
Georgos
Gotis
Ilya Nagibin

Information Security, master's level (120 credits)
2017

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering


**ABSTRACT**

The Swedish National Archives are in charge of the management of Common Specifications (CS). CS are generic metadata specifications that provide structure and markup when transferring digital information between information systems and to electronic archives. As of now there is no CS for electronic healthcare records (EHR). Organizations around Sweden have developed their own specifications for transferring healthcare information. In addition to that, there are comprehensive international EHR metadata standards established. The Swedish National Archives have commissioned a study of EHR metadata specifications and standards to aid in the development of the CS.

A Delphi study was conducted, including respondents from major archiving organizations in Sweden, to identify necessary metadata categories when exchanging EHRs. The data was analyzed considering the international EHR metadata standards HL7 CDA2 and CEN/ISO EN13606, as well as digital preservation metadata categories. The results were a set of metadata categories necessary to include in a CS. In addition, a subset of suggested mandatory metadata categories is proposed and a list of implications for practice. Clinical codes, auditing, and separating metadata related to different contexts are a sample size of the implications.

The results were evaluated in an interview with the Swedish National Archives, as well as Sydarkivera. Three criteria for evaluating the results were proposed, being that the results had to consider a common terminology as well as be based on a metadata standard and Swedish metadata specifications for EHRs. The interview revealed that the results satisfied these criterions, except for requiring a study on one additional user environment of EHRs.

*Key words: Common Specifications, Metadata, Electronic Healthcare Records, HL7 CDA2, CEN/ISO EN13606, Digital Preservation, Metadata category.*
PREFACE

We would like to thank everyone who have been involved in making this study possible. It has been a pleasure and a rewarding experience conducting this research.
# Table of Content

1 Introduction .................................................................................................................. 1
   1.1 Swedish electronic healthcare records .................................................................... 1
   1.2 Understanding metadata ......................................................................................... 2
   1.3 Common Specifications ......................................................................................... 2
   1.4 Purpose and problem ............................................................................................ 2
      1.4.1 Research Question ......................................................................................... 3
2 Key concepts ................................................................................................................ 4
   2.1 Digital preservation ............................................................................................. 4
      2.1.1 Metadata ....................................................................................................... 4
   2.2 Electronic healthcare records .............................................................................. 6
      2.2.1 Integration ..................................................................................................... 6
      2.2.2 Electronic healthcare record Standards ......................................................... 7
   2.3 Research gap ........................................................................................................ 11
3 Research method ........................................................................................................ 12
   3.1 The Delphi study approach .................................................................................. 12
      3.1.1 The Delphi panel ......................................................................................... 12
      3.1.2 Data gathering & analysis ........................................................................... 13
   3.2 Evaluation of results ........................................................................................... 14
   3.3 Scope .................................................................................................................... 15
   3.4 Methodology reflections ...................................................................................... 16
      3.4.1 The Delphi panel ......................................................................................... 16
      3.4.2 Non-ranking Delphi variant ......................................................................... 17
      3.4.3 Questionnaire 1 ......................................................................................... 17
      3.4.4 The evaluation interview ............................................................................ 17
      3.4.5 Alternative methodology approaches ......................................................... 18
4 Results & analysis ....................................................................................................... 19
   4.1 Delphi round 1 ..................................................................................................... 19
   4.2 Delphi round 2 .................................................................................................... 24
      4.2.1 Respondent comments .............................................................................. 25
   4.3 The evaluation Interview ..................................................................................... 27
      4.3.2 Applicability ............................................................................................... 29
5 Discussion ................................................................................................................... 30
   5.1 Metadata in different contexts ............................................................................. 30
      5.1.1 EHR-specific metadata .............................................................................. 30
      5.1.2 Digital preservation metadata .................................................................... 31
   5.2 EHR or EMR ........................................................................................................... 32
1 INTRODUCTION

The Swedish National Archives are part of the Swedish Ministry of Culture. Their functions are determined by the Parliament and the Swedish government. The Swedish National Archives are responsible for supervision of other public authority archives in Sweden. They also preserve records for the future and digitalize some of the holdings. Most of the records in the archives can be accessed by the public. Private individual’s records and businesses are also preserved in the archives (“Om oss,” n.d.). The Swedish National Archives develops and manages standard specifications for transferring information called Common Specifications (CS). The Swedish National Archives vision for information management is that archiving, finding, and reusing the information should be easy regardless of where and how it is stored.

During the spring of 2016, the Swedish Association of Local Authorities and Regions (SALAR) assigned Sydarkivera the project of developing a CS for electronic healthcare records (EHR). Sydarkivera is a local authority operating in the south of Sweden. They are assigned to be handling the archival services for their confederates regarding IT-based business systems, Governance of communal archives, counseling and support for document management as well as for archives and archival functions. The CS department in the Swedish National Archives has the role of tending the proposal from Sydarkivera.

This report has been commissioned by the Swedish National Archives with the aim to study the EHR-landscape including both international standards as well as national initiatives to aid in the development of a CS for EHRs.

This report has been commissioned by The Swedish National Archives to study EHR metadata specifications and standards. The results will aid in the development of the CS for EHRs. The report is structured as follows: Introduction, Key Concepts, Methodology, Results & Analysis, and Discussion that ends with a conclusion.

1.1 SWEDISH ELECTRONIC HEALTHCARE RECORDS

Swedish EHRs may only belong to one patient. The EHR can include different types of data and media such as test results, videos, images etc. The notes in the EHR can come from different healthcare institutions. The notes can also be written and signed by persons with different roles (e.g. doctor, nurse, or assistants). A note can for example be drugs prescribed, given treatments, or observations. The purpose of the EHRs is to enhance the quality of the medical care. It is also a legal document that can be used for research. After an EHR have been transferred for archiving it must be preserved for at least 10 years. It must also be preserved for at least 10 years after the date of latest insert (“Frågor och svar om patientjournaler,” n.d.).

The National Board of Health and Welfare (“Frågor och svar om patientjournaler,” 2017) suggests that EHRs should include:

- A patient identity or identification
- Records of previous received healthcare
- Records of diagnosis and reason for actions
- Records of what's been done and what's planned
- Records of information given to the patient regarding treatment alternatives as well as possibilities for new medical assessment
- Information about who wrote notes in the EHR as well as the date and time.
- Written consents
- Patient’s wishes regarding treatment
- Traceability information of medical devices prescribed, extradited, or transferred to the patient
- Issued certificates, referrals, as well as other incoming and outgoing data
- The planned healthcare for the patient

If following information is present or occurs, it must also be included in the EHR in addition to
the suggestions above:

- Current state of health and medical assessments
- Prescription of drugs and other treatments
- Cause of prescription of drugs
- Examination findings
- Hypersensitivity to certain drugs or substances
- Infections
- Discharge notes and summaries of completed treatment
- Vaccine batch number or similar identification of vaccine

### 1.2 Understanding Metadata

Metadata is commonly referred to as data about data, e.g. name, features, creation, and topic relating to one information entity. It comes in many forms and is universal in information systems. Metadata enables information to be discovered, shared and recorded in systems and is key to the functionality. For preservation of digital content, there are different types of metadata and they support different use cases in information systems. The types are Descriptive metadata, Administrative metadata, Structural metadata, and Markup languages. Descriptive metadata is used for discovery and is perhaps the most common type. Administrative metadata is used for decoding, rendering and long-term management of files as well as adding intellectual properties to the content by including technical, preservation, and rights metadata. Structural metadata is used for relating parts of resources to one another. Markup languages are used for mixing metadata and content together, as well as adding semantic features to it. Interoperability, digital-object management, preservation, and navigation of information all relies on metadata to function and to be achieved (Riley, 2017).

Metadata is only useful for people and software if it is understandable. Metadata is only fully understandable by people and software if the same metadata vocabulary is used. XML metadata vocabularies, also known as schemas, element sets, or formats, defines elements together with their attributes as well as in what order and how many times they may appear. These metadata vocabularies can be formally standardized by organizations such as ISO, NISO or W3C (Riley, 2017). Most metadata vocabularies consist of optional fields and a subset of mandatory fields. The difference between optional fields and mandatory fields is that the mandatory fields demand the intended information from the user. Additional fields can typically also be added for extensibility (McGreal & Roberts, 2001).

### 1.3 Common Specifications

CS are well defined formats for exchanging information between different IT-systems. Their intended use is:

- Aid in the tendering of e-services for e-archives and e-records considering local authorities and regions in Sweden
- Exchange of information between business systems
- Transfer of information to an e-archive, as well as to a final archive

CS were developed with the purpose of defining how the information should be described and how it should be structured when exchanged. It should be noted that they are specifications, not established standards. They are also supposed to refer to other independent standards to the extent possible, whilst being customized to cohere with the Swedish context. The CS consists of three parts, being a Specification, a Supplement, and an XML-Schema. The Specification is simply a description of the CS in text. The Supplement contains pictures, value lists etc. Lastly, the XML-Schema is a set of rules for data types and can be used for validating them.

### 1.4 Purpose and Problem

Per the Swedish National Archives (2016), the goals of the CS are to:

- simplify the development, procurement, and implementation of unitary solutions
lower the costs
provide opportunities for facilitating discovery and reuse of organizational information

The CS has a wide area of use and with their support digital information can be exchanged in a structured and standardized way, both internal and externally. As of now there is no CS for EHRs in Sweden. By developing and publishing a CS for EHRs nation-wide in Sweden, organizations can cohere to the same metadata vocabulary when transferring the documents. Some possible results are enhanced interoperability between the information systems and simplified transfer to the e-archives. As of now, organizations around Sweden have developed their own metadata specification solutions for healthcare information. These solutions must be studied to make the CS generic. The CS should also be based on standards to the extent possible. In the field of Digital Preservation, the concept of metadata categories is introduced. A metadata category example is descriptive metadata including chunks of metadata elements with the same purpose, being discovery and retrieval. As of now, there are no metadata categories for EHRs. The focus will be to identify metadata categories for EHRs, like the concept of categories for preserving digital content. The purpose of this report is to determine the most necessary metadata categories when structuring EHRs for transfer. “Necessary” is here defined as the categories of relevance to the development of the CS. Thus, the characteristics of the CS defines what will be considered as necessary. The characteristics of a CS is that it should be interoperable, generic in terms of the Swedish user base, and support preservation purposes. Thus, a necessary metadata category is a category required when transferring EHR content to an e-archive operating within Sweden.

1.4.1 Research Question
What metadata categories are necessary to include in a Common Specification for Swedish electronic healthcare records?
2 Key concepts

The literature review consists only of peer-reviewed material. Two database search engines were used, being Primo, provided by the Luleå University of Technology Library, and Google Scholar. The keywords used for finding the articles were primarily “digital preservation”, “electronic healthcare records”, and “metadata standards”. In Key concepts, metadata is initially reviewed in the field of Digital Preservation. Metadata is then further reviewed to understand the purpose of its existence and the origins of the metadata standards. Then, EHRs are reviewed and the concepts of integration and terminology systems are introduced. Two metadata standard specifications for exchange of EHRs is reviewed after that. Lastly, the identified research gap is presented.

2.1 Digital preservation

Digital preservation is the policies, actions and strategies performed on content over time to ensure accurate rendering despite media failures and technological change (Delaney & Jong, 2015). It consists of a digital life-cycle process which includes data acquisition, ingest, metadata creation, storage, preservation management, and access (Gracy & Kahn, 2012)(Delaney & Jong, 2015). This applies to both created and re-formatted content (Delaney & Jong, 2015).

Standards and guidelines exist for defining levels of digital preservation services. Open Archival Information System (OAIS) is a Reference model with guiding principles for long-term digital preservation, developed by the Consultative Committee for Space Data Systems (Woodyard, 2002)(Delaney & Jong, 2015). OAIS categorizes information required for preservation as Packaging Information, Content Information including Representation Information and Preservation Description Information (Woodyard, 2002). These categories simply state how and where the bits are stored, how to interpret the bits into data, and how to interpret the data as information (Woodyard, 2002).

Per Delaney and Jong (2015), there are two key concepts to digital preservation, being integrity and authenticity. Integrity means that the content is not corrupted over the timespan of the preservation, and authenticity means that the content is what it claims to be. Integrity and authenticity are ensured by the strategies, actions, and workflows that the content goes through as well as the systematic metadata registration of the content during its whole life-cycle (Delaney & Jong 2015). Metadata provides users a way to manage digital objects and can be used for auditing in terms of tracking the history of the object and providing proof of the origin of the source, which is important for the life-span of the object (Qarabolaq, Inallou, Hafezi, & Tabaei, 2013)(Gracy & Kahn, 2012). Metadata in digital preservation is essentially needed for ensuring accessibility long-term (Woodyard, 2002).

2.1.1 Metadata

In the context of digital resources, metadata is used for description and discovery (Woodyard, 2002). Olson D, (2009) presents the metadata definition given by the National Information Standards Organization (NISO):

“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”.

In this definition, users can be both information systems and humans. Metadata can essentially be used to described totally structured resources as well as unstructured information in terms of digital objects (Qarabolaq et al., 2013). For example, it can be applied to both text-based material and three-dimensional objects such as video, audio, websites, PowerPoint presentations etc. (Olson D, 2009). Some examples of metadata can per Olson D, (2009) be:

- The information contained in the head element of a web page
- Technical details about photos from a digital camera such as pixels, width, and height
- Artist and genre of an audio music file
- Text document properties such as file format, creator and location

Metadata must store technical details about the file, structure, how to use it, audit information about what actions have been made to it, proof of authenticity, and rights as well as responsibilities for performing actions on the object (Woodyard 2002)(Troselius & Sundqvist, 2012)(Olson D, 2009)(Ma, 2006). The object also needs to be understood when shared across platforms, e.g. being interoperable (Olson D, 2009)(Troselius & Sundqvist, 2012).
2.1.1 Categories

Metadata in digital preservation can be categorized into administrative, technical, structural and descriptive metadata (Olson D, 2009) (Woodyard, 2002) (Troselius & Sundqvist, 2012). Administrative metadata stores rights and preservation metadata which is information about access, creation, and management (Olson D, 2009) (Otto, 2014) (Troselius & Sundqvist, 2012). Administrative metadata is critical to the preservation of digital resources and have been cited to be the biggest hindrance for robust long-term preservation (Otto, 2014). Technical metadata can in some cases also be included in the administrative category (Olson D, 2009) (Troselius & Sundqvist, 2012). Technical metadata is information about how the digital object was generated or transferred, for example embedded camera settings (Olson D, 2009). Structural metadata contains information about how the object is to be compiled (Olson D, 2009) (Troselius & Sundqvist, 2012). Descriptive metadata stores data related to identification and retrieval (Troselius & Sundqvist, 2012). Examples can be genres, author(s), titles, and subjects (Olson D, 2009).

