Guideline CITS eHealth1

Guideline for the E-ARK Content Information Type Specification for Patient Medical Records

DILCIS BOARD

Date: 31.08.2021

Version: 1.0.0

1. Preface

1.1 Aim of the specification

This document is one of several related specifications which aim to provide a common set of usage descriptions of international standards for packaging digital information for archiving purposes. These specifications are based on common, international standards for transmitting, describing and preserving digital data. They also utilise the Reference Model for an Open Archival Information System (OAIS), which has Information Packages as its foundation. Familiarity with the core functional entities of OAIS is a prerequisite for understanding the specifications.

The specifications are designed to help data creators, software developers, and digital archives to tackle the challenge of short-, medium- and long-term data management and reuse in a sustainable, authentic, cost-efficient, manageable and interoperable way. A visualisation of the current specification network can be seen here:

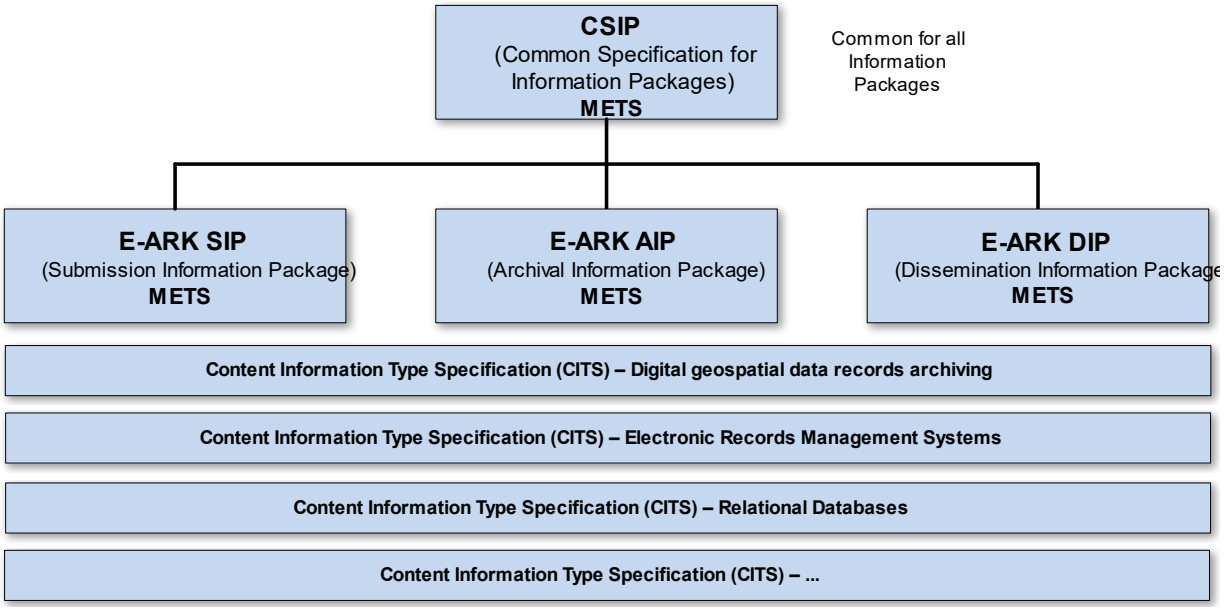

Figure I: Diagram showing E-ARK specification dependency hierarchy. Note that the image only shows a selection of the published CITS and isn't an exhaustive list.

Specification	Aim and Goals
Common Specification for Information Packages	This document introduces the concept of a Common Specification for Information Packages (CSIP). Its three main purposes are to: • Establish a common understanding of the requirements, which need to be met in order to achieve interoperability of Information Packages. • Establish a common base for the development of more specific Information Package definitions and tools within the digital preservation community.

Specification	Aim and Goals
	• Propose the details of an XML-based implementation of the requirements using, to the largest possible extent, standards which are widely used in international digital preservation. Ultimately, the goal of the Common Specification is to reach a level of interoperability between all Information Packages so that tools implementing the Common Specification can be adopted by institutions without the need for further modifications or adaptations.
E-ARK SIP	The main aims of this specification are to: • Define a general structure for a Submission Information Package format suitable for a wide variety of archival scenarios, e.g. document and image collections, databases or geographical data. • Enhance interoperability between Producers and Archives. • Recommend best practices regarding metadata, content and structure of Submission Information Packages.
E-ARK AIP	The main aims of this specification are to: • Define a generic structure of the AIP format suitable for a wide variety of data types, such as document and image collections, archival records, databases or geographical data. • Recommend a set of metadata related to the structural and the preservation aspects of the AIP as implemented by the eArchiving Reference Implementation (earkweb). • Ensure the format is suitable to store large quantities of data.
E-ARK DIP	The main aims of this specification are to: • Define a generic structure of the DIP format suitable for a wide variety of archival records, such as document and image collections, databases or geographical data. • Recommend a set of metadata related to the structural and access aspects of the DIP.
Content Information Type Specifications	The main aim and goal of a Content Information Type Specification is to: • Define, in technical terms, how data and metadata must be formatted and placed within a CSIP Information Package in order to achieve interoperability in exchanging specific Content Information. The number of possible Content Information Type Specifications is unlimited. For a list of existing Content Information Type Specifications see the DILCIS Board webpage (DILCIS Board, http://dilcis.eu/).

1.2 Organisational support

This specification is maintained by the Digital Information LifeCycle Interoperability Standards Board (DILCIS Board, http://dilcis.eu/). The role of the DILCIS Board is to enhance and maintain the draft specifications developed in the European Archival Records and Knowledge Preservation Project (E-ARK project, http://eark-project.com/), which concluded in January 2017. The Board consists of eight members, but no restriction is placed on the number of participants taking part in the work. All Board documents and specifications are stored in

GitHub (https://github.com/DILCISBoard/), while published versions are made available on the Board webpage. The DILCIS Board have been responsible for providing the core specifications to the Connecting Europe Facility eArchiving Building Block https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving/.

