TECHNICAL SPECIFICATION

ISO/TS 29585

First edition
2010-05-15

Health informatics — Deployment of a clinical data warehouse

Informatique de santé — Déploiement d'un entrepôt des données cliniques

Reference number
ISO/TS 29585:2010(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years in order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn. If the ISO/PAS or ISO/TS is confirmed, it is reviewed again after a further three years, at which time it must either be transformed into an International Standard or be withdrawn.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 29585 was prepared by Technical Committee ISO/TC 215, *Health informatics*.

Introduction

This Technical Specification furthers the work of ISO/TR 22221 by providing implementation guidance for a clinical data warehouse and describing general considerations of development and deployment, issues and applications of data aggregation and data modelling, and architecture and technology approaches.

The role of the clinical data warehouse is to enable data analyses in support of effective policies and decision-making, to improve quality of care, to improve health services organizations, as well as to influence learning and research. It will have relevance to both developing and more established health systems. It will enable meaningful comparison of programmes and outcomes.

Although data warehouse technologies are becoming increasingly used in non-healthcare sectors, their use in health is still at an early stage. ISO/TR 22221 had a primary goal of underpinning a coherent approach to the diverse and multi-stakeholder perspectives of secondary use of data from various health system sources. This Technical Specification is intended to have pragmatic relevance by indicating best practice in setting up a clinical data warehouse and in using it from data abstraction and architectural perspectives. The clinical data warehouse is distinguished by the complexity of the interactions of data and hence the challenges to provide adequate methods for evaluating process and outcomes of care for different populations and sub-populations. Currently such knowledge is relatively fragmented and it is too early to be integrated into an International Standard. A Technical Specification will however benefit progression to an International Standard by aligning emerging best practice from different international experience.

The clinical data warehouse is also, in health informatics, the place of the intersection of health services delivery, organization and epidemiological expertise concerned with adequate and effective data abstraction and presentation for different decision-making contexts as presented in ISO/TR 22221. Good use of the clinical data warehouse will depend on furthering common approaches to frequently used data abstractions that concern analysis of care delivery and organization. Effective data warehouse deployment will be enabled by promoting good practice in furnishing dynamically-accessible, interpretable data combinations, which will depend on showing the relationship between clinical and health system need and the architectural properties of the data warehouse.

This technical specification complements the ISO 13606 series in that competent extended use of data beyond immediate care delivery depends on the effective organization of the original source data.

Health informatics — Deployment of a clinical data warehouse

IMPORTANT — The electronic file of this document contains colours which are considered to be useful for the correct understanding of the document. Users should therefore consider printing this document using a colour printer.

1 Scope

This Technical Specification has three sections, 1) general considerations of design and deployment, 2) data aggregation and data modelling and 3) architecture and technology, and is intended to provide an overall set of guidelines for clinical data warehouse deployment supported by useful descriptions concerning different data aggregation and modelling approaches as well as particular aspects of information architecture that contribute to successful deployment. The first section is of particular interest to healthcare decision-makers, including information technology managers, of requirements and procedures that support successful clinical data warehouse deployment. The second section supports the understanding, choice, instigation and evaluation of methods that ensure reliable selection and aggregation of primary data for adequate compilation and presentation to support decisions – this section is of particular interest to statisticians, epidemiologists, healthcare evaluation specialists and others. Section three is of particular interest to informaticians concerned with efficient architectures, data mining methods, dynamic data querying and visualization for clinical data warehouses.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/TR 22221, *Health informatics — Good principles and practices for a clinical data warehouse*

ISO/TS 25237, *Health informatics — Pseudonymization*

ISO 27799, *Health informatics — Information security management in health using ISO/IEC 27002*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
clinical data repository
CDR
operational data store that holds and manages clinical data collected from service encounters at point of service locations

NOTE Data from a CDR can be fed to the EHR for that client, such that the CDR is recognised as a source system for the EHR. The CDR can be used to trigger alerts in real time.

3.2
clinical data warehouse
CDW
grouping of data accessible by a single data management system, possibly of diverse sources, pertaining to a health system or sub-system and enabling secondary data analysis for questions relevant to understanding the functioning of that health system, and hence supporting proper maintenance and improvement of that health system

NOTE A CDW tends not to be used in real time. However, depending on the rapidity of transfer of data to the data warehouse, and data integrity, near real-time applications are not excluded.

