Semantic HealthNet
Semantic Interoperability for Health

Deliverable 3.3

Report to clinical stakeholders on the creation and implementation of domain-specific semantic interoperability resources


Call: FP7-ICT-2011-7

Grant agreement for: Network of Excellence (NoE)

Project acronym: SemanticHealthNet

Project full title: Semantic Interoperability for Health Network

Grant agreement no.: 288408

Budget: 3.222.380 EURO

Funding: 2.945.364 EURO

Start: 01.12.2011 - End: 31.05.2015

Website: www.semantichealthnet.eu

Coordinators:
In this deliverable, we take a broad look at semantic interoperability and try to bring the perspective of clinicians and patients into focus. Lessons learned and recommendations stem from the experiences in two use cases in the SemanticHealthNet (SHN) Project, one on heart failure and one on cardiovascular health. Different approaches are discussed to link heterogeneous terminologies, clinical models, and ontologies. The target audience for this deliverable are the leaders of the learned medical societies. Clinical involvement and even clinical governance will be essential to get the ambition of semantic interoperability back on track. The insight grows that the roadmap to clinical operability is a roadmap for a long and cumbersome quest. There are no quick solutions, and multiple dangerous dead-end streets. This deliverable condenses the experiences from the use cases and the work of partners and experts in this Network of Excellence for Semantic Interoperability in Heath, from 2012 to 2015. We hope that this deliverable will contribute to a successful start of a dedicated Centre of Excellence in Semantic Interoperability as a part of the newly formed European Institute for Innovation Through Health Data. The deliverable is intended for a general public, but more technical aspects are expanded in 13 annexes, written by SHN experts.
Table of Content

Summary of SemanticHealthNet SHN D3.3 ...........................................7

1 Introduction ..................................................................................9
  1.1 Background ..............................................................................9
  1.2 Introducing the project SemanticHealthNet (SHN) ......................9
  1.3 Objectives of this deliverable ...................................................10
  1.4 Target audience for this deliverable .........................................10

2 General method for the comprehensive development process of
the basic public semantic resources for a medical subdomain ......13
  2.1 Focus of the general method proposed ....................................13
    Focus on data entry ......................................................................13
    Focus on human – machine communication ..................................13
    Focus on common, standardized approaches .................................15
    Focus on the difference between human language and machine language .................................15
    Focus on the patient perspective .................................................16
    Focus on the collaborative and iterative nature of the development process .........................16
    Focus on resources and not on tools ...........................................17
    Focus on sustainability of meaning .............................................17
  2.2 Disciplines involved ...............................................................18
    Disciplines related to the domain .................................................18
    Disciplines related to ICT ..........................................................18
  2.3 Describing the chain of public semantic resources to be build ..19
  2.4 Methods for lexical resources (end-user terminologies) ..........20
    Unilingual lexicons of words and phrases for end-users ...............20
    Multilingual lexicons ................................................................20
  2.5 Methods for conceptual resources (reference terminologies) ........20
    National and domain specific reference terminologies ..................20
    Integrated interface terminologies ..............................................21
    International Reference Terminologies .......................................23
    Translations of International Reference Terminologies ..................24
    Mappings between International Reference terminologies .............24
  2.6 Methods for architectural resources (clinical models) ............24
    Basic clinical models ...............................................................25
3 Lessons learned and recommendations ................................................................. 35

3.1 Patient involvement ............................................................................................. 35
3.2 Involvement of clinicians .................................................................................... 35
3.3 Need for multilingualism .................................................................................... 36
3.4 Need for a lexical approach and interface terminologies ................................... 37
3.5 No single choice for terminologies ..................................................................... 37
3.6 A clear choice for a systematic development process based on at least one of the standardized clinical models and reference archetypes ............................................. 38
3.7 The importance of context .................................................................................. 40
3.8 The importance of semantic web techniques ...................................................... 41
3.9 Sustainability of public resources ........................................................................ 42
3.10 Discrepancy between expectation and progress ............................................... 43
3.11 The need for a governing Institute .................................................................... 43

4 Approach to testing ............................................................................................... 45

5 Approach to validation by stakeholders ................................................................. 48

6 Annexes .................................................................................................................. 49

6.1 Medical terminologies for patients ...................................................................... 50
6.2 Approach to building multilingual terminological resources using the ISO-standards Terminological Markup Framework (TMF) and Lexical Markup Framework (LMF). A use case for heat failure .................................................................................. 50
6.3 Multilingual management of conceptual and lexical resources. A comparison of Bioportal and HeTOP (Health Terminology/Ontology Portal). ............................................................................. 50
6.4 Mapping between international medical terminologies ..................................... 50
6.5 Technical management of multilingual semantic resources ............................... 51
6.6 Strategic development of structured clinical models for medical subdomains ............. 51
6.7 Experiences from the SemanticHealthNet Use Cases ........................................51
6.8 Position Paper: SemanticHealthNet Semantic Interoperability Processes and Semantic Resources ................................................................. 51
6.9 Terminologies and ontologies in medicine ....................................................... 51
6.10 Perspectives of patients with long term health conditions ............................. 51
6.11 Integrating what matters to patients into health records based on the ICF: Examining the utility of interRAI to operationalize the ICF Disability Set as a case in point .................................................. 52
6.12 Cross Continental Medical Communication.................................................. 52
6.13 Adoption of interoperable EHRs: barriers, challenges, incentives.................. 52
List of figures

Figure 1. Deterioration of data quality migrating through systems from sender to receiver. .......... 13
Figure 2. Scope of semantic interoperability. ................................................................................. 14
Figure 3. Difference between words and concepts........................................................................ 15
Figure 4. Resources and tools........................................................................................................ 17
Figure 5. Resources needed by leaders of learned medical societies............................................. 18
Figure 6. Structure of a multilingual interface terminology system............................................... 22
Figure 7. The outpatient (discharge) letter and the patient Summary (SUMEHR): two (electronic) key documents in the Health Care System ........................................................................ 28
Figure 8. Extensions to the Patient Summary for a heart failure patient ...................................... 29
Figure 9. Data-types in two dimensional, object-oriented, and triplet-oriented data recording....... 30
Figure 10. Resource Description Framework (RDF) triplets.............................................................. 30
Figure 11. Representation of reality with universals and particulars. ............................................. 32
Figure 12. Summary of clinical information modeling process steps. ............................................ 39
Figure 13. Summary of reference models and their clinical information model artefacts.............. 40
Figure 14. Diagram for SHN typology of resources ..................................................................... 41
Figure 15. SHN semantic-driven interoperability architecture....................................................... 42
Figure 16. Testing semantic interoperability with quality indicators in Health Care..................... 46
Summary of SemanticHealthNet SHN D3.3

In this deliverable, we take a broad look at semantic interoperability and try to bring the perspective of clinicians and patients into focus.

Lessons learned and recommendations stem from the experiences in two use cases in the SemanticHealthNet (SHN) Project, one on heart failure and on cardiovascular health.

Different approaches are discussed to link heterogeneous terminologies, clinical models, and ontologies.

The target audience for this deliverable are the leaders of the learned medical societies. Clinical involvement and even clinical governance will be essential to get the ambition of semantic interoperability back on track. The insight grows that the roadmap to clinical operability is a roadmap for a long and cumbersome quest. There are no quick solutions, and multiple dangerous dead-end streets.

This deliverable condenses the experiences from the use cases and the work of partners and experts in this Network of Excellence for Semantic Interoperability in Heath, from 2012 to 2015.

We hope that this deliverable will contribute to a successful start of a dedicated Centre of Excellence in Semantic Interoperability as a part of the newly formed European Institute for Innovation Through Health Data.

The deliverable is intended for a general public, but more technical aspects are expanded in 13 annexes, written by SHN experts.
1 Introduction

1.1 Background

The background for this work is the rapid evolution of health informatics in the world, and in Europe in particular. All European countries are setting up e-Health Platforms, closely linked to the public social security systems. Demands of citizens sound louder. Patients are more and more empowered. Caring for patients is no longer an activity of individuals but a collective process, involving multiple care givers, from different disciplines and different health care settings. The demand for continuity of care is increasing and spanning a person’s lifetime. Increasingly, genetic information becomes available for diagnostics and even treatment, leading to additional information explosion. Numerous standards are published to govern and guide electronic communication and storage of medical information. The internet has become ubiquitous. The sophistication with which information can be stored, protected, published, analysed, and put at work has grown exponentially, with the increase of computer power and the advent of new semantic techniques.

In healthcare this technical evolution has coincided with a full reorganisation of the distribution of scientific knowledge. The number of clinical trials has grown exponentially in the second half of the 20th century. From the early nineties on, the transformation of this knowledge base into systematic reviews has been taken on in a systematic approach by the successful creation of the Cochrane Collaboration. The concept of Evidence-Based Medicine was rapidly adapted in medical education, medical practice and health care policy. Medical societies learned to harness the art and science of producing practice guidelines, containing answers to specific clinical questions, in the form of weak or strong specific recommendations, based on weak or strong evidence, duly referenced and easily accessible. The dissemination and implementation of these Evidence-based medical information became much easier through the internet, with the creation of guideline platforms and decision support systems in many countries, many of which have gained an international audience, and some target the patient.