1.3 Authors & Revision History

A full list of contributors to this specification, as well as the revision history, can be found in the Postface material..

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 Context

1.1 Purpose

The purpose of this guideline is to further explain and describe the eHealth1 Content Information Type Specification (CITS).

1.2 Scope

The eHealth1 CITS builds on work done by the Directorate of Health and National Health Archive in Norway (NHA)[1] to create an archive of patient medical records and the eArchiving specifications described above. Using the Norwegian case as a starting point, the specification limits its scope to the use cases, data sources, and data submission methods as defined within that project. These are not the only possible use cases, data sources or submission methods that exist, and the specification can and should be extended to include others over time. The qualification and prioritisation of these will be driven by identified actual use cases discovered through the eHealth1 CITS review and eArchiving outreach program and specification development enabled by ongoing funding availability for continual review and extension.

The current specification draft makes the following assumptions:

- A case for creating an eHealth archive includes the incorporation of a backlog of physical and digital patient records.
- An eHealth archive concerns the Complete Patient Medical Records for deceased patients within a given jurisdiction. Note that the jurisdiction does not imply that a Central Health Archive must be national or federal. Many health administrations are organised at a state or region level, and the specification is equally valid for this scenario. Note that there are significant potential benefits for using standard for archiving Patient Medical Records if complied with by all regional administrations within a federation. This can also apply to environments where there are private healthcare providers, and a controlling administration is creating a Central Health Archive.
- Source systems submitting records to the archive are healthcare providers' (i.e. distributed) Electronic Medical Record (EMR) systems.
- The use cases for an eHealth archive are described in section 2.3.

2 Informative sections

2.1 Elements of an eHealth Archive

This section describes information systems and data models in eHealth and could contribute as source records for an eHealth archive. Note that the eHealth1 specification was written with the Norwegian

[1] https://ehelse.no/standarder/epj-standard-del-5-arkivuttrekk

Health Archive as a base and only considers submission by distributed Electronic Medical Record (EMR) systems.

2.1.1 Physical and Electronic Patient Records and Electronic Medical Record Systems

A Patient Medical Record can be defined as: "a collection or compilation of recorded information about a patient in connection with healthcare, the patient record is the principal repository for information concerning a patient's health care."[2] Before the widespread implementation of Electronic Medical Record (EMR) systems, the recording of patient health records was paper and film-based (plus additional materials such as images, video, audio).

Electronic Medical Records (EMRs) are digital versions of paper or film records. A healthcare provider may have a single EMR system for all of its patient records. For larger organisations, there can be fragmentation because of specialisation or organisational sub-division, and a patient's total medical record at that organisation may be constituted from many subsidiary systems. Potentially, a considerable amount of these patient records exists at healthcare providers and within centralised organisations because of legal remits to store the records for extended periods.

A Complete Patient Medical Record may contain information sourced from several different organisations' systems (e.g. different hospitals, specialist healthcare providers, primary healthcare providers). Viewed from an academic perspective, the information in each of these organisations constitutes an archive (or several archives). In creating a Central Health Archive, a healthcare provider must make separate extractions from each system for the patients to be included in a delivery and aggregate them before submission to the central archive.

Creating a Central Health Archive can encompass the digitisation and preservation of physical records and the collection and preservation of electronic records from EMR systems. Generally, a patient's aggregated medical record is not complete until there are no new additions (i.e. when the individual has died), consequently a health archive will consist only of records for patients who are known or believed to be deceased).

2.1.2 Electronic Medical Record and Electronic Health Record Systems

The terms "electronic medical record" and "electronic health record" (or "EMR" and "EHR") can be used interchangeably. However, the difference between the two terms is quite significant, especially in archiving standards.

EMR is the older term, and early EMRs were 'medical' in nature; they were for use by clinicians primarily for diagnosis and treatment. Because of a lack of available standards when EMR systems were first developed, the information in EMRs does not travel easily out of a healthcare provider. In fact, the patient's record might have to be printed out and delivered by mail to specialists or other care team members. In that regard, EMRs are not much better than paper records.

Electronic health records (EHRs) focus on the patient's total health – going beyond standard clinical data collected in the provider's office and inclusive of a broader view on a patient's care. EHRs are designed to reach out beyond the health organisation that originally collects and compiles the information. They are built to share information with other health care providers, such as laboratories and specialists, so they contain information from all the clinicians involved in the patient's care. The

[2] Institute of Medicine (US) Committee on Improving the Patient Record; Dick RS, Steen EB, Detmer DE, editors. The Computer-Based Patient Record: Revised Edition: An Essential Technology for Health Care. Washington (DC): National Academies Press (US); 1997. 1, Introduction. Available from: https://www.ncbi.nlm.nih.gov/books/NBK233055/

National Alliance for Health Information Technology stated that EHR data "can be created, managed, and consulted by authorised clinicians and staff across more than one healthcare organisation."[3]

The information moves with the patient – to the specialist, the hospital, the nursing home, or even across a region or country. In comparing the differences between record types, HIMSS[4] Analytics stated that "the EHR represents the ability to easily share medical information among stakeholders and to have a patient's information follow him or her through the various modalities of care engaged by that individual." EHRs are designed to be accessed by all people involved in the patient's care – including the patients themselves. Indeed, that is an explicit expectation in the so-called "Stage 1" definition of "meaningful use" of EHRs.

