion very new statements about SIOP were proposed and approved by the participants (see chapter 4 conclusion)

Among them 3 were selected to assess the EU and international e-Health initiatives as semantic specificity items in deliverable 1.2 in addition to the common conceptual framework of Semantic HEALTH as defined in deliverable 1.1:

1. **The first semantic specificity item is to start such experiences from use cases or scenarios**
2. **The second one is to link the field of direct care to the patient to public health and secondary use**
3. **The third one is to integrate terminologies with ontology and multilingualism.**

The most important issues and problems to deal with during the roadmap are

1. **Full semantic Interoperability SHALL NOT BE a generic characteristic of e-health initiatives: it shall be restricted to the smaller number of problems creating the most trouble**
2. **The first priority of SIOP is to develop a public health layer or a public health record (for a population of 1 or 2 millions) on the top clinical information (meta system)**
3. **To insure the semantic character of interoperability it is essential to develop this century ontology driven tools (human investment, semantic web and social computing approaches, open process and open products) and to optimise the satisfaction of multi lingual needs**
4. **On the midterm and under the assumption that it is a priority problem as defined in 1 it is necessary to modularise Snomed (scalability) and to reengineer Snomed with ontology:**
5. **Large scale pilot demonstrators are the only way to convince real health care professionals**

## 8.2 Organisation

### 1.1 Location:

CLARION HOTEL COPENHAGEN  
Molestien 11  
DK 2450 KOBENHAVN SV

### 1.2 Time:

09:30-17:00

### 1.3 Agenda

9:30	Welcome, Introductions (P.Zanstra)
9:45	Setting overall goal of the workshop (J.M Rodrigues)
10:00-10:45	WP4. The clinical setting / Electronic Health Record (D.Kalra)
11:00-11:45	WP5. Public Health and secondary use of clinical data (B Ustun)
12:00-12:30	Lunch
12:30-13:15	WP3. Socio-economic issues(P Zanstra)
13:30-14:15	WP2. Technological issues(P Lewalle)
14:30-15:15	WP6. Biomedical ontologies (A Rector)
15:30-17:00	Concluding discussion(J.M Rodrigues)

## 8.3 Requirements for the workshop

### 8.3.1 Project summary

To efficiently implement e-Health to meet the rising needs of mobile citizens, patients and providers, the fragmented interoperability initiatives now distributed throughout different EU and Member State programmes must come together. They must be coordinated with the increasing need to link clinical data to information from basic biological sciences and evidence of best clinical practice.

Considering the need for interoperability at the Member State and cross-border level of the European Union – as expressed in the EU e-Health Action Plan – and for global interoperability – as represented by WHO – *it is necessary to embark on a process that will prompt the divergent initiatives to join forces for the benefit of all citizens.*

This *SemanticHEALTH* SSA develops a European and global roadmap for deployment and research in health-ICT, focusing on semantic interoperability issues of e-Health systems and infrastructures. The roadmap will be based on consensus of the research community, and validated by stakeholders, industry and Member State health authorities. It

- identifies key short-term (2-5 years) and medium-term (4-10 years) needs to achieve semantic interoperability of e-Health systems (including issues of nomenclatures presently in use, classifications, terminologies, ontologies, EHR and messaging models, public health and secondary uses, and decision support, their relationships, mapping needs, limitations)
- analyses unsolved issues arising in the context of realistic approaches to priority clinical and public health settings (reflecting on models of use, benefits expected, concrete application experience and lessons learned; relevance of open source model)
- takes account of the impact of non-technological (health policy, legal, socio-economic) aspects
- reflects and integrates results of related FP6 (*eHealth ERA*, *i2-Health* and other) studies.

The consortium and associated experts represent centres of excellence from four continents and the WHO.

### 8.3.2 Workshop Background information

This project is about an R&D roadmap. Recommendations will be about immediate solutions, and areas requiring further study, standardisation, and alignment. It is always tempting to dwell on discussions on possible solutions. The eHealth community has proven to be good at that.

