

Iris Institute Strategy  
for future development and  
maintenance of Iris in light of the  
upcoming ICD-11

# Strategy Plan



6 April 2020

Version: 1.0

Strategy Plan .....	1
Strategic Goals .....	3
Business strategy .....	3
Status quo .....	3
Mission of the Iris Institute .....	4
Relationships of the Iris Institute with other organizations working on cause of death statistics .....	4
Business requirements to be addressed .....	4
Performance issues (throughput).....	5
A higher rate of automated coding is desired for the future .....	5
Management strategy .....	6
Developing ways of working .....	6
Funding/Resources/Stakeholders.....	6
Project management and timing .....	6
Technical strategy.....	7
Iris Software: Adaptation of Iris software for ICD-11 and Adaptation of MUSE algorithm for ICD-11 .....	7
Software security.....	7
Backward compatibility .....	7
Transition of unicausal and multicausal Tables .....	8
How complete do we want to have the tables?.....	8
Adaptation of Table Editor for ICD-11 .....	8
Implementation of Table Editor for multicausal rules.....	8
Organization of the technical work.....	8
Input to ICD-11 from the Iris Core Group .....	9
Electronic death certification (eDC).....	9
Maintenance of the Strategy Plan .....	9
Appendix: List of abbreviations and used terms.....	10

## Strategic Goals

For the future of Iris in light of the upcoming ICD-11 (International Classification of Diseases 11<sup>th</sup> Revision) the Iris Core Group decided to outline the strategy for Iris for the near, mid-term and long-term future. The strategy aligns to the main priorities that have driven the Iris development from the beginning and will do so in the future:

- Iris is a software-tool to achieve **better quality data** on cause of death statistics by following the ICD rules and implementing them in a transparent way
- **Long-term maintenance** of this international endeavor is enabled through input from the Iris Core Group members as well as by voluntary contributions (financial contributions or contributions of collaborative work) from Iris Users and Supporters
- Iris makes use of ICD in an **economic and flexible framework**, adapting to new technology and classification versions as need arises
- **Transparency** on the way Iris works and how it is maintained is key
- Interaction with third parties (all kind of Institutions) on the Iris software is encouraged and possibilities to exchange experiences on the use of Iris foster its fitness for purpose

Based on these priorities the main goals for the upcoming years are to:

- redevelop Iris for the use of ICD-11 by making use of the new features and technologies the ICD-11 framework provides
- consider new technologies for the software redevelopment to stay up-to-date with the evolving IT landscape and allow flexible use of Iris in a digital environment, still reusing proved and tested features as much as necessary
- enhance the skills, knowledge and capabilities in the Iris Core Group and enlarge the team of experts on the Iris system, to enable smooth ICD-10 to ICD-11 transition and future software development
- ensure that all technical solutions make the coding process and application of the ICD rules transparent
- exchange with users and third parties to achieve best results in the endeavor of the redesign

The Iris Institute promotes comprehensive coding of causes of death, based on death registration, according to the ICD-10 (and subsequently ICD-11) system. Other ICD coding scenarios (such as hospital activity) are outside its scope. Alternative approaches to recording or classifying causes of death (such as Verbal Autopsy) are outside the core mission, although collaboration on such issues with other stakeholders may be possible if priorities and resources allow.

## Business strategy

### Status quo

The Iris Institute emerged from an international cooperation for the deployment, maintenance and development of the Iris software, an electronic system for automated coding of causes of death.

The increasing international interest in the Iris software has made it necessary to provide an institutional foundation to secure its supply and support. The cooperating partners formed the Iris Core Group and agreed on an administrative regulation that secured the possibility of a long-term maintenance of Iris through the provision of a secretariat at DIMDI and the commitment of each partner country to spend at least 250 hours per year on the maintenance and development of the Iris software. Currently 6 countries are members of the Iris Core Group with two additional countries being in the process of joining it in a candidature position.

The maintenance supported by the member countries of the Iris Core Group is generally limited to the core versions and functionality and does not cover implementation in individual countries. However, the Core Group may agree to include in its efforts, if resources allow, support for innovative methods of deployment or other developments at national level where these would provide benefit to the Iris users as a whole.

In order to give users the opportunity to contribute to the maintenance and update of the Iris system, and thereby guarantee the long-term availability of the Iris software, a support model was put in place. Contributing to the Iris development by agreeing to financially support Iris enables countries to contribute to the availability and up-to-date software that is adapted to ICD updates and software developments. All financial support received by the Iris Institute will be used to maintain, update and further evolve the Iris software. Currently 12 countries support the Iris Institute through the Support Model.