The benefits of EHR systems to patient care mean that the trajectory for healthcare worldwide is towards national EHR systems. The complexity and lack of standards in existing systems mean that realisation is difficult and expensive. Adoption is hence not yet widespread. Implementations of EHR systems can also rely on summary patient data gathered using standardised clinical documents (such as HL7 CDAs). This means that extractions from EHR systems may sometimes only yield patient summary data, not the Complete Patient Medical Record.

2.1.3 Living patient medical record archives

The definition of a health archive as a repository for Complete Patient Medical Records and hence for the use case of deceased patients only may require revision. Use cases have been observed at regional health archives where records on living patients are aggregated to be accessed between healthcare providers. This is an additional use case to those described by the Norwegian Health Archive and overlap with one of the possible objectives of a centralised Electronic Health Record system (EHR). If the packaging of the submissions is performed according to the specification (i.e. to eArchiving standards), then it is fair to treat these as archives and extend the specification. Also, as the specification allows for multiple submissions per patient from different source healthcare providers and aggregation is not performed before ingest to the archive (as is the case in Norway), then incremental submissions through the patient's life should be acceptable.

An important additional consideration for the management and accessibility of live datasets is that of patient confidentiality. The CSIP makes provision for the inclusion of rights data in packages, but further work needs to be done on standards for patient confidentiality in health records and inclusion of references to specific standards in the eHealth1 CITS.

2.1.4 Health registries

Brooke and the World Health Organisation[5] define registries in health information systems as a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way to serve a predetermined purpose.

An example of this is with cancer registries which are used to collect information about cases and treatment paths of cancer diagnoses for research purposes, which are then aggregated at national and international levels. The eHealth2 CITS considers the special case of Cancer Registries with regard to submission to an eArchive.

[3] https://www.healthcareusability.com/article/terminology-hit-emr-ehr

[4] HIMSS Analytics: himssanalytics.org

[5] https://apps.who.int/iris/handle/10665/36936

2.1.5 Clinical Document Architecture

A clinical document is a printed or electronic record that provides evidence of medical care. The most common standard for electronic clinical documents is the HL7 Clinical Document Architecture. In *A basic view of CDA*[6] , CDAs are described as:

2. **Intuitively clinical documents**

3. **The historical form of human-readable healthcare records**

4. **A mix of discrete data and free-flowing narrative**

5. **Always (at least theoretically) attested (i.e. signed)**

CDAs are XML encoded documents that are both machine and human-readable and use standard vocabularies, metadata schemas or resources such as HL7 FHIR (See 4.1.1). A CDA can reference other digital objects (such as images or sound files, etc.) and provide an exchange model for patient medical information. CDAs are signed (attested), which means that provenance and authenticity are managed as part of the patient care process. A series of CDAs could achieve a complete Patient Medical Record.

HL7 CDA is not the only standard; for example, the eHealth Digital Service Infrastructure (eHDSI or eHealth DSI) is the initial deployment and operation of cross-border health data exchange services under the Connecting Europe Facility[7]. It defines a document framework or Clinical Document Architecture (CDA) for sharing medical data across borders (Patient Summary).

A CDA is an XML encoded version of a comprehensive or summary Patient Medical Record. Although machine-readable, it will include only structured data that has been specified or is available from the source (EMR and EHR) systems. CDAs are a standardised way of transmitting patient medical data between local or centralised systems

CDAs are used as submission standards for aggregation within centralised EHR systems. Two possible incremental use cases have been observed which could be considered by the eHealth1 specification:

1. Creation of an archive as a long-term repository for outputs from an EHR system. EHR systems are generally online systems, and hence the data storage requirement for these systems is expensive and consistently grows over time. A need will emerge for the archiving of EHR records into an eArchive for the purposes of long-term preservation and value for research purposes.

2. Use of CDA standards as inputs to a centralised Health Archive. This is not ideal from the perspective of content structure but potentially is a lot simpler from the perspective of obtaining submissions from an EMR system that already supports CDA production.

[6] https://www.hl7.org.uk/wp-content/uploads/HL7UK_Media/Documents/Technical/A-basic-view-of-CDA-v3.doc

[7] https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home

2.2 Use cases for a Central Health Archive

According to the health archive regulation, the mission of the Norwegian National Health Archive (NHA)[8] is to:

a) receive and preserve patient archives from public and private hospitals, and
b) to disseminate health information for researchers and the patients next of kin in compliance with regulations and confidentiality acts.

There is no limit to the age of the records to be presented to the NHA from hospitals and so can consist of physical and electronic patient records. Other standards could apply elsewhere, and the scope of a Central Health Archive will be determined through local legislation.

The Norwegian regulation envisions two possible use cases for the archive when built, which are to:

a) provide records to next of kin in compliance with open information regulation.

b) harvest the vast amount of historical healthcare-related data within the archive for medical research.

To achieve the use case a), it is necessary to ensure that the specification allows for access to all records pertaining to a single patient, regardless of the submitting institution.

Use case b), requires that the specification allows for ingestion of digitised records and the ingestion of extracts from EMR systems for all deceased patients and that sufficient metadata is provided to enable searches across the archive to create cohorts supporting medical research. Metadata regarding Patient Administrative Information and Patient Clinical Information may be encoded in EMR systems or may have to be entered at a digitisation stage. The scope of the metadata to be included in the archive is therefore very much a determination for the local and national organisations based on the existing records, resources available, standards, etc.