This workshop is NOT about proposing solutions, but finding ways in which these can most effectively be identified, developed and adopted in the near to medium term. It is about identifying the key processes that should be in place to indeed enable 'useful and useable' methods of sensible communication between people and machines in a multi-cultural and multi-lingual context. As a roadmap is needed for planning a journey, also this discussion has to pay attention to the order in which identified processes need to be in place. 'What comes first', 'What do we not understand sufficiently', 'What are the barriers', are typical questions to ask.

This project addresses applications in two major focal points, i.e. 'The clinical setting / Electronic Health Record', and 'Public Health and secondary use of clinical data'. We intend to start the discussion from this user end. From there we will deal with the necessary Technological and Socio-economic issues that need to be resolved.

At the end of the day we will wrap up loose ends and upcoming new issues of the WP's discussed earlier on. Finally we will synthesize and prioritize the suggestions of the day.

As a preparation for the discussants, the following pages contain small descriptions of each of the relevant Work Packages, and a listing of critical issues identified so far. Are these the real problem areas, are we complete (we know we aren't), what is the urgency. The WP descriptions follow the order of the agenda.

## 8.4 Discussion paper for Work package 4 EHR

This pre-workshop paper outlines some of the issues and challenges that may need endorsement, investment or funded research towards achieving semantic coherence and interoperability of health record information in support of direct patient care (including quality, governance and risk management), and the secondary uses of EHR data for service management, public health, research and education.

### *Objectives of Semantic Health Work package 4*

- to identify the requirements for semantic coherence in direct patient care and personal health records
- to evaluate the existing formalisms, standards and tools used in (or by) clinical and EHR systems
- to identify the challenges faced in real settings in achieving semantic interoperability
- to prioritise key changes that could make a difference to realising semantic interoperability in the clinical workplace

### **What are the semantic interoperability issues relating to the EHR?**

*What are the situations (scenarios) in which the sharing or processing of EHR data requires explicit and interoperable interpretation of clinical meaning?* On the one hand, human interpretation of clinical notes and correspondence on paper has for decades been sufficient to meet most shared care needs without the use of formalised record structures or terminology. On the other hand, prescribing systems require access to comprehensive allergy, diagnosis and past medication data from the EHR in a processable form in order to offer safety alerts to the prescriber. Other scenarios include the use of electronic guidelines and care pathway systems, and clinical audit systems.

*What are the gaps in the current capability of EHR systems and tools such as terminology and guideline systems to enable these scenarios? Are current systems good enough? Will SNOMED-CT and its capability to combine terms into compound expressions be sufficient to ensure that clinical meaning is unambiguously interpretable across systems and countries? Have existing products or forthcoming standards already addressed outstanding gaps? Has each eHealth programme a suitable strategy to ensure safe record sharing within its national network?*

*Are there particular interoperability issues that do not yet have a sufficiently complete solution, or where theoretical solutions cannot achieve a sufficient momentum to be put into practice?*

### *Further issues for debate:*

1. "Have eHealth programmes pushed the healthcare IT industry away from the needs of clinical practice and stifled the piloting of innovative products, in favour of evidence-free pipe-dreams?"
2. "Have HL7 version 3 and SNOMED-CT each promised more semantic coherence and interoperability than they can reasonably yet deliver?"

## 8.4.1 Key questions regarding Clinical settings/ EHR

Significant progress has been made in many areas of health informatics that contribute to achieving EHR semantic interoperability, and other areas are recognised as needing further quite focussed (i.e. not blue sky) research. These have been grouped according to the kind of action that is perhaps now needed in order to deliver a useful solution. For each one, the key questions are:

- what are the barriers to making further progress?
- what approaches might be taken to accelerate progress?

### 1. Areas needing adoption

(by national eHealth programmes, and therefore by industry, and ideally internationally)

- standardised sharing of clinical data structure specifications – e.g. to agree to use archetypes
- to agree on a generic model for EHR communications e.g. CEN/ISO 13606-1

### 2. Areas needing wide-scale evaluations

(results exist, but need refinement and real clinical use, to determine best practice)

- archetype design and sharing
- SNOMED-CT sub-setting and term co-ordination
- privacy and access control for EHRs

### 3. Areas needing international standardisation

(needs facilitation and funding)

- Data types
- Generic EHR access services

### 4. Areas needing investment

(business cases not strong enough for industry, but products needed: maybe sponsored open source?)