## **Mission of the Iris Institute**

The mission of the Iris Institute is to support the development of the Iris software, in order to promote standardized electronic coding of cause of death statistics throughout the world, and to share resources for this international enterprise. In doing so, the Iris Institute is committed to following the ICD classification and instructions on coding strictly, according to the guidance of WHO and the MRG (Mortality Reference Group).

In order to allow users of the Iris software to interact with the Iris Institute, its members and other users, the Iris Institute holds annual meetings for exchange of information and for joint discussion on Iris developments. As well, the Iris Institute assists users in organizing training when starting with implementation. All of the Iris Institute activities are on a not-for-profit basis.

## **Relationships of the Iris Institute with other organizations working on cause of death statistics**

In order to achieve its mission the Iris Institute works with other key players, mainly through cross-membership of Iris Core Group members:

- World Health Organisation: As the World Health Organisation (WHO) is responsible for the development, publication and implementation of the ICD, the Iris Institute collaborates with WHO, mainly through its Family of International Classifications Network (WHO-FIC).
- The WHO-FIC Network is composed of national collaborating centres and the WHO classifications team. Its Mortality Reference Group (MRG) advises WHO on the development of the ICD for its use in mortality statistics. The Informatics and Terminology Committee (ITC) is discussing standardized electronic applications for the use of the Family of International Classifications. Both groups are informed on lessons learned from the implementation of ICD into the Iris Software and its application on a regular basis.
- Eurostat: The European Statistical Agency has funded Iris development in its early stages substantially and therefore enabled the establishment of Iris as de-facto standard software for coding of cause of death statistics. Members of the Iris Core Group regularly exchange experiences on the use and development of the Iris with the relevant bodies at Eurostat.

## **Business requirements to be addressed**

The Iris software is a system that needs updating and maintenance, just like any other IT-System. Especially the updating of ICD requires an annual cycle of new versions. With ICD-11 being approved and the implementation date set from January 1<sup>st</sup> 2022 onwards new challenges for the Iris software arise. As well, the basis of the Iris software was set in 2003 and the infrastructure of Iris is still based on this Version, even though with many modernizations and adaptations.

Therefore, the next few years necessitate intensive work on the Iris software requiring a solid business and management framework.

Challenges are for example:

- Modernization of the technical design of Iris to meet user needs for high coding accuracy and throughput and make effective use of latest developments in programming methods, data science and IT platforms will require a phase of conceptualizing and evaluation.
- Development of ways of working that will enhance the sustainability of Iris over a long period, has the potential to scale up technical effort, and allows the successful delivery of new developments, taking into account good practices in project management and software production.
- Support of the introduction of ICD-11 by creating a valid version of Iris that can handle ICD-11 coding algorithms and logic. Considering the potential of new features of ICD-11, like cluster coding or the coding tool, this version will be developed in a more flexible framework to allow countries to use Iris

in a way that legal circumstances require (e.g. integration of stand-alone coding tool vs. use of ICD-11 API, use of Iris as a web service, integration of Iris in a CRVS system starting with electronic coding and ending a fully digital dataflow in an almost fully automated Iris coding with output of underlying and multiple cause codes).

- Encouragement of and support to countries adopting Iris assisting in the organisation of training, by regular meetings for international exchange on Iris and by information to the relevant stakeholders e.g. by presentations on Iris or through publications.

Due to the requirement to continue routine processing of death certificates in all countries using Iris with ICD-10, the ICD-10 version of Iris should be supported at least until an ICD-11 version is in place. For a few following years, while countries will prepare to move to ICD-11, support for the Iris with ICD-10 will be reduced to an absolute minimum and then ended. Major changes to tables should no longer be performed as the updating of ICD-10 has ended. Nevertheless, correction of errors should be guaranteed for the coming years.

## Performance issues (throughput)

The Iris software should be thought to be able to code in one day several hundred thousand-death certificates in order to allow for powerful testing and as well for real time coding of large amounts of certificates. With the interaction of ICD-11 tools, the complexity is expected to increase with simultaneous increase of computational requirements.

### A higher rate of automated coding is desired for the future

In the past, Iris has always rejected a series of certificates due to varying reasons. General rejects of Iris have been implemented for cases where the Iris algorithm was not sufficiently developed and the cases are of high public interest.