2.3 User needs, use cases and information dissemination structures

A look at the current information systems for disseminating patient medical information provides insights into the user groups and cases that the systems serve. The following table summarises these user groups, user needs and how they are/can be met with existing system and information structures:

User Group	Needs, record structure	Rationale	Possible platform
1.Patients	**Access to all of own health record (single patient record)** Longitudinal record	**Transparency** Governance of own healthcare	Centralised EHR
2.Practitioners	**Access to all of a patient's health**	Improving patient healthcare outcomes	Centralised EHR

[8] https://ehandbok.arkivverket.no/folder/92

	record (single patient record) **Access to health records across borders** Longitudinal Record		CDAs
3.Next of Kin	**Access to health records after death (single patient record)** Longitudinal Record	**Transparency** **Family history as a factor in own or family healthcare** Litigation	**Centralised EHR** Centralised Health Archive
4.Researchers	**Access to selectable cohorts of patient data or global data trends based on:** **Dates** **Institution names and IDs** **Diagnoses, treatments, therapies** **Complex cohorts based on disease, condition, location etc** **Rich selection metadata and access functions** Cross patient data extractions based on rich metadata at the patient, care provider and clinical metadata level	Support research for improved healthcare outcomes	**Centralised health registries (specific domain)** Centralised Health Archive (general)

Table 1: User needs and use cases

As can be seen, the user needs for groups 1 to 3, for which many centralised systems are designed, have a core of common characteristics that are being met through the deployment of centralised EHR systems and CDA architectures. If long-term benefits are to be delivered from patient medical data for

research, the information structures of these systems need to be enhanced with researcher needs in mind.

3 Metadata and mapping

3.1 Data aggregations in eHealth1

The names of aggregation levels within an archive and represented within an archival package (IP) will depend on the agreements between data producers (Creators) and archives. EAD3 has defined a set of values (class, collection, file, fonds, item, otherlevel, recordgrp, series, subfonds, subgrp, subseries) for that purpose, allowing other values to be used in addition if they are defined as “otherlevel”. However, even though the aggregation levels in this context could be described in this way, the EAD template for archival description is considered broadly unsuitable for a Central Patient Health Archive.

A Central Patient Health Archive has a single purpose and may be instituted as a stand-alone entity or as a sub-entity within a larger institution (e.g. National Archive or Health Authority). Therefore, the overall aggregation of a health archive is implicit (it is an aggregation of patient medical records), and further aggregation levels must be defined that suit the use cases for navigation within the archive and for the way the archive is populated.

Patient data will most likely be submitted by hospitals or other healthcare providers in periodic batches, consisting of multiple patient records. Patient Medical Records will be submitted to a Central Health Archive either when a patient is known to have died or after a set period when it is not feasible that a patient is still alive. Depending upon the availability of a national death register, the accessibility and responsiveness to such a register and the periodic batching of archival extracts at healthcare providers, it cannot be expected that individual patient submissions from multiple creators will be at all coordinated. Aggregation of a total patient record at the archive before submission into a preservation system is deemed in the eHealth1 specification to be impractical.

The proposed structure for the aggregations of submissions of Patient Medical Records is as shown in the data model in Figure 4. As patient data is likely to be submitted in batches, each submission package will contain information from multiple patients. It is likely that the archive will split these submissions on receipt to create patient-specific archival information packages (AIPs) to simplify the dissemination process. In this context, the submission package can be considered a submission information collection (SIC) or collation of SIPs compiled to simplify extraction and transmission. However, in this specification, the term SIP means both a submission package for a single patient record or a submission package containing multiple patient records.

The levels of the aggregation in an eHealth1 package are as follows:

- **Patient**: An individual who has received healthcare at any number of healthcare providers who is described by Patient Personal Information. Each patient will be identified through a unique identifier (ID) provided from the source EMR system. This unique ID connects the Patient Personal Information and the Patient Medical Record in the information package.
- **Case**: A Patient Medical Record can be structured in various ways, which may be dictated by national standards or guidance or local practice. A Patient’s total medical record will consist of multiple individual thematic Cases which may be concerned with particular medical conditions, periods or treatments. The proposed aggregation allows for flexibility in this

grouping. These Cases will be held in one healthcare provider's local archive and may contain several Sub-cases and/or Documents with associated Data Files.

- **Sub-case**: A Sub-case is an allowable component consisting of a set of Documents and Data Files nested below a Case. Sub-cases may originate in departments within a large hospital or may be related to a different diagnosis to other Sub-cases. A Sub-case may have common (to the Case) or specific metadata.
- **Document**: A Document is a component that may consist of multiple related Data Files with common metadata; for example, a Document may be a PDF file together with associated attachments, or there may be a Document and a separate signature sheet. A document can be considered to be an entity that is approved/signed as a whole.
- **Data File**: A Data file is a component that contains data and has an associated MIME file type. A Data File can be a single bit stream or can encapsulate bit streams and attributes according to a standard such as a DICOM or MP4, in which case it will have a recognised MIME file type. A Data File, which is a container for multiple byte streams and metadata, can be included in the package as a Data File or can be unpacked and included as separate Byte Streams and described by metadata within METS. It is expected that containers such as DICOM and MP4 files will be submitted unaltered in Submission Information Packages (SIPs). Any decision to unpack them is part of a preservation plan at the archive.
- **Byte Stream**: A Byte Stream is a component that contains data, has an associated MIME file type and is encapsulated in a container such as MP4, DICOM or Matroska. Each Byte Stream has its own associated metadata, such as technical metadata, but is generally only accessible with specialised tools (such as ffprobe for video container formats).

3.2 Examples of different patient record submissions

With the flexibility of the structure of the eHealth1 archival package and the differences that are likely to be found in making Patient Medical Record extractions from disparate EMR systems, there can be expected to be different cases for the extraction of records.

3.2.1 Example 1: The entire archived Patient Medical Record as one file (document)

In this example, the extraction of a Patient's Medical Record consists of one unstructured file in, for example, PDF format, which contains a complete extract from an EMR system. In such a case, an Archived Patient Medical Record will consist of one Case containing one Document and one Data File (see Figure 1).

