



WORLD HEALTH ORGANIZATION

WHO/GPE/CAS/C/02.36
Distr.: LIMITED
ENGLISH ONLY

**MEETING OF HEADS OF WHO COLLABORATING CENTRES
FOR THE CLASSIFICATION OF DISEASES**

Brisbane, Queensland, Australia
14-19th October 2002

**Title: Development of ICD-10-CA and CCI in a Database
Format**

Authors: WHO Collaborating Center for the Family of International
Classifications for North America

Diane Caulfeil, B.Sc. P&OT, MBA
Consultant Classifications
Canadian Institute for Health Information
Ottawa, Ontario, Canada

Mea Renahan, BScPT, MBA, CHE
Manager Classifications
Canadian Institute for Health Information
Toronto, Ontario, Canada

Purpose: For information and discussion

Abstract:

The Canadian Institute for Health Information (CIHI) developed two electronic databases, the ICD-10-CA (Canadian enhancement) and the Canadian Classification of Health Interventions (CCI). These databases were published in CD-ROM format in April 2001. They were developed to coordinate and standardize authoring and editing activities, manage the data repository, and automate the publishing process. The CD-ROM provides a simple and fast method for health record administrators to search, retrieve and report on the more than 34,000 ICD-10-CA and CCI codes electronically. The copy code utility reduces the risk for transposition errors in code abstraction. The data architecture encompasses both classification code sets. These are represented in over 30 tables. Most tables contain code components of the classification system and supporting documentation, such as, index terms, cross-referencing, notes, objects and images. Other tables contain referencing information used in other departments in the organization.

One data file can be used to generate a document in a variety of media, including paper, CD-ROM, and the Web. Information products generated from the CIHI data repository can be easily distributed to other organizations needing to exchange documents. XML compliant Information Products can be viewed by many of the latest Browsers on the Internet.

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

The purpose of this paper is to describe the development process for creating the International Statistical Classification of Diseases and Related Health Problems – 10th Revision – Canadian enhancement (ICD-10-CA) and the Canadian Classification of Health Interventions (CCI) in a database format. The CCICD-10-CA application supports the two separate classification systems in electronic format.

History

The development of the electronic product for ICD-10-CA and CCI has gone through several iterations. – the initial prototypes, versions 1.0, 1.1 and 1.2 were MS Word documents on a CD-ROM. The first version of the ICD-10 database was loaded from the Word Perfect documents (ASCII version) provided by the WHO to which CIHI made the enhancements to make it ICD-10-CA. The CCI database was developed by Canadian Institute for Health Information (CIHI). The first production of ICD-10-CA and CCI on a database was Version 2000 using the MS SQL Server database. This was updated as Version 2001 and made available to facilities across Canada for implementation in March 2001 as mandated by the Federal and Provincial governments. Version 2003 is currently under development using the Oracle database and will be released in March 2003.

With the MS SQL database, much of the information that is integral to the classifications was generated manually rather than generated automatically from the database application. This impacted version control and record management. Text formatting information (e.g. special characters, bullets and braces) was not previously stored in the database and was corrected at the time of publication. This prompted the re-engineering of the database from SQL to Oracle.

FOLIO and the Infobases

To develop a CD-ROM and generate infobases, the CIHI began discussions with a Systems Production organization early in 2000 concerning the development of an Information Product Delivery System (IPDS) for ICD-10-CA and CCI. The major parts of the delivery system were 1) Pre-processing, 2) CD-ROM and 3) Print Product. Much of the work that was required in developing the databases and populating it with ICD-10-CA and CCI Information Products had already been completed.

Prior to the development of the databases, the production operations were impacted by editorial and typesetting cycles that were inefficient due to paper-based information processing. Time and money was being spent on formatting documents rather than developing intellectual content. CIHI undertook to restructure the core data set, manage it within relational databases, automate the workflow of editing and maintaining the content of information products. This allowed changes to be made directly to the database through edit forms and other related tools. The IPDS initiative ensured that design and development costs were directed toward building an automated delivery system with extensive reuse value. The IPDS of choice ensured that one data file could be used to generate a document in a variety of delivery media, including paper, CD-ROM and the Web.

IPDS pre-processing tools were created using Open Source programming languages, ensuring that changes could be made easily and efficiently. The IPDS was database-independent which allowed for a future move of the information holdings from MS SQL to Oracle. These tools also allowed for the generation of valid XML documents that could be readily viewed by many of the latest Browsers on the Internet and easily distributed to other organizations. The generation of the Information Products was performed through scripts converting valid XML data to proprietary coding models for output to paper or CD-ROM.

Folio Views 4.2 and FrameMaker were the software applications used. Folio supported the development tools and security features required. It operated in Windows 95/98/2000/NT and Windows 3.X and offered printing, file exporting, viewing, navigation, hypertext linking, search and personalizing options. Personalized annotations on the infobases allowed the user to find key pieces of information through coloured highlights or to add information through 'sticky notes'.

Since Folio Infobases are prepared from structured, tagged ASCII files, much of the infobase development could be prepared in SGML and XML tagged formats. Folio allowed for the building of specialized extensions into the program interface and specialized search templates could be customized searching for complex material. From Frame's context and rule-based layout scripts, complex page and paragraph layouts, graphics and tables were able to be applied.

