

IRIS: a language independent coding system for mortality data

G rard Pavillon, Lars Age Johansson, Donna Glenn,
Stefanie Weber, Bj rn Witting, Sam Notzon



Abstract Iris is a software for mortality coding using ICD. Iris includes components from MMDS, the NCHS coding software considered as a standard. This software has been designed so that it can be adapted and used by any country. Iris is the product of a long international collaboration aiming at improving and facilitating the quality and comparability of mortality statistics.

Introduction

The fourth plenary meeting of the International Collaborative Effort (ICE) on Automation, organized by the US National Center for Health Statistics (NCHS) in May 2008 will present Iris, a system for international use aiming at improving international quality and comparability of mortality data. It is a major step for this longstanding project and a perfect illustration of "Owning and Sharing" at the international level. This software has been developed through a longstanding collaboration between several countries including France, Germany, Hungary, Italy, Spain, Sweden, and the United States. Eurostat, the statistical office of the European Community has also supported this project from the beginning. The final product will be available before the ICE meeting in May 2008. The software will be available free of charge. It will be open source software for the Iris user's group only, in order to maintain international standardisation of the algorithms. Iris is based on the de facto international standard, MMDS, the NCHS coding system already used by most Anglophone countries. Two main observations led the Iris user's group to invest resources in this project:

- 1) Mortality data are widely used to describe and compare health characteristics between countries, yet between-country comparisons show large discrepancies in the results of manual mortality coding. The International Classification of diseases (ICD), published and updated by the World Health Organisation (WHO), is used by most countries to code causes of death. The ICD is now more than one century old and the last revision of this classification (ICD-10), reflecting the increasing complexity of medicine, is so complex that only classification experts can use it properly.
- 2) Automated Coding Systems standardise mortality coding. However existing systems are language dependent, i.e. they include specific language processing in their algorithms. For instance MMDS is specific to English and is difficult to use with other languages.

Methods & Materials

The first attempt involved rewriting MMDS to separate the language component from the algorithms. This task proved to be very complex due to the difficulty of synchronizing separate updates of two systems. In the end, it proved impossible to maintain compatibility between the new system and MMDS.

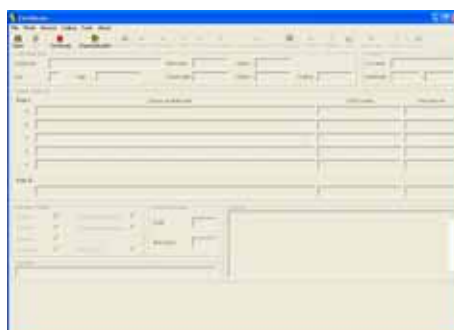
The Iris project is based on a different concept. Two modules of MMDS, MICAR and ACME, are used as components of a language independent system. These two modules deal with coding only and do not include language aspects. Nevertheless MICAR and ACME require specific coding, different and more complex than the strict coding based on ICD rules and guidelines. A specific front-end module has been developed in order to cope with this problem. This module accepts standard ICD coding and maps the codes so that they comply with the more complex MICAR and ACME coding.

The resulting system – Iris - attains the desired goals:

- 1) It is language independent because linguistic aspects depend only on the dictionary provided;
- 2) It is MMDS compatible since it uses MICAR and ACME modules as components;
- 3) It is compatible with ICD rules and guidelines.

Furthermore, specific features have been developed to help Iris users:

Interactivity. Iris has been designed to cover the whole process from coding to underlying cause selection in one step. Iris also allows batch processing for large files of death records. Iris main screen has been designed for the coder:



Language standardisation. An integrated tool in Iris allows pre-processing of language expressions in order to reduce the size of the dictionaries.

National adaptations. Original Iris display screens are in English, but the program has been written so that specific national adaptations can be made.

Explanations. On request, Iris can provide explanations for each step of the process. This aspect is essential when coding is complex. It also allows the use of Iris as a pedagogical tool to teach cause of death coding.

Results

Several tests of Iris have been carried out:

- MMDS and Iris compatibility: Hungary and Sweden
- Iris and manual coding: Spain
- Language standardisation: Germany

The English international version of Iris is nearing completion and will be available before the ICE meeting in May 2008. A German version will be implemented in 2008 in several Federal states. A Swedish and a French version will be developed in 2008.

Iris software has been referenced by APP, the Agency for Software protection. It will be licensed as free software. Modifications to the source code will be limited to the Iris user's group to keep Iris standardised.

Conclusions

Iris is a language-independent software system available to the international community. This system is compatible with the MMDS software already used by many countries. The main objective is to increase mortality data quality and comparability at the international level. This software system will be presented during the fourth ICE meeting in May 2008.