2.1.1.2 Standards & schemes

A metadata standard/schema provides syntax rules for how to construct values, describe content, and address what needs to be included (Olson D, 2009). The quality of digital resources is enhanced when more information is known about them since their preservation and management can be carried out more effectively (Otto, 2014) (Pahuja, 2011). Easy production of metadata records is crucial because of the wide variety of digital information sources. Per Otto (2014), management, gathering, as well as recording of metadata has always been expensive and the landscape is complex because:

- There are standards that can be applied to all formats, but aren’t specific enough for any one.
- There are standards intended for specific formats that lacks the corresponding metadata for the original source material.
- There are standards that do include source metadata, but are only applicable to one format.
- Boundaries between metadata types are not stated.
- Standards that are not designed for certain use cases are used extensively for those cases nonetheless.

Metadata standards have been crafted to satisfy the necessity of effective metadata gathering, creation, and management as well as for improving quality (Qarabolaq et al., 2013). Some of those standards available today are Dublin Core, PREMIS, and METS. Their primary use is for libraries and museums (Qarabolaq et al., 2013) (Steiner & Koch, 2015) (Delaney & Jong, 2015).

The Dublin Core Metadata Element Set (DCMES), also known as Dublin Core (DC) was developed in the mid-1990s and initiated the development of current metadata standards (Olson D, 2009) (Ahronheim, 1998). A set of categories that could be used to describe electronic resources was requested and the development involved professionals from archives, libraries, indexing, and computer science. The resulting elements were optional, repeatable and 15 in total: title, publisher, language, creator, rights, date, description, relation, identifier, contributor, source, coverage, subject, type and format (Olson D, 2009). The DC elements have been further expanded and other standards have also been developed since the mid-90s (Olson D, 2009).

Preservation Metadata: Implementation Strategies (PREMIS) purpose is to record metadata which is related to information resource maintenance. PREMIS aims to support preservation, understanding, validity and identification of information sources. It is based on XML and includes elements of descriptive metadata, technical metadata, as well as information about regulations of the set and information about agent. PREMIS is technically neutral, meaning that no specialized archiving technology, preservation alternative, or database architecture are required (Qarabolaq et al., 2013). PREMIS consists of five main elements, being Intellectual entity, Object, Event, Agent, and Rights.

Intellectual entity is content considered as an independent intellectual unit, that may include other entities as well. The intellectual unit is for managing or describing the content (Coppens et al., 2015) (Qarabolaq et al., 2013). Objects are the entity or entities in digital form (Coppens et al., 2015) (Qarabolaq et al., 2013). Event is a recording of events that includes at minimum one information source or agent that is documented (Qarabolaq et al., 2013). An event is an action that impacts the object or an agent (Coppens et al., 2015). Example elements of Event are identifier, type, time and history, detailed description, related agents, and resources involved as well as their roles. Agents are connected with events of an object and can be software, agency, person or organization that is involved in the events during the lifecycle of the source or changes in its rights. Rights are simply the rights and permissions for an
agent or an object (Coppens et al., 2015). Example elements of Rights are an identifier, description of the right, rights basis description, permitted cases, validity time interval, resources, and an agent or an interfering factor as well as its role (Qarabolaq et al., 2013).

Metadata Encoding Transmission Standard (METS) is an initiative of the Digital Library Federation and is maintained by Library of Congress (Cantara, 2005). The standard is a framework based on an XML-schema that supports interoperability, scalability and long-term preservation (Kowal & Martyn, 2009)(Cantara, 2005). METS is used for transmitting digital packages across networks and for storage in digital repositories (Delve, Wilson, & Anderson, 2015). A METS document consists of four main components being descriptive metadata, administrative metadata, file inventory, and structural map. Administrative metadata also has four optional sub-components being rights, preservation, source and technical metadata. The descriptive and administrative metadata components are optional whilst the file inventory and structural map are required by the schema. The file inventory consists of all the files associated with the object (both internal and external) and they can be grouped into different areas. The structural map describes the object with hierarchical tree structure, which links metadata to content files (Eden, 2002).

2.2 ELECTRONIC HEALTHCARE RECORDS

EHR is described by the Centers for Medicare & Medicaid Services (CMS) as an electronic version of a patient’s medical history that is maintained by the provider over time (Albert, 2013). Kohli & Tan (2016) refers to as vehicles for improved communication. EHRs must have guaranteed availability, integrity and confidentiality and follow various legislations (Ruotsalainen & Manning, 2007). An EHR may include laboratory tests, diagnostic imaging reports, observations, treatments, therapies, drugs administered, patient identifying information, legal permissions, and allergies (Eichelberg, Aden, Riesmeier, Dogac, & Laleci, 2005) (Jardim, 2013). Attention to EHRs was increased as soon as information systems deployed automatic functionalities for patient registration, order of clinical tests, transmission of test results, etc. (Kohli & Tan, 2016). They should not be confused with earlier initiatives of storing patient records in electronic formats, being electronic medical records (EMR). EMRs are only patient records within one institution, whilst EHR includes healthcare from all institutions, family medical history, and diet history (Kohli & Tan, 2016). Furthermore, EHRs are not simply scanned versions of paper charts. In many cases, they contain more metadata than data itself (Albert, 2013). Metadata for EHRs includes handwritten notes for the specific patient as well as an audit trail of access. One can also look on metadata for EHRs as evidence since it provides the record’s origins, context, authenticity, reliability, and distribution (Albert, 2013). Due to legal requirements, responsibility of documents also must be included (Dogac et al., 2011). Garde, Knaup, Hovenga, Heard, (2007) summarize the EHR characteristics as:

● An EHR is patient-centered, relating only to one individual’s care
● An EHR is longitudinal, meaning that it’s timeframe is long-term (birth to death if possible)
● An EHR is comprehensive, meaning that it includes a record of all events no matter institution, provider, or specialty
● An EHR is prospective, meaning that not only historical events are included, but also instructions, plans, goals, orders and evaluations

2.2.1 Integration

Integration is presented by Kohli & Tan (2016) to be one of the two primary thematic areas of EHR in which the IS discipline can contribute. Integration in this context is the requirement of interoperable patient records across different systems and devices. Interoperability means that two or more systems, components, or applications can exchange and use information. Put into the context of EHRs, ISO adds to the definition that the content should be communicated in an effective manner without any compromises (Kohli & Tan, 2016). The healthcare sector in many countries is a fragmented mix of public and private sector providers, creating difficulties for integration (Kohli & Tan, 2016)(Jardim, 2013). The development of non-standardized communication architectures has created an interoperability gap. Interoperability is a lingering issue that challenges the development and widespread use of EHRs (Kohli & Tan, 2016). The same data can be structured in many ways, even if the same standard is used. Since EHR documents are longitudinal, there are a lot of different document types depending on events (e.g. health examinations, observations, tests etc.) (Dogac et al., 2011). Applying standards for structure and markup is the
international strategy to close the interoperability gap (Pilar Muñoz, Jesús D. Trigo, Ignacio Martínez, Adolfo Muñoz, Javier Escayola, & José García, 2011)(Kohli & Tan, 2016)(Jardim, 2013). Per Kohli & Tan (2016), EHR data integration standards involve:

- a common unique patient identifier - this is required for identification of patients to all their health data records and should therefore be common in all of them (Dogac et al., 2011).
- a messaging standard - this provides syntactic interoperability, e.g. transport of messages across systems, by defining data formats and syntax for exchange of data.
- a data encoding standard - this provides semantic interoperability, e.g. consistent and correct interpretation of messages, by handling the content meaning for human interpretation.

Eichelberg et al. (2005) argues that the only way to solve the interoperability problem is to address the semantic interoperability, as there will always be some healthcare institutes using incompatible EHR standards. To increase consistency, contradictory knowledge between systems during exchange of information must be addressed (Lin, Vreeman, McDonald, & Huff, 2012). The most commonly used reference information exchange standard is HL7 RIM and it has been serving as the data modeling and exchange standard for more than two decades (Kohli & Tan, 2016) (Bhartiya, Mehrotra, & Girdhar, 2016). However, it has its limitations in interoperability between systems. The CEN/ISO EN13606 standard, as well as HL7 CDA, was designed with the purpose of achieving semantic interoperability in EHR communication (Kohli & Tan, 2016)(Eichelberg et al., 2005). Per Kohli & Tan (2016), in addition to harmonization of reference information exchange standards, terminology systems are also a key challenge for semantic interoperability.

2.2.1.1 Terminology systems
There are several international terminology systems. SNOMED CT, The Systematized Nomenclature of Medicine Clinical Terms, is a comprehensive clinical terminology implemented as a standard by the International Health Terminology Standards Development Organization (IHTSDO) member countries (Kohli & Tan, 2016). SNOMED CT can be used to describe procedures and clinical findings (Ochs et al., 2015). It uses numerical identifiers for the clinical concepts with human readable term(s) as descriptions bound to each (Monsen et al., 2016). SNOMED CT has a hierarchical structure with 19 top-level hierarchies which covers different healthcare topics (López-García & Schulz, 2016). It is used by many countries as medical terminology and has incorporated other standards such as the International Classification of Diseases (ICD-9 and ICD-10) as subsets (Monsen et al., 2016)(Ochs et al., 2015).

When looking at clinical and laboratory observations, LOINC is a universal terminology for such reports. With more than 65,000 codes in the current version, LOINC has users in 143 countries and has growing international adaptation. It has been adopted by public as well as private organizations and the desire for translating LOINC into other languages than English is growing as well (Lin, Vreeman, McDonald, & Huff, 2012)(Vreeman, Chiaravalloti, Hook, & McDonald, 2012)(Kroth, Daneshvari, Harris, Vreeman, & Edgar, 2012).

In 2001 International Classification of Functioning, Disability, and Health (ICF) was approved by World Health Assembly(WHO), and is another terminology system intended for a specific type of information. By using ICF it is possible to describe the impact of health conditions on the individual’s functioning. ICF functions as a conceptual framework and a classification system that binds components with descriptions (Escorpizo & Bemis-Dougherty, 2015). ICF helps to describe interactions between person and society, impact of disability on daily living (Yarar, Cavlak, & Başakçı Çalık, 2016)(Ross, Case, & Leung, 2016). ICF consist of three main parts, being body function and structure, activity, as well as participation (Yarar, Cavlak, & Başakçı Çalık, 2016)(Ross, Case, & Leung, 2016).

Kroth et al. (2012) does note that despite all the standards, powerful tools, and perceived benefits of using recognized standard terminologies for coding biomedical data sets, there is limited success when compared to the size of and number of data sets that has the possibility to benefit from terminology standardization.

2.2.2 Electronic healthcare record Standards
A survey and analysis of EHR standards done in 2005 revealed that DICOM SR was the specification that had been stable for the longest time and was most advanced. DICOM is however focusing on medical images and the protocol as well as encoding is complex, making it hard to implement compared to XML-encoding standards. In terms of a
long-term comprehensive solution, the CEN/ISO EN13606 and HL7 CDA were mentioned to be strong candidates (Eichelberg et al., 2005).

2.2.2.1 HL7 CDA2
Clinical Document Architecture (CDA) is since year 2000 an American standard developed by Health Level Seven (HL7). There are two releases, the second being approved as a standard since 2005. The difference between the releases are that the first release can only derive the header from the Reference Information Model (RIM), whilst the second release can also derive the body (Dolin RH et al., 2006). CDA is a document markup standard for clinical documents encoded in XML that specifies the structure and semantics, to aid in the process of exchange. The CDA document can include multimedia content such as image, text, sound and others. CDA ensures structure consistency and can be readable by both systems and humans (Jardim, 2013). CDA documents can be transferred within a message, and can exist independently, outside the transferring message (Dolin RH et al., 2006).

2.2.2.1.1 Structure
CDA is based on the HL7 RIM and the HL7 Version 3 (V3) messaging standard. The messages are essentially XML documents and can be validated against relevant XML schemas (Jardim, 2013). CDA uses the HL7 V3 data types and derives its machine process able meaning from the RIM. The data types and the RIM are mechanisms that enables incorporation of CDA-documents in other clinical systems (Dolin RH et al., 2006).

2.2.2.1.1.1 Terminology
The HL7 V3 vocabulary is based on SNOMED CT and LOINC. A “coding strength” can be set to be with or without any extensions. The only CDA components without extensions are those in the stated value set. However, when extensions are allowed, stated values outside the set can be used. CDA components can also be post-coordinated, combining data types with terminology concepts creating a coordinated expression (Dolin RH et al., 2006).

2.2.2.1.2 Exchange format
CDA documents can be exchanged with any transport solution so long as:

- all the components that are integral to its state of wholeness is included
- content requiring rendering or associated files such as style sheets are included
- critical metadata regarding the CDA instance required for document management is included

There is no need to change any references inside the CDA-document and there are no restrictions relating to document structure for the receiver. Once the exchange package is received, the components can be placed in directories of their choosing (Dolin RH et al., 2006).

2.2.2.1.2.1 Document components
A CDA document has a Clinical document element with a Header and a Body. The purpose of the Header is to describe the document’s context. The purpose of the Body is to contain the clinical report (Dolin RH et al., 2006). The header and body parts will be further explained below.

2.2.2.1.2.1.1 Header
The header of the CDA document stores descriptive, historical, and security related information of the document. Some example data is the patient, the encounter, the involved providers, superseded parent documents as well as authentication. The header also stores exchange-related information. Moreover, it sets the context in clinical terms and document hierarchy by storing identification of the document and identification the document type (e.g. code relating to a clinical concept). Security-related information is also included in the header. For example, confidentiality status and other contextual components can be set and propagate to the Body. Those values can be overwritten for parts in the Body. Some examples of other contextual components are the human language used and time of document creation. Lastly, the header provides an extensive definition of participants such as author, authenticator, and performer in terms of encounters and legalities. Although the roles are often filled by one individual, clarity is added to less common scenarios where there are different authors, participants, and authenticators (Dolin RH et al., 2006). All the clinical components of the header are presented in Figure 1.
2.2.2.1.2.2 Body

The body contains the actual clinical report. The body can either be unstructured and represented by a non-XML body class, or be fully structured. The non-XML body class provides the option to only wrap a header to a non-xml document. The unstructured body provides easier adoption of the standard, since one can start by adding the header and then structure the body later (Dolin RH et al., 2006).