- archetype & template editors
- terminology servers and browsers

### 5. Areas needing further (focussed) research

- Archetype indexing (ontology)
- term binding to archetypes and record structures
- linking EHR data to educational materials and clinical evidence, to enable consumer engagement and support health professional training
- Evaluations of citizen and clinical acceptance of shared EHRs

## 8.5 Discussion paper for WP 5: Public Health and secondary use

Public Health Information Systems compile various information elements stemming out from the health services and related sources to produce meaningful information and knowledge about health at population level. The "information rubrics" commonly used in Public Health as indicators are:

- Demographic Information
  - Birth registration (e.g. live births, stillbirth etc)
  - Mortality (e.g. causes of death, etc)
  - Morbidity (e.g. disease statistics, hospital discharge summaries, etc)
  - Disability (e.g. rates of impairments, activity limitations etc)
  - Risk Factors (e.g. smoking, obesity, environmental exposures etc)
- Health Care facilities
  - Health Settings (e.g. type, facility, human resources etc)
- Health System Parameters
  - Coverage (e.g. key interventions: immunization, antenatal care, eyesight etc).
  - Patient Safety (e.g. adverse reactions, errors, etc)
- Financial
  - Costs
  - DRGs or similar casemix systems
  - Payment systems ( e.g. insurance vs. Out-of-Pocket expenses etc)

These indicators are complex computed variables which are usually derived from population level aggregation of different individual level health data .

(For example, birth rates are calculated by reported live births (excluding stillbirths) divided by the total number of persons in the population which is obtained in a census.)

Various agencies, including the World Health Organization, have used some of these information rubrics either alone or jointly in public health information systems, there is not, however, a common agreed conceptual model and standardization for a comprehensive public health information system at global level. This lack of agreed framework results in unintentional variability in data collection which hinders the use of these data for comparisons or representation of health situation in countries or populations.

(For example, various countries may use different definitions of a live birth which not only affects birth rates rate but also infant mortality rates. Similarly the denominator upon which such a rate is calculated has to be defined as a standard (e.g. simultaneous, mid year population etc).

**In summary**, this work package will explore the following areas:

- a. existing defined or de facto standards for public health indicators used by national and international bodies stemming out from individual data rubrics.
- b. procedures used to compute these indicators ( e.g. aggregation from individual data)
- c. technical and legal requirements for a comprehensive information model allowing aggregation of data using anonymous ID-removed individual health records.

**Semantic Health Project Work Package 5 initial expected outputs are:**

- a Linking the public health indicators with a **global norm and standard development activity** for overall health information systems which integrates with the European / Global standard systems such as EC, CEN, eHSCG, ISO...
- b Identifying the **functionalities** of the health information systems useful for public health such as query functions (e.g. rates for mortality, morbidity, disability, risk factors ,alarm functions for international health regulations and patient safety, etc.)
- c Creating the **information models in public health** with proper building blocks (terminological definitions, standards for data and metadata, underlying ontology and knowledge representation) and integrating them into the overall health information systems.

#### **Deliverables:**

Report 1: Assessment of Semantic Congruence between key HIS indicators (Month 12): *Evaluation of possible mechanisms of use of data at individual level for public health indicators at hand.*

Report 2: Recommendations for Roadmap Priority actions: Priorities for SiOp from Public Health Perspective (Month 15): *evaluation of key inputs for global surveillance, international health regulations, patient safety and other global health issues.*

Report 3: Barriers, approaches and research priorities for PH and secondary uses (Month 16): *Identification of key actions for implementation focusing on barriers and priorities for research*

These reports will be developed and shared through an "International Web-based Platform" in which a network of experts and stakeholders could contribute to the roadmap development in line with essential public health functions and key international tasks on global public health.

### **8.5.1 Key questions regarding Public Health and secondary use**

Regarding the Work package 5 please try to respond to the issues listed below as well as any other points you deem appropriate.