Frequent reasons for rejects in the Iris system are:

- Unrecognized text of conditions [new methods such as NLP (Natural Language Processing) and AI (Artificial Intelligence) can be tested to improve recognition of spelling and other errors]
- Surgery and medical procedures (and external causes): These cases mainly lacked enough detail on the certificates and the rules in ICD did require precise details. (A simplification of the rules should be discussed)
- Maybe-Rejects: rejects where the coder was requested to check the certificate for additional information as the Iris logic could not determine the correct underlying cause due to ambiguities in the logic. Often Maybe-Rejects were placed for cases of high public interest. (It is hoped that with ICD-11 most can be solved by using postcoordination)

The main principles that should be followed while increasing the rate of automatic coding are:

- Machine learning methods are powerful tools to increase the rate of automated coding (potentially near 100% coding), and recent promising results should be taken into consideration. However, such methods are evolving very fast, and are still needing high level technical skills and specific IT infrastructures to be implemented. Further investigations should be conducted to adopt common acceptable methods. These tools could also improve the quality of the input data, e.g. by developing tools that interact directly with the medical certifier (checking when obvious inconsistencies are declared for example).
- The method chosen should be fully transparent. In a statistical perspective, it does not necessarily mean that the rules applied have to be explicitly formulated for each record, currently it is not the case for all human coded records. If machine learning methods are used, then the method, the data used to build the algorithm, and the performances in accurate coding should be clearly described.
- Specifically regarding the underlying cause identification step, explicit rules methods (also called expert systems, MUSE being the current one for Iris) should be used as a first stage in the coding process. The coding engine should be able to code the highest possible rate. Thus, changes in coding rules could directly be applied and under control.

## Management strategy

### Developing ways of working

With the initial development of Iris a small team of programmers and epidemiologists have worked in a flexible and requirement-driven way. For the redevelopment, now a bigger team of developers and epidemiologists will have to work together in a more structured and organized way. Therefore, a clear process has to be established from the get-go. The structure will have to consider as well the sustainability of the software after the redevelopment so the described ways of working will not only be project driven but as well will have to consider the time after the redevelopment, which might lack the availability of the same amount of resources as the redevelopment.

- Introduction of a stable and easy to use centralized and secured source code management (e.g. version-control system like Git)
- Introduction of test-driven development (automated unit tests)
- Investigating the need for CI/CD tools (continuous integration, continuous delivery)
- Clear structure for the documentation and description of source code, table structure and table maintenance

### Funding/Resources/Stakeholders

It is essential to establish a way of collecting funding for the redevelopment as soon as possible. The amount of funding necessary for the redevelopment was roughly estimated in the pre-study for the redevelopment.

Possible sources of funding are:

- International organizations interested in the collection of standardized and coded Cause of death data according to ICD (WHO, Eurostat, OECD, other organizations)
- Countries using Iris now and want to continue using Iris in the future with ICD-11 (existing Iris supporters, other countries using Iris)
- Countries not using Iris yet, but considering starting with Iris at the time of the introduction of ICD-11
- Funded projects working on establishment or improvement of CRVS-systems in countries
- Research projects interested in the use of standardized coded cause of death data for research purposes

In order to allow best results experts and relevant expert groups should be addressed and requested to engage in the redevelopment of Iris from the get-go. Mainly the two relevant working groups from the WHO-FIC-Network, the MRG and the ITC, have to be included in discussions to allow best alignment of Iris with other ICD-11-tooling. As countries will use a set of ICD-11-Tools together, the IT-infrastructure should be able to make use of the tools in the same environment.

### Project management and timing

For the redevelopment of Iris, a project management framework needs to be in place to manage and update the project plan, the progress of development, the testing phase and the documentation for sustainability. Especially if the possibility of a distributed development through in-kind contributions and contracted work is considered, a strong project management with clear reporting and management regulations needs to be established.

The Iris Core Group will develop a (estimate of) timeline for the coming years based on the results of the pre-study, including the cycle for development, testing, feedback and adaptation. As a result of the Evaluation Project, a list of activities with estimated time per task was proposed.

Initially the project management will coordinate the development of functional specifications followed by detailed specifications. A dedicated development team will be constituted in parallel, with a documentation, a technical and functional checking procedure.

Regular reporting to the Iris Core Group, the Users Group and the public will allow for management of expectations to timeline and implementation dates.

## Technical strategy

### Iris Software: Adaptation of Iris software for ICD-11 and Adaptation of MUSE algorithm for ICD-11

The redevelopment of Iris will start 15 years after the initial development of Iris started. New technologies have emerged and need to be considered especially as the IT-Infrastructure in which the Iris software is deployed has changed substantially since the first development of Iris.