The Structured body class include one or more Sections, which may each include one narrative. The section class is where the header can propagate to the body. The contextual components (e.g. confidentiality, language, participants) in the header are modelled with exact copies in the sections, creating the propagation. If sections have other contextual specifications, those values can be added by changing the context control code (Dolin RH et al., 2006). The sections have an identification, label, a code from the LOINC value set representing what kind of section it is, and a text (which is the narrative). The narrative content is wrapped by the Text element and requires to be readable by humans. In addition to the Text element, other elements can be added to encode as much as possible from the Narrative to enhance further computer processing. Per Dolin RH et al. (2006), some examples of these elements are:

- **Observation metadata** - includes vital signs, blood pressure, examinations, allergies, and other clinical information. These are added as codes from the terminology.
- **Procedure metadata** - includes effective time of treatments, the status, priorities, the actual method of treatment etc.
- **Prescription metadata (SubstanceAdministration and Supply)** - includes medicine, the dosage, status, the consumption time, the provider of the medicine and the patient.

All the clinical document components of the body are presented in Figure 2. CDA provides further flexibility in that it can be extended with additional XML-elements and entries can refer to external objects (Dolin RH et al., 2006).
2.2.2.2 CEN/ISO EN13606

The CEN/ISO EN13606 standard was developed by CEN/TC251 and uses the language Archetype Definition Language (ADL) instead of XML (Pilar Muñoz et al., 2011)(Martínez-Costa et al., 2010). The standard supports interoperability and structures information in the EHR, such as medications, medical history, progress notes, patient demographics, etc. (Pilar Muñoz et al., 2011). Cen/ISO EN13606 consists of 5 parts: Reference archetypes and Terms Lists, Security, Interface Specification, Reference model, and Archetype model. (Pilar Muñoz et al., 2011). Reference Archetypes and Term lists establishes set of coded terms for different components of the Reference model (Pilar Muñoz et al., 2011). Security defines access policies and privileges (Pilar Muñoz et al., 2011). Interface Specification is used for requesting extracts, archetypes or audit logs (Pilar Muñoz et al., 2011).

CEN/ISO EN13606 is based on dual model approach, which consists of Reference and Archetype models (Martínez-Costa et al., 2010)(Pilar Muñoz et al., 2011)(American Psychological Assoc.). It builds on a concept of dynamic knowledge, represented as archetypes, and static information, represented as a reference model. The reference model can for example contain demographics for actors such as organizations, patients, healthcare providers as well as devices. The actors are provided unique identifiers, which are used to provide the possibility to transfer information anonymously. Content added to archetypes, such as documentation of a healthcare event, does not affect the reference model (Pilar Muñoz et al., 2011).

2.2.2.2.1 Reference model

The Reference model is the provision of information. It structures information in a hierarchy of elements including Folders, Compositions, Sections, Entries, Elements, and Clusters being presented in Figure 3 (Martínez-Costa et al., 2010) (Pilar Muñoz et al., 2011). EHR_Extract contains part of or whole clinical record of a patient (American Psychological Assoc.). Folder organizes high level EHR parts such as episode of care, compartments of care, etc. Composition can include record documentation session e. g. test results, reports, etc. or a clinical encounter. Section contains clinical headings such as subjective symptoms, findings, treatment, etc. (Pilar Muñoz et al., 2011). Entry is concrete and has no specialization (Martínez-Costa et al., 2010). An Entry can include a clinical statement, that can be a measurement, symptom, etc. (Pilar Muñoz et al., 2011). Cluster can be used for organization of tables, time series, etc. Element is the leaf node of EHR hierarchy in CEN/ISO EN13606 for a single value. Reference model has capabilities to add version number to the information stored and it supports auditing functionalities by making it possible to log what was modified/requested when and by whom (Pilar Muñoz et al., 2011).
2.2.2.2 Archetype model

Archetype model is the provision of knowledge (Pilar Muñoz et al., 2011). It represents clinical concepts and serves as clinical guide (Martínez-Costa et al., 2010). Information such as the patient’s blood pressure and body weight can as an example be stored in archetypes. The Archetype model includes three parts: description, ontology and constraints. Description can include additional metadata, if needed. Ontology binds archetype nodes to healthcare terms. Standards such as SNOMED-CT can for example be added as terminology (Pilar Muñoz et al., 2011). Archetype constraints specify the hierarchical schema (Pilar Muñoz et al., 2011). Several archetypes can be part of one archetype (Martínez-Costa et al., 2010).

2.3 RESEARCH GAP

The metadata categories for digital preservation are based on metadata standards that are primarily used for library and museum content. The metadata standards for structuring EHR content when transferred is, based on this conducted literature review, mainly concerned with interoperability. It is uncertain if the EHR metadata standards support structure for content being transferred to an e-archive. Additionally, there are no metadata categories for EHR content.
3 RESEARCH METHOD

Research method is divided into four main sections, being the Delphi study approach, Evaluation of results, Scope, Methodology reflections. The Delphi study approach contains information about the expert panel, the questionnaires for the rounds, as well as the data analysis method. Evaluation of results explains the logic behind the interview performed with the Swedish National Archives and Sydarkivera. Scope is where delimitations are discussed and motivated. The final section, Methodology reflections, contains reflections regarding the validity and reliability of the Delphi study and the interview as well as alternative methodology approaches.

3.1 THE DELPHI STUDY APPROACH

Delphi study is a method for collecting data by structuring a group communication process with experts to deal with a complex problem (Okoli & Pawlowski, 2004). A criterion for Delphi study research questions and aims is that they should have a direct bearing on informing decision making, policy, or practice (Brady, 2015). The results of this study will be used in the development of the CS for EHRs. Another motive for using the Delphi study approach for this research was the opportunity of access to experts through collaboration with the Swedish National Archives. Delphi studies also bridges the geographical gap by providing means for collecting the data electronically.

Furthermore, advantages of the Delphi study are that it’s flexible in that it can be used for both qualitative and quantitative data sources and it does not require highly specialized knowledge to be conducted (Brady, 2015). Delphi studies minimizes power dynamics by letting the participants contribute without any knowledge of the other participants (Brady, 2015)(Williams PL & Webb C, 1994). Another advantage of the participant-anonymity is that direct confrontation of the experts is avoided through the controlled interaction of the Delphi study (Okoli & Pawlowski, 2004). Direct confrontation does not promote independent thought and gradual formation of a considered opinion (Okoli & Pawlowski, 2004).

3.1.1 The Delphi panel

In a Delphi study, a group of experts is used to gather judgmental information and produce a reliable consensus (Okoli & Pawlowski, 2004). Through collaboration with the Swedish National Archives and Sydarkivera, contact information of various information managers around Sweden was provided. These persons were requested to provide references to individuals who possess knowledge of archiving and exchange of information between information systems. It was meriting if that knowledge included structuring EHRs. Even though statistical power is not dependent, it is recommended to use 10-18 experts on a Delphi panel (Okoli & Pawlowski, 2004). In the initial request, six individuals were contacted being both information managers and archivists. They were requested to participate as well as to refer to other individuals possessing the knowledge demanded. Out of these six individuals, two answered. In the other attempt, another request of participation was sent out to six individuals, including two from the previous list that didn’t answer. Out of these six individuals, only one answered. However, this individual provided six additional individuals including one that was already in the list of participants. Out of the other five, two accepted to participate. Finally, one individual gave contact after being referred to the study. The goal was to include as many regions and major EHR organizations (archives and EHR coordinators) of Sweden as possible, to get a more comprehensive view of how EHRs are being structured when transferred. This goal guided the reasoning behind contacting some of the individuals multiple times.

To summarize, 15 individuals were contacted firsthand, four accepted participation, and one gave contact after reference. Two of these five participants were senior members of the same organization. They therefore assigned one to participate instead. A total of four experts participated in this Delphi study. Two of those experts represented the view of their organization. These organizations have developed their own specifications for exchanging EHRs to archives. Their specifications are both used on a large-scale. More information about the experts is presented below:

- Respondent 1 - This respondent is an archivist at the regional archive of Västmanland (a region of Sweden). Currently works with archiving EHR-databases to R7e-archive. R7e-archive will be further explained below. This respondent has so far successfully archived two smaller EHR-systems and one patient-administrative system.
● Respondent 2 - This respondent is Chief of archiving in the county of Sörmland and part of the R7 steering committee as their representant. R7e-archive is a collaboration of 10 regions of Sweden and covers about 4 million citizens. It is managed by a central administration staffed with representatives from the province councils (“Så här fungerar det,” n.d.). R7 has developed a specification for exchanging EHRs to their digital archive based on xml and it is in use by the provinces. Respondent 2 is representing the view of the R7e-archive organization.

● Respondent 3 - This respondent, like Respondent 2, is representing the view of an organization. This organization is the Regional Archive of Skåne. This digital archive mainly consists of liquidated EHR-systems. Their digital archive has been up-and-running since 2011 and currently stores 2,7 TB of data from over 37 archived information systems. Currently there are more than 44 million healthcare documents, over 2 million dentist documents, as well as close to 3 million obstetrics documents in these terabytes of data. Apart from healthcare-related documents there are also financial documents, contracts, commissions, and staff-administrative documents. Their stored documents are based on xml-schemas. The schemas are in some cases based on CS, and in other cases specific schemas that suits their organizational goals. In the study, the questionnaires were each answered by a different respondent. These individuals are the information manager and an information architect within the organization and have been working for several years with the digital archive.

● Respondent 4 - This respondent is an information architect at the province of Stockholm. The province of Stockholm has a digital archive in service since 2008 for EHRs. They are also implementing another digital archive for administrative information. Respondent 4 has since 2011 governed and developed specifications, information models as well as xml-schemas for digital archives. In addition to that, the respondent has been involved with the development of some of the CS. The specification in use for EHRs by the province of Stockholm is based on NI (provided by the National board of health and welfare), a service contract provided by Inera, as well as their experiences from their other digital archives. Inera is a noncommercial organization that develops IT-services for healthcare and is owned by SALAR, as well as all the counties, local authorities, and regions of Sweden.

3.1.2 Data gathering & analysis

The communication process in a Delphi study uses a series of questionnaires with controlled opinion feedback. Per Okoli & Pawlowski (2004) the communication process should provide some:

- feedback of contributions from the experts
- assessment of the view/judgement of the group
- opportunity for the individuals to revise their views
- anonymity for the individual responses

Delphi studies typically consists of three rounds of data collection, where the initial questionnaire is based on the literature, the second questionnaire lets the participants provide feedback on the responses of the first round, and the final questionnaire developed from the data of the previous rounds is used to find a consensus (Brady, 2015)(Fink, Kosecoff, Chassin, & Brook, 1984)(Vernon W, 2009). If consensus is not found, rounds of data collection may continue until it is reached (Brady, 2015)(Fink, Kosecoff, Chassin, & Brook, 1984).

The Delphi study conducted consisted of two rounds in which the participants answered questionnaires via email communication. The initial questionnaire was not based on the literature. Instead, the results from the questionnaire were combined with the literature when forming the second questionnaire for the second Delphi round. The reason for doing so was that the EHR standards are too comprehensive and large (in terms of metadata elements) to be meaningfully summarized into a Delphi questionnaire. Additionally, the conductors of this study did not possess the knowledge about the most necessary metadata categories with the purpose of exchange in the EHR standards, making it difficult to provide a list of categories for the respondents to rate/rank from. By giving the respondents the possibility to provide the metadata categories, there is also a smaller chance of the conductors influencing the results (Mahajan, 1976). The first questionnaire was pilot tested before initiation. The rounds had an initial timeframe of five working days with a reminder set two days, including the last day, before the deadline date.
However, due to holiday the second questionnaire had to be extended by one week to provide sufficient time for the respondents.

Thematic analysis was conducted iteratively as the rounds of data was gathered. The analysis was carried out separately and the results were discussed afterwards to increase the validity (Bengtsson, 2016). The analysis between the rounds had a timeframe of five working days as well. The goal of the Delphi study was to identify the metadata categories that are necessary to include in a specification for structure and markup of EHR information when transferred to another information system, most importantly to an e-archive. To get a more comprehensive understanding of what metadata category respondents are proposing, they were asked to briefly describe it with relating metadata elements. This would also provide more clarity when defining the metadata categories. Both Delphi rounds had the goal to identify as many categories as possible and not to reduce or exclude any categories. Consensus in this context is therefore that the higher majority (three out of the total four) of the respondents had no more categories to add.

3.1.2.1 Questionnaire 1
The first questionnaire consisted of six questions (see Appendix 1). Four of the questions had the purpose of gathering a more detailed background information of the participant. The two remaining questions inquired the respondents to propose the most necessary metadata categories when exchanging EHRs, both between business systems and to e-archives. The respondents were asked to propose at least 5 or more categories for both contexts and to briefly describe their proposed categories and if possible to include examples of metadata elements.

The metadata categories with their examples of metadata elements suggested by the respondents were compiled into unified groups of data. These unified groups were then put in comparison to one another to find similarities and differences regarding their purpose and context. The result of this comparison was nine unified groups of data. These were the identified metadata categories and they were given descriptive names and examples of metadata elements based on the phrasing of the respondents and the terms used by the EHR standards as well as the metadata categories for digital preservation. As a last step, these nine categories were compared to the metadata categories and standards in the literature, to determine similarities, differences, and non-mentioned categories.

3.1.2.2 Questionnaire 2
The second questionnaire presented a merged list of the metadata categories proposed by the respondents in the first round and the metadata categories from the literature which were not suggested. The questionnaire requested the respondents to state any missing metadata categories, and if stated to describe them briefly with examples of metadata elements (see Appendix 2). The major rigor control in Delphi studies is the use of consensus in determination of data validity as well as the ability of participants to extend and revise data during the study (Brady, 2015). The first question of the second round provided the respondents the means to review and, if needed, change their opinion. A second question was also added that asked the respondents to determine which of the categories that should be mandatory as well as optional, with brief motivations as to why. The purpose of the last question was to provide some clarity regarding mandatory and optional metadata fields.

Three out of the four respondents had no additional metadata categories to add. It was determined that the necessary metadata categories in the list were those that are necessary to include in a CS for EHRs and additional rounds in the Delphi would not change the results in a meaningful way. Therefore, the Delphi study had reached a consensus regarding the research question.

