



ISO Language Codes

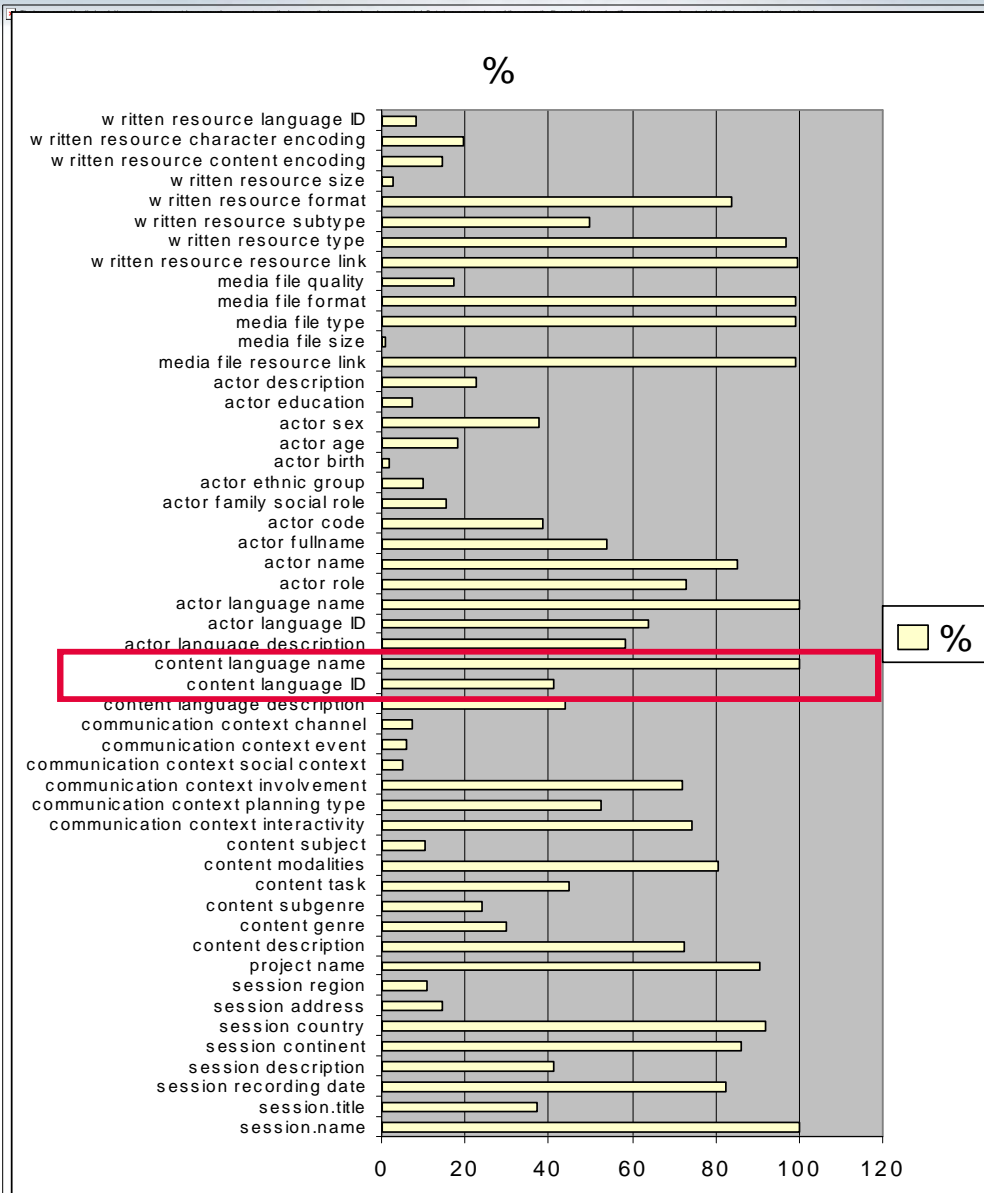
Peter Wittenburg, Håvard Hjulstad
CLARIN / Max-Planck Institute for Psycholinguistics
Chair ISO TC 37