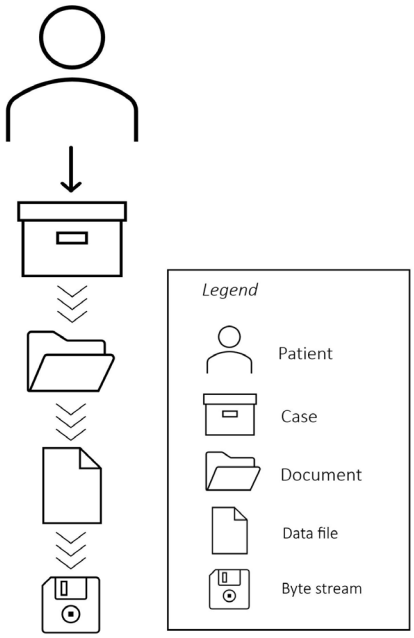

Figure 1: Archived Patient Medical Record as one file.

3.2.2 Example 2: The archive Patient Medical Record as a set of thematic files (Documents)

In this example, extraction of the Patient's Medical Record consists of a set of unstructured files, typically PDF documents where each file includes all of the information within a subject/theme that reflects the organisation of information in the current system. In this example, an archived Patient Medical Record would consist of several Cases, each containing one document, each containing one data file (see Figure 2).

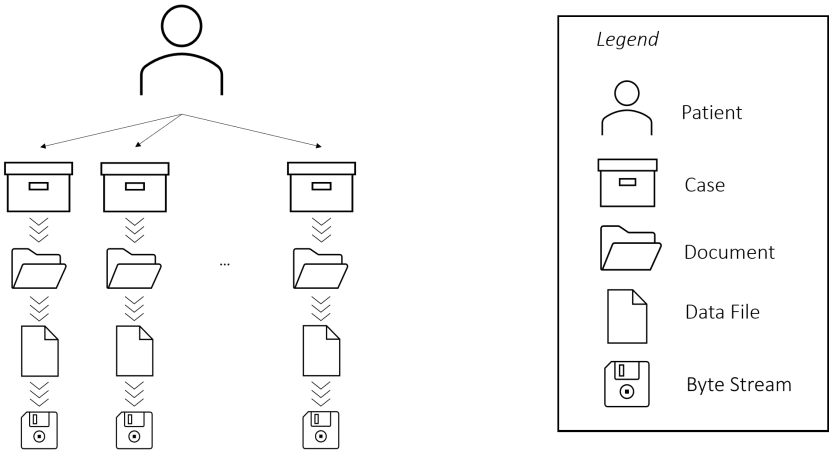

Figure 2: Archived Patient Medical Record as a set of thematic files.

3.2.3 Example 3: The archived Patient Medical Record as a set of documents per case

In this example, extraction of the Patient's Medical Record consists of a set of unstructured files which can be documents, images, videos, DICOM files, etc., and where each Data File may be related to other Data Files within a Document which can be associated with each other within a Case or a Sub-case (see Figure 3).

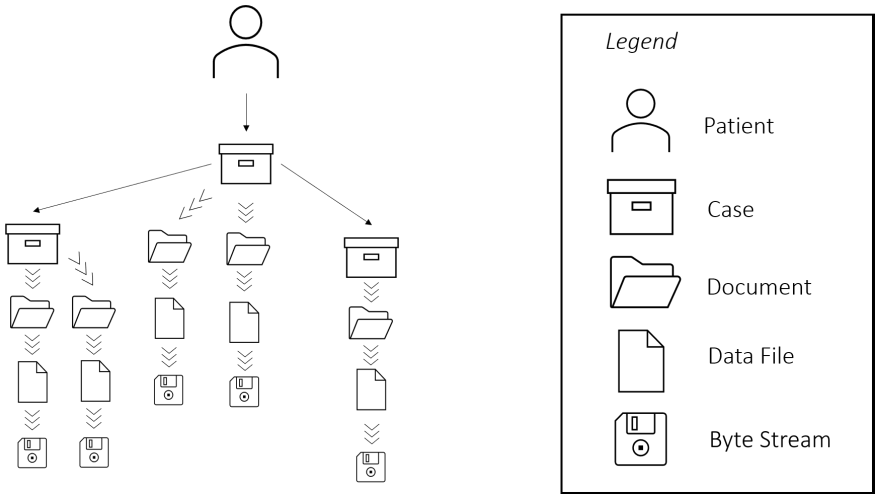

Figure 3: Archived Patient Medical Record as a set of documents per case or sub-case.

3.3 Data model

As described, the eHealth1 CITS is patient-centric (i.e. the submission packages contain groups of patient cases from a single care provider, grouped by patient). Archival Packages (AIPs) can contain single patient or multiple patient data at the discretion of the archive. For simplicity and security in providing individual, complete health records to next of kin, the Norwegian Health Archive chose for each AIP to only contain data from one patient. The data model below shows this patient centricity and the relationship between the different elements of the package.

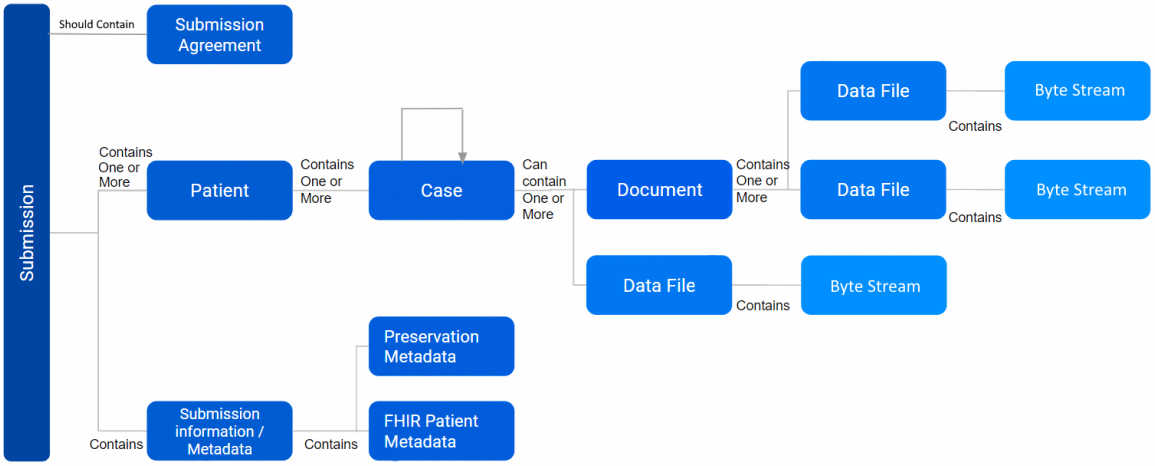

Figure 4: Data model.