3.3
dashboard
user interface based on predetermined reports, indicators and data fields, upon which the end user can apply filters and graphical display methods to answer predetermined business questions and which is suited to regular use with minimal training

3.4
data dictionary
database used for data that refer to the use and structure of other data, i.e. a database for the storage of metadata

3.5
data mart
subject area of interest within the **data warehouse** (3.6)

EXAMPLE An inpatient data mart.

NOTE Data marts can also exist as a standalone database tuned for query and analysis, independent of a data warehouse.

3.6
data warehouse
subject-oriented, integrated, time-variant and non-volatile collection of data

NOTE See Reference [5].

3.7
data warehouse dimension
subject-oriented, often hierarchical business relevant grouping of data

3.8
drill down
exploration of multidimensional data which makes it possible to move down from one level of detail to the next depending on the granularity of data

EXAMPLE Number of patients by departments and/or by services.

3.9
episode of care
identifiable grouping of healthcare-related activities characterized by the entity relationship between the subject of care and a healthcare provider, such grouping determined by the healthcare provider

[ISO/TS 18308:2004, definition 3.23]

3.10
health indicator
single summary measure, most often expressed in quantitative terms, that represents a key dimension of health status, the healthcare system, or related factors

NOTE A health indicator is informative and also sensitive to variations over time and across jurisdictions.

[ISO/TS 21667:2004, definition 3.1]

3.11
metadata
information stored in the data dictionary which describes the content of a document

NOTE In a data warehouse context, metadata are data structure, constraints, types, formats, authorizations, privileges, relationships, distinct values, value frequencies, keywords, interpretative notes and users of the database sources loaded in the data warehouse and the data warehouse itself. Metadata help users, developers and administrators manage and interpret information.

3.12
master data management
enablement of a program that provides for an organization's data definitions, source locations, ownership and maintenance rules

3.13
online analytical processing
OLAP
set of applications developed for facilitating the collection, analysis and reporting of multidimensional data

NOTE See Reference [7].

3.14
organization
group of people who have their own structure rules and culture in order to work together to achieve goals and/or to provide services through processes, equipment and technology, etc.

3.15
performance indicator
measure that supports evaluation of an aspect of performance and its change over time

3.16
persistent data
data in a final form intended as a permanent record, such that any subsequent modification is recorded together with the original data

3.17
roll up
method of regrouping and aggregating multidimensional data to move up the hierarchy into larger units

EXAMPLE Weekly count of patients aggregated by quarter or by year.

3.18
secondary data use
expression sometimes employed to describe the use of data for additional purposes other than the primary reason for their collection, adding value to these data

3.19
star schema
dimensional modelling concept that refers to a collection of fact and dimension tables

3.20
widget
standalone visualization component (e.g. a heat map, gauge or geographic map) that can be integrated with a data warehouse source and presented in an end-user dashboard

NOTE Custom widgets can be developed using a business intelligence vendor software development kit (SDK) and managed via the widget library.

4 Abbreviated terms

— DICOM[SM] Digital Imaging and Communications in Medicine

— EHR Electronic Health Record

— HL7 Health Level 7

— ICD© International Classification of Diseases

— LOINC© Logical Observation Identifiers, Names and Codes

— SNOMED CT© Systematized Nomenclature of Medicine — Clinical Terms

5 Principle

The roles and capacities of operational databases and informational databases (data warehouses) are complementary. An operational database is designed to perform transactions in real time such as adding, changing or deleting patient data, or displaying current data for immediate care decision making. It has a limited capacity for data analysis and is focused on online support for care delivery. The exploitation of already existing and persistent data for other purposes, sometimes referred to as secondary use of data, typically involves data aggregation and/or linkage from multiple data sources. The concept of a clinical data warehouse here is an application of the notion of data warehouse (that is the bringing together of data relevant to the functioning of an enterprise), for clinical purposes understood in the broadest sense, including the ensemble of healthcare system factors that can influence patient care. Emerging issues such as semantic interoperability with research databases in the fundamental sciences are not considered in this Technical Specification.