Contents



- State using language codes
- Standards ready for use
- new ISO plans
 - integration of all codes
 - ISO concept database as basis
 - decision structure

State in LRT



- MPI tools such as IMDI support Ethnologue list (now ISO 639-3)
- but statistics on 27000 records...
 - language name for 100%
 - language code for 40%
 - language names differed
 - typos, variants
- why?
 - people know "their" name
 - people are lazy looking up
 - people are unwilling
 - people are ...

Why should we care?



- everyone would agree that we need a unique ID for each spoken language and probably even beyond (dialects, extinct languages, etc)
- why: we and in particular machines should know where we talk about
- AND: linguists need >7000 language codes to be specific enough
- can't we just use the language names?
- of course not:
 - same things in real life are called different by different people
 - different things are called the same
- and in addition: who dares to decide what the "truth" is
- language names are societal and political issues and there are a variety of names
- numbers and codes are mostly not so sensitive

What has been done?



- ISO 639-1 2 letter codes 136 languages 1988/2002
ISO 639-1:2002 *Codes for the representation of names of languages — Part 1: Alpha-2 code*
 - ISO 639-2 3 letter codes 460 languages 1998
ISO 639-2:1998 *Codes for the representation of names of languages — Part 2: Alpha-3 code*
 - ISO 639-3 3 letter codes ~6000 languages 2007
ISO 639-3:2007 *Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages*
-
- the message must be: start using ISO 639 codes now

639-1 (alpha-2)



Codes for the representation of names of languages (Library of Congress) - Mozilla Firefox

Menu File View Tools Help New Window New Tab Duplicate Tab Duplicate Tab To New Window On/Off AiOS-Toolbar IE V

Back Forward Reload Stop Home <http://www.loc.gov/standards/iso639-2/php/langcodes-search.php>

Most Visited

ISO 639 - Language Codes x Codes for the representation of n... x

[Library of Congress](#) >> [Standards](#)

ISO 639-2 Registration Authority

-- Select Alpha-2 code --

(none)
aa
ab
ae
af
ak
am
an
ar
as
av
ay
az
ba
be
bg
bh
bi

-- Select Alpha-2 code --

[ISO 639-2 HOME](#) - [Code List](#) - [Changes to Codes](#)
[ISO 639 Joint Advisory Committee](#) - [ISO 639-1 Registration Authority](#)

or Search the ISO Language Codes

code. Where two codes are provided (21 languages total), the bibliographic code is given second. ISO 639-1 is the alpha-2 code. Multiple codes for the same synonyms.

3 code:

2 code:

Search by Keywords (by codes or by language names in English or French):

Search for in

Full Text

Comments on this document: iso639-2@loc.gov

[ISO 639-2 HOME](#) - [Code List](#) - [Changes to Codes](#)
[ISO 639 Joint Advisory Committee](#) - [ISO 639-1 Registration Authority](#)

[Library of Congress](#) >> [Standards](#)

[Contact Us](#)
November 7, 2008

639-2 (alpha-3)



Codes for the representation of names of languages (Library of Congress) - Mozilla Firefox

Menu File View Tools Help New Window New Tab Duplicate Tab Duplicate Tab To New Window On/Off AiOS-Toolbar IE V

Back Forward Reload Stop Home http://www.loc.gov/standards/iso639-2/php/langcodes-search.php

Most Visited

ISO 639 - Language Codes

-- Select Alpha-3 code --

- [none]
- aar
- abk
- ace
- ach
- ada
- ady
- afa
- afh
- afr
- ain
- aka
- akk
- sqi
- ale
- alg
- alt
- amh
- ang

-- Select Alpha-3 code -- Go

Browse the Index by Alpha-2 code:

-- Select Alpha-2 code -- Go

Registration Authority

[ISO 639-2 HOME - Code List - Changes to Codes](#)
[ISO 639 Joint Advisory Committee - ISO 639-1 Registration Authority](#)

or Search the ISO Language Codes

code. Where two codes are provided (21 languages total), the bibliographic code is given second. ISO 639-1 is the alpha-2 code. Multiple codes for the same synonyms.