The eHealth1 allows for multiple patient submissions within each package (SIP, DIP or AIP) for bulk transfer from the submitting organisation or for aggregation within the archive.

4 Standards used

The specifications for the information packages are built upon several standards described in this section and in the Guideline for Information Packages.

4.1 eHealth standards and use in the eHealth1 specification

Controlled vocabularies and coding provide a standardised way for the unambiguous recording of health data. Most EMR and all EHR systems will hold coded data concerning Patient Cases that can be extracted as metadata for the Patient Medical Record and will use an international standard such as ICD or SNOMED. Data can be recorded in standardised formats (such as HL7 FHIR) or to a local format which is specified by the health aArchive and referenced within a Submission Agreement.

4.1.1 HL7 FHIR[9]

Fast Healthcare Interoperability Resources (FHIR, pronounced “fire”)[10] is a standard describing data formats and elements (known as ‘resources’) and an application programming interface (API) for exchanging electronic health records (EHR). The standard was created by the Health Level Seven International (HL7) healthcare standards organisation.

Its goals are to facilitate interoperation between legacy health care systems, to make it easy to provide health care information to health care providers and individuals on a wide variety of devices

9

10 https://www.hl7.org/fhir/summary.html

from computers to tablets to mobile phones and to allow third-party application developers to provide medical applications which can be easily integrated into existing systems.[11]

FHIR provides resources that can be used for the standardised description of Patient Personal Information and Patient Clinical Information, which reference controlled vocabulary and coding standards such as ICD and SNOMED. The use of FHIR is suggested within eHealth1, but local standards for encoding metadata are allowable if specified elsewhere and referenced within a Submission Agreement.

4.1.2 HL7 Clinical Document Architecture

HL7 CDA provides a standard for the organisation of material within clinical documents for exchange between systems. By using XML, the HL7 v3 standard and coded vocabularies, the CDA facilitates the exchange of both machine and human-readable documents, enabling electronic processing for decision support, etc., whilst being easily retrieved and used by the people who need them (HL7 UK, 2018).

4.1.3 ICD[12]

The International Classification of Diseases is the foundation for identifying health trends and statistics globally and the international standard for reporting diseases and health conditions. It is the diagnostic classification standard for all clinical and research purposes. ICD defines the universe of diseases, disorders, injuries and other related health conditions, listed in a comprehensive, hierarchical fashion that allows for:

– easy storage, retrieval and analysis of health information for evidence-based decision-making;

– sharing and comparing health information between hospitals, regions, settings and countries; and

– data comparisons in the same location across different periods.

ICD is mapped from other standards such as HL7 FHIR and will be part of the process used by many institutions to record Patient Clinical Information. The use of international standards such as ICD within supplied clinical metadata is encouraged but will be limited by their use within the source EMR or EHR system.

4.1.4 SNOMED[13]

SNOMED CT or SNOMED Clinical Terms is a systematically organised computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting. SNOMED CT is considered the most comprehensive, multilingual clinical healthcare terminology in the world. The primary purpose of SNOMED CT is to encode the meanings that are used in health information and to support the effective clinical recording of data to improve patient care. SNOMED CT provides the general core terminology for electronic health records.[14]

SNOMED CT is mapped from other standards such as HL7 FHIR and will be part of the process used by many institutions to record Patient Clinical Information. The use of international standards such as

[11] https://en.wikipedia.org/wiki/Fast_Healthcare_Interoperability_Resources
[12] https://www.who.int/classifications/icd/en/
[13] http://www.snomed.org
[14] https://en.wikipedia.org/wiki/SNOMED_CT#:~:text=SNOMED%20CT%20or%20SNOMED%20Clinical,in%20clinical%20documentation%20and%20reporting.

SNOMED CT within supplied clinical metadata is encouraged but will be limited by their use within the source EMR or EHR system.

4.1.5 DICOM[15]

Digital Imaging and Communications in Medicine (DICOM) is the standard for communicating and managing medical imaging information and related data. A DICOM file is a file that encapsulates attributes and bit streams (image, video, etc.) and has embedded patient personal information and IDs. DICOM files have a recognised MIME file type. Extraction of DICOM files from specialised EMR systems for inclusion in Patient Medical Records should present no problem, but it is essential to ensure that patient IDs in DICOM files match those in archival package Patient Personal Information.

4.1.6 eHealth DSI (eHealth Digital Service Infrastructure)[16]

The eHealth Digital Service Infrastructure (eHDSI or eHealth DSI) is the initial deployment and operation of cross-border health data exchange services under the Connecting Europe Facility (CEF). It defines a document framework or Clinical Document Architecture (CDA) for sharing medical data across borders (Patient Summary). E-ARK eHealth1 considers the totality of a Patient Medical Record. The eHDSI is too limited in scope to be useful in this context, eHDSI aims to specify an interchangeable derivation and extract of a Patient Medical Record. In contrast, the E-ARK eHealth1 CITs aims to preserve the Patient Medical Record in its entirety.