Progress in the domain of medical documentation, however, has been slow and hampered by lack of standards, abundance of standards, language barriers, inability to integrate user’s needs in technical design, rapidly evolving technology without due consideration for legacy conversion, compartmentalisation of the medical field into sub-disciplines and health care settings, and last but not least, evasion strategies of clinicians against increasing control and McDonaldization of society.  

1.2 Introducing the project SemanticHealthNet (SHN)

SemanticHealthNet (SHN) faces the challenge of improving the semantic interoperability of clinical information. It assumes that several standards and proprietary implementations for representing the content of electronic health records (EHRs) will co-exist for a long time.

The approach targets the whole range of health-related information about all medical domains.

---

SHN has the task to overview all types of resources needed for semantic operability, such as linguistic resources, clinical terminologies, clinical data models and overarching ontologies. As practical exemplars, SHN has set the focus on chronic heart failure and cardiovascular prevention.

The intention is to develop for sub-disciplines in medicine a structured methodology to create and maintain domain-specific resources to assure correct representation of a specialized discipline, while preserving semantic interoperability with the resources developed for general purpose, for neighbouring disciplines (including allied health personnel), and for patients.

It was agreed that the chain of resources to be created in semantic interoperable systems should be able to capture the contextual situations in which information is produced, and at the same time be useful for use at three levels:

- optimal clinical documentation for clinical purposes (primary use of information)
- exploitation of the data for internal secondary use (internal audit, information retrieval, decision support, communication with patients)
- exploitation of the data for external secondary use: research (clinical trials, epidemiological surveys) and public health policy (e.g. construction of quality indicators)

In the first Work Stream of the SHN project, the first two work packages aimed at producing domain-specific clinical models. In WP1, it was foreseen to produce specific conceptual clinical models to capture clinical information in a patient summary for heart failure patients, useful for decision support systems based on structured guidelines. In addition, in WP2 a conceptual clinical model was made for the representation of risk factors (smoking cessation status) to be used in public health policy for cardiovascular prevention.

In the third work package (WP3) the aim is to learn from these two experiments described above and to develop a general methodology for other sub-disciplines to develop, test and validate a chain of domain-specific resources, including logical clinical model representations from the conceptual models from WP1 and WP2, useful for supporting semantic interoperability within the discipline and across disciplines.

1.3 Objectives of this deliverable

1. To formulate a general methodology for the comprehensive development process of a set of public semantic resources, needed for the full integration of a clinical subdomain in the modern health care environment of eHealth.
2. To propose testing criteria for the efficacy of the use of these resources in applications, in terms of semantic interoperability.
3. To validate the proposed development process with the relevant clinical stakeholders.

1.4 Target audience for this deliverable

In this document we address primarily the leaders of the learned medical societies. In the recent past, these persons have already build Cochrane review groups, guideline committees, specialized digital libraries, quality indicators, point-of-care platforms, decision support systems, spe-
cialised registers, international audits, sentinel practices, and clinical trial platforms. They now face a new challenge to participate in the creation of semantic resources, which will strengthen the communication between their colleagues and make the most of technical and societal evolutions. They are beginning to grasp the importance of semantic interoperability to enhance the quality of processes of healthcare. It is up to the leaders of the learned medical societies to convince their fellow clinicians of the importance of the extra effort for sound medical documentation, to help them choose the right tools for this, to provide the right semantic resources, to assure a return on investment for the clinicians, and finally, to create the environment of trust to enable correct clinical documentation.

We hope that health policy makers, leaders of standard development organisations, specialists in Information and Communication Technology, vendors of medical software, and procurers of large health care organisations will look over the shoulder of the clinicians, to read this document, too.

However pressing the deadlines of policy makers and e-health engineers are, this process of creation of the right semantic resources (able to represent the clinical reality) is a formidable task that will require serious involvement of the clinicians, intense cooperation with professionals of unfamiliar expertise, and above all time, to test, reiterate the development of the design, and test again.

The challenge is comparable to historical examples of collective intellectual achievements:

- The making of the Oxford English Dictionary\(^2\)
- The making of the “Encyclopédie Universelle”\(^3\)

2 General method for the comprehensive development process of the basic public semantic resources for a medical subdomain

2.1 Focus of the general method proposed

Focus on data entry

Quality of documentation and the extent to which the initially recorded information can be used in applications and secondary use downstream depends on the skills of the person responsible for input and the quality of the support provided by the input system.

![Figure 1. Deterioration of data quality migrating through systems from sender to receiver.](image)

There are of course a number of mechanisms that can preserve and even enhance information, through the power of aggregation of data. But it must be clearly stated that coding free text is a projection of a world of expressiveness to another world of expressiveness, usually much smaller, and thus inevitably produces a loss of expressiveness. The type of loss and the consequences of this loss is context-dependant and must be assessed in each situation.

A castle cannot be built on sand, and hence, due attention is needed to the data entry process and the support of the person who puts the data in, be it a health care provider or a patient.

For this reason, we focus on the creation of semantic resources facilitating input of clinical data by humans in computers. Because only if that process is functional and effective, the communication from computer to computer and back to humans will be of high standard and will improve communication of usable information.

Focus on human – machine communication

Interoperability is the ability of a system or a product to work with other systems or products without special effort on the part of the customer. Traditionally, in the communication between computers the distinction is made between technical (or syntactic) interoperability and semantic interoperability.
Technical interoperability is the mere ability to exchange data among systems. Semantic interoperability is the ability of computer systems to exchange data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inference, knowledge discovery, and data federation between information systems.

An ultimate test of the quality of semantic interoperability of systems is to examine its power to support the communication between humans, mediated by information systems. The functionality for human communication should be the ultimate goal.

Machine-mediated human communication occurs within one physician, consulting his medical record to see the notes from a previous consultation. It occurs between healthcare providers from a same practice, caring for the same patient at different points in time. Between the general practitioner and the specialist, between one specialist and another specialist. Between the various health professionals caring for the patients, including allied health personnel. Between the clinician and the researcher or the healthcare policy maker. There is also the communication between the patient and the physician, via the Electronic Health Record of the physician or the Personal Health Record of the patient.

Figure 2. Scope of semantic interoperability.

Use cases to test interoperability of computer systems should be set up in each of these pairs of communication.
Focus on common, standardized approaches

Interoperability stems from the willingness to agree to interact with common approaches. Therefore, in this document, the use of international reference terminologies, as well as the use of ISO and CEN and other standards for communication of health data will be stressed. We need common approaches in the way we deal with concepts and structure of information. Standards are available, and the problem is rather in their abundance then in their paucity. Despite this, too often, information is still encapsulated in closed, proprietary systems.

In this document, four types of semantic resources will be described: lexical, conceptual, architectural, and ontological resources. Each of these type of resources can and should be constructed on the basis of available specific standards for terminologies, for information architecture and for knowledge representation.

Focus on the difference between human language and machine language

Human language can be agile, poetic and ambiguous. Machine language must be precise, rigorous and logic. Words and phrases in human language may have different meanings. A concept in a machine language must be clear-cut and represented by one, and only one preferred term in a meta-language. For practical reasons, this meta-language is mostly English, building an extra barrier for non-native English speakers.

![Figure 3. Difference between words and concepts.](Image)

The word terminology is often used as a container concept for lexical terminologies and conceptual (or reference) terminologies.
Lexical terminologies (e.g. a dictionary) store words and phrases. Their structure must be able to catch all the subtleties of human language, synonyms, multiple meanings, deflections, etc.

For multilingual lexical terminologies there is an ISO standard to guide the structure of these resources (LMF Lexical Markup Framework ISO 24613). An example of application is BabelNet. BabelNet is both a multilingual encyclopedic dictionary, with lexicographic and encyclopedic coverage of terms, and an ontology which connects concepts and named entities in a very large network of semantic relations, made up of 14 million nodes, called Babel synsets. ⁴ ⁵

Conceptual terminologies (also called reference terminologies) are collections of concepts pertaining to a scientific domain, identified by a code and a one preferred term, with the concept defined by scope notes or by description in a formal language.

For multilingual reference terminologies, there is also an ISO standard (TMF Terminological Markup Framework ISO 16642). This standard has been less applied in concrete projects, despite the fact that there are multiple translations of international reference terminologies, such as ICD, ICPC, ATC, and SNOMED-CT. ISO 16642 is designed to support the development and use of computer applications for terminological data and the exchange of such data between different applications.

**Focus on the patient perspective**

Semantic interoperability is not only important for the health care professionals and their mutual communication. It should support patients, too, and their communication with health care professionals and vice versa. There are numerous consequences. The difference between lay language and professional language must be accounted for in the construction of semantic resources (lexical and conceptual terminologies). For the construction and alignment of architectural models for electronic health records and patient health records, the patient’s perspective must be integrated in the early design phase of such models. Adequate methods of patient involvement must be deployed to assure that this is more than lip service to the patient.