With the IPDS all tasks for developing the CD-ROM product were entirely at the disposal of CIHI system administrators. Data repository material could be transformed into Folio flat file formats. The Folio flat files were compiled into Product Infobases, bound and prepared for distribution. Also, the installation software was generated and a data mirror was prepared for mastering on a CD-ROM.

First steps included extracting the data from the SQL Server database and transforming that data into an intermediate XML document. Application tools generated the complete CD-ROM product from the XML. FrameMaker allowed for the entire corpus of ICD-10-CA and CCI publications to be printed-to-paper using the same XML source document. Content components were handled through the databases. Open source coding allowed for new processes and delivery models to evolve.

The FOLIO Views Builder, Publisher and InstallShield Installer software allowed for the generation of the complete CD-ROM product. Folio search interfaces were prepared to help the user search for codes across specific sections of the product. The Installer built the data archive, the English/French installation interfaces and compiled the executables.

A bridge through an XML DTD (Document Type Definition) was created between the information in the data repository and the data used to generate the paper and CD-ROM products. This ensured that the data extracted from the data repository was validated prior to its final transformation to one of the output formats.

All non-tabular components, such as front and back matter and tables, not stored in the database, were converted to XML from MS Word or Excel. The Folio document

file was transformed from the XML and then combined with the definition file to produce the ICD-10-CA and CCI infobases. The definition files defined the structural level, formatting and layout, navigation and linking attributes, field structure and the inclusion of search templates and graphics.

Oracle, XML and Java

With the new Oracle database, CIHI is able to fully manage the complete maintenance of the content of the system and generate the CCI and ICD-10-CA documents in the different formats without relying on external sources. The goal of the new Oracle database was to include all information required for publication within the database itself including all associated text and the front and back matter of the publication. Previously these documents had to be converted manually at publication time. A customized XML editor now manages the large blocks of text in the front and back matter of these documents. The XML editor allows CIHI to create and maintain tables to be inserted in the CCI and ICD-10-CA documents as a formatted block of text. Specific strings of texts can be found through a search function in the XML editor making it easier to locate the content to be modified. The Include and Exclude XML fragments have reference ID tags that identify the ICD-10-CA and CCI category codes referenced, so they can be validated against the database. This also allows the reference ID within a block of XML text to be searchable for reporting purposes.

A custom Java application is used to manage the category codes and associated information. Java forms allow the system users to create, modify and disable category codes, maintain code validation edits, conversions and CCI attribute and category code components. The Java application allows the users to see all data associated with a code in a formatted representation similar to the final product. This means after editing the XML fragments the user can preview the code with all include, exclude, notes and graphics hyperlink information clearly visible on the screen. With the XML editor and the Java application, CIHI can generate the documents with a limited amount of manual editing.

Version control identifies when changes were made to a code. When a record is created, disabled or modified the system updates the record to indicate in what version the change took effect. Each version of the ICD-10-CA and CCI documents contains a summary of the errata and enhancements. It lists the changes from the previous version. The version code fields are used by the system to generate a summary of changes when a new edition of the classification system is generated. The errata and enhancements report is manually edited before being released to clients in the field. Disabled codes are also included in the summary of changes.

The Java application ensures that mandatory fields are populated when a category code is modified. These fields include the different English and French description fields, plus the status code and version control information. All text and description fields, notes, 'see also' references, component codes, attribute code descriptions, index terms, front and back matter ICD-10-CA and CCI are generated in English and French. For the CCI codes, the system populates extent, location, status and mode of delivery attributes and validates the codes.

The indices for the ICD-10-CA and CCI documents are maintained through the Java application. Entries in an index act as jump links to the codes in the documents.

The database index terms are updated separately one at a time in the Java application. At publication time each English term must have a link to a corresponding French term in the French Index.

The system prints a red maple leaf logo adjacent to an ICD-10-CA category code if it is a Canadian enhancement. The system creates a hot link between the dagger and asterisk codes and displays the appropriate characters adjacent to the codes. The Java application also supports French accents as well as Greek or Nordic/Germanic characters.

A customized XML editor has edit checks built in to validate that references can only be made to valid codes in the database. Since the XML editor does not have any edit checks built in to it, reports are run to identify potential problems with the data that is maintained by the system and the impact of the changes on other tables. Any data discrepancies are addressed before generating the ICD-10-CA and CCI documents. Some of the reports are Category Code Completeness Check, Component List Report, Disabled Code Check Report, Modified Code Check Report and Translation Check Report.

Summary

In summary, the decision to move from the SQL Server database to Oracle was based on a variety of functional, labour and time saving rationales. The new Oracle database formats the includes and excludes in a block of XML data while in the old SQL database each line of the include was stored in a separate record and had to be formatted after it was exported to ASCII by an outside source. The brackets and braces in the includes can now be maintained in XML rather than having the external company generate them. The front and back matter was previously converted from MS Word documents. They are now stored in XML in the database making maintenance easier and eliminating the need to convert. Storing raw XML structured data in the database makes the process of publication much simpler, since less formatting is required, and it is easy to create the complete XML document from the database. All level of codes are now stored in one table, where previously "codes" were stored in one table and all other level of code (Chapter to rubric) were stored in a separate table. It is now easier to retrieve the data. The model has been simplified. There are about half as many tables with the Oracle database as there were with the SQL database. Graphics can be stored within the database and maintained by the application. Audit trail information was added to the new database, as well as error report generation; for example, a report now indicated all references to a code and ensures that they are removed before a code can be disabled. And finally, the Oracle database has allowed the French data to be populated into the database simultaneously with the English data.