ISO/TC 37/SC 2/WG 1 **N208**
2011-06-07

TC 37 – Terminology and other language and content resources
SC 2 – Terminographical and lexicographical working methods
WG 1 – Language coding – *Convener:* **Håvard Hjulstad**

L2/11-235

"Next generation" ISO 639

1 Background

This document is intended to support the discussion at the 2011-06 meeting of ISO/TC 37/SC 2/WG 1 relating to the development of ISO 639 toward a database-based version.

The planning so far has assumed that there will be one linear document published as a "regular" International Standard. That document will be based on ISO 639-4 and the linear parts of the other parts of the standard. Document "ISO 639-0" is intended as a starting-point for that document. It has been referred to as "Part 0" and "new Part 4". It will probably end up being just "ISO 639" with no part designation. Document **N204** is the latest version of "ISO 639-0".

The plan has been that all tables of all current parts of ISO 639 be merged into one table, which will be represented in one database. That database is intended to be an integral part of ISO 639. At the same time the database will need to support the language coding development process.

The hope and expectation has been that the ISO Concept Database (CDB) would be developed in a way that could support the development process as well as the final publication. However, it is now likely that this will not be possible in the near future. The CDB will be suitable for publication of the "final" version of ISO 639 tables, but all stages of the development of the standard will need to be hosted in a different database environment.

2 Data model

A data model needs to be built that will handle both the final publication of the code tables and the development processes, including balloting. The development needs to follow the procedures that are specified in the ISO/IEC Directives for "Standards as databases".

There are many types of information that could be useful in relation to language coding, including (more or less obviously)

- identifiers (alpha-2, alpha-3, alpha-4, alpha-x);
- names in any number of languages, including multiple names in any one language;
- linguistic hierarchy and typology;
- geographical area of use (this could be specified through country coding and country subdivision coding, but also by means of geographical coordinates);
- temporal information (through dates and "classification" like "ancient", "extinct", etc.);
- information about writing system;
- status of language in relation to government, education, etc.;
- etc.

And things that relate to the process, including

- ongoing and past ballots;
- change request submission portal;
- publishing results of ballot processes;
- maintenance of finalized standard.

The purpose of ISO 639 shall not be to collect all kinds of information about languages. Much information is available in other sources, which should be linked and not duplicated. Should ISO 639 aim to serve as a "portal" to "everything that the world knows about all languages", or should the core purpose of assigning language identifiers be also the only purpose?

3 Governance structure

A main key is to find a host for the database and a secretariat that is needed to support the standard. This could be an organization with a linguistic purpose or a standardization organization. There are good arguments to

support both views. Long-term funding is needed, as well as sufficient level of linguistic expertise to ensure a high-quality operation.

The “Standards as databases” procedures specify Validation Team (VT) and Maintenance Team (MT). In the case of ISO 639 the VT will be persons appointed by National Standards Bodies. The expertise for the VT cannot be stipulated by ISO 639 “rules”. On the other hand the MT needs to be a broadly composed scientific body that can provide reliable information for the VT to base its decisions on. It will be important to make sure that the MT attracts the “correct” people. It needs to “make sense” and serve a clear purpose for experts to serve on the MT.

ISO/TC 37 may not have sufficient visibility in all linguistic circles to be the only basis for the new ISO 639 MT. Since most of the relevant expertise is to be found in academic environments, it may make sense to liaise with academic linguistic networks to find the best possible composition.

The experts will have specific linguistic regions as their area of expertise. They should not be “bothered” with issues that are entirely outside their area of interest. Consultation with experts should be “regulated” according to relevant criteria. On the other hand, an expert who hasn’t heard from us for a long time will lose interest. It will be a challenge for find the correct balance.

4 Conclusion

There are a number of things that will need to be done more or less simultaneously:

1. the data model needs to be finalized;
2. a host organization for database and secretariat needs to be found;
3. all ISO 639 data needs to be merged, and from that point maintained in one database;
4. the Maintenance Team and the secretariat need to be up and running.