**Focus on the collaborative and iterative nature of the development process**

Building new clinical models for domain-specific applications requires cooperation between developers and end-users. End-users are busy clinicians, mostly unfamiliar with the technical aspects of data-entry and record structures. Every medical discipline should select and train among its members a number of clinicians, who can acquire the skills and liberate time to be involved in the development

---


process of crucial resources (and applications), from the very early design phase, till the proof of concept phase.

**Focus on resources and not on tools**

This document focusses on semantic resources: terminologies, clinical models and ontologies. For health care, most of this resources could be in the public domain, financed by the health care system or supported by some collective funding mechanism. Maintenance of such resources can be so costly that it does not fit in private business plans, or must rely on interactive cooperation via wiki-like processes on the web, involving the clinicians and experts of the various medical disciplines.

Tools or software applications working with these public semantic resources may be developed as open source, or by commercial vendors, building EHR solutions for primary care and hospital setting.

![Figure 4. Resources and tools.](image)

Good software development is also at its best when it is user-driven and working in an iterative process. For applications however, competition based on software functionality (and not on the quality of the resources) might be crucial to drive innovation.

**Focus on sustainability of meaning**

An important challenge, as expressed above about data migration, is the need to ensure sustainability of semantics across systems (interoperability) but also across time. This is rarely discussed but represent the same type of complexity than between systems.
2.2 Disciplines involved

Leaders of the learned societies may have already organised Cochrane Review Groups for their field, guideline committees, platforms for the dissemination of Evidence-Based Practice, disease registers, quality indicators, etc. Now they need to embark on a quest to build, complete and maintain a chain of public semantic resources for semantic interoperability in their medical sub-discipline.

On that quest they will meet a collection of colleagues in health care and other strange bedfellows who give advice on technical issues.

Disciplines related to the domain

Whatever the nature of the specialized sub-discipline is, its leaders will have to engage in discussions with other clinicians: the general medical disciplines (general practitioners, paediatricians, geriatricians), specialists from neighbouring subdomains, nurses and other allied health personnel active in the subdomain. Inevitably, talks will also be needed with public health specialists and epidemiologists. Finally, a dialogue must be established on these issues with patients, suffering from diseases pertinent to the subdomain.

Disciplines related to ICT

Things get more complicated when it has to come to a dialogue with the Information and Communication experts, or the engineers of e-Health. All of the sudden, the leaders of learned medical societies are confronted with beforehand unknown professions. Computational linguists can help them to
harvest the words and phrases of their profession, and to cherish this collection as necessary prerequisite for more formal steps ahead. Health informatics experts and data modellers expect them to engage in difficult discussions on the structure of the medical records, asking them for precise clinical specifications of routine medical actions, of which the sequence and logic is questioned along the way. Knowledge engineers and ontologists cut up the expertise of the discipline along totally new lines, challenging the clinicians to a deep and new introspection of the discipline.

Things would be easier in case these different expert professions would easily talk to each other. Unfortunately, it sometimes seems as if they live on a different planet and talk different languages, while sharing the same words but with a different meaning, also and especially when talking about terminology.  

### 2.3 Describing the chain of public semantic resources to be build

Four types of public semantic resources can be distinguished: lexical, conceptual, architectural and ontological resources. Lexical semantic resources (or end-user terminologies) capture words and phrases, pertaining to the medical subdomain, possibly linked to bigger Natural Language Processing systems. They are language-dependent and can be linked to conceptual resources and even ontologies.

Conceptual resources (or reference terminologies) may be national and in the local language. They organise a collection of concepts relevant to the subdomain or to a nation and can be linked to or be a subset from an international reference terminology. For medical documentation, large nomenclatures such as SNOMED-CT and LOINC, serve as international reference terminologies. Another type of international reference terminologies are classification systems. There is the Family of World Health Organisation (WHO) classifications: the International Classification of Primary Care, The International Classification of Diseases (ICD), The Anatomical Chemical Therapeutic Classification (ATC), The International Classification of Functionality (ICF). A famous international reference terminology is Medial Subject Headings (MeSH), a controlled vocabulary or thesaurus for information retrieval in Medline, the bibliographic database of the US National Library of Medicine. This Institution also host the Universal Medical Language System. (UMLS).

Architectural resources are information models for basic electronic medical records or extensions thereof. It also refers to information models for the communication of the clinical data, such as in HL7 messages. There are 3 competing systems, each with their own history, but all recognised open source specifications or standards for clinical documentation: OpenEHR, HL7 CDA, HL7 version 3 messaging, HL7 FHIR, and EN13606. These are all based on the two level modeling principle: baseline models for generic system versus messaging functions and detailed models for the granular coded clinical data. Hence, most of these approaches use a generic reference (information) model and sets of small clinical models (templates, archetypes, Detailed Clinical Models) that represent collections of specific clinical concepts.

---

Ontological resources are formalized representations of the knowledge in a scientific field, and can come in the form of ontologies or structural guidelines, or decision rules.

2.4 Methods for lexical resources (end-user terminologies)

Unilingual lexicons of words and phrases for end-users

In the interface between humans and Information Technology systems, language is of utmost important. Physicians express themselves with words and phrases in a specific language, and these need to be entered by the actors or captured by the systems. In that process words and phrases will be transformed into machine-readable concepts.

The medical language is extremely rich and many extensive dictionaries are available. When looking at the content of medical records, especially when taken into account current patient history, it becomes clear that medical language is not a sub-language, but a supra language. It uses all the common words of a language, and adds large sets of specialized words.

In recent years multilingual lexicons based on semantic principles were developed to collect words and phrases and their possible multiple meanings and synonyms. These resources use the ISO standard Lexical Markup Framework (ISO 24613) and its LEMON application and allow also connection to powerful Natural Language Processing applications. ⁷

Semantic techniques can be used to harvest definitions for the different meanings from the World Wide Web. These definitions can be transformed in formal logical statements, turning them in stepping stones for automated ontology building applications. This in turn can help the linkage of these lexical resources to the concepts in reference terminologies.

Multilingual lexicons

Creation of unilingual domain-specific lexicons can also be undertaken as a multilingual collaborative effort, by using applications such as BabelNET and WORDNET, and multilingual term-extraction techniques.

2.5 Methods for conceptual resources (reference terminologies)

National and domain specific reference terminologies

A number of European countries are multilingual within the state. They will have to develop their local multilingual reference terminology.

There is an ISO standard for the multilingual reference terminologies (managing concepts of machine readable terminology systems) : ISO Terminology Markup Framework (ISO 16642). In Annex 2, more details are given on this standard.

⁷ http://babelnet.org/
Individual countries can develop a core set of concepts to be dealt with on a national way by a termino-
logy centre, but by choosing a multilingual approach, profit from international cooperation.

Medical sub-disciplines may want to create their own core set of concepts, possibly partly based on a
subset of a large international reference terminology such as SNOMED-CT, but probably augmented
with specific concepts not (yet) found in SNOMED-CT, but needed for the discipline, or linking to oth-
er domain-relevant terminologies. This is the very reason why the CIMI group has negotiated with
IHTSDO to formally handle extensions that are needed based on identified concepts, required in the
clinical models, but absent in SNOMED-CT.

It is important for a nation or a medical domain to have their own core set of concepts, managed
independently of but linked to (multiple) international systems. Keeping a firm grip on this pivotal
resource is essential, both for nations and for specific medical domains. For medical domains it is
preferred that this is managed on the European level by the different European Scientific Societies.
For national reference terminologies the level is nation, but of course, intense international collab-
oration is possible in the selection of the core set of concepts and in the mapping to international no-
menclatures, classifications, and thesauri.

**Integrated interface terminologies**

A more complex approach to terminology has been advocated by Roosenbloom, insisting on the cre-
ation of interface terminologies, which start from the language used by the end-users.8

One of the basics premises of realistic ontology development is to start from words and phrases (ut-
terances) used by patients, physicians and scientists.9

The idea is that an ideal interface terminology is a hybrid system, consisting of a lexical resource in
the language of the user and a reference terminology in a meta-language (English). This interface
terminology is a resource that permits applications to harness both the power of Natural Language
Processing and the linking and mapping to and between multiple international classifications. Intellig-
ent Medical Objects offers a specific interface terminology for use in electronic health records,
which maps to ICD 10 and SNOMED-CT among others.10

A pilot project for this approach was recently performed to use both ISO standards to extract termi-
nology from a guideline on heart failure in Dutch and French and to link the words and phrases and
the terms to several nomenclatures, classifications, and Thesauri (ICPC, ICD, SNOMED-CT, MeSH and
UMLS).11

---

10 [https://www.e-imo.com/](https://www.e-imo.com/)
Such an interface terminology would link to reference terminologies in the following way: During the clinical data registration process, the data is coded using the interface terminology in which each term has a unique code and which is served to the user interface for data entry. From this, the mapping to multiple reference terminologies is done supporting reasoning and reporting.\footnote{Kottke TE, Baechler CJ. An Algorithm That Identifies Coronary and Heart Failure Events in the Electronic Health Record. Prev Chronic Dis 2013;10:120097. DOI: http://dx.doi.org/10.5888/pcd10.120097}
International Reference Terminologies

Classification systems

Traditionally, classification systems from the World Health Organisation (WHO) were used to support data entry on mortality, morbidity, symptoms, reasons for encounter and procedures. The International Classification of Diseases (ICD) is used for mortality and morbidity documentation, and is a longstanding tool for epidemiological and public health policy research. The International Classification of Primary Care (ICPC) is used in general practice for episode-oriented documentation of reasons for encounter, symptoms, prevalent diseases, and procedures.