3.2 Evaluation of Results
When all the data was gathered and analyzed from the Delphi study, an interview was conducted with the contact persons from the Swedish National Archives and Sydarkivera. The interviewed person from the Swedish National archives is a commissioner and metadata-expert that has about 15 years of experience with metadata specifications and CS. The interviewed person from Sydarkivera is a preservation strategist that has been working in projects for electronic preservation on a regional as well as national level for several years. The purpose of the interview was to determine if the metadata categories would be useful for the development of the CS for EHRs, e.g. if there is evidence to suggest that they are necessary to include. The semi-structured format was determined to be most
suitable for the interview since qualitative data was the intended result and the questions that needed to be answered were known in advance, (Rothe JP, Ozegovic D, & Carroll LJ, 2009)(Scheibelhofer, 2008). An interview guide with open-ended questions was constructed beforehand (see Appendix 3). The guide also included a checklist of the interview purpose, format, ethical matters, given permissions etc.

The interview was structured into five phases: Preface, Background Information, Evaluation Criterions, Presenting the Results, and Evaluation.

The Preface informed the interviewee that the gathered data would only be processed for the study. Consent was inquired and received for recording the interview and the wishes for anonymity were made allowances for.

The Background Information phase consisted of easy and soft questions to provide a smooth start of the interview. These questions also provided background information of metadata specifications that supports the validity of their answers.

The Evaluation Criterions phase inquired the target to state what information that is significant to gather prior to the development of a CS for EHRs. In stronger light of the research question, the targets were also asked to state what information a list of metadata categories, constructed for a development of a CS, should be based on. By asking these two questions, a basis for evaluation was created.

Presenting the Results was the phase in which the interview targets were exposed to a summary of the gathered data from the Delphi study. The summary included the source of the data gathered, the table of metadata categories produced, information about which of the categories that should be mandatory as well as optional, and lastly a list of implications for practice concluded from comments given by the respondents in the Delphi study. The respondents were given the time they wished for to comprehend the information and was provided the possibility of explanations if needed.

The last phase, Evaluation, consisted of two additional questions, this time for evaluating the results. The first question inquired the targets to, considering the basis for evaluation provided in the third phase of the interview, state if those criterions were fulfilled. The second question inquired the targets to state how the results may aid in the development of the CS for EHRs.

The interview was recorded and transcribed to lower the chances of missing valuable points. This also contributed to a more relaxed atmosphere because the dialog and the discussions were given all the attention since there was no obligation to take notes (Whiting LS, 2008).

3.3 Scope

The gathered data has been derived from respondents working in Swedish organizations only. International organizations have not been contacted or considered. However, international EHR metadata standards have been reviewed. The review was restricted to two EHR standards, CEN/ISO EN13606 and HL7 CDA2. Other candidates that were considered but not extensively included were DICOM, XDS IHE, and OpenEHR.

DICOM is a standard for exchanging medical images. However, DICOM added an extension to include other clinical data as well, named DICOM Structured Reporting (SR). DICOM SR has a complex technical specification and vendors in the past had no experience with the complex protocol and binary encoding rules, resulting in limited acceptance outside the medical imaging sector (Eichelberg et al., 2005).

IHE XDS, Cross-Enterprise Document Sharing, is a standard for facilitating sharing of documents. The standard only specifies metadata for locating documents, not the actual content. The standard is therefore not sufficient as a stand-alone solution, but can be used in combination of any other EHR standard that specifies the content information format (Eichelberg et al., 2005).

OpenEHR is the foundation that currently maintains the Good European Health Record (GEHR) initiative that introduced the archetype concept (e.g. the dual model approach). The formal language for expressing archetypes
(ADL) was also introduced by openEHR. Conceptually, HL7 CDA2 and DICOM SR templates are very like archetypes (Eichelberg et al., 2005).

CEN/ISO EN13606 and HL7 CDA were chosen because they are presented standards by ANSI and ISO. They are also comprehensive solutions that considers different types of media as well as structure and markup of clinical documents with the purpose of exchange.

3.4 METHODOLOGY REFLECTIONS

3.4.1 The Delphi panel

Although statistical power is not dependent in a Delphi Study, a total of 10-18 experts has been a recommended number of participants (Okoli & Pawlowski, 2004). The size of the panel in this study was a total of four respondents, which is significantly lower than the recommended number. However, two of the respondents were representing their organizations, meaning that even though the respondent count is lower, the number of experts that may have participated in forming the answers provided by the two respondents may bring the count to the recommended sum. The most important goal was to gather information from as many parts of Sweden as possible, to get a more comprehensive view of how EHRs are currently being structured when exchanged. The CS should be based upon existing standards and practices to the extent possible, which further supports that argument. The specifications which the respondents possess knowledge of, are used for archiving EHRs in 11 out of the 21 regions of Sweden. Thus, the gathered data is based on specifications used for archiving EHRs in more than half of the regions in Sweden.

The regions Blekinge, Gotland, Halland, Värmland, Västra Götaland, Jämtland Härjedalen, Norrbotten, Kalmar, Västerbotten, and Kronoberg are not directly included. Large organizations and regions were the main targets during the gathering of experts to the Delphi panel. The largest regions with a significant margin, judging by population, are Stockholm, Västra Götaland, and Skåne in that order (“Folkmängd i riket, län och kommuner 31 december 2016 och befolkningsförändringar 2016,” n.d.). Potential respondents working in the region of Västra Götaland were contacted multiple times but gave no responses. Although two out of the three largest regions were included in the study, gathering data from respondents working in the region of Västra Götaland could have made impacts on the results.

In terms of organizations, three were identified as strong candidates, being R7e-archives, Inera, and eHälsomyndigheten. They were strong candidates because they, on a regional and national level, develop and provide healthcare e-services that require interoperability to function properly. To clarify, their services assemble healthcare information provided by different actors to centralize it. Thus, these organizations possessed significant knowledge for addressing the research question. Inera, which provides e-services on a national level, was contacted multiple times with no responses (“Tjänster,” n.d.). eHälsomyndigheten, whom are responsible for IT-functions and registries used by providers and pharmacies to prescribe as well as dispatch pharmaceuticals, was also contacted without any response (“English, eHälsomyndigheten,” n.d.). Delphi panel respondents from Inera and eHälsomyndigheten could have made an impact on the results. That being said, one of the respondents did possess knowledge of a service contract provided by Inera, which to a small degree includes their reasoning in the results.

Out of the regions that were not included in the Delphi study, Halland, apart from Västra Götaland, were directly contacted with no response. However, there is no evidence of how many organizations and regions that were contacted, since all the contacted individuals were requested to further contact individuals of interest and potential respondents. Since one out of the respondents participating in the study was not directly contacted, there is reason to assume that regions were second handedly contacted. However, including any more of the smaller regions of Sweden were assumed to have no further impact on the results.

3.4.1.1 Organizations as respondents

Typically, respondents answer as individuals in Delphi studies. Their expert judgement of the issue at hand is challenged by one another, with the purpose of reaching a consensus. In this study, two of the respondents were answering on behalf of their organizations. It is unclear if the data given by these respondents is based on their own
or several individuals’ knowledge. If the data given by these respondents was based on several individuals’ knowledge, a concern is that their own individual expert judgement was demoted. However, by having several experts dealing with the questionnaires, a stronger understanding may have been the promoted. Furthermore, that scenario would also promote more comprehensive answers.

3.4.2 Non-ranking Delphi variant
Traditionally, the main aspect of Delphi studies is that the respondents rank/rate the content presented. Per By de Loë, Melnychuk, Murray, and Plummer (2016), even if two people give same ratings, they both might have had different reasons for that. Generic metadata specifications, such as CS, should have considered all the various information and metadata in which the specification was developed for. There would have been no benefits of excluding certain categories of metadata due to low ratings. Even though the nature of the RQ is to identify the most necessary metadata categories, a rating system would only exclude necessary categories, not promote them. Rather than rating the categories, the respondents were challenged to suggest and reason for those that were necessary to include in a generic specification. They were also given the chance to suggest and reason for excluding the categories, or including additional ones, in the second round. Thus, exclusion of metadata categories would have been based on reasoning instead of rating.

That being said, metadata specifications typically consist of a subset of mandatory fields (McGreal & Roberts, 2001). The Swedish National Archives also expressed the relevance identifying the lowest common denominator. Thus, a rating system was included for the reason of identifying what categories that should be mandatory in a metadata specification for EHRs. The rating system would therefore not exclude any necessary metadata categories, only separate those that could be considered as mandatory fields from those that could be considered as optional.

3.4.3 Questionnaire 1
Nature of Delphi studies is clear short questions for the respondents. The questions were rather vague in that the metadata category concept was not extensively defined with examples. The reason for not explicitly defining the concept was that it may influence the answers by restricting the respondents. The main concern with the questionnaire was if the respondents would understand what metadata category stood for. A pilot test was done with the contacts from the Swedish National Archives and Sydarkivera. This pilot test, apart from some minor considerations, gave no indication of obscure questions or terms. That being said, the respondents did answer in different levels of abstraction. There is however no evidence to suggest that the phrasing and questions of the questionnaire were the reason for that. A question being interpreted differently by respondents is only natural. Besides that, none of the answers given were outside the scope of the questions. The respondents had the chance to ask for clarification regarding the questions in both Delphi rounds as well.

The first questionnaire separated the question of metadata categories that are necessary to include into two different contexts, being in the exchange between business systems and exchange to e-archives. The purpose of separating the questions to the contexts was to clarify if there were any differences. However, only half of the respondents provided answers to question regarding the context of the exchange between business systems. All the participants are working with the archival of EHRs, which may suggest that their knowledge is tied to that setting only. The CS developed by the Swedish National Archives can be used in both contexts, which suggests that their CS for EHRs should do the same. The contact person from the Swedish National Archives stated that if a specification can be used to transfer documents to an archive, it can also be used for transfer between information systems. The argument given was that transfers to archives only add metadata related to digital preservation. Thus, the proposed metadata categories were all compiled into the same list.

3.4.4 The evaluation interview
The interview with the Swedish National Archives and Sydarkivera was conducted in one session. The main concern was that one of the interview targets might influence the other. Conducting the interviews in two sessions and individually could have provided more data. Since they had different backgrounds, the evaluation would have been explicitly based on two different viewpoints. However, by conducting the interview in one session, the targets had the opportunity to discuss the questions and results presented, which may have led to a more comprehensive and accurate evaluation.
The interview agenda was the only information provided to the interview targets beforehand. The exact questions and the results that were to be presented were kept secret. If the interview targets would have been exposed to the questions and results to be presented beforehand, the evaluation may have been influenced. Instead, the evaluation criteria were established early in the interview, before the results were presented. Evaluation criteria would not have been useful if they were influenced by the results which they were to be applied to. Since the criteria were only based on the knowledge and judgement of the interview targets and not the results of this study, a more genuine evaluation could be made.

3.4.5 Alternative methodology approaches

3.4.5.1 Case study

Another possible approach for conducting this research was Case Study. This would require a case where metadata specifications for EHRs could be studied. Examples of cases may be existing EHR archives or the IT departments of healthcare organizations. In the case of an existing archive, data could be collected regarding how the EHRs are being preserved, with focus on the metadata specification in use for transfers. In the case of a healthcare organization IT department, data could be collected regarding EHRs and their structure. The EHR structure could then be used to identify coherent standards supporting the requirements.

Case study can be used to understand a situation in great depth by studying it in its natural setting (Leedy & Ormrod, 2015)(De Massis & Kotlar, 2014). The major limitation for the case study design is that the findings are hard to generalize (Leedy & Ormrod, 2015). One of the main characteristics of CS is that they should be as generic as possible. A way to deal with this limitation is to increase the number of cases studied, which is called a Collective case study (Leedy & Ormrod, 2015)(Eisenhardt & Graebner, 2007). A comparison can also be made between the literature and the case, or from case to case. As an example, two hospitals might transfer their EHRs differently. Conducting a case study for the two, comparing their solutions, could have provided relevant contribution for addressing the RQ.

3.4.5.2 Interviews

An interview was conducted to evaluate the results from the Delphi study. However, interviews could have also been used as the main data gathering methodology. Interviews could have been conducted on targets with extensive knowledge of the metadata specifications for EHRs, such as the respondents participating in the Delphi Study. The interviews would have been semi-structured with the overall goal of addressing metadata specifications for EHRs. Presumably, the interview targets would have answered the questions considering the metadata specification they had developed or possessed extensive knowledge of. In that setting, the benefits of controlled feedback are lost. The analysts are the only individuals interpreting the data and challenging the suggestions given. Other individuals with similar knowledge would not have been able to challenge their suggestions and the interview target would not have been able to change their view. In the Delphi study, the experts were challenged with the opinions of considered equals (in terms of knowledge of metadata) and were provided opportunity for reconsideration.
The results and analysis is divided into three sections, Delphi round 1, Delphi round 2, and Interview. Delphi round 1 presents the gathered data from the first round of the Delphi study as well as the analysis process for the construction of the metadata categories. Delphi round 2 presents the gathered data from the second and final Delphi round. In addition, the section summarizes a list of implications for practice. The last section, Interview, presents the evaluation given by the Swedish National Archives and Sydarkivera on the results. The analysis of their evaluation is presented the same way as the analysis is presented in the previous sections.