#### **6. Is the programme of work comprehensive?**

- a. Do they represent the **key issues** as desired and/or what needs to be done in this area?
- b. Are the components/dimensions **clearly identified/formulated**? Any suggestions for reformulation, renaming, etc?
- c. Are there any **missing elements** that may be included?

- d. **Overall consideration** regarding its coverage, its **substantive nature** ( i.e. light - heavy for the project deliverable audience ? e.g. too technical, or needs introduction, clarification of context with examples, needs, policy relevance etc)
- 7. Are there any critical factors that need to be considered in creation of a roadmap for the future work including 7<sup>th</sup> FP**
- e. **Content-wise**: linkages to other entities/projects (e.g. patient summaries, translation, patient safety, bio-surveillance, ...)
  - f. **Process-wise**: actions, consultations, referrals, meetings ...
  - g. Any important **"lessons-learnt"** from past applications to avoid mistakes and redundancy
  - h. consideration for **future applications** - important developments in the pipeline and how the project work may be useful?
  - i. Are there any **other sector examples** in this area (e.g. in education, agriculture, meteorology, geographical information systems or others) where the semantic interoperability issues could be of value in addressing the issues?
- 8. In terms of deliverables:**
- j. Your **prioritization** for the identified deliverables
    - i. in terms of time sequence - what needs to be done first?
    - ii. In terms of relevance to original objectives
    - iii. In terms of importance
  
  - k. Suggestions in the **format** of the deliverable:
    - i. Paper
    - ii. Web site
    - iii. Other: computer programme, discussion forum, sharepoint, etc.
  
  - l. Future work:
    - i. Suggestions for linking the deliverables to **future applications**

## 8.6 Discussion paper for WP 3: Socio-economic Factors

The objectives of this work-package are:

- to identify at the health systems level key health policy, legal and other socio-economic issues of relevance for the wider context of semantic interoperability.
- to identify at the organisational level socio-economic enablers and obstacles that help or hamper implementing meaningful interoperating systems.
- to delineate the research and implementation related issues and tasks in improving interoperability at the institutional, regional, national and trans-national level.
- to synthesise and identify initial / intermediate actions needed and milestones for the short term (2-5 years) and the long term (4-10 years) for a European and global Roadmap on socio-economic issues.

### Approach/Methods:

To achieve these objectives, white and grey literature, output from other projects and further materials will be collected. If appropriate additional structured expert and stakeholder interviews will be held. These activities shall lead to a preliminary Position Paper presenting an analysis and synthesis of initial needs. A workshop involving the research community will be organised to validate and prioritise areas of future work on the basis of highest expected short term impact. Participants are expected to produce pre- and post workshop recommendations. The workshop materials and outcomes will be synthesized in a final deliverable Recommendations to health authorities, and stakeholders and particularly the research community and programme managers will be derived on key actions and areas to enable effective, meaningful interoperability.

### 8.6.1 Key questions regarding Socio-economic Factors

*What is the problem?*

- Problem space not sufficiently detailed in terms of objectives, timelines, languages/cultures, dissimilar health delivery systems.

*What do we know of proposed solutions?*

- Coping with highly dynamic systems.
- Imprecise estimates of size of investment.
- Unrealistic estimates of potential return on investments.
- Synchronisation of implementation within institutions/ regions/ nations.
- Synchronisation of updating within languages/ institutions/ regions/ nations.

*What are the existing and emerging legal consequences*

- Liabilities for incomplete, imprecise representation of terminological knowledge , and hence unsafe data expression/derivation.
- Legal restrictions on recording/exchanging certain types of information.

*What does it mean to practitioners?*

- Registration dividend
- Effectiveness of data entry technologies
- Unclear short/long term impact on business processes

*What is at stake between solution providers?*

- 'Trespassing' of developers on domains outside their own responsibility (e.g. terminology/archetype/record architecture)

*One monolithic solution provider?*

- Coping with diversity between nations
- Transformation of one single development institute into effective network of national collaborating centres.
- ...

## 8.7 Discussion paper for WP 2: Technical issues

### 8.7.1 General observation

Health Informatics products are developed to meet specific needs: recording health care at individual level, retrieving information, transferring information (messaging systems, e.g. HL7) across applications, sharing information with patients (health promotion, disease prevention, informed consent) in heterogeneous levels of understanding, across linguistic and cultural barriers.