Other examples are the International Classification of Functioning, Disability and Health (ICF), focussed on the impact of diseases\(^\text{13}\) and the Anatomical Therapeutic Chemical Classification (ATC) for medicinal products, focussed on medicinal products.\(^\text{14}\)

Historically, the bulk of medical data has been classified with these classifications. It is likely that these systems will continue to be used in the next decade, so mapping to any new reference terminology to be developed for semantic interoperability will be necessary, at least for legacy conversion reasons.

Nomenclatures

For precise recording in the medical record, also a granular nomenclature is needed, which covers the whole domain of medicine. SNOMED-CT evolved as a reference terminology from the American SNOP (Systemized Nomenclature for Pathology) and the British Read-codes. It has incorporated over the years the terms of many classifications systems and even named entity lists. The governance of SNOMED-CT moved to the global level, by the creation of the International Health Terminology Standards Development Organisation (IHTSDO).\(^\text{15}\) A sophisticated formal ontology layer is under constant development.

SNOMED-CT is now adopted on many e-Health platforms across the world as a single solution for interface terminology, clinical model and ontology needs.

In a number of projects, core sets of SNOMED-CT have been created, in an attempt to shield the users for complexity beyond routine use.

In the EU project ASSESS CT the aim is to assess the usefulness and feasibility of SNOMED-CT at a European Level. HL7 is partner of this project.

Logical Observation Identifiers Names and Codes (LOINC) is a nomenclature for identifying medical laboratory observations. It is developed since 1994 by the Regenstrief Institute, a US non-profit medical research organization.\(^\text{16}\)

\(^\text{13}\) http://www.who.int/classifications/en/
\(^\text{14}\) http://www.whocc.no/atc_ddd_index/
\(^\text{15}\) http://www.ihtsdo.org/
\(^\text{16}\) http://loinc.org/
Bibliographic thesauri

In science, one of the most important bibliographic index is the thesaurus (in the sense of a controlled vocabulary for information retrieval) Medical Subject Headings (MeSH), a tool for searching the bibliographic databases Medline and Pubmed from the US National Library of Medicine.

In the background of Mesh there is the Unified Medical Language System® (UMLS®). This set of resources integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records. It is a complex and sophisticated system, providing a metathesaurus, a SPECIALIST lexicon and lexical tools, and a Semantic Network.  

Translations of International Reference Terminologies

In the non-English speaking countries where SNOMED-CT has been adopted, translations of preferred terms for concepts from English into the target language (usually on core sets of SNOMED-CT) have been performed. These translations are kept under the umbrella of IHTSDO.

Also other international classifications and the thesaurus MeSH has been translated in a number of languages. In Annex 3, a state of the art of systematically maintained translations of medical terminologies is given.

Mappings between International Reference terminologies

A bewildering set of mappings has been developed between the nomenclatures, classifications and thesauri, often with the UMLS system in the pivotal role. In Annex 4 an update of the overview of mappings and methods for mapping is given.

2.6 Methods for architectural resources (clinical models)

When a physician documents his/her actions in the consultation room, words and terms are used to describe the situation, and then must be converted to a standardized term to be written down into the medical record. The right term must be chosen. But also the right choice of the right place to put the term in the medical record is important.

The medical record has a complex structure with a number of items such as reason for encounter, problem list, allergies, family history, processes and procedures, lab results, etc.

When the result of a recording of the blood pressure needs to be documented, implicit information needs to be made explicit in the computer systems. Who made the recording, with what purpose, in what circumstances? To capture the context of information in the medical record, structure and modelling is needed.

17 http://www.nlm.nih.gov/research/umls/
Modelling a medical record is a cumbersome process, where clinicians and clinical modellers (health informatics experts) work together to decide on the overall structure of the medical record, on the different clinical entities that form the nodes of a model (called clinical entities or data categories), and on the value sets of the data categories chosen. Different words have been used to describe these structures, such as archetypes, templates, detailed clinical models.

The European Institute for Health Records (EuroRec) defines archetypes as follows: EHR Archetypes are a formal, rigorous and standardised (interoperable) specification for an agreed consensus or best practice representation of a clinical data structure within an electronic health record. Archetypes are translations/transformations of Detailed Clinical Models into technical structures to be used by IT-systems. Archetypes and the associated Reference Model also take care of documentation/versioning aspects of data. Archetypes are used to instantiate data in IT-systems.

The standard ISO TS 13972 specifies characteristics and processes for Detailed Clinical Models (DCM). It states that each clinical model must have medical knowledge, bibliographical information, data element specification and code binding to terminologies. Each data element must have a name, definition, code, code system, data type, and where applicable a unit or value set enumeration. In this standard the Detailed Clinical Model (DCM) is defined as a logical model designed to express one or more clinical concepts and their context in a standardized and reusable manner, specifying the requirements for clinical information as a discrete set of logical clinical data elements.

The demands put on a clinical model are heavy. The structure must be able to reflect the reality of medical practice and the context and the processes in which the patient is seen. The content must be communicable in an semantic interoperable way with other systems. The content must survive transitions of old, obsolete computer systems to newer, modern systems (legacy conversion). Anyone who has seen the dawn of the personal computer and migrated along with increasing processing power from DOS platforms in successive versions to Windows platforms in successive versions knows that this is a formidable challenge. And yet, choosing the right clinical model may achieve all these goals to some extent.

That is because the expertise of numerous brilliant experts and the rigorous work of Standard Developing Organisations (SDOs) has gone into the creation of ISO and CEN norms. The existence of these standards help software vendors to harness the immense complexity of producing Electronic Health Care Record applications.

**Basic clinical models**

There are various standards or specifications for the production of medical records systems: OpenEHR, HL7, and EN 13606.

- **openEHR** is an open source set of specifications defining a health information reference model, a language for building 'clinical models', or archetypes, which are separate from the software, and a query language. The architecture makes use of external health terminologies. It serves as a specification for the electronic health records.
- **HL7** (Version 3 Clinical Document Architecture (CDA®)) is a document markup standard that specifies the structure and semantics of "clinical documents" for the purpose of exchange between healthcare providers and patients. It defines a clinical document as having the fol-
lowing six characteristics: 1) Persistence, 2) Stewardship, 3) Potential for authentication, 4) Context, 5) Wholeness and 6) Human readability. The HL7 CDA is based on the HL7 v3 Reference Information Model (RIM) and the content of the documents is specified in the form of clinical statements or HL7 templates. There is also HL7 V3 messaging and HL7 FHIR (Fast Healthcare Information Resources), which takes the best of HL7 CDA and HL7 v3, while supporting mobile technology.

- EN 13606 (health informatics - Electronic Health Record Communication) is an international standard for the definition of a rigorous and stable information architecture for communicating part (definition of data objects to be used in EHR-system interfaces) or all of the Electronic Health Record (EHR) of a single subject of care (patient), preserving the original clinical meaning intended by the author, and reflecting the confidentiality of that data as intended by the author and patient. EN13606 is by design both for communication between EHR-systems and also for use in interfaces inside EHR-systems. EN13606 defines facilities for the Patient Mandate (Access control to detailed data).

In the SemanticHealthNet Project, these three standards have been compared to each other, and to the clinical models proposed in the European eP SOS project on Patient Summary and Medication list, with the conclusion that well-formed information in an electronic health record, pertaining to either one of these models can migrate back and forward between EHRs without substantial loss of information (see the analysis of clinical and technical specifications of a heart failure patient summary in SHN Deliverables 1.2 and 1.3).

One of the reasons why these three standards for Electronic Health Records are similar is because several high level ISO standards exist with sound ground rules for information architects:

- ISO/HL7 10781 the Electronic Health Record System Functional Model describe a large set of functional requirements in various sections, up to the level of conformance statements against which the EHR System can be evaluated. It serves as a frame around the clinical models, where clinical models can reside in the various sections such as vital signs, medication, public health reporting.
- ISO 13803 defines the logical structure of an electronic health record, with a focus on long-term storage, clinical use, aggregate use, workflow and continuity support etc. This serves as the overall organizing principle for clinical models.
- ISO 13940 system of concepts for continuity of care describes a conceptual model of health care, in which actors, goals, processes and more are defined and modeled. This serves as the larger context for most clinical models.