4.1  DELPHI ROUND 1

Figure 4 presents a table of answers derived from the results of the first questionnaire. This table was derived from analysis of the respondent answers and is a collection of the meaningful units.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient identifier</td>
<td>3/4</td>
</tr>
<tr>
<td>Written consent</td>
<td>3/4</td>
</tr>
<tr>
<td>Confidentiality information</td>
<td>1/4</td>
</tr>
<tr>
<td>General information about patient (name etc.)</td>
<td>1/4</td>
</tr>
<tr>
<td>Organizational unit</td>
<td>4/4</td>
</tr>
<tr>
<td>Descriptive information about originator including operation time</td>
<td>1/4</td>
</tr>
<tr>
<td>Descriptive information about healthcare visits</td>
<td>2/4</td>
</tr>
<tr>
<td>Clinical codes (diagnosis, treatments, prescribed drugs, dosage, conditions)</td>
<td>3/4</td>
</tr>
<tr>
<td>Diagnosis descriptive info (type of diagnosis, bi-diagnosis), text + code</td>
<td>1/4</td>
</tr>
<tr>
<td>Information about drugs</td>
<td>2/4</td>
</tr>
<tr>
<td>Operational time of journal</td>
<td>1/4</td>
</tr>
<tr>
<td>Information about medical devices</td>
<td>1/4</td>
</tr>
<tr>
<td>Patient Record type (referral, notes, prescription etc.)</td>
<td>2/4</td>
</tr>
<tr>
<td>Information about the journal note context, e.g. stage of treatment.</td>
<td>1/4</td>
</tr>
<tr>
<td>Form of care (institutional, non-institutional)</td>
<td>1/4</td>
</tr>
<tr>
<td>Treatment results</td>
<td>1/4</td>
</tr>
<tr>
<td>Measurements/analysis (type, value)</td>
<td>1/4</td>
</tr>
<tr>
<td>Observations (hypersensitivity, treatment, infection/severe illness, description)</td>
<td>1/4</td>
</tr>
<tr>
<td>Provider (s)</td>
<td>2/4</td>
</tr>
<tr>
<td>Date of events</td>
<td>2/4</td>
</tr>
<tr>
<td>Prescription/Administration drugs/vaccine (prescription reason (text + code), name of drug (text + code), dosage)</td>
<td>1/4</td>
</tr>
<tr>
<td>Archivist</td>
<td>4/4</td>
</tr>
<tr>
<td>Journal post (electronic or scanned paper)</td>
<td>3/4</td>
</tr>
</tbody>
</table>
By further analyzing the table of answers, unified groups were constructed through compilation of the meaningful units. To clarify, where the respondents suggested metadata regarding the organization and the organizational unit that provided healthcare to the specific patient, these suggestions are gathered into one group of unified information. Figure 5 is a graphic picture that sheds some light to the process of compiling the groups. Reasoning and analysis for each individual unit is presented further below.

**Information about the patient**

<table>
<thead>
<tr>
<th></th>
<th>1/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient identifier</td>
<td>3/4</td>
</tr>
<tr>
<td>Written consent</td>
<td>3/4</td>
</tr>
<tr>
<td>General information about patient (name etc.)</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Information about the patient was heavily suggested by the respondents to be a necessary category to include. CEN/ISO EN13606 considers patients as components and they are provided demographic information as well as unique identifiers in the Reference model (Pilar Muñoz et al., 2011). In HL7 CDA2 the patient must be included in the header with other management information (Dolin RH et al., 2006). Written consent is not explicitly mentioned in the EHR standards, but can be included in rights metadata. Rights metadata is part of the Administrative metadata category in digital preservation (Olson D, 2009)(Otto, 2014)(Troselius & Sundqvist, 2012).
Information about health state

| Diagnosis descriptive info (type of diagnosis (main diagnosis, bi-diagnosis), text + code | 1/4 |
| Observations (hypersensitivity, treatment, infection/severe illness, description) | 1/4 |

Information about health state consists of clinical information about the patient such as diagnosis, hypersensitivities, infections, or illness. CEN/ISO EN13606 has clinical information stored in the Reference model under the composition hierarchy element, either as clinical headings or as clinical statements (Pilar Muñoz et al., 2011). In HL7 CDA2 clinical information is presented as observation metadata, but the clinical document itself may also reveal some clinical information through the descriptive information in the header (Dolin RH et al., 2006).

Information about treatments

| Treatment results | 1/4 |
| Measurements/analysis (type, value) | 1/4 |

Information about treatments should include a history of all the treatments, measurements, analysis etc. performed on the patient. Treatments, test results, measurements, findings and such are all part of the composition element in the CEN/ISO EN13606 standard (Pilar Muñoz et al., 2011). In HL7 CDA2 treatments are phrased as procedure, and the method of treatment is presented with status and effective time etc. (Dolin RH et al., 2006).

Information about drugs and aid tools

| Information about drugs | 2/4 |
| Information about medical devices | 1/4 |
| Prescription/Administration drugs/vaccine (prescription reason, name of drug, dosage) | 1/4 |

Information about drugs and aid tools consists of all the medicine and medical devices prescribed to the patient including the reason for prescription and dosage. The Reference model in CEN/ISO EN13606 is the provision of information that provides identifiers for medical devices, among others (Pilar Muñoz et al., 2011). HL7 CDA2 has two components, Supply (for products) and SubstanceAdministration (for consumables). Regardless, the consumable or the product are structured under ManufacturedProduct where the providing organization and the product are coupled up (Dolin RH et al., 2006).

Information about journal post

| Responsible editor | 1/4 |
| Patient Record type (referral, notes, prescription etc.) | 2/4 |
| Information about the journal note context, e.g. stage of treatment. | 1/4 |

Information about journal post is the notes/records about the patient inside the EHR from the healthcare encounters. These notes would be structured with descriptive information such as the type of note, who the responsible editor is, and the context in which the note was done. CEN/ISO EN13606 provides the hierarchy of elements in the Reference model. The composition element is the actual clinical encounter and sections with entries are clinical headings and statements, e.g. notes (Pilar Muñoz et al., 2011). The Header in HL7 CDA2 is where all the descriptive information is structured. The clinical document type is set here, as well as an extensive definition of
participants, where the responsible editor is named author. Furthermore, depending on how much of the clinical document that is structured, the body contains the actual clinical report/note (Dolin RH et al., 2006).

**Information about organization**

<table>
<thead>
<tr>
<th>Organization including unit</th>
<th>4/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information about organization</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Information about organization is where all the information about the healthcare providing organization is stored. The respondents were united in that the organizational unit was necessary to include in a CS for EHRs. In CEN/ISO EN13606 demographic information can be added for organizations and be given a unique identifier (Pilar Muñoz et al., 2011). Organization is one of the major components in the HL7 CDA2 standard. It can both be assigned in the header as a participant of the clinical document, as well as being referenced to in terms of substance or supply administration (Dolin RH et al., 2006).

**Information about healthcare event**

<table>
<thead>
<tr>
<th>Descriptive information about healthcare visits</th>
<th>2/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form of care (institutional, non-institutional)</td>
<td>1/4</td>
</tr>
<tr>
<td>Date of events</td>
<td>2/4</td>
</tr>
</tbody>
</table>

Information about healthcare event is the descriptive information about healthcare encounters that includes the date and form of care. The difference between this category and Information about journal post, is that notes in the EHR does not necessarily have to be originating from a healthcare encounter. However, in CEN/ISO EN13606 there is no distinction between the two. The composition element can take form at a single clinical encounter, a test result, a report and any other documentation session (Pilar Muñoz et al., 2011). In HL7 CDA2 the header is where the contextual and descriptive information is structured, there included the date (Dolin RH et al., 2006).

**Information about healthcare provider**

<table>
<thead>
<tr>
<th>Provider(s)</th>
<th>2/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare provider identification</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Information about healthcare providers structures information about the provider to make distinctions between doctors, nurses, nursing assistants etc. CEN/ISO EN13606 provides demographics as well as identifiers for healthcare providers (Pilar Muñoz et al., 2011). HL7 CDA2 assigns participants in the clinical document header, and those participants are referenced to the major component Person in the standard (Dolin RH et al., 2006).

**Preservation information**

<table>
<thead>
<tr>
<th>Archivist</th>
<th>4/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal post (electronic or scanned paper)</td>
<td>3/4</td>
</tr>
<tr>
<td>Info about who is delivering the journal</td>
<td>1/4</td>
</tr>
<tr>
<td>Archivist notes</td>
<td>1/4</td>
</tr>
</tbody>
</table>
Preservation information includes the information tied to the archival of an EHR. Archival of clinical documents must include management-related information such as the operational time, the originator, who is delivering the journal as it can be an outside archival organization, as well as the archivist and notes from that person. HL7 CDA2 and CEN/ISO EN13606 does provide some capabilities that resembles preservation-related metadata, mainly for accessibility. Security in terms of confidentiality and privileges is included in both. In the concept of digital preservation metadata categories, preservation metadata is included in the administrative metadata category. Administrative metadata stores information about rights, access, creation, and management (Olson D, 2009)(Otto, 2014)(Troselius & Sundqvist, 2012). While these examples of information are put in broad terms, there is a strong resemblance to this unified group of metadata, particularly regarding creation and management. The category Descriptive metadata in digital preservation also stores information about author(s), and depending on the context, the archivist may be the author (Olson D, 2009).

**Technical information**

- Technical metadata (files, file formats, encoding etc.)
- Source of data (source system information such as name, version, status)

Technical information includes technical characteristics of the electronic document. In addition to files, file formats, and encoding suggestions were made to include information about the source system. What’s motivating the suggestion is enhanced authenticity and availability. Technical metadata is not explicitly presented by the EHR standards, but CEN/ISO EN13606 does include versions of entries and documents as part of the audit functionality (Pilar Muñoz et al., 2011). In the digital preservation metadata categories, technical metadata is presented to be information about how the digital object was generated or transferred (Olson D, 2009). It is also stated that the technical metadata category may in some cases be included in administrative metadata, mentioned in the preservation information unified group explanation previously.

**Exception: Clinical codes**

Clinical codes (diagnosis, treatments, prescribed drugs, dosage, conditions)

Clinical codes were mentioned by three out of the four participants. Clinical codes are used as representation for clinical information such as diagnosis, conditions, test results etc. They are apparent in the EHR standards as well. In CEN/ISO EN13606, the main part “Reference Archetype and Terms lists” includes a set of coded terms for different components of the Reference model. Here, SNOMED-CT can as an example be incorporated (Pilar Muñoz et al., 2011). HL7 CDA2 communicates through the HL7 V3 vocabulary, which includes the coding system SNOMED-CT as well as LOINC (Dolin RH et al., 2006). Clinical codes were not put into one metadata category since they are the terminology that can be used for all clinical information.

**Metadata categories from the literature:**

<table>
<thead>
<tr>
<th>Narrative</th>
<th>The healthcare visit event in clear text.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural metadata</td>
<td>How the object is compiled</td>
</tr>
</tbody>
</table>
Metadata categories from the literature, that were not explicitly mentioned by the respondents, are three in total. A key part in HL7 CDA2 when structuring clinical reports is the “narrative”. This is the Text element inside the Section that provides human readable information about the healthcare event (Dolin RH et al., 2006). The last two metadata categories, Structural metadata was identified in the metadata categories for digital preservation, and audit information is an important functionality for both preservation and management of sensitive digital content. Structural metadata category contains information about how the object is to be compiled (Olson D, 2009) (Troselius & Sundqvist, 2012). Two key concepts when preserving digital content are integrity and authenticity (Delaney & Jong, 2015). Audit information metadata for digital content in the context of preservation is essential and widely argued for in the literature (Woodyard 2002) (Troselius & Sundqvist, 2012) (Olson D, 2009) (Ma, 2006). The HL7 CDA2 specification provides auditing capabilities through one of its main components, Parent document, in which superseded documents are kept for historical as well as auditing purposes. CEN/ISO EN13606 incorporates audit information through the Reference model. The Reference model audit information class has capabilities to store information about modified/requested, when and by whom (Pilar Muñoz et al., 2011).

Figure 6 represents a merged list of all the metadata categories presented by the respondents as well as those that can be identified in HL7 CDA2, CEN/ISO EN13606, and for digital preservation. Note that Preservation information has not been marked for the EHR standards, since evidence of access metadata was the only metadata identified that could belong to that category.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Example</th>
<th>Respondents</th>
<th>HL7 CDA2</th>
<th>EN13606</th>
<th>Digital preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information about patient</td>
<td>Identification, name, contact information, given consent</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about health status</td>
<td>Diagnosis, bi-diagnosis, allergies, vaccinations, hypersensitivities</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about treatments</td>
<td>Lab results, measurements and analyses, dates and time, healthcare providers</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about drugs and Aids</td>
<td>Drug description, medical equipment, prescriptions, dosage</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about journal post</td>
<td>Responsible journal editor, dates and time, electronic or scanned type, type of journal post e.g. relevant note, prescription etc</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about healthcare provider</td>
<td>Identification and descriptive information</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about organization</td>
<td>Identification code, organization, organizational anti-fraud and compliance limits</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Information about healthcare event</td>
<td>The healthcare visit event in class text</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Structural metadata</td>
<td>How the object is compiled</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Technical Information</td>
<td>File, file format, encoding and source system information such as name, version, status etc</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

4.2 DELPHI ROUND 2

Three out of the four respondents had no additional metadata categories to add. An additional category was suggested by one of the respondent, and will be presented further below. In regards to the second question, one of the respondents answered that they didn’t possess the knowledge to see the consequences of having categories mandatory or optional. However, comments were given by the respondent and those were considered in the analysis. In Figure 7 the respondents’ answers regarding mandatory and nonmandatory metadata categories are summarized. The definition of mandatory in this context is not whether the categories should be included in a specification or not since all the categories were determined to be necessary to include. Mandatory in this context should rather determine if the specification should force the inclusion of the information under those categories.

The respondents were unanimous in that the categories Information about patient, Information about journal post, Information about healthcare provider, Information about organization, Preservation information, and Technical information should be mandatory in a specification for structuring EHRs when sent to an archive. Two out of the three respondents suggested that the category Information about healthcare event should be mandatory. The categories Information about health state, Information about treatments, Information about drugs and aid tools, Audit information, and Structural metadata only had one suggestion for mandatory status respectively. The category Narrative was not considered mandatory by any of the respondents.
4.2.1 Respondent comments

4.2.1.1 The additional metadata category
One respondent argued for a missing metadata category, or rather argued for splitting one of the categories in two. Treatments should not be structured the same if they are performed by another healthcare facility via referral.

R4 quote:

“Treatments/examinations performed by doctors in the healthcare institution concerned should be separated from treatments/examinations performed on other places via referral.”

Moreover, R4 presented the category as:

“Information about referral/healthcare request - date and time for request, problem, executor (example: laboratory) should exist in addition to the actual result.”

4.2.1.2 The mandatory content dilemma
When EHRs are sent to an e-archive they are often older than 10 years. Also, some source systems used by providers are specialized, such as lab-systems, which means clinical documents from these systems only consist of specific information. By setting some categories of metadata as mandatory, fictive information may be forced into those fields. Older systems were not structured the same way as modern EHR systems are, making it unreasonable to have certain demands on that content. Quotes from the respondents are:

R1 - “I know that if you add mandatory fields such as point in time for a specific category it can force fictive dates in order to be able to write EHR information that is missing records of the healthcare contact moment.”