As well, the digitalization of the death certificate is progressing and many countries have now started to implement electronic death certification, allowing for interaction with the certifier at the time of certification. This enables changes to the Iris processing logic and hopefully will allow for better source data for the coding of cause of death data. These innovations should be considered, including:

For dictionary and language processing:

- More advanced NLP tools, possibly in interconnection with the WHO API for ICD-11
- Self-building dictionary and standardization applying AI (Artificial Intelligence), based on the ICD-11 foundation. (Experts, supervisors or authorized person associates a rejected term to a foundation element and Iris includes the term in the dictionary and learns if there are non-essential modifications and grammatical variations.)
- Language processing mechanism should be pluggable and customizable. Beyond these new revolutionary tools respectively, the existing Regex based standardization with dictionary matching approach could be available as plug-in option for text processing.

For interface to certifier:

- API to directly dialog with the medical certifier, in order to highlight inconsistencies or insufficient information given at the time of certification
- Explore machine learning tools that would learn from validated data, suggest coding results to the coders and give a confidence in the coding propositions

WHO coding tool is referred to as a future proof and strategic mechanism which integration with Iris is considered highly significant. Therefore it is substantial to start testing its capabilities for medical text coding in automated coding.

### Software security

Software security is still a major issue for the Iris development. All countries have legal rules of confidentiality protecting mortality data, with national variations, as well as different approaches to protecting the security and integrity of their data and IT systems. With possible API interfaces, the value of protecting personal data processed in the Iris system needs to be highlighted and necessitates special attention. Using a local instance of the WHO API on local servers can be one solution.

### Backward compatibility

Concerning the backward comparability with the previous dictionary, it would be essential to provide some tools for facilitating the transition of the dictionary to ICD-11 for each country that has developed a separate dictionary. Still, as the maintenance of a specific language dictionary is outside the realm of the Iris Institute, resources on this issue can only be invested through separate country-specific projects. One option could be to provide a tool for attributing a foundation URI to each dictionary entry (via the coding tool).

Backward compatibility is an essential asset for the users of Iris, including the Iris Core Group member countries. There are two ways to achieve backwards compatibility:

- a. To start from the software as is and only adapt where absolutely necessary
- b. To develop new software independently of the backward compatibility first, looking for the best target, and then address backward compatibility through thorough testing.

It is likely that the production and coding process will change because of the new design of ICD-11. That is why the change to ICD-11 is the right time to push structural changes, so option b seems to be preferable from a theoretical point of view. For example, when applying automatic coding method, the

coders may more frequently have to access to very specific sets of death certificates. Currently the way the databases and the interface are organized is not optimized for pointing easily to such specific records.

## Transition of unicausal and multicausal Tables

The translation of tables is resource intensive; nevertheless there is still the need to have the tables in the current format since the ICD is still based on rules with decision nodes. Resources might be reduced by translating only the decision nodes most frequently used.

A new approach to develop ICD-11-decision tables can be informed by Artificial Intelligence (AI) methodology based on a huge amount of coded cause of death data where decisions were taken in the past by human coders or by automated systems. However, it will certainly be inefficient for rare cases or cases impacted by changes in coding rules.

There are ways to optimize the combinations of rules and AI methods, but this field is still under research and development.

### How complete do we want to have the tables?

The Iris Core Group is aware that the work on the tables is continuous. Redeveloped tables should guarantee at least the coding of the most frequent cases (the threshold of 95% certificates was chosen as the target in the Evaluation Report of the prestudy).

It should be of interest to open the decision tables more to the medical and scientific community, with possibilities to receive suggestions of modifications, once the tables have reached a more or less stable state. With a (hopefully) increasing amount of electronic certificates, holding the capacity to reduce ambiguities and undefined terminology at the time of data entry, better quality of information might reduce the workload on the tables as they often have been addressed whenever badly certified cases were not handled by the Iris system. The focus of which (groups of) codes will need to be translated might be impacted. Still, as many countries yet have to tackle this endeavor, the near future might still necessitate the Iris system to handle such cases as well.

## Adaptation of Table Editor for ICD-11

The Decision Table Editor is a powerful tool developed by the Iris Core Group, namely by the Italian members of the Group. The structure of the Decision Table Editor is usable also for ICD-11 with some additions and corrections such as the use of post-coordination. Nevertheless, the software is becoming old, and evolutionary maintenance has been scarce. For this reason, it would be useful to address a major revision in view of ICD-11.

It would be preferable if new tooling for the maintenance of the decision tables can be linked more closely to the outcome of the MRG, and serve as the place for a reproducible and clear documentation of coding changes, so probably the tool has to be thought differently.