For assuring the interoperability, a sound standardized structure of the Electronic Health Record is very important. But the physician has to interact with that structure through an interface, and struggle to get the information gathered from the patient into the system.

Patients can be quite chaotic in presenting their symptoms, signs and reasons for encounter. Physicians try to structure the rather short time span of the consultation by attempting to impose some structure to the encounter with the patient. They must work their way through processes of diagnosis from differential diagnoses, over work hypotheses to confirmed diagnoses. The techniques for conducting a consultation with the patient must reflect in the structure of the documentation, too. An early example of this approach was developed by Weed in the sixties, by proposing the Problem-
oriented Medical Record, and the structuring of progress notes under the SOAP headings (Subjective data, Objective data, Assessment, Planning).\textsuperscript{18,19} The approach is still valid for modern electronic records.\textsuperscript{20} There is a tension between expressivity of clinical notes and structured clinical documentation, and this requires attention from the clinical model developers.\textsuperscript{21}

When there is a great variety in the strategy for running a consultation and for documenting observations among physicians, while using the same model of electronic Health Record, this must inevitably lead to problems with interoperability, in case the modellers are not aware of these realities. On the other hand it is true that a well-structured electronic record can help the physician to organise his workflow. However, rigid models may impose a mode of working which can be perceived as unnatural to many physicians, and might lead to misinterpretation of information in systems.

**Specialised clinical models**

Electronic health records are kept in different settings for health care. Communication between settings often arises in the form of key documents or subsets from the record. When the information migrates through these documents in the health care system, it becomes crucial to add context information, and to position the information within the health care process. Also, rules about cooperation between health care settings and responsibilities to provide reliable and timely information are needed. This requires coordination in the development of electronic medical records in different settings and in the development of the clinical model of the different communication documents.


\textsuperscript{19} Weed LL. Medical Records, Medical Education and Patient Care; The Problem-Oriented Record as a Basic Tool. Cleveland, USA: Case Western Reserve University Press; 1969.


In a well-organised public health care system, with universal coverage, the function of primary care is crucial. The electronic health care record of the General Practitioner plays a pivotal role in the flows of information.

Recently, the idea that this EHR needs a structured Patient Summary, where the wealth of information inside the medical record is regularly summarised and structured for outside scrutiny under conditions of privacy. This Patient Summary can be placed on a protected site on the web in the e-health platform. It contains the basic elements of the EHR, and might contain a number of patient-specific extensions. Example are a heart failure extension for the Patient Summary of a patient with heart failure, including documentation of advance care planning, or cardiovascular status, with possibly subsets for smoking cessation status.
A great number of detailed clinical models can be built inside the Electronic Medical Record or as extensions of a basic Patient Summary. The use of extensions is the particular approach applied in HL7 FHIR.

To assure and preserve interoperability, a systematic approach to the creation and coordination of such clinical models is needed. In Annex 6, the basic principles of creation of clinical models are explained.

It is generally recognised that an agile approach to rapid development of new clinical models is needed. In deliverable 5.2, the results of a meeting of SemanticHealthNet with the CIMI consortium, created to assure flexibility and respect for standards in health informatics are presented.

Much is expected from new clinical models created with involvement of patients, with regard to Patient Reported Outcomes, patients’ preference for information and for shared decision making, and for patients’ care goals.

**Nodes, knots and clinical entities within clinical models**

Information can be structured in many ways: from narrative clinical statements to the simple excel sheet or two-dimensional database to more complex class-oriented data management systems, or the new way to store information in the semantic web in triples.

---

In the atomic structure of the information in clinical models, data are represented in rows and columns, in fields and variables, data categories and domain values, in classes and attributes, in triplets.

<table>
<thead>
<tr>
<th>Two-dimensional Data recording</th>
<th>Object-oriented data recording</th>
<th>Triplet-oriented data-recording</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excel files with collums and rows</td>
<td>DBMS systems with tables, data categories and domain values</td>
<td>RDF format with subject, predicate and object</td>
</tr>
<tr>
<td>Databases with records and fields</td>
<td>UML models and XML schema's with classes and attributes</td>
<td>Linked Open Data on semantic web</td>
</tr>
</tbody>
</table>

Figure 9. Data-types in two dimensional, object-oriented, and triplet-oriented data recording.

Triplet-oriented data-recording is a basic method of the semantic web. Information is stored as a basic structure of subject, predicate, and object. This representation in graphs has major advantages, such as being multi-dimensional, but also being able, such as with N3 notation, to carry rules along with data representation.

Figure 10. Resource Description Framework (RDF) triplets.
The good news in this bewildering complexity of information representation is that migration (at least from complex to simple systems) is possible and can even be seamless. This is supported by the earlier analysis of various clinical model approaches.\(^{23, 24}\)

However, each domination of knowledge representation has its own way of naming the nodes and knots of the clinical model. The general term for these nodes and knots is: clinical entities or data element (ISO TS 13972: 2015).

In addition, clinical entities for the same element of structure in different clinical models to represent the same topic (e.g. smoking cessation status) can have different wordings and different value sets. It is another source of Babylonian confusion.

**Libraries of clinical entities/data elements/data categories) and value sets**

In the world of the computational linguists, a first attempt to solve the problem of discrepancy in naming entities in information models has been elaborated. ISO-CAT\(^{25}\) is a registry of data categories and domain values (or phrased differently: clinical entities/data elements/data categories and value sets), that can be used to build terminological models. It is a voluntary initiative, counting on the discipline of designers to harmonise and recycle the bits and bytes of modelling information. It may become an extremely powerful approach, when multilingual management and terminology binding is associated to these efforts.\(^{26}\)

**Translation of clinical entity names and value sets**

In the European program EpSOS\(^{27}\), a number of value sets for description of medication have been translated in a number of European languages. These multilingual resources should be available on the web in a standard format.

**Terminology binding to clinical entity names and value sets**

The next and final set is that the names of the clinical entities (or data elements or data categories) and every element of a value set (or collection of domain values) is bound to (possibly multiple) international reference terminologies and their codification (ISO TS 13972)\(^{28}\).

---


\(^{25}\) http://www.isocat.org/


\(^{27}\) http://www.epsos.eu

\(^{28}\) ISO TS 13972: Health informatics — Detailed clinical models, characteristics and processes. Geneva, ISO.
2.7 Methods for knowledge resources (ontologies, guidelines, decision rules)

Ontologies

Ontology is the philosophical study of the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations. In information science, ontologies are resources for formal naming and definition of the types, properties, and interrelationships of the concepts\(^{29}\) that really or fundamentally exist for a particular domain of discourse. The fields of artificial intelligence, the Semantic Web, systems engineering, software engineering, biomedical informatics, library science, and information architecture all create ontologies to limit complexity and to organize information (free after Wikipedia).

![Figure 11. Representation of reality with universals and particulars.](image)

Leaders of the learned medical societies will have to participate in the construction and maintenance of an ontology for their domain, because the quality of that ontology will be of utmost importance for the quality of information management as a whole.\(^{30}\) This effort will imply a coordinated action with other medical disciplines.\(^{31}\) In Annex 9, more information is given on the recent developments in top-down and bottom up ontologies. The ontology resources will be very important to make the other terminological and architectural resources work properly with interoperable applications. Creating and maintaining ontologies is a cumbersome, invisible endeavour, but of primary importance to the achievement of the many promises of semantic interoperability. The participation of clinicians in this work again is crucial for the quality of the end result.


Guidelines, knowledge databases, decision support

Increasingly, the recorded medical information is confronted with external information from knowledge databases. The guidelines produced by the learned societies are transformed in knowledge databases, which can be consulted during the interaction at the point of care, or reshaped into structured scripts that can interact with the data from the EHR and produce alerts, reminders or specific guidance, to be seen by the physician during interaction with the patient.

It is a tremendous task to produce internationally accepted knowledge databases, certified for being Evidence-based, and capable to interact with numerous proprietary record systems. Some commercial vendors create their own internal knowledge databases within their applications, to assure much needed functionality in medication safety. Although some of these systems may currently work in as satisfactory way, it is a tremendous task to keep the ever evolving knowledge behind the scene up to date and reliable. It is unlikely that in the long run this business model will be sustainable. Moreover, it is important to guarantee the independence of the producers of content for the knowledge databases.

Much is expected from Computerised Decision Support Systems (CDDS) but their effectiveness has yet to be demonstrated on hard outcomes. Lack of interoperability is often mentioned as a reason for unfulfilled expectations. If the clinical decision making functions or system would deploy the same clinical models as the EHR systems and messages apply, part of this problem might be overcome. Without a shared semantic format to express data in a uniform way each CDS-system will be encounter proprietary and local, data definitions. Each implemented CDS-system will then be a unique one.