R4 - “The e-archive we have at the regional archive in Stockholm is a “final” archive which means that the information is often older than 10 years when it comes to us. In that setting you cannot have too high demands on the information content since the older medical systems were not structured the same way as today’s EHR systems.”

R2 - “The lab-system only includes a part of the information concerning the patient which for example means that the category Information about drugs and aid tools and Information about health state etc. cannot be made mandatory. There are more examples of “specialized systems”.

To summarize, Information about health state, Information about treatments, as well as Information about drugs and aid tools are the categories that per the respondent shouldn't be mandatory because of specialized systems do not include this type of information.
R3’s way of looking at mandatory and nonmandatory content is that some metadata categories are mandatory, and the ones that are not should provide content nonetheless, being either the intended content or substituted with fixed values. Quotation from R3:

“What we mean by saying non-mandatory information: if the information exists in the system it should be included. If the information is missing it should be substituted with fixed values such as name missing in the source system.”

R4 further argues that the information that makes it easier to identify the records should be mandatory must be included as description for information classes. For example, the information class “diagnosis” may have the added information “heart flare”.

R4 - “health state, treatments, drugs etc. there should always be a description of which type of information (information class) ex. “note”, “observation” as well as an additional level which described the information, ex. note type “reception note”, referral type “pathology”, drug “panodil”, diagnosis “heart flare” etc. That is, information that makes it easier to identify the records should be mandatory.”

4.2.1.3 Separating metadata
Some metadata may be specific to the healthcare organization, and some metadata may be specific to the e-archive. A good idea would be to separate the metadata relating to the two contexts.

R4 - “The confidentiality does not necessarily have to be the same at the healthcare provider as in the receiving archive, it could be good to try and separate the organizational metadata in the source system and the archival metadata which has to be present in the e-archive.”

4.2.1.4 Metadata that should be present at all times
Two of the respondents argued for some metadata categories being tied to each other. The metadata categories in mind for these arguments are Information about patient, Information about journal post, information about the healthcare provider, information about the healthcare event, and information about the organization. Quotes from the respondents are:

R2 - “All patient-related information must be tied to dates and time, journal editor, the provider, and the unit/reception/clinic”

R4 - “point in time/timeframe of the healthcare contact and point in time of the note/event as well as organization/archivist should always exist for all journal posts in an EHR”.

To summarize what has been said, EHR content should always be tied to the specific patient and the organizational unit. Healthcare visits should, in addition to information about the patient and organizational unit, also include the point in time, the concerned providers, as well as the journal editor.

4.2.1.5 Clinical codes
A specification should have opportunities to add clinical codes. Clinical codes provide possibilities for search and compilations if structured, specified and presented in human readable language. They can be added for diagnosis, drugs, procedures, and more. However, clinical codes shouldn’t be mandatory due to absence in older medical systems.

R4 - “There should be opportunities to also add codes, set of codes etc. for diagnosis, drugs, procedures and more but this should however not be mandatory because it is not used in older medical systems.”

4.2.1.6 Metadata categories from the literature
The non-mentioned categories from the EHR standards and digital preservation metadata categories where either only selected by one respondent to be mandatory, or none. The category Narrative was the one not considered mandatory by any of the respondents. However, when looking at EHR-systems instead of specifications, Narrative was suggested to be a requirement. Quote from R3:
“Narrative - Should be mandatory in new EHR-systems. There is lack of this in older ones.”

The categories fetched from the concept of digital preservation metadata categories where the ones that were selected as mandatory by one respectively. Those categories were structural metadata and audit information. Per one of the respondents, audit information should be saved for five years. Audit logs enhance integrity, which is one of two key concepts of digital preservation (Delaney & Jong 2015)(Woodyard 2002)(Troselius & Sundqvist, 2012)(Olson D, 2009)(Ma, 2006). Technical details, rights as well as responsibilities for performing actions due to legal requirements are additional must-have metadata for preserving digital content (Dogac et al., 2011)(Woodyard 2002)(Troselius & Sundqvist, 2012)(Olson D, 2009)(Ma, 2006). Quote from R3:

“All user-logs should be saved for 5 years (who was using the data, what this person did and when (dates), why you used it can be included in some cases.”

4.2.1.7 Implications for practice

The comments given by the respondents are here summarized into a list of implications for practice.

- Opportunities for adding clinical codes should exist.
- Implications related to mandatory categories are that:
  - Clinical codes should not be mandatory since older medical systems may not provide it.
  - Having mandatory fields in the specification may force fictive information in some user cases.
  - You cannot have too high demands on information content since the source may be an older medical system that was not structured the same way as today’s EHR systems.
  - There are specialized systems producing medical documents that does not provide all the information that is provided structure for, which may affect the decisions of mandatory categories.
  - When designing a specification, information that makes it easier to identify the information that should be mandatory was suggested to be included. For example, information type “drugs” and additional level describing the drug “aspirin”.
- Some information may be present in more than one metadata category. For example, the organizational unit, the provider, date and times, the editor, etc. should be included in all types of journal posts in an EHR.
- It may be beneficial to separate metadata related to the source system from metadata related to the archival of the EHR.
- It may be beneficial to separate metadata for treatments done by the provider at the organization where the EHR is established from treatments done via referral. Consequently, an additional metadata category for treatments done via referral was suggested.
- User logs must be saved for at least five years, and include who is using the data, what the user did, date and time of the occurrence, and in some cases the reason for usage.

4.3 The evaluation interview

A summary with citations of the interview is presented below. The representative from the Swedish National Archives is referred to as Interview target 1 (IT1), and the representative from Sydarkivera is referred to as Interview target 2 (IT2). The theme of the interview is discussion and evaluation of the results from the Delphi study. The results from the Delphi study were the table of metadata categories, some lists of the considered mandatory and optional categories, and a list of implications for practice compiled of comments from the respondents.

4.3.1.1 Criterion 1: International standard

When information is gathered to aid in the development of a CS, the fundamental criterion is that the information is based on a standard, preferably on an international one. A standard provides a vast information mass which can be modified by adding or removing parts that are useful or unsuitable.

IT1: “Basically, the most important, the foundation, is, the CS builds upon some kind of, preferably it should be based on some type of standard. Preferably also an international standard since we work a lot with EU.”
Separating metadata relating to the source system from metadata relating to archiving, which is included in the “implications for practice”, is one of the requirements of the CS. More specifically, the requirement is that the metadata relating to the preservation of objects should always be based on other standards.

IT2: “Preservation Information, technical information, descriptive information, there are other standards.”

The standards HL7 CDA2 and CEN/ISO EN13606 were included in the Delphi study, whom are both comprehensive and international. The HL7 RIM, which is used by the CDA2 standard, has been serving as the exchange standard for more than two decades (Kohli & Tan, 2016) (Bhartiya, Mehrotra, & Girdhar, 2016). Both CDA2 and CEN/ISO EN13606 was later developed to achieve semantic interoperability in the communication (Kohli & Tan, 2016)(Eichelberg et al., 2005).

When looking at IT2’s argumentations, some of the metadata categories of the digital preservation field was brought forward. As IT2 argues, there are standards specific to digital preservation, and some examples are PREMIS, METS, and DC (Qarabolaq et al., 2013)(Steiner & Koch, 2015)(Delaney & Jong, 2015). The concept of digital preservation metadata categories, which involves those standards, have been included in the Delphi study.

4.3.1.2  Criterion 2: Swedish user base
With a standard as a basis, a second criterion for the information gathered is that it should be put into comparison with Swedish specifications to extract similarities and differences. The goal is not to conclude how all the users in Sweden structure their information, as there will always be providers structuring information their own way. The goal is to establish if it is the same type of information that the users work with, or if there is any variety.

IT1: “Thus, the most important thing is not how everyone structure the information really, the important thing is, is it different information that you use?”

The basis of users are exclusively regional healthcare and archival-oriented organizations. The Social Service Department in Sweden oversee EHRs that are specified for the elder care and school health services. These EHRs might have some specific information that must be considered. Therefore, a study may be necessary to do on the municipal-level of Sweden, to determine if there is any information variety in their EHRs. However, the elder care may be the primary concern, since the school health services traditionally have used the same products as the primary healthcare of Sweden.

IT1: “Is there any difference if you look at the municipalities elderly care, for example, where they use EHR systems, because I know they do, and they need something specific, because they have a pretty specific area they work with.”

That being said, the school health service is specialized into growth curves of height and weight as well as back inspections. There should however be no differences in the information variety, as these specializations are existent in the primary healthcare nonetheless.

IT1: “There is no difference to the information, really, only what’s in it.”

Since the same data can be structured in many ways, the development of non-standardized communication architectures has resulted in issues for interoperability (Dogac et al., 2011). Application of standards for structure and markup has been the international strategy to address interoperability issues (Pilar Muñoz et al., 2011)(Kohli & Tan, 2016)(Jardim, 2013). However, IT1 argues that it’s key that the information being structured is the same, not how you structure it. Semantic interoperability, which per Kohli & Tan (2016) should be involved in EHR standards, was brought forward. Moreover, Eichelberg et al. (2005) also states that there will always be healthcare institutes using incompatible EHR standards, and that one should address the semantic interoperability to solve interoperability issues.

4.3.1.3  Criterion 3: Common terminology
The final criterion is that the same language is used. The terms and their definitions should be the same no matter where they are used. Even though the same terms are used by all users, it may still be beneficial to have common definitions for them. Clinical codes are the equivalent to terminology in the gathered data. As commented, it was
concurred that they shouldn’t be mandatory. Older EHR systems must be considered, which may not support the clinical codes.

IT1: “...what is also very important, i don’t really know what it looks like on this page, but it is that we speak the same language. That “this” means the same thing here in Boden as it does in Malmö.”

IT1 talks about semantic interoperability again, and this time in more detail. Semantic interoperability is heavily argued for being the most crucial issue to address in the literature (Kohli & Tan, 2016)(Eichelberg et al., 2005). SNOMED CT is a comprehensive clinical terminology system implemented internationally (Kohli & Tan, 2016). There are also specified clinical terminology systems, such as LOINC and ICF. LOINC is internationally implemented as well, and is specified for laboratory observations (Lin, Vreeman, McDonald, & Huff, 2012)(Vreeman, Chiaravalloti, Hook, & McDonald, 2012)(Kroth, Daneshvari, Harris, Vreeman, & Edgar, 2012). ICF, approved as a standard by WHO, is specified to condition impacts (Escorpizo & Bemis-Dougherty, 2015). Kroth et al. (2012) does however argue that standard terminologies for data sets has limited success, when compared to the size and number of data sets that may benefit from them.

4.3.2 Applicability
When the Swedish National Archives initiate the development of a CS, they start with collecting information about metadata categories, what information they should contain, and which ones that should be mandatory. The concluded results were considered to be the foundation work of a CS and an important insertion to the coming development of the CS for EHRs. In addition, the results should reduce the amount of work required for assessing and comparing EHR specifications in the field. Lastly, the results may also be a source of inspiration for the individuals who are working with EHR specifications specified for different types of information.

IT1:
“As I see it, this is the fundamental work, it’s like a start of a CS.”
“This, I think, is an important addition to the development of an CS, or will be.”

IT2:
“It should simplify the EHR work, i.e. to look at the EHRs in other areas, and to compare...”
Discussion is divided into six sections, being Metadata in different contexts, EHR or EMR, Audit trail, Scrutinizing metadata examples, The additional metadata category, and Conclusion. Metadata in different contexts discusses the differences of healthcare data and data related to archiving. EHR or EMR reviews the definitions and characteristics of healthcare documents in light of the results. Date and time as well as roles metadata are discussed considering auditing in the section Audit trail. The metadata examples for the suggested categories are being discussed based on oppositional comments given by respondents in the section Scrutinizing metadata examples. The additional metadata category is being presented in the following section, and finally the Conclusions of this report is presented with contributions and suggestions for future research.

5.1 Metadata in different contexts

R4 argued for differences between metadata in different EHR settings. Confidentiality does not necessarily have to be the same at the healthcare provider source system as in the receiving archive. The interview with the Swedish National Archives and Sydarkivera revealed that separating metadata related to the source system from metadata related to the archive is in fact a requirement for the CS. Furthermore, the metadata related to archiving should always be based on other standards that exist for those purposes. Since the concept of digital preservation metadata categories is included in the Delphi study, the requirement is, to a lesser degree, included in the results. For further discussion, what metadata is specific for the healthcare organization, and what metadata is specific for the receiving archive?

5.1.1 EHR-specific metadata

There are no indications that the EHR standards separate metadata from the source system from metadata that is required when the EHR is sent to an archive. In fact, the sole purpose of the EHR standards seems to be interoperability between healthcare information systems (Kohli & Tan, 2016)(Eichelberg et al., 2005). However, the goal of the CS is twofold, being to provide support for both interoperability and archival purposes. Therefore, the EHR standards can only provide suggestions for half of the intended functionality, being the EHR specific metadata. On the most fundamental level, an EHR standard must include a common unique patient identifier, a messaging standard, and a data encoding standard (Kohli & Tan, 2016). A common unique patient identifier was heavily argued for by the respondents as well. What appears to be the most common messaging standard is XML, which is used by both the respondents and the HL7 CDA2 standard. A data encoding standard is required for semantic interoperability, e.g. correct interpretation of messages for the human user (Kohli & Tan, 2016). Terminology systems and reference information exchange standards are used to address the issue of contradictory knowledge between information systems, making them consistent in their transfer of information (Lin, Vreeman, McDonald, & Huff, 2012). Eichelberg et al. (2005) argued that semantic interoperability is the only way to solve the interoperability problem, since there will always be organizations with incompatible EHR standards. If the knowledge comes from one place, everybody can adopt it and talk the same “language”. This has been referred to as “clinical codes” by the respondents. SNOMED CT is a comprehensive clinical terminology used by both HL7 CDA2 and CEN/ISO EN13606. The National Board of Health and Welfare in Sweden are developing and ensuring the quality of NI, which is being adopted by healthcare organizations around Sweden (“Snomed CT,” n.d.). NI, The National Information Structure, is a generic Reference model that provides standardized structure and terminology. SNOMED CT, among other terminology systems, is the source of the clinical terminology used by NI for describing the detailed clinical content (“Nationell informationsstruktur,” n.d.).