5 Glossary

Name	Description
Archival Creator	Organisation unit or individual that creates records and/or manages records during their active use.
Case or Patient Case	Type of component consisting of a set of Documents and/or Sub-cases. This is represented in the specification as a folder that sits within the data directory of a representation (which in this case is a Patient's Medical Record). A Case is an aggregation of individual patient records related to one patient and which are related in a way that is defined by national standards, guidance or local practice. A Patient's Medical Record will consist of multiple individual thematic Cases which may be concerned with particular medical conditions, periods or treatments.
Central Health Archive	An organisation within a national or regional jurisdiction with a (usually legal) remit to create an archive of Patient Medical Records for people who have received primary or secondary healthcare in the jurisdiction. The Central Health Archive will be populated with Patient Medical Records from multiple

[15] https://www.dicomstandard.org/current/

[16] https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home

	healthcare providers in the jurisdiction, drawn from local patient health archives (e.g. a hospital archive).
Component	In this standard: meaningful, logically delimited, and uniquely identifiable information that may be subject to treatment in manual and/or automated processes. This standard operates with four generic components types: Case, Document, Data File and Byte Stream.
Data File	A component which contains data and has an associated MIME file type. A Data File can encapsulate multiple bit streams and metadata according to a standard such as a DICOM but must have a recognised MIME file type. A Data File may comprise one or more subsidiary Byte Streams; for example, an MP4 file might contain separate audio and video streams, each of which has its own associated metadata.
Death Register	National system which records deaths within the jurisdiction.
Document	A single or group of related Data Files with common metadata. For example, a Document may consist of a PDF file together with associated attachments or a word file with a separate image signature sheet. A document can be considered to be an entity that is approved/signed as a whole by a practitioner.
General EMR System	Electronic Medical Record system intended for documentation of all forms of healthcare. Note: large scale healthcare providers may have a primary general-purpose EMR system but can also have several distributed general-purpose EMR systems serving parts of the organisation that operate as separate sub-services.
Healthcare Provider	An organisation providing primary or secondary healthcare. Can be general in scope or specialised, public or private.
Local Patient Health Archive	An archive of physical or electronic Patient Medical Records within a Healthcare Provider or group of Healthcare Providers. A Patient Medical Record will usually be transferred to an archive either when the patient is known to have died or after a set number of years have passed since its creation exceeding normal life expectancy.
Patient Clinical Information	Structured patient clinical data related to Cases such as diagnoses, procedures, medication, allergies, etc.
Patient Medical Record	Collection or compilation of recorded information about a patient in connection with healthcare. Note: a Patient Medical Record may contain information in digital form and/or information recorded on other media such as paper or film. In this specification, Patient Medical Records are assumed to be digital, where the content may be born digital and/or digitised from physical records.

Patient Medical Record Extraction	Extract from a Local Health Archive to hand off to the Central Health Archive. All Patient Medical Record Extractions should be under a Submission Agreement.
Patient Personal Information	Demographics and other administrative information about an individual receiving care or other health-related services. For example, as can be described using the resource FHIR.Patient. Information will include but not be limited to name, patient ID(s), administrative gender, date of birth, date of death, address(es).
Specialised EMR System	Electronic Medical Record system specially adapted for documentation of a type of specialised healthcare or integrated with a specialised device. Examples: food/maternity system, gastrosystem, laboratory system, etc.
Sub-case	Type of component consisting of a set of thematically related Data Files also related to a Case. Sub-cases are represented in the specification as folders that sit within a Case.
Submission Agreement	The agreement reached between an archive and the submission producer that specifies a submission format (eHealth1 CITS), and any other arrangements needed, for the data submission Session. Any special conditions on patient confidentiality could be specified in the submission agreement.
Submitting Organisation	Name of the organisation submitting the package to the archive.
Complete Patient Medical Record	The sum of the submissions of patient Records made for an individual.

Table 2: Glossary

6 Explanations of specific eHealth1 requirements

6.1 Placement of data in an eHealth1 package

As submitted by hospitals or healthcare providers, patient data will likely be periodically extracted from source systems and sent in batches. The eHealth1 specification allows for the inclusion of multiple patients per package, so that these batches can be transmitted in a single submission. The number of patients included in each AIP is then a matter for local implementation; the decision in Norway at NHA was for each AIP to consist of data from a single Patient and a single Submitting Organisation.

Each Patient Medical Record is placed in a single representation within the representations folder of the package. The ID of each representation should follow instructions for naming of representation folders in CSIPTR10 to have a string name that is unique within the package scope. It is also suggested that this name should include the unique ID of each patient.

Each representation should contain a METS file at its root (Representation METS). The folder structure should follow that defined by the CSIP and must have a 'Data' folder. If clinical descriptive metadata is to be supplied, the representation structure must include a folder '/metadata/descriptive' and any descriptive metadata files must be placed in this folder.

6.2 Root METS file

6.2.1 Root METS root element

The content category <mets@TYPE> attribute is set to the value "OTHER", and the other content category <mets/@csip:OTHERTYPE> attribute is set to the value "Patient Medical Records."

The Content information type specification referenced by the <mets/@csip:CONTENTINFORMATIONTYPE> attribute is set to the value provided in the vocabulary for requriment CSIP4 found at, http://earkcsip.dilcis.eu/schema/CSIPVocabularyContentInformationType.xml . In the time of the publication of this document the value is "citsehpj_v1_0". ".

The METS profile <mets/@PROFILE> attribute is set to the value "https://citsehealth1.dilcis.eu/profile/E-ARK-eHealth1-ROOT.xml" which is a specific METS profile for eHealth1 referenced at the DILCIS website.