Quality indicators

Increasingly, faithful follow up of vital recommendations in evidence-based guidelines is a seen as good and expected clinical practice, provided the guidelines are developed in a rigorous, evidence-based, and independent way. Public health policy makers may want to reward health care workers adhering to the guidelines and maybe intervene among health care workers who deviate from the

---

guidelines, without justified motivation. Health care workers and their scientific organisations may want to promote high quality work and therefore develop quality indicators to inform clinicians of areas for improvement, or to confirm the quality of their work. Quality indicators may be derived from re-used existing data or from data that is recorded specifically for reporting.

In the field of pharmacotherapy for the elderly, many explicit criteria of potentially inappropriate prescribing have been developed. The existing lists of these explicit criteria have been reviewed in a recent European Science Fund expert meeting. The experts recommended that a number of these criteria can evolve through scientific research to the status of quality indicator, and that some of these quality indicators can be operated on a large scale in electronic Health records, provided semantic interoperability standards are respected. Similar developments occur in other fields of medicine on the European Level.

3 Lessons learned and recommendations

3.1 Patient involvement

In the SemanticHealthNet Project patient involvement was addressed in several ways. Focus groups were held with patients to discuss the completeness of the clinical model of the heart failure summary, as conceived by heart failure specialists (See SHN D3.2). In Annex 10, the relevance of the International Classification of Functioning, Disability and Health (ICF) for patients was explored. In Annex 11, the worldwide used data-collection system for monitoring patients with long term health conditions (Resident Assessment Instrument –interRAI) was cross-linked with the terms of ICF.

Finally, due emphasis was given to the importance of incorporating lay language in vocabularies for consumers and directly into national reference terminologies (interface terminologies) in Annex 1.

Lessons learned are that involvement of patient requires a commitment of the leadership of a development team and a systematic approach. The effort is reworded by

- much richer clinical models
- alignment of clinical models for Electronic Health Records, kept by the health care providers, and Patient Health Records, kept by the patient
- terminological resources that help to abolish hurdles in communication between lay men and professionals.
- alliance with patients representatives, who can be strong supporters of semantic interoperability, once the advantages to patients become clear.

3.2 Involvement of clinicians

Clinicians are busy professionals, absorbed in daily work. It is not easy for them to step out from daily routine and engage in reflexion on the nature of their profession and technical discussions with ICT-experts. Development teams must reach out to clinicians to drag them into their projects. It is their responsibility that this is done in the most efficient way, and more than lip-service to user friendliness. Leaders of learned medical societies must select among their members the unsung heroes who are interested in these dry matters of classification, terminology, ontology and ICT. These emissaries need training and support, and funding of their time off from clinical work. The most likely candidates will be found among the colleagues who are already involved in medical education, research, guideline development and dissemination.

It is difficult to tell which incentive will be the most crucial for each of the medical domains, to motivate their clinicians to adopt strong interoperable data-entry systems.

As a general rule, one may say that for clinicians the strongest arguments lay with the enhancement of quality of their own medical documentation inside the group of colleagues working together. In addition, the internal secondary uses may be the strongest drivers towards interoperability. Physicians may be found ready to put extra effort into high quality documentation if that is rewarded with better decision support (the push model with alerts and reminders) or with better access to practical information at the point of care (the pull model, with info-buttons, guidelines, and digital libraries). Therefore, the binding of diagnostic codes in the medical record also to the MeSH keywords can be
very motivating, because it supports easy access to digital libraries and the basic scientific literature, for anyone who wishes to escape from cookbook medicine, or answer very specific clinical questions.

Sophisticated documentation coupled to multilingual terminologies can facilitating communication with the patient through linguistic support, and multilingual communication with specialists, treating our patients travelling in foreign countries.

Internal systems for quality assurance can also be powerful drivers.

The external secondary use of medical information is of course also important for clinicians but in a more indirect way. Researchers, companies, and ultimately the patient will profit from the hard everyday labour of correct and extensive medical documentation.\(^38\)

Another form of secondary use is for the health authorities who can build quality control systems and even quality-based financing, based on information from the electronic health record. Those kind of applications may not always motivate clinicians, and even instigate bias and unintended effects.

In Annex 7, the experiences of key players in the use cases of SHN with regard to involvement of clinicians are described. Also, ISO TS 13972 gives explicit guidance in how to involve clinicians in clinical modelling. In SHN deliverable D3.2 a systematic approach to physician involvement in clinical modelling is given\(^39\).

### 3.3 Need for multilingualism

Computer programming, ICT, and semantic interoperability are issues that are discussed and developed in English, most of the time. The international reference terminologies and most of the ontologies developed up today are also in English. Yet, the world is multilingual, especially in Europe.

Although health care professionals of the world are expected to master the English language for retrieval of scientific information, they will work, train, and engage in continuing education, preferably in their mother tongue language. It is their fundamental right to do so\(^40\). In non-English speaking countries, patients will interact with the health care system in their own language.

In the world of EBM, international sets of structured guidelines (e.g. Duodecim Guidelines from Finland, available in Finnish and English) have been translated in German, French and Dutch. When guidelines are available in the local language, this increases the chances that they are consulted by non-native English medical professionals. This set of translated guidelines is intended for distribution on dissemination platform, closely linked through e-Health to the Electronic Medical Records, for point-

---


of-care information gathering, based on the codification of routine medical documentation. Bringing this kind of implementation strategies alive necessitates the intricate interplay of translations of guidelines, decision rules, keywords, and terms in the medical record in the local language and in English. Evidence may have no borders, but there is a language barrier to overcome, before internationally renowned knowledge databases will be able to address a local market for decision support.

### 3.4 Need for a lexical approach and interface terminologies

Both patients and physicians are rightly very keen on the richness and expressivity that natural language permits. They may be reluctant to the reductionism of coding systems, standardized data elements and value sets that force structured data entry. When it is inevitable to interact with coding systems and such data elements, they will expect maximal support by underlying technology.

Given the experience of SemanticHealthNet is seems imperative that work on clinical models and terminology binding is preceded by a basic preliminary creation of a domain-specific lexicon, through methods of term extraction and harvesting the words and phrases of both healthcare workers and patients.

There are multiple advantages for the use of proper standards (Lemon en LMF) to create these lexical resources: connection to Natural Language Processing capabilities, alignment of lay and professional term vocabularies, exploitation of semantic web techniques in the production, easy maintenance and publication of these resources.

### 3.5 No single choice for terminologies

The terminology problem seems to be the Achilles heel of semantic interoperability.

Progress in interface terminologies seems to be slow. There is in many countries a lot of expertise in computational linguistics, but the leaders of the e-Health platform do not seem to be aware or do not have the patience for making mid-long term investment in Natural Language Processing and lexical resources.

Reference terminologies abound. SNOMED-CT has taken a dominant position, and collaborates with many nations and other terminological systems, evolving in a global SNOMED INTERNATIONAL organisation.

The WHO has a whole family of classifications. The most important one is the International Classification of Diseases, now tangled up in version successions, but still the basis for epidemiological morbidity and mortality statistics. Global support for the other classifications is weaker and their existence might be under threat.

---

Allied health associations (nurses, physiotherapists, etc.) are also working on domain-specific nomenclatures and classification systems.

SNOMED-CT is undoubtedly the most granular, and most extensive nomenclature, its ontological layer is in evolution, and still dealing with redundancy from the integration of most of the terms of most of the other terminologies. Whether SNOMED-CT can also replace pure taxonomies and classification systems (often ergonomic constructs for practicing physicians) for classifying findings in epidemiological research remains to be proven. In current reviews of studies on the effect of decision support systems, most systems did not work (yet) with SNOMED-CT.44

It is likely that for a long time coming several combinations of reference terminologies will co-exist. These combinations may be different in function of the domain, or in function of the semantic types (signs, symptoms, reasons for encounter).

Mapping between international reference terminologies may contribute partly to solutions but will not solve a number of fundamental problems.

Terminology binding of clinical entities within clinical models is best performed on a multi-terminology basis. This will contribute to the transparency and scope of applicability of medical documentation.

A too simplistic approach to multilingualism consists in translating a core set of English preferred terms from an international reference terminology to the local language. There is no escape for every reasonably big nation with an official language other than English then to go all the way from harvesting words and phrases in the mother tongue, selecting concepts relevant to the medical culture of the country, and then bridging to codes of a number of international reference terminologies.

There is no need for huge and unmanageable systems. On the other hand it is unlikely that KISS systems (keep it simple and stupid) will do the trick. A selective, sustainable but sophisticated (KISSS) solution is needed.

3.6 A clear choice for a systematic development process based on at least one of the standardized clinical models and reference archetypes

Basic clinical standards for Electronic Health Records

Several similar and interoperable clinical standards exist for EHRs. SHN has reviewed OpenEHR, HL7 CDA, and EN13606 in their representation of the heart failure case. Further, specifically for clinical models, the ISO TS 13972 has been developed.