5.1.1.1 Terminology systems

Apart from SNOMED CT, which provides codes for medical terminology, there are other code frameworks that are more specialized into one or a few healthcare concepts. LOINC, ICF and ICD are examples of specialized code frameworks internationally adopted for laboratory data, conditions and disabilities, and diseases (Lin, Vreeman, McDonald, & Huff, 2012)(Vreeman, Chiaravalloti, Hook, & McDonald, 2012)(Kroth, Daneshvari, Harris, Vreeman, & Edgar, 2012)(Yarar, Cavlak, & Başakçı Çalık, 2016)(Ross, Case, & Leung, 2016). The use of terminology and codes is a big part of interoperability for EHR systems, because codes can be used as metadata for clinical terms such as
diseases, prescribed drugs, etc. to make sure that information is unambiguous and will be interpreted as it is intended. In other words, it enhances metadata. Free text is overused in Swedish EHR systems, which hinders a more rapid implementation of NI around Sweden (“Snomed CT,” n.d.). This suggests that the Swedish EHRs, like many other countries, lacks semantic interoperability.

Terminology systems is one of three criteria proposed by the Swedish National Archives that must be considered prior to the development of a CS. Terminology systems are basically a language used by both computers and humans. The codes, which the computer interprets, resembles a clinical term, which the human interprets. Despite all the synonyms of a clinical term, for example migraine being referred to as sick headache, the human and computer users will understand it if the same code is used. Clinical codes may also clears up confusions regarding if the patient is diagnosed with migraine, or is suffering from headache due to other causes unknown (by providing different codes for the two). To conclude, if all parts are using the same terminology system, misinterpretations are less likely to occur. The Swedish initiative NI is being implemented gradually around Sweden and will be the go-to reference model for Swedish terminology and structure. NI includes the terminology systems/frameworks SNOMED CT, LOINC, ICF, ICD and more. Thus, when developing a CS for EHRs, NI must be considered for interoperability reasons.

5.1.2 Digital preservation metadata

Metadata for digital preservation is commonly referred to the categories of administrative, technical, structural, and descriptive metadata (Olson D, 2009) (Woodyard, 2002)(Troselius & Sundqvist, 2012). Although being abstract, these categories promote discovery and preservation of content over long time periods. Legislative concerns are also included, being in the rights metadata associated with administrative metadata. However, how are the rights of the EHR affected when transferred from a healthcare organization to an archive? Ultimately, the documents belong to the patient. When they are being created, the responsibility of the document lies on the healthcare providing organization. When the document is sent to an e-archive for preservation, the responsibility for the document most likely changes to that archiving organization. These metadata complexities during change of possession must be considered when creating a CS.

How much of the digital preservation metadata categories can be identified in the EHR standards? To begin with, Descriptive metadata is data that helps in with identification and retrieval of the document (Troselius & Sundqvist, 2012). There are clear indicators of descriptive metadata being existent in the EHR standards, judging by the document number identifications, roles such as author and editor, as well as document type relating to clinical concepts (Dolin RH et al., 2006) (Pilar Muñoz et al., 2011). Administrative metadata, is metadata about rights, access, and management (Olson D, 2009) (Otto, 2014) (Troselius & Sundqvist, 2012). The analysis of the first Delphi round revealed some equivalents of administrative metadata in the EHR standards, being access capabilities. In addition to access metadata, the date of creation and another management metadata can be identified. Technical metadata, which can be included in administrative metadata in the form of preservation metadata but has been separated here for perspicuity, is metadata from the creation of the object, e.g. how it was generated or transferred (Olson D, 2009). Technical metadata seems to be outside of the scope of the EHR standards, as there are no clear indicators for existence in the literature. Finally, Structural metadata is data about how objects are to be compiled (Olson D, 2009)(Troselius & Sundqvist, 2012). Document versioning and relationships are existent in both EHR standards, which suggests that structural metadata is present.

There is evidence to suggest that the digital preservation metadata categories are implicitly incorporated in the EHR standards, except from the technical metadata. The respondents, whom are archivists and are working with or have developed specifications for structuring EHRs when sent to an e-archive, made suggestions for digital preservation specific metadata and were united in that both the created metadata categories “preservation metadata” and “technical metadata” from the Delphi study includes information that is mandatory to be included when structuring an EHR for archival purposes. That being said, the interview with the Swedish National Archives and Sydarkivera revealed that it is required that the CS base the metadata relating to archiving on metadata standards for those purposes. DC, PREMIS, and METS are examples of metadata standards for the preservation and exchange of digital content. However, these standards were developed mainly for digital content in museums, libraries, and universities (Qarabolaq et al., 2013)(Steiner & Koch, 2015)(Delaney & Jong, 2015). If EHRs have other requirements that are not
considered by these metadata standards, modifications may have to be made. Whether the metadata standards for digital preservation provides all the metadata necessary for archiving EHRs or not is outside of the scope of this research.

5.2 EHR OR EMR

EHRs are patient centered and comprehensive. There has been some confusion in the literature regarding EHRs and EMRs. EMRs, electronic medical records, are medical records of a patient from only one healthcare facility, whilst EHRs includes records from all facilities, making it more comprehensive (Kohli & Tan, 2016). Per the National Board of Health and Welfare, a Swedish EHR must include all healthcare documents and reports that are received by the healthcare facility or have been created there. This suggests that one patient may have multiple EMRs, and depending on if the information have been shared, some of these EMRs may include information from other facilities, making them partly EHRs. Per the respondents, there are strong indications that the healthcare information they archive is EMRs, as one respondent suggested that there are specialized systems providing information which does not include all the information that a comprehensive specification may provide structure for. However, by the nature of archiving, many of the documents sent to the archive may be from a time where even electronic documents were not as established as they are today. There is therefore a chance that the respondents view of healthcare documents is influenced by the older standards of storing healthcare information. Nevertheless, even though the Swedish healthcare information is not always stored in EHRs, as defined by (Kohli & Tan, 2016), that may be the overall goal since information is shared among healthcare institutes and “partly EHRs” are existent. The Swedish initiative NPÖ, National Patient Overview, summarizes a patient’s healthcare information from all sources if consents by the patient has been given (“Nationell patientöversikt (NPÖ),” n.d.). NPÖ, which is provided by Inera, seem to be the proper equivalent for EHRs. However, a CS must consider that the healthcare information may not always be structured to the degree that NPÖ provides. In addition, apart from providing structure for the healthcare information content, the CS must provide structure for archival purposes as well.

The Swedish National Archives suggest that, in regards to the criterion that a CS for EHRs should be based on a Swedish user base, the goal is to determine if the users are working with the same type of information, or if there is any variety. The regions of Sweden are responsible for duties common for large geographical areas, and most importantly the healthcare. However, the municipalities are responsible for a large part of the public service. Among other local tasks, the public service in the municipalities includes care of children, elders, and people suffering from disabilities (“Styra Sverige - så fungerar det,” n.d.). The healthcare related public service in municipalities deals with EHRs as well. Although there is no indication that these EHRs would contain any other type of information, since this user base was not included in the Delphi study this statement cannot be claimed. Additionally, specialized healthcare information systems may have the same consequence, apart from not providing as much information, being that they may bring forward new information.

5.3 AUDIT TRAIL

There are different roles in healthcare which can be bound to an EHR. The Delphi study participants mentioned such roles as archivist, patient, organization, responsible editor and healthcare provider. Different types of notes in Swedish EHRs can be added as well as signed by healthcare providers with different roles (“Frågor och svar om patientjournaler,” n.d.). It can be an assistant writing a note, a nurse adding an observation, or a doctor adding a prescription. For example, if the main doctor is not present, or if the documentation is postponed to a later occasion, another person might do the documentation. The person documenting the healthcare event must therefore be separated from the provider participating in the event. The date and time of the event must also be separated from the date and time of documentation. These capabilities are required for auditing purposes. User-logs (who was using the data, what this person did and when) are required to be saved for 5 years, per one of the respondents. Auditing supports the key concepts of Digital preservation and EHRs, being integrity and authenticity (Delaney & Jong, 2015)(Ruotsalainen & Manning, 2007). Integrity and authenticity is enhanced through auditing since access rules can be set to the different roles and logs can be used as evidence in the case of corrupted content. The National Board of Health and Welfare also suggests that auditing should be present in all EHRs (“Frågor och svar om patientjournaler,” 2017).
5.4 Scrutinizing metadata examples
The metadata examples presented by the metadata categories are based on the data gathered in the first round of the Delphi and the information about the EHR standards. Two of the respondents were opposing that some of the metadata examples are necessary to include in multiple metadata categories. The argument was that patient-related information must be tied to dates and time, journal editor, provider, and unit/reception/clinic. In addition to that, all types of journal posts in an EHR should always have metadata attached about patient, journal editor, provider, healthcare type, timeframe or date, and unit/reception/clinic. Preservation information and Technical information should always exist as well. Lastly, metadata categories such as treatments, health state, and drugs should also have descriptive information relating to information classes. The respondent argued that this would enhance detection and retrieval of the information and should be mandatory.

These arguments suggest that the concerned respondents’ interpreters the metadata categories as detached from one another. The goal of the metadata categories is to gather matching metadata under one topic. However, the metadata categories should complement one another to reduce redundant metadata. For example, the healthcare provider included in one part of an EHR, say a treatment, should not necessarily be defined there. The provider could instead be added with an identifier, and be defined elsewhere. This, in addition to reducing redundant data, also promotes a clear structure. Perhaps, this was the argument of the concerned respondents. However, these are design considerations and how the metadata categories are to be used is ultimately up to the developer(s).

5.5 The additional metadata category
A consideration should be made whether to distinguish treatments which are performed at the hospital/healthcare center and those performed by some other institution or organization through a referral. The proposed metadata category was Information about referral/healthcare request, with date and time for the request, the executor and the actual result as example metadata.

In many countries, the healthcare sector is a fragmented mix of public and private providers with non-standardized as well as standardized communication (Kohli & Tan, 2016) (Jardim, 2013). In addition to that, the same data can be structured in different ways even if the same standard is used (Dogac et al., 2011). When treatments are provided via referral, the healthcare information relating to the treatment may be produced in a different organization. When all the organizations do not cohere to the same way of structuring the content, e.g. when there is lack of semantic interoperability, the communication will not be consistent and reliable (Lin, Vreeman, McDonald, & Huff, 2012). If the content of treatments and referral treatments may or may not be structured differently when produced, it makes sense to not have the same requirements for both. Separating treatments done via referral from treatments done in the main healthcare organization is something that should be considered when creating a CS for EHRs.

5.6 Conclusion
The question underlying this study is:

**What metadata categories are necessary to include in a Common Specification for Swedish electronic healthcare records?**

CS must be based on a metadata standard, a Swedish user base, and support a common terminology. A Delphi study was conducted including respondents from organizations that archives healthcare documents for the Swedish population in more than half of the regions. The Delphi study used two international metadata standards for EHRs (HL7 CDA2 and CEN/ISO EN13606) and the digital preservation metadata categories as supporting literature. The results were 13 metadata categories suggested to be necessary to include in a CS for EHRs. Those are:

1. Information about patient
2. Information about Journal post
3. Information about healthcare provider
4. Information about organization
5. Preservation information
6. Technical information
7. Information about healthcare event
8. Information about health state
9. Information about treatments
10. Information about drugs and aid tools
11. Audit information
12. Structural metadata
13. Narrative

The metadata categories should reflect the variety of information existent in Swedish EHRs. A consideration should be made whether to distinguish treatments which are performed at the hospital/healthcare center and those performed by some other institution or organization through a referral. One should also separate the date and time of the actual healthcare event from the date and time of documentation. Additionally, the healthcare provider participating may not be the person documenting the healthcare event. Thus, a specification should support both user cases. Information presented as examples in one category may be related to multiple categories. How the metadata categories are to be used when designing a CS is however outside the scope of this study.

Metadata specifications generally have a subset of mandatory fields, as in fields forcing information in the specification. Metadata categories 1-6 includes information that is mandatory when structuring the content of an EHR. A strong candidate for being mandatory is the seventh category. Metadata categories 8-12 are up for consideration whether they should be mandatory but category number 13 should not. There are some implications related to mandatory categories. Older medical systems and specialized systems may not provide the information forced by mandatory fields in a specification. Users must add fictive information in those cases. Even though there should be opportunities for adding clinical codes and descriptive information for mandatory categories, one cannot have too high demands on information content provided by the older medical systems.

The CS must provide support for both interoperability between healthcare information systems and digital preservation. Metadata used for structuring the content of an EHR should be separated from metadata used for archival purposes. The Swedish National Archives also demand that the different contexts should be based on metadata standards for their purposes. The metadata categories from the results considers both contexts but does not separate them.

The third criterion, regarding terminology, is the concept of semantic interoperability. Swedish EHRs, like many other countries, lacks semantic interoperability. The National Board of Health and Welfare are currently implementing the Reference model NI around Sweden. NI includes large international terminology systems such as SNOMED CT, ICD, ICF etc. and must be considered for achieving semantic interoperability.

5.6.1 Contributions
The results of this study will be an insertion to the project of developing the CS for EHRs in Sweden. They resemble the fundamental work of a CS and will facilitate the review and comparison of EHRs in other areas. They also contribute to the field of Digital preservation, by suggesting metadata categories for EHR content being transferred to an e-archive. Additionally, the metadata categories in the field of Digital preservation have been studied in terms of their applicability on EHR content. Lastly, contributions have been made to the EHR metadata standards in their support for preservation of healthcare documents.

5.6.1.1 Disclaimer
The results may have a more practical relevance to the Swedish National Archives, than a research relevance. The results are based on a lower number than the praxis of respondents and may only reflect what the respondents view as necessary based on the metadata specifications they have developed and work with themselves. Lastly, the EHR metadata standards may provide further functionality, related to digital preservation, not being presented in the literature.
5.6.2 Suggestions for future research
Since a CS is required to base the archival structure on metadata standards for those purposes, a more thorough investigation may be necessary to determine if EHRs have any specific requirements when being archived. Even though digital preservation metadata standards were included in this study, they were not evaluated in regards to their applicability on EHRs. They were also treated as a concept, not as individual standards. Some of the digital preservation standards may be suitable for different purposes, for example discovery or transfer, when preserving EHRs.

This study focused on metadata categories rather than all possible metadata elements. The EHR metadata standards were not reviewed on a deeper level. Thus, the standards may provide further functionality and structure that was not considered in this study. A thorough review and comparison of the EHR metadata standards would provide more information about them and how information is being structured internationally.

Some relevant organizations in Sweden such as Inera, eHälsomyndigheten, and the region of Västra Götaland were not included in the study. Stakeholders dealing with EHR metadata on the municipal level for elder care and pediatrics were not included either. To determine if there is any information variety from the content provided by those EHRs, they must be studied as well.