6.2.2 Root METS header element

Due to the presence of personal data in the package, eHealth1 requires that there must be a reference to a Submission Agreement and that the attribute <mets/Hdr/altRecordID@TYPE> must have the value "SUBMISSIONAGREEMENT".

In order to positively identify the organisation that originally created the Patient Records (rather than simply submitted them) eHealth1 requires an Archival Creator Agent reference via the element <mets/Hdr/agent/> with the attribute <mets/Hdr/agent@TYPE> set to "ORGANISATION" and <mets/Hdr/name> holding the organisation's name, a note <metsHdr/agent/note> containing an organisation ID as used within the locality and the attribute <metsHdr/agent/@csip:NOTETYPE> set to value "IDENTIFICATIONCODE".

6.2.3 Root METS descriptive metadata section (element dmdsec)

There must be a reference to the Patient Personal Information held in the metadata/descriptive folder of the package, referenced by locators within the <dmdSec/mdRef> element and with the attribute <dmdSec/mdRef/@MDTYPE> other set to "OTHER" and <dmdSec/mdref/@OTHERMDTYPE> set to the type of metadata used for Patient Personal Information. For example, "FHIR.Patient" if this resource is used.

6.2.4 Root METS file metadata section (element fileSec)

The transferred Patient Medical Records are placed in representation folder(s), each described with its own representation METS document. eHealth1 makes the inclusion of a file section (fileSec) in the METS file mandatory as Patient Medical Records are required to be included in representations, the METS documents for which must be referenced from the root METS file section.

All documentation pertaining to the package should be referenced from one or more file groups with the <mets/fileSec/fileGrp/@USE> attribute value ”Documentation” and any documentation pertaining to the transferred content is referenced within the representation METS. This requirement has been adjusted from CSIP to specify that although the documentation file group is mandatory, all documentation relating to the transferred content (Patient Medical Records) should be held in the representation Documentation folder and file group.

The Content information type specification referenced by the <mets/@csip:CONTENTINFORMATIONTYPE> attribute is set to the value provided in the vocabulary for requriment CSIP62 found at, http://earkcsip.dilcis.eu/schema/CSIPVocabularyContentInformationType.xml . In the time of the punlication of this document the value is “citsehpj_v1_0”

6.2.5 Root METS structural map (element structMap)

There must be a discrete `div` element for each Patient Medical Record. This requirement makes allowance for multiple Patient Medical Records within a package (e.g. in a SIP). The eHealth1 specification makes it mandatory for each of these Patient Medical Records to be represented by a separate ‘div’ element within the representations division.

6.3 Representation METS

6.3.1 Representation METS root element

The mets/@OBJID attribute is mandatory. Its value is a string identifier for the METS document. For a representation level METS document, this value records the name of the representation folder, which should include the unique ID of the patient. The Common and Package specifications do not specify separately particular requirements for the representation METS document. As representations and hence the representation METS are mandatory in the eHealth1 it references parts of the specification that require further or different detail for the representation METS.

6.3.2 Representation METS descriptive metadata section (element dmdSec)

Used to reference Case Patient Clinical Information held in the metadata/descriptive folder of the representation.

There is one dmdSec present for each descriptive metadata file located in the “/metadata” section of the representation. The descriptive metadata section is used to reference Patient Clinical Information held in the metadata folder of the representation. This requirement is mandatory in the eHealth1 CITS as the requirement for inclusion of Patient Clinical Information is also mandatory and is described in descriptive metadata files in the representation.

6.3.3 Representation METS file section (element fileSec)

Representation of the Patient Case structural hierarchy is only possible if the file section (fileSec) is present in the representation. The representation file groups contain the file elements which describe the digital objects. The hierarchical structure of the Patient Medical Records within eHealth1 requires that digital objects (groups of files that form a single intellectual entity) can be described through the structural map (structMap) element. Thus, file groups are defined for each Document element which will be within a Case or Case/Sub-case structure and the individual file or byte stream elements listed within those groups. For these filegroups the attribute <fileSec/fileGrp/@csip:CONTENTINFOREMATIONTYPE> is set with the value “citsehpj_v1_0”.

If administrative metadata has been provided at a representation level (i.e. there are rights and/or digital provenance metadata specific to the Patient Case information), then the fileSec/filGrp/@ADMID attribute refers to the administrative metadata section (amdSec) of the representation METS by ID. This is the administrative data for the file group defined above, (i.e. the Patient Medical Records).

6.3.4 Representation METS structural map (structMap element)

Each representation METS file must include ONE structural map (structMap) element to describe the Patient Case structure described above.

The attribute <structMap/@LABEL> of the structMap is set to value “eHealth1”

Within eHealth1, Patient Cases must be held within data folders within a single minimum representation and described in the structural map within a single sub-division (/div). There are no files within the data division itself, only from the structure defined below it.

Through a structure of sub-divisions, the Patient Case structures described above are defined in the structural map. (i.e. Case/Document/File or Case/Sub-case/Document/Data File structures). Data files are always contained within Documents.

All file groups containing content described in the package are referenced via file group identifiers to the file section element of the METS file. There is one file group reference per group of files (a Document).

7 Postface

AUTHOR(S)	
Name(s)	Organisation(s)
Stephen Mackey	Piql AS

REVIEWER(S)	
Name(s)	Organisation(s)
Karin Bredenberg	Kommunalförbundet Sydarkivera
Jaime Kaminski	Highbury R&D

Project co-funded by the European Commission within the ICT Policy Support Programme

Dissemination Level		
P	**Public**	x
C	**Confidential, only for members of the Consortium and the Commission Services**	

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
V1.0	03/02/2021	Stephen Mackey	Piql AS	Draft for review
V1.0	20/07/2021	Stephen Mackey	Piql AS	Revised Draft
V1.0	31/08/2021	Stephen Mackey	Piql AS	Publication of version 1.

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.