## Implementation of Table Editor for multicausal rules

On the one hand, the use of a software for the table maintenance, besides to the obvious advantages, reduces flexibility of intervention. So, the editor for multicausal tables should be developed when the multicausal tables structure reaches sufficient stability ideally with reduced complexity compared to ICD-10. On the other hand, currently the multicausal decision tables are kept in a flat table format where table maintenance is very error-prone. Therefore, the Iris ICD-11-project needs to consider in due time which way of maintaining the multicausal tables it will choose and then implement a tooling environment in line with the unicausal decision table editor.

As with the decision tables for unicausal coding, the availability of the tables for the interested public might be a way to enhance its content through feedback mechanisms from users and researchers.

## Organization of the technical work

Two key questions have to be addressed by the Iris Core Group in order to address the long-term maintenance of Iris as well as the redevelopment of Iris for modernization and for inclusion of ICD-11:

- How is it possible to scale up the technical effort so that the time of the existing experts is used most efficiently, while knowledge is shared so that a wider group can assist in supporting tasks (e.g. testing, making minor changes to code, documentation) and in developing the application of new technologies?



- As testing of new software is always the key issue to a seamless implementation and Iris is implemented in a variety of different settings, an important question to answer is: How can the testing be organized?

## **Input to ICD-11 from the Iris Core Group**

Since the development of the Iris software valuable lessons have been learned for the general collection of data on mortality statistics. Most coding rules and coding logic steps from ICD-10 have been carried over to ICD-11. With the implementation and the results of the pre-project (Evaluation Project for integration of ICD-11 into Iris) some coding steps seem to be unnecessarily complex and do not alter results significantly. The Iris Core Group is planning to propose simplifying some complex ICD instructions. The elaboration of the proposals is currently in progress and will be introduced to the WHO Reference Groups and Committees through the regular channels.

## **Electronic death certification (eDC)**

In almost every country, which is engaging in the Iris User group, electronic certification is an important issue. A few countries are already very far with this and receive most of their death certificates electronically. Others are working on it, often struggling to get it done. While the Iris Institute is not the platform for electronic certification, the development itself is important to take into consideration because of its possible implications for coding in ICD-11 with Iris. It is possible that more and hopefully better data will be available when electronic certification will make up a greater part of the causes of death certificates that are being coded for our statistics. It would therefore be advisable to keep a close eye on the developments in this area in the different countries and to keep track of the progress with the electronic certification in order to properly estimate its effect on the coding in Iris in ICD-11 and to be able to make adjustments for this when necessary.

## **Maintenance of the Strategy Plan**

This plan will be agreed by the Iris Core Group to guide the detailed planning and activities of the Iris Institute, such as the Development Paper and project management plans. It should be possible to validate that all plans and activities are in line with the strategy.

The strategic direction as a whole is expected to be constant for at least five years. However, the Strategy Plan should be reviewed annually to ensure that new circumstances are taken into account, and updated as outstanding questions are answered.

## Appendix: List of abbreviations and used terms

Table 1 List of abbreviations and terms with explanation

Abbreviation/Term	Explanation
AI	Artificial Intelligence
API	Application programming interface: Allows programmatic access to the International Classification of Diseases (ICD). It is an HTTP based REST API. You may use this site to access up to date documentation on using the API as well as managing the keys needed for using the API. ( <a href="https://icd.who.int/icdapi">https://icd.who.int/icdapi</a> )
CI/CD	Continuous integration/continuous delivery: Is a method to frequently deliver apps to customers by introducing automation into the stages of app development. The main concepts attributed to CI/CD are continuous integration, continuous delivery, and continuous deployment.
DIMDI	German Institute of Medical Documentation and Information
eDC	Electronic Death Certificate
Eurostat	Statistical Office of the European Union
Git	It is free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency
ICD	International Classification of Diseases
ICD-10	International Classification of Diseases 10th Revision ( <a href="https://icd.who.int/browse10/2019/en">https://icd.who.int/browse10/2019/en</a> )
ICD-11	International Classification of Diseases 11th Revision ( <a href="https://icd.who.int/en">https://icd.who.int/en</a> )
Iris	Is an automatic system for coding multiple causes of death and for the selection of the underlying cause of death
Iris Core Group	Developer group of the Iris software
Iris Institute	An international cooperation for the deployment, maintenance and development of the Iris software, an electronic system for automated coding of causes of death
ITC	Informatics and Terminology Committee (develops policies on electronic standards used for WHO-FIC classifications and related terminologies and promotes their international implementation)
MRG	Mortality Reference Group (improves international comparability of mortality data by establishing standardized application of the ICD)
MUSE	Multicausal and Unicausal Selection Engine (coding kernel of the Iris software)
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
WHO	World Health Organisation
WHO-FIC	World Health Organisation – Family of International Classification