This study was conducted in Sweden with Swedish stakeholders. It has not been determined whether Sweden differs from other countries in regards to EHR metadata. Whether the results are applicable internationally or only reflects the metadata structure of EHRs in Sweden is a suggestion for further research.


Om Sydarkivera. (2014, December 3). Retrieved from https://sydarkivera.se/om/


8 APPENDICES

8.1 APPENDIX 1 - DELPHI ROUND 1 TEMPLATE

English version:

Questions to be answered:

What is your name?

What is your line of duty?

What is your experience with exchange-formats for exchange between business system and/or to e-archives.

What is your experience with exchange-formats for electronic healthcare records or other healthcare related information?

Suggest the most necessary metadata categories (5 or more) for exchange of EHRs between business systems. Motivate briefly, if possible with relating metadata elements.

Suggest the most necessary metadata categories (5 or more) for exchange of EHRs to e-archives. Motivate briefly, if possible with relating metadata elements.

Your answer will be confidential and used for the purpose of this study. See the attached file for description of the Swedish EHR information requirements.

Swedish version:

Frågor att besvara:

1. Vad heter Du?

2. Vad arbetar Du med?

3. Vad har Du för erfarenhet med utbytesformat både mellan system och/eller till e-arkiv?

4. Vad har Du för erfarenhet med utbytesformat för patientjournaler eller annan hälsorelaterad information?
Den metadata som söks i följande frågor avser både kompletterande information samt att strukturera innehållet i journalen.

5. Föreslå de nödvändigaste (minst 5) metadata-kategorierna för utbyte av patientjournaler mellan verksamhetssystem. Motivera kortfattat, gärna med tillhörande metadata-element.


För beskrivning informationskraven för patientjournaler, se bifogad fil. All responses will be kept confidential and no results will be displayed at the individual level.

8.2 Appendix 2 - Delphi Round 2 Template

English version:

The necessary metadata categories in a specification for transfer of EHRs between business systems and/or to an e-archive was asked for in the last questionnaire. The following list is a compilation of identified metadata categories from your answers:

- Information about the patient - Identification, name, contact information, given consent.
- Information about health state - Diagnoses, bi-diagnoses, allergies, vaccinations, hypersensitivities, observations.
- Information about treatments - Lab results, measurements and analyses, dates and time, healthcare provider.
- Information about drugs and aid tools - Drug description, medical equipment, prescriptions, dosage.
- Information about journal post - Responsible journal editor (journalförare), dates and time, electronic or scanned type, type of journal post e.g. referral, note, prescription etc.
- Information about healthcare provider - Identification and descriptive information.
- Information about organization - Identification code, organization, organizational unit/reception/clinic.
- Information about healthcare event - Institutional or noninstitutional healthcare, dates and time, healthcare provider.
- Narrative - The healthcare visit event in clear text.
- Preservation information - Archivist, Archivist notes, related files and its metadata, regulations and confidentiality, the journal operation time, author, operation time, etc.
- Audit information - Requests, changes made, approved requests, dates and time etc.
- Structural metadata - How the object is compiled
- Technical information - Files, file formats, encoding and source system information such as name, version, status etc.

Questions to be answered:

1. Are there any metadata category missing in the list? If so, briefly describe this category with examples of metadata elements.
2. Which of the categories should be mandatory and which should be optional? Briefly motivate why (do not forget to include categories from Question 1 if existent).

Swedish version:
I det föregående frågeformuläret tillfrågades de nödvändigaste metadata-kategorierna i en specifikation för överföring av patientjournaler mellan verksamhetssystem och/eller till e-arkiv av Er. Efter sammanställning av svaren kunde följande kategorier av metadata identifieras:

- **Information om patient** - Identifikation, namn, kontaktuppgifter och samtycken.
- **Information om hälsotillstånd** - Diagnoser, bidiagnoser, allergier, vacciner, överkänsligheter, observationer.
- **Information om behandlingar** - Labresultat, mätresultat och analyser, datum och tider samt vårdgivare.
- **Information om läkemedel och hjälpmedel** - Läkemedelsbeskrivningar, medicintekniska hjälpmedel, doseringar och ordinationer.
- **Information om journalpost** - Ansvarig journalförare, datum och tider, elektronisk eller skannad, typ av journalpost som remiss, anteckning och ordination.
- **Information om vårdgivare** - Identifikation och deskriptiva uppgifter.
- **Information om organisation** - Identifikationskod, organisation samt organisatoriska enheten/mottagning/klinik.
- **Information om vårdtillfällen** - Öppenvård eller slutenvård, datum och tider samt deltagande vårdgivare.
- **Narrativ** - Vårdtillfället i klartext.
- **Information för bevarande** - Arkivbildare, anteckningar från arkivbildare, relaterade filer och deras metadata, sekretessuppgifter, journalens verksamhetstid, upphovsman samt verksamhetstiden.
- **Information gällande granskning** - Begäran av tillgång, ändringar, godkända begäran, datum och tider etc.
- **Strukturell metadata** - Hur journalens alla delar är sammanställda.
- **Teknisk information** - Filer, filformat, kodning samt information om källsystemet som namn, version, status etc.

Frågor att besvaras:
1. Saknas det någon kategori av metadata i listan? Om ja, beskriv denna kategori kortfattat med exempel på tillhörande metadata-element.
2. Vilka av kategorierna bör vara obligatoriska eller valfria inom specifikationen? Motivera kortfattat varför (glöm ej att även inkludera kategorier från fråga 1 om så förekommer).

8.3 APPENDIX 3 - INTERVIEW GUIDE

English version:

RQ: What metadata categories are necessary to include in a Common Specification for Swedish electronic healthcare records?

Preface
A request for approval is made for recording the interview. The interview target is informed that their answers will only be processed for the purpose of the study. The interview target is asked if they wish for anonymity. Prior to the interview start, the agenda is briefly presented.
Background information

1. What is your line of duty?

2. How many years of experience do you have in the line of the profession?

3. What experience do you have with metadata specifications?

4. What experience do you have with Common Specifications?

Evaluation criterions

5. What information is significant when gathering data for the purpose of aiding in a development of a Common Specification (such as the CS for EHRs)?

6. If categories of metadata were produced for the development of a Common Specification for EHRs, what should they have been based on?

Presentation of Results

The data is gathered from individuals with several years of experience in archival of EHRs in Swedish contexts. It's based on specifications used for archiving EHRs in 11 out of the total 21 regions of Sweden, including Stockholm, Skåne, Dalarna, Sömland, Gävleborg, Västmanland, Västernorrland, Jönköping, Uppsala, Värmland, and Örebro. The data is also based on the two largest international standards (HL7 CDA2 and CEN/ISO EN13606) for exchange of EHRs. The results brought out are 13 categories of metadata as well as a list of implications for practice to consider when developing a metadata specification for exchange of EHRs.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Example information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information about the patient</td>
<td>Identification, name, contact information, given consent.</td>
</tr>
<tr>
<td>Information about health state</td>
<td>Diagnoses, bi-diagnoses, allergies, vaccinations, hypersensitivities, observations.</td>
</tr>
<tr>
<td>Information about treatments</td>
<td>Lab results, measurements and analyses, dates and time, healthcare provider.</td>
</tr>
<tr>
<td>Information about drugs and aid tools</td>
<td>Drug description, medical equipment, prescriptions, dosage.</td>
</tr>
<tr>
<td>Information about journal post</td>
<td>Responsible journal editor (journalförare), dates and time, electronic or scanned type, type of journal post e.g. referral, note, prescription etc.</td>
</tr>
<tr>
<td>Information about healthcare provider</td>
<td>Identification and descriptive information.</td>
</tr>
<tr>
<td>Information about organization</td>
<td>Identification code, organization, organizational unit/reception/clinic.</td>
</tr>
<tr>
<td>Information about healthcare event</td>
<td>Institutional or noninstitutional healthcare, dates and time, healthcare provider.</td>
</tr>
<tr>
<td>Narrative</td>
<td>The healthcare visit event in clear text.</td>
</tr>
<tr>
<td>Preservation information</td>
<td>Archivist, Archivist notes, related files and its metadata, regulations and confidentiality, the journal operation time, author, operation time, etc.</td>
</tr>
<tr>
<td>Audit information</td>
<td>Requests, changes made, approved requests, dates and time etc.</td>
</tr>
<tr>
<td>Structural metadata</td>
<td>How the object is compiled</td>
</tr>
<tr>
<td>Technical information</td>
<td>Files, file formats, encoding and source system information such as name, version, status etc.</td>
</tr>
</tbody>
</table>

Categories which should be mandatory:
- Information about patient
- Information about Journal post
- Information about healthcare provider
- Information about organization
- Preservation information
- Technical information

A strong candidate for being a mandatory category is:
- Information about healthcare event

Category “Narrative” should be optional.

The rest of the categories are up for consideration whether they should be optional or mandatory:
- Information about health state
- Information about treatments
- Information about drugs and aid tools
- Audit information
- Structural metadata

Implications for practice:
- Opportunities for adding clinical codes should exist.
- Implications related to mandatory categories:
  - Clinical codes should not be mandatory since older medical systems may not provide it.
  - Having mandatory fields in the specification may force fictive information in some user cases.
○ You cannot have too high demands on information content since the source may be older medical system that was not structured the same way as today’s EHR systems.
○ There are specialized systems producing medical documents that does not provide all the information that is provided structure for, which may affect the decisions of mandatory categories.
○ When designing a specification, information that makes it easier to identify the information that should be mandatory was suggested to be included. For example, information type “drugs” and additional level describing the drug “aspirin”.

- Some information may be present in more than one metadata category. For example, the organizational unit, the provider, date and times, the editor, etc. should be included in all types of journal posts in an EHR.
- It may be beneficial to separate metadata related to the source system from metadata related to the archival of the EHR.
- It may be beneficial to separate metadata for treatments done by the provider at the organization where the EHR is established from treatments done via referral. Consequently, an additional metadata category for treatments done via referral was suggested, alternatively modifications can be made to the metadata category “Information about treatments”.
- User logs must be saved for at least five years, and include who as using the data, what the user did, when the event occurred, and in some cases the reason for usage.

Evaluation

7. Does the presented results fulfill the criterions previously stated (Question 5)?

8. How can the presented results be used in the development of a Common Specification for EHRs?

Swedish version:

RQ: Vilka metadatakategorier är nödvändiga att inkludera i en Förvaltningsgemensam Specifikation för patientjournaler?

Förord


Bakgrundsinformation:

1. Vad arbetar Du med?

2. Hur många år av erfarenhet har Du inom yrket?

3. Vilka erfarenheter har Du av metadata-specifikationer?

4. Vilka erfarenheter har Du av Förvaltningsgemensamma Specifikationer?
Utvärderingskriterier

5. Vad för information är viktig att ta reda på under en insamling av information ämnad att stödja ett utvecklande av en FGS (som FGS för patientjournaler)?

6. Om kategorier av metadata skulle tagits fram för utvecklandet av en FGS för patientjournaler, vad borde dessa vara baserade på?

Presentation av resultat


<table>
<thead>
<tr>
<th>Kategori</th>
<th>Exempel på information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information om patient</td>
<td>Identifikation, namn, kontaktuppgifter, samtycken.</td>
</tr>
<tr>
<td>Information om journalpost</td>
<td>Ansvarig journalförare, datum och tider, elektronisk eller skannad, typ av journalpost som remiss, anteckning, ordination etc.</td>
</tr>
<tr>
<td>Information om organisation</td>
<td>Identifikationskod, Organisation, organisatoriska enheten/mottagning/klinik.</td>
</tr>
<tr>
<td>Information om vårdtillfällen</td>
<td>Öppenvård eller slutenvård, datum och tider, vårdgivare.</td>
</tr>
<tr>
<td>Information om hälsotillstånd</td>
<td>Diagnoser, bi-diagnoser, observationer, allergier, vacciner, överkänsligheter.</td>
</tr>
<tr>
<td>Information om läkemedel och hjälpmedel</td>
<td>Läkemedelsbeskrivningar, medicintekniska hjälpmedel, doseringar, ordinationer.</td>
</tr>
<tr>
<td>Information om behandlingar</td>
<td>Labresultat, mätresultat och analyser, datum och tider, vårdgivare.</td>
</tr>
<tr>
<td>Information om vårdgivare</td>
<td>Identifikation och deskriptiva uppgifter.</td>
</tr>
<tr>
<td>Information för bevarande</td>
<td>Arkivbildare, anteckningar från arkivbildare, relaterade filer och deras metadata, sekreteressuppgifter, journalens verksamhetsid, upphovsman samt verksamhetsstiden etc.</td>
</tr>
<tr>
<td>Teknisk information</td>
<td>Filer, filformat, kodning samt källsystemsinformation som namn, version, status etc.</td>
</tr>
</tbody>
</table>
De kategorierna som borde vara obligatoriska:

- Information om patient
- Information om journalpost
- Information om vårdgivare
- Information om organisation
- Information för bevarande
- Teknisk information

En stark kandidat för att vara obligatorisk är:

- Information om vårdtillfällen

Kategorien “Narrativ” borde vara valfri.

Ett övervägande kan göras om de resterande kategorierna borde vara valfria eller obligatoriska:

- Information om hälsotillstånd
- Information om behandlingar
- Information om läkemedel och hjälpmedel
- Information gällande granskning
- Strukturell metadata

Implikationer för tillämpning:

- Det skall finnas möjlighet att använda kliniska kodverk, men det skall inte vara obligatoriskt då det inte finns i gamla journalsystem.
- Metadatakategorierna bör samverka då patienten, organisatoriska enheten, vårdgivaren, tid och datum, journalförare, osv. bör finnas för alla typer av journalposter.
- Det kan vara fördelaktigt att separera metadata relaterad till källsystemet från metadata relaterat till arkivering.
- Det kan vara fördelaktigt att separera metadata för behandlingar utförda av vårdgivare på vårdenheten där journalen är upprättad från behandlingar utförda via remiss. Därför kan en till metadatakategori för remissbehandlingar adderas, alternativt modifiera metadatakategorien “Information om behandlingar”.

Utvärdering

7. Uppfyller det presenterade resultatet utvärderingskriterierna tidigare nämnda (fråga 5)?

8. Hur kan det presenterade resultatet användas i utvecklandet av en FGS för patientjournaler?