National certification bodies (but more importantly the payors and Health care users) should demand from software vendors to comply with at least one of them, and no longer accept proprietary clinical models, incapable of archiving data in a standardized and two-level way. Development of new clini-

---

Systematic approach to clinical modelling

ISO TS 13972 gives explicit guidance on development and governance of Detailed Clinical Models. This approach is following the well known ISO 9001 approach for quality management.

In the framework of the SemanticHealthProject, guidance for structuring the development process of domain-specific models has been issued (see Figure 12).45

The principles of constructing Detailed Clinical Models have been reviewed by Goossen46 and are further described in Annex 6.

International attempts to cooperation


A number of international initiatives have been taken to foster cooperation in the development of clinical models between business, Standard Development Organisations, and academia, on an international level. SemanticHealthNet organised in 2014 an international workshop with the Clinical Information Modeling Initiative (CIMI). Another working group within HL7 is Fast Healthcare Interoperability Resources (FHIR, pronounced "Fire"). This group defines a set of "Resources" that represent granular clinical concepts. The resources can be managed in isolation, or aggregated into complex documents.

To represent a variety of clinical models, the Clinical Element Model was designed. The existence of this bewildering array of convergence initiatives testifies of the need of vendors, health authorities, and clinicians to cooperate and get more coherence in the technical aspects of health informatics. In fig. 13, a attempt is made to assemble the artefacts of these similar approaches into one schema.

Figure 13. Summary of reference models and their clinical information model artefacts

3.7 The importance of context

New clinical models should embrace the multi-morbidity aspects of a greying population, with patients attending multiple caregivers, evolving through disease trajectories. Attention to patient-reported outcomes, patient preferences and care goals are new elements to be dealt with.

The new ISO standard System of concepts to support continuity of care (CONTSYS / EN 13940) (described in Annex 8) offers an approach to integrate elements of context and process to the medical record, at the point of care, during the process of medical documentation, in a multidisciplinary way.

---

48 http://informatics.mayo.edu/CIMI/index.php/Main_Page
49 http://www.hl7.org/fhir/
50 www.clinicalelement.com/
The basic idea is that the data-entry systems need to be developed in such a sophisticated way that context information, which usually is implicit, should be made explicit and attached to the data at the point of care. It is difficult to reconstruct contextual information, when medical information has migrated to other systems. In Annex 6 some aspects of different kinds of contexts that are expressed.

In Annex 8, the vision of the consortium EN 13606 Association on semantic operability and interpretability is formulated.

### 3.8 The importance of semantic web techniques

Many of the applications in Health Care will work through the ubiquitous world wide web. They will increasingly use new semantic web techniques, allowing much more powerful and deep information management. Therefore, it is important that the public multilingual terminological resources are published in the format of the semantic web (Linked Open Data), which is facilitated by the use of standards for multilingual lexical and reference terminologies. It will make them accessible to the tools for querying and reasoning with information through logical inference.

Semantic techniques can also be used to facilitate migration of information through nearly compatible computer systems and electronic medical record models, but the use of semantic web techniques to represent the complex health care records is still experimental.

![Figure 14. Diagram for SHN typology of resources.](image)

In SemanticHealthNet, Work Package 4 focussed on this issue, and produced an impressive array of deliverables. In Annex 9, this approach is further explained.

The basic idea is here that information from Electronic Health Records, constructed with one of the basic standards, or even with local, proprietary schema’s, constitutes a first layer of heterogeneous
data. This layer can be supplemented with new layers (semantic mapping using common ontology patterns, and semantic mediator using SNOMED-CT with core ontology set, to produce virtual homogeneous data, that can be accessed by applications.

![SHN semantic-driven interoperability architecture](http://www.semantichealthnet.eu/index.cfm/deliverables/)

In the foreseeable future, a combination of this kind of retrospective approach to remediate for lack of interoperability and greater sophistication at the point of care start of data-entry will be needed.

### 3.9 Sustainability of public resources

As language, medical culture and knowledge and attitudes constantly evolve, these public resources (lexical, conceptual, architectural and ontological) will have to be maintained. Public bodies will govern their maintenance, in cooperation with experts, clinicians and patients, working together in a web-based, interactive editorial environment. It is important that national health authorities and scientific medical associations maintain a sense of ownership of these resources. A good mix of public funding, private initiative, support of scientific associations, collaborative crowd-sourcing among medical volunteers and patient organisations might be able to sustain these resources.

In SHN several deliverables address the issue of sustainability of public resources and business applications.\(^2\)

Annex 8 from EN13606 Association describes the Semantic Interoperability Stack of artefacts and standards that need to be maintained.

In annex 5, an technical description is given for a platform that is able to create, maintain, and host a domain-specific or national reference terminology.

### 3.10 Discrepancy between expectation and progress

One of the most intense cause of frustration is a bad situation that is beginning to improve. Expectations are high, and progress is slow. There are no quick fixes for all the problems, hurdles and barriers, and research-driven progress will take time. An illustration of a difficult but tenacious approach to inter-continental cooperation is given by the experiences of the Trillium Bridge project, where patient summaries from the European epSOS project are compared with patient summaries of the “meaningful use” approach of the US (see Annex 12). In Annex 13, an overview is given of the barriers and incentives to intensify the use of interoperable medical records.

### 3.11 The need for a governing Institute

The challenge of developing semantic interoperability resources, as described in this deliverable, cannot be addressed by anyone stakeholder group, but by multiple stakeholders coming together. This includes clinicians from different specialties and professions, and different countries, health informatics experts, standards developers, developers of health ICT products, representatives from public health, health ministries and clinical research, and many others. At present there is no coordinating forum to unite the stakeholders and to facilitate consensus on a focus (interoperability use case) or multiple high priority foci so that they can work constructively on tackling tractable elements of the total semantic interoperability landscape.

The European Institute for Innovation through Health Data is being formed as one of the key sustainable entities arising from SemanticHealthNet, in collaboration with the Electronic Health Records for Clinical Research (EHR4CR) project and several other European projects and initiatives supported by the European Commission.

Its vision is to become the European organization of reference for guiding and catalyzing the best, most efficient and trustworthy uses of health data and interoperability, for optimizing health and knowledge discovery.

This Institute has been established in recognition that there is a need to tackle areas of challenge in the successful scaling up of innovations that critically rely on high quality and interoperable health data, to sustain and propagate the results of health ICT research, and to specifically address obstacles to using health data that are not being addressed by other current initiatives. It has been formed after wide consultation and engagement of many stakeholders to fill a recognised gap, to develop products and services that can help to maximise the value obtained by all stakeholders from health data, to support innovations in health maintenance, health care delivery and in knowledge discovery. It will importantly bring multiple stakeholder groups together in order to ensure that future solutions serve their collective needs and can be readily adopted affordably and at scale.
The Institute is being established as a European not for profit body, registered in Belgium. It will be governed by its member stakeholders, public and private, through an elected Board and officers. It will be financed by a mixture of membership subscriptions, fees from providing services such as certification and accreditation, specific project grants and other income from education, training and expert advisory roles.
4 Approach to testing

In the SHN project, an iterative process of development of domain-specific clinical models and ontologies is proposed (using the Semantic Interoperability Stack proposed in Annex 8), where the draft models and ontologies are brought back to stakeholders and end-users for cycles of comments and refinement. However, at the intermediate and final stages of the process, there must be a moment of truth, by formal external testing of the results.

There are several levels in the definition of semantic interoperability, and hence, there are also several levels of testing.\(^{53,54}\)

First, there is the quality of technical communication between two different computer systems, without looking at human interface issues.

Second, there is the quality of communication taken the whole chain of interactions, including the interactions of the human entering data and the human receiving data.

As the complexity of the communication increases, other forms of test will be needed. Communication can be between systems with similar or different reference terminologies, clinical models, or ontologies.

The starting point is the actor in health care who is making medical documentation in computer systems for clinical purposes. On the receiving side, functionality can differ from clinical purpose, decision support, epidemiological research, clinical trials, health policy. In principle, each of these functionalities need to be tested separately.

Once the interaction of the end-user (at the sending or receiving site) is taken into account, aspects of degree of professional specialisation (GP, allied health personnel, specialist, super-specialist) come in to scope. In addition, the end-user (on both sides) can be a lay person (the patient or his/her caregiver). Finally, to add to the complexity, communication can be in different languages.

For testing methods, one is referred to ISO 25000 series for software testing. ISO/HL7 10781 holds a comprehensive set of EHR system requirements, including those for semantic interoperability, and reuse of data. The methodology of certification can be used, where a list of functionalities is checked for satisfying execution.

One can engage in process evaluation: How much time and effort was needed? Where did things go wrong? There is outcome evaluation: How much misunderstanding? To what extent were the objectives achieved? What were the (unexpected) benefits and (unexpected) harms? Finally, the satisfaction of the user can be tested: Was the sender satisfied? Was the receiver satisfied?