Many of the products partly need the same information elements. The analytical breakdown in data element categories has proved to be a real challenge to ensure that the truly significant data elements are selected (sufficient and necessary comprehensiveness and accuracy for the purpose in point).

Much less has been achieved to ensure that the content of those categories is actually understood the same way by all message recipients. Not only must this be achieved for exchanges among humans, who can intelligently make up for inadequacies by calling on their experience and contextual knowledge (way of thinking of originator)

For each particular purpose, several implementations are usually developed over time, in order to improve focus, accuracy or specificity. The reasons for the proliferation of such systems should be elicited.

### 8.7.2 Objectives of WP2

- to develop a comprehensive definition of and elaborate necessary conditions for semantic interoperability in order to validate the analytical conceptual framework under which pending research issues could be envisaged
- to analyze and synthesize core generic and technological semantic interoperability issues and research challenges by benchmarking European and global research experience and needs
- to synthesize and identify initial / intermediate milestones for a roadmap for research on technological issues

### 8.7.3 Key questions regarding Technical issues

A. What are the semantic interoperability issues that need to be confronted?

#### Utility:

*What are the features of information that need to be captured to ensure that messages sent across the spectrum of HIS components are understood and valued as if they had been produced within that system component, i.e. are completely equivalent in meaning. To what extent would you consider that a fully integrated system can possibly cover such different areas as financing (NHA, DRG), epidemiology, bed side care, variety for HC providers (doctors, nurses, etc.), research?*

#### Etiology:

*Why, in your opinion, do current HIS address components separately with little regard for neighbouring or related systems? Is it for lack of a proper conceptual framework? Is it due to the overwhelming complexity of the system? Is it due to exacerbated competition (between*

promoters of individual HIS components, domains, political entities)? Does it reflect different thinking among different schools?

**Standards:**

*What are the areas of semantic interoperability that should be standardized and which one should be given more flexibility without jeopardizing the consistency of the whole? To what extent should ambiguity of the human language be allowed to be expressed in a common system? Should underlying standards be limited to the structure of information? How would you see a compromise developed between controlled vocabularies (which many recommend as a necessary evil, and other reject as an unbearable constraints) and natural language processing?*

**Formalism:**

*To what degree of formalism would you see HIS must go to capture the true intended meaning of messages across the total HIS spectrum. What criteria should be met for formal static and dynamic representations? To what extent is it possible to ensure the real time evolution of such representation to match the evolution of medical and health related knowledge?*

**Multilingualism:**

*How do you see communication across languages can be overcome? To what extent is translation better or worse than machine-driven representation of clinical terms? Of patient records? Of other HI items?*

**Success stories:**

*What are the best success stories that can show the way to go? What else has been attempted and failed? For what reason?*

**B. Remedial measures and driving forces**

What would you consider, in order of priority, are the key development areas that would provide the best return on possible investments?

Why, in your opinion, have they not (sufficiently) been recognized as such?

What would key areas of semantic interoperability be that would require standardization? Would you see that as an incentive or deterrent for system developers and application developers?

What would be the optimum criteria for catalyzing efforts in the field of semantic interoperability? Is there sufficient agreement on the nature of the problems to be solved? If so, why doesn't it work better? If not, what can be done to improve the situation?

Are research centres equipped to address those issues successfully? What is the best platform? Conferencing? Collaborative workspaces? Distributed environments, including grid technology applications?

What areas would need further research? What kind of investment would you think could be required to do that (in monetary and human resource terms)? How would you break down the whole research area package in order to ensure short-term return on investments that could benefit both health professionals (care and non-care categories) and citizen patients?

## **8.8 Discussion paper for WP 6: Biomedical Ontologies**

We outline here key issues related to terminologies, classifications, coding systems and ontologies for interoperability in healthcare and biomedical research and prime issues.