Deployment of a clinical data warehouse

ISO/TR 22221 provides an informative description of the uses and principles of implementation of a CDW, including an overview of the issues that are further developed and addressed in this Technical Specification. A CDW allows many perspectives of use and is therefore of interest to many categories of stakeholders. The activities of CDW use in ISO/TR 22221 are considered under the headings of:

— quality assurance and care delivery;

— evaluation and innovation of health procedures and technologies;

— disease surveillance, epidemiology, and public health;

— planning and policy;

— knowledge discovery;

— education.

These titles give insight into the increasing relevance of the CDW to several aspects of the health system and its mission of effective healthcare.

6 General considerations of deployment of a clinical data warehouse

6.1 Overview

This subclause guides the setting-up, deployment and ongoing management of a clinical data warehouse. It should be of use to an array of management and deployment stakeholders by articulating the range of considerations pertinent for successful planning, project management and ongoing governance, as well as describing key issues of data sources and quality, choice of architecture and maintenance of privacy.

There is a distinction between an operational electronic health record system, designed to support direct patient care, as opposed to a clinical data warehouse which will typically combine data from a number of sources and/or organizations for analytical purposes, with a diverse group of users accessing highly sensitive personal information, where use of this information is governed by multiple pieces of legislation and policy. Sometimes the term “secondary uses” is used for the latter application.

A clinical data warehouse is typically used for many purposes such as planning, management, research, audit and public health. The intention is that information is automatically collected or abstracted from operational electronic health record systems and then organized and maintained for subsequent reporting. The range of reporting applications can be very wide, however, and as illustrated in Figure 1 the different purposes have different characteristics.

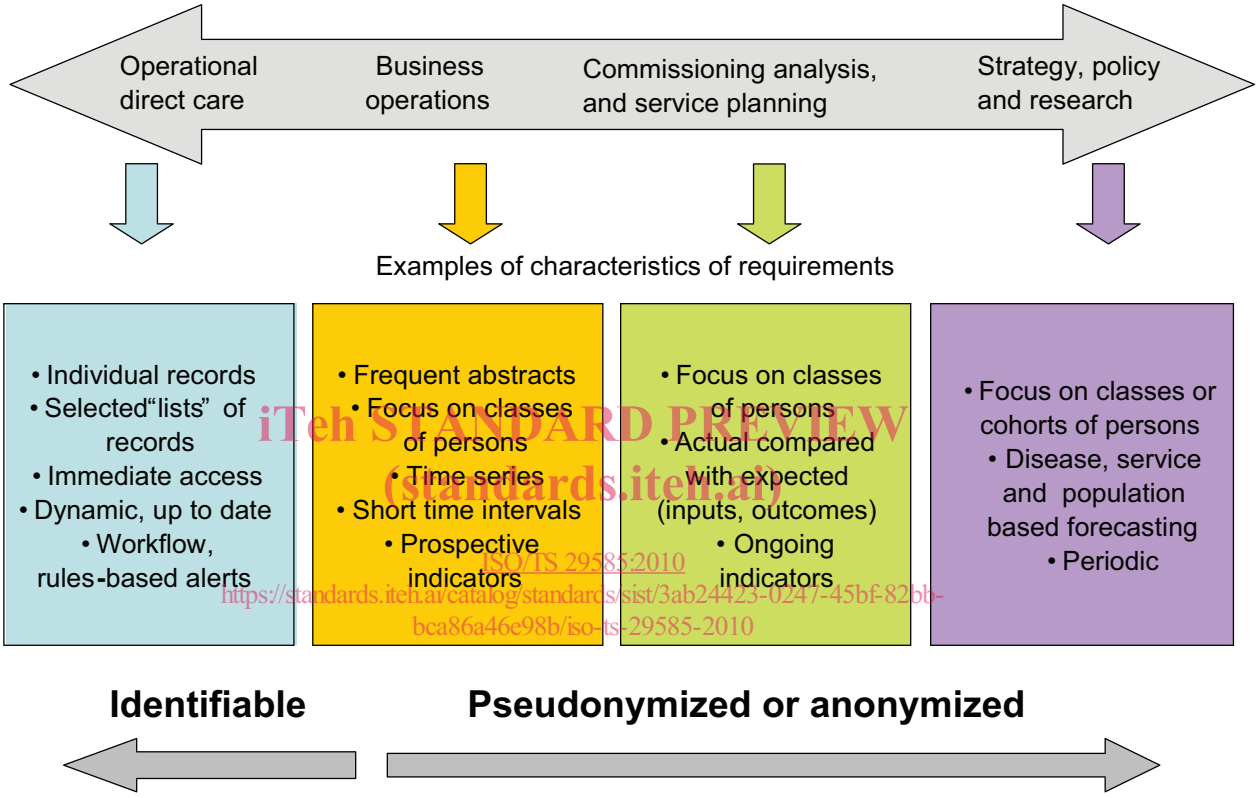

Figure 1 — Different types of information use

One of the issues, therefore, is to consider where to start. This subclause aims, therefore, to provide advice for those considering the development of a clinical data warehouse, to assist in:

— understanding the requirements;

— clarifying the scope;

— planning and implementation issues;

— design considerations;

— data and metadata;

— security and privacy.

6.2 Requirements

6.2.1 Overview

Consideration is given to three levels of clinical data warehouses: national (global); regional; local. National uses might be for statistical collection and comparison purposes including at a global level; regional might (depending on the country) be state, province or regional health organizations; local might mean individual organizations or hospitals.

Figure 2 illustrates how data may be collected at local level and the level of granularity within existing fields can be abstracted and summarised for use at regional and national levels. See Reference [37]. Subclauses 6.2.2 to 6.2.8 consider potential uses for a clinical data warehouse at national, regional and local levels. While it is often appropriate to have CDWs at each of these levels, each of which is attuned to the particular information analysis and reporting requirements of the sponsoring organization, a coordinated strategy for developing and populating the CDWs recognises there is a great deal of commonality in the underlying source data.

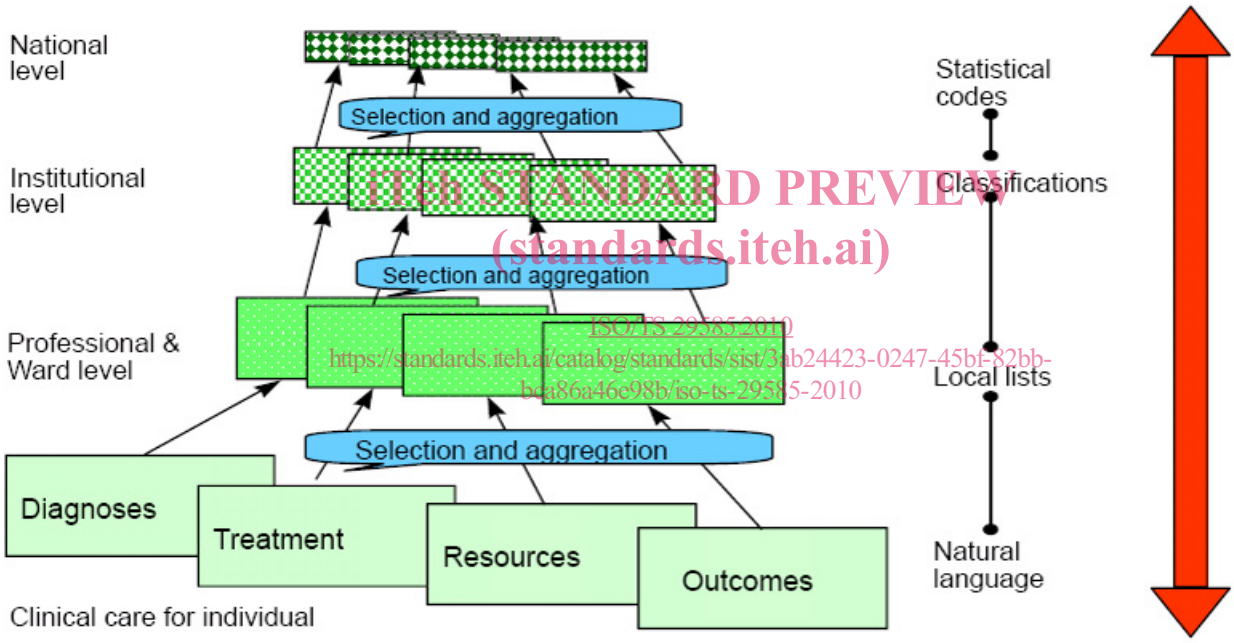

Figure 2 — Levels of clinical data aggregation

6.2.2 Users

There is a large range of potential users, which might include:

— national or regional government;

— government agencies, e.g. analysis and reporting centres;

— regulators;

— international organizations, e.g. World Health Organization (WHO);

— professional bodies, e.g. colleges;

— medical research and education;

— local care organizations, e.g. health providers;

— local government, e.g. environment, education, housing;

— other commercial users, e.g. pharmaceutical companies.