Sophisticated techniques exist for the observation of end-users interacting with computer systems, with analysis of both process and outcome.\(^{55}\)

---

\(^{53}\) Dolin RH, Alschuler L. Approaching semantic interoperability in Health Level Seven. J Am Med Inform Assoc 2011;18:96-103

\(^{54}\) In : Sicilia MA, Balazote Interoperability in Healthcare Information Systems: Standards, Management and technology, Medical InformationScience reference (an imprint of IGI Global), 2010

---
A more complex testing method is the comparative study of secondary use applications, using the same data migrating through different systems. Recently, the impact of data quality on the use of quality indicators has been studied. It is possible to turn this around and to use established quality indicators to test the quality and semantic interoperability of clinical data migrating through systems; in Fig. 16 an example is given with quality indicators for heart failure treatment.

Figure 16. Testing semantic interoperability with quality indicators in Health Care

There are a number of frameworks to examine the quality dimensions of medical documentation in health care. These frameworks can ultimately be tested on a sequential set of patient records, stored in one system, and then tested again, after migration to another system. It is an approach to test high level of semantic interoperability and/or interpretability.

---

Finally, the Antilope network, a group of international eHealth stakeholder organisations and national competence centres, has released a final set of documents that provide the best available guidance and reference for advancing eHealth interoperability in Europe. They include:

- a refinement of the eHealth European Interoperability Framework,
- recommendations for quality management in interoperability testing,
- an overview of eHealth interoperability testing tools, and
- a description of relevant quality label or certification processes.

These documents have been reviewed and validated in the course of 2014 by national and international experts, stakeholders and policymakers in numerous consultations, discussions and ten regional workshops all over Europe.\textsuperscript{64}

\textsuperscript{64} www.antilope-project.eu.
5 Approach to validation by stakeholders

The proposal of methods to create and maintain domain-specific clinical models and ontologies needs to be reviewed by relevant stakeholders, such as patients, general practitioners, general cardiologists, developers of structured guidelines, and developers of decision supports systems, software vendors, epidemiologists, health policy makers and clinical trial researchers.

On April 29, 2015, a Clinical Advisory Board Meeting was held, to discuss the present deliverable. The minutes are available on the SHN web site.\textsuperscript{55}

This final deliverable will now be circulated to leaders of learned medical societies and to the other stakeholders.

\textsuperscript{55} www.semantichealthnet.eu
6  Annexes

Preamble to the annexes

In this deliverable, we have gathered 13 annexes, which are extensions to paragraphs in the body of the document.

Most of the annexes (8 of 13) are educational summaries of the state-of-the-art for a particular aspect of semantic representation (Annex 1 to 6, 9, and 13); one is a compilation of experiences of key actors in the two use cases of SHN project (Annex 7); one is a position paper (Annex 8); two are reports of specific research projects, commissioned within this project (Annex 10 and 11); one is a report of another European Project (Annex 12).

We provided 5 educational annexes on terminological resources. Annex 1 focuses on terminologies for laymen. Annex 3 deals presents a French Initiative for multilingual management of medical terminologies. Annex 4 focuses on mapping between medical terminologies. Annex 2 deals with the ISO standards for multilingual lexical and conceptual terminologies, while Annex 5 outlines a tool for an interactive web site for the creation and maintenance of a reference terminology.

The educational Annex 9 deals with bottom-up and top-down ontologies; Annex 6 with the development of clinical models.

Annex 7 is a compilation of reflections and lessons learned from the use cases of the SHN project. It is a summary of the source material from those work packages that has informed the body of this deliverable.

Annex 8 is a position paper of the EN 13606 Association on the importance of the new standard for continuity of care CONTSYS, and its relevance for documentation of contextual information. It is an overview of the standards based approach to semantic interoperability.

Annex 10 and Annex 11 are of a different nature and present original work related to the International Classification of Functioning (ICF), focusing on the perspective of patients.

Annex 12 reports on the European / US project Trillium Bridge, on migration of clinical information to and from European to US patients summaries.

Finally, Annex 13 considers the barriers and incentives needed to scale up the adoption and use of interoperable eHealth solutions.

All of the above annexes are targeted at helping clinicians, although not only clinicians, to gain an understanding of the different aspects and approaches within this complicated area of health informatics.
6.1 Medical terminologies for patients

Elena Cardillo

In this Annex, the state of the art of consumer-oriented medical terminologies is described, together with the tools to evaluate readability of medical texts. An approach is presented to the development of a multilingual consumer-oriented vocabulary, covering the most prevalent concepts in medical communication. Finally, integration of such a vocabulary into health information systems and linking to multiple pertinent international reference terminologies (International Classification of Primary Care, International Classification of Diseases, SNOMED-CT) is discussed.


Joseph Roumier

In this Annex, the author describes the existing ISO-standards for multilingual terminologies, both lexical (Lexical Markup Framework LMF) and conceptual (Terminological Markup TMF; ISO-CAT (the international resource of data categories and domain values for linguistic resources); and Linked Open Data as a technology to publish semantic interoperability resources on the semantic web. The structure is described of a proof-of-concept multilingual interface terminology for heart failure.

6.3 Multilingual management of conceptual and lexical resources. A comparison of Bioportal and HeTOP (Health Terminology/Ontology Portal).

Stefan Darmoni, Julien Grosjean

In this Annex, the internet portal HETOP for multilingual terminologies and ontologies is described, and compared to the portal BIOPORTAL, with regard to content, functionality, user interface, and technical aspects. Potential applications of the HETOP site for semantic interoperability and the provision of lexical and reference terminologies are highlighted.

6.4 Mapping between international medical terminologies

Elena Cardillo

In this Annex, the author first describes the available methods for mapping between international medical terminologies, including ontological alignment and the use of semantic web technologies. In addition, existing mapping efforts are reviewed.
6.5 Technical management of multilingual semantic resources

Olivier Latignies

In this Annex, the author describes the architecture of an interactive website for the collaborative, crowd-sourced production and maintenance of interface terminologies as resources for semantic interoperability. The architecture has a dual structure, with a server for the data in an object-oriented and in a semantic framework, and permits the publication of the resource in Linked Open Data.

6.6 Strategic development of structured clinical models for medical subdomains

William Goossens

In this Annex, several initiatives for structuring clinical models are described, with their usefulness for diverging goals in health care, multi-disciplinary collaboration between medical and allied health subdomains, preservation of contextual information, and support for archiving and legacy conversion.

6.7 Experiences from the SemanticHealthNet Use Cases

Ian McNicoll, Charlie McCay, Tony Austin, Shanghua Sun, Gerard Freriks, James Cunningham

In this annex, we collected the experiences of key players in the use cases of the SHN project. The first Use Case was focused on the development of a clinical model for the outpatient letter for heart failure. The second Use Case dealt with the secondary use of clinical data for public health research in cardiovascular disease.


Gerard Freriks

In this Annex the EN13606 Association explains its position on generic EHR-Systems, its actors, its context, its components (HISA), and its processes (CONTSYS) and elaborates on the lessons learnt in SemanticHealthNet.

6.9 Terminologies and ontologies in medicine

Catalina Martinez Costa, Stefan Schulz

In this Annex, existing top-level ontologies (with special emphasis on BioTopLite 2) are described, as well as the relationship between ontologies and clinical models. Furthermore, an thorough description is given of the SemanticHealthNet Semantic Interoperability Framework is given, with special
attention to ontological framework, content ontology patterns and SHN semantic-driven interoperability architecture.

6.10 Perspectives of patients with long term health conditions

Paul Rastall, Darren Wooldridge, Iain Carpenter, Jan Hoogewerf, Birgit Prodinge

In this Annex, the authors describe a process of iterative development and incremental refinement to develop greater understanding of patients’ experiences of long term health problems (i.e. Heart failure) and their interactions with healthcare systems and providers. An exploration was made of the suitability of the International Classification of Functioning (WHO-ICF) to represent patient’s perspectives.

6.11 Integrating what matters to patients into health records based on the ICF: Examining the utility of interRAI to operationalize the ICF Disability Set as a case in point

Birgit Prodinge, Paul Rastall, Darren Wooldridge, Iain Carpenter, Jan Hoogewerf

In this Annex, the authors examine the suitability of the interRAI instrument (a data collection tool for geriatric assessment) to serve as a source of clinical data for the operationalization of the newly developed ICF Rehabilitation Set (based on the World Health Organisation’s International Classification of Functioning).

6.12 Cross Continental Medical Communication

Catherine Chronaki

In this Annex, the key findings from the Trillium Bridge project are summarized, with respect to cross-continentaal medical communication of Electronic Health Record Summaries, in an effort to elicit observations relevant to clinical stakeholders and perhaps engage them into solving some of the easier problems.

6.13 Adoption of interoperable EHRs: barriers, challenges, incentives

Veli Stroetman, Dino Motti, Dipak Kalra

In this Annex, the existing literature on barriers and facilitators for the adoption by health care providers of semantically interoperable Electronic Health Records is reviewed, from a European and an American perspective. Suggestions for breaking the barriers are listed.