### **8.8.1 Objectives of Semantic Health Workpackage 6**

- to identify the requirements for terminologies and ontologies to support achievable semantic coherence in direct patient care and personal health records
- to examine existing resources and formalisms with respect to concrete objectives
- to examine the requirements for specifying interface between terminology and EHR specifications.
- to identify the issues concerning quality and safety requiring further action, particularly with respect to major existing resources such as SNOMED.
- to identify issues in linking to biomedical research resources including publication resources and PubMed/UMLS on the one hand and major molecular biological resources such as the Open Biomedical Ontologies Consortium's resources, most critically the Gene Ontology and with the National Cancer Institute's Enterprise Vocabulary Initiative.
- to identify issues in internationalisation of terminologies
- to clarify the relationship between terminologies, classifications, coding systems and ontologies.
- to identify opportunities for language technology for both input and output
- to prioritise key changes that could make a difference to realising semantic interoperability in the clinical workplace

### **8.8.2 Key questions to be addressed for Workpackage 6 Ontology and terminologies**

#### **A. What are the semantic interoperability issues relating to the Ontologies?**

*1 What are the situations (scenarios) in which the sharing or processing of clinical data requires explicit and interoperable interpretation of clinical meaning and standard codes and ontologies?*

What are the requirements and scenarios for use of ontologies and terminologies for communication ?

- a) as controlled vocabularies;
- b) as browsing exploratory tools;
- c) for retrieval of information for secondary uses;
- d) for use in clinical decision making and monitoring.

*2 Can we establish sufficient terminologies/ontologies that they can be used reliably?*

What actions and investment are required to improve the quality of existing terminologies to the point where they can be relied on for interoperability?

What is the granularity at which we can achieve realistic interoperability at

- a) technical and
- b) human factors levels?

What is the level of detail which can realistically be collected in various situations?

Where are there acute risks of misinterpretation or miscommunication involving coding and ontologies.

*3 How can the phenotypic information in the EHR be integrated with basic biological information for clinical research?*

There has been an explosion in bioinformatics, genetic, and genomic databases

*4 Further issues for debate:*

1. For what purposes is SNOMED currently fit? Given known limitations in the hierarchical structure is it more cost effective to seek gradual improvements or a major overhaul before adoption?
2. Can users ever achieve sufficient inter-rater reliability in the use of complex coding systems that the data will be of value, irregardless of the technical virtues or faults of the coding system?
3. Is the best way to collect information for standard epidemiological purposes using ICD to do so directly or via point of patient contact information systems?
4. What organisational structures for the SNOMED SDO would be required to give confidence in:
  - a) the quality,
  - b) responsiveness, and
  - c) adequate European content.
- 5) Are we using the last century's tools for this century's problems? Are current ontology and semantic methods in healthcare out of date given the rapid advances in Web technology and the Semantic Web?

**B. What are the present barriers to the development or adoption of semantic interoperability solutions?**

1. Areas needing development and testing

Interface and Binding between EHRS and coding systems. Major progress has been made recently but needs to be tested on a wider scale.

2) Areas needing capacity building and training

Modern terminology and ontology methods require a serious commitment to understanding the logical foundations by a core of developers. Staff with the needed skills are very limited. Training takes time, and in some cases requires significant “unlearning”.

Can the required base of people with the necessary skills be built?

### 3. Areas needing wide-scale evaluations

(results exist, but need refinement and real clinical use, to determine best practice)

- Practical use of SNOMED-CT in clinical systems
- Inter-rater reliability and data quality related to various terminologies
- SNOMED-CT internationalisation
- Effort required for the quality assurance of SNOMED or other large terminologies/ontologies.

### 3. Areas needing international standardisation

(needs facilitation and funding)

- Generic terminology access services

### 4. Areas needing investment

(business cases not strong enough for industry, but products needed: maybe sponsored open source?)

- ontology and terminology development environments
- terminology servers and browsers
- quality assurance

### 5. Areas needing further (focussed) research

- Human Factors and user interface issues related to the use of ontologies,

a) for system configuration/localisation staff,

b) end users

- Natural language interfaces to clinical systems, both for  
input (text understanding/extraction/mining) and  
output (text and dialogue generation)

- More expressive ontology/terminology formalisms compatible with other emerging standards, particularly the W3Cs standards.

- term binding to archetypes and record structures