6.2.3 Requirements for local healthcare provider organizations

This subclause identifies typical requirements for healthcare providers, both as individual organizations and as part of a group or network of care service providers. Similar requirements might exist at a district or community level.

For individual providers, requirements might include:

— care service planning, monitoring and review analysis via

  — demand and capacity;

  — quality (access to services, etc.);

  — costs, efficiency and productivity;

  — outcomes and effectiveness, including clinical or care audit;

  — benchmarking and comparison with "peer" service providers;

  — strategic service planning;

— financial and contract management via

  — calculation of local costs (for reference cost comparison);

  — income estimation;

  — assignment of care provided to contracts;

  — monitoring of income against plans (budgets) and costs.

For provider networks, requirements might also include:

— care service planning, monitoring and review (as above but requires analysis of patient activity data which might have inputs from several networked providers);

— capability for spanning organizational and geographical boundaries where appropriate (e.g. cancer networks).

These functions require the capability to receive, manage, link and analyse data extracts from existing systems. Typically, this will require standards to be defined for these local extracts. The following functional components may be required:

— functionality to produce mandatory datasets;

— functionality to enable users to select and extract data from their operational systems (all elements of a patient's record);

— functionality to manage/store these extracts and combine them (via linkage) with data extracted from other systems;

— functionality to enable analysis and reporting of these data and to provide users with access to other specialist analysis tools;

— production of standard reports (both scheduled and ad hoc).

6.2.4 Regional requirements

This subclause identifies possible requirements for commissioners, insurers or purchasers of healthcare services for a large geographic area. In some countries the purchasers may be government organizations or insurance companies operating at regional level.

The requirements might include:

— public health and planning via

  — analysis of the prevalence of risk factors and disease;

  — analysis of the incidence of disease/conditions;

  — early outbreak detection, bio-terrorism surveillance;

  — analysis of the demand for care services;

  — analysis of the outcome of care services (in both the short term and the longer term);

  — analysis of access to services;

  — analysis of preventative care;

  — evaluation of the quality, efficiency and outcomes of traditional and alternative health service providers;

  — evaluation of the efficiency and effectiveness of alternative care approaches, service models and configurations, including alternative primary care provision and prevention services;

  — programme monitoring;

  — allocation of services in relation to health needs, resource inputs, access, quality, etc.;

— commissioning and contracting via

  — management of patient access to services, including achievement of access targets for pathways which span care providers;

  — expenditure forecasting;

  — allocation of resources;

  — assignment of care received against agreed contracts with healthcare providers;

  — reimbursement, contract and grant management;

  — monitoring of expenditure against plans (budgets) and provider service levels.

The functionality required includes the capability to receive, manage, link and analyse data extracts from relevant healthcare providers. The data requirements covering non-acute care events and settings might not be covered by universal standards. It will typically be necessary to have the ability to maintain local "reference data" relating to contracts, budgets, etc. and geographical analysis capability to support planning and purchasing activities.

6.2.5 National requirements

There are many potential uses for a national or supra-regional data warehouse:

— business and performance management;

— capacity and demand planning, commissioning linked to reimbursement;

— improving productivity, possibly through comparative analysis and benchmarking;

— evaluation of health procedures and technologies;

— standards and performance monitoring, looking at clinical indicators or other performance benchmarks, such as used by a regulator;

— monitoring and evaluation and international reporting;

— public health information, including screening, surveillance and epidemiology;

— research and development, including longitudinal studies and the monitoring of outcomes and effectiveness;

— support for broader health issues, including social care;

— international data comparisons.

6.2.6 Non-functional requirements

The requirements also include "non-functional" aspects. All of these features are important for the effectiveness of any CDW. They include:

— data volumes – capacity and scalability;

— timeliness of source data;

— timeliness of reporting feedback (which for local organizations might need to be close to real time, whereas for regional or national organizations reporting could be daily, weekly or monthly);

— data validation and data quality;

— robustness and resilience (e.g. fail over capability);

— strong security and privacy safeguards;

— deidentification services such as pseudonymization;

— usability and the flexibility to address the analytic needs of diverse stakeholders;

— performance requirements.