-3 code: Search by Keywords (by codes or by language names in English or French):

Search for in

Full Text Go

Go

Comments on this document: iso639-2@loc.gov

[ISO 639-2 HOME - Code List - Changes to Codes](#)
[ISO 639 Joint Advisory Committee - ISO 639-1 Registration Authority](#)

[Contact Us](#)
November 7, 2008

[Library of Congress >> Standards](#)

639-3 (alpha-3 for ...)



ISO 639 code tables - Mozilla Firefox

Menu File View Tools Help New Window New Tab Duplicate Tab Duplicate Tab To New Window On/Off AiOS-Toolbar IE View IE Tab Showcase All Windows

Back Forward Reload Stop Home <http://www.sil.org/iso639-3/codes.asp> ☆ 63

Most Visited

SIL HOME ABOUT SIL SITE MAP SEARCH CONTACT US

ISO 639 Code Tables

This page offers a combined view of the language code tables of ISO 639 parts 1, 2, and 3. Select just the elements of the Part 1, 2, or 3 code, or show the set of code elements sorted by name. Viewing by name will enable you to browse for any name associated with a specific identifier, including an inverted form of a name (e.g., the code element id=[aaq] "Eastern Abnaki" will also be found under "Abnaki, Eastern"). The elements may also be ordered by [scope of denotation](#) or [type of language](#). The "more" link provides further documentation on what the code element denotes. In the case of a macrolanguage, this includes a listing of its individual member languages.

View by:

639-3 639-2 639-1 Name Scope Type Starting with **A**

See also: [ISO 639-2 Registration Authority — Library of Congress](#) (which is the primary authority for Parts 1 and 2 of the standard and is the definitive source in the case of a discrepancy between that site and this).

See also [Retired Code Elements Index](#)

639-3	639-2	639-1	Language Name	Scope	Type	
aaa			Ghotuo	Individual	Living	more ...
aab			Alumu-Tesu	Individual	Living	more ...
aac			Ari	Individual	Living	more ...
aad			Amal	Individual	Living	more ...
aae			Arbëreshë Albanian	Individual	Living	more ...
aaf			Aranadan	Individual	Living	more ...
aag			Ambrak	Individual	Living	more ...
aah			Abu' Arapesh	Individual	Living	more ...
aai			Arifama-Miniafia	Individual	Living	more ...

Who owns what?



- ISO 639 is “owned” by ISO/TC 37
(*Terminology and other language and content resources*)
- For ISO 639-2 there is a “joint ownership” with ISO/TC 46
(*Information and documentation*)
- Registration Authorities until now:
 - ISO 639-1: Infoterm, Austria
 - ISO 639-2: Library of Congress, USA
 - ISO 639-3: SIL International, USA (former Ethnologue)
- Some statements taken from Anthony Aristar (LinguistList)
 - work on Ethnologue for 50 years
 - goal: coverage of languages still in use (incl. Sanskrit)
 - EMELD added ancient languages
 - well developed/emerged rule systems

Not yet the end or?



- let's cite Anthony again
 - not always obvious what the evidence is/was
 - some languages should not be there – others should be there
 - dialects versus languages issues
 - wrong information associated (location, population, etc)
- nevertheless great resource and improvement
- however political dispute
 - ISO needs to be based on broad agreements
 - ISO 639-3 as it is, is not accepted in large areas of the world
- therefore
 - ISO started a whole new initiative to overcome limitations and to adhere to ISO principles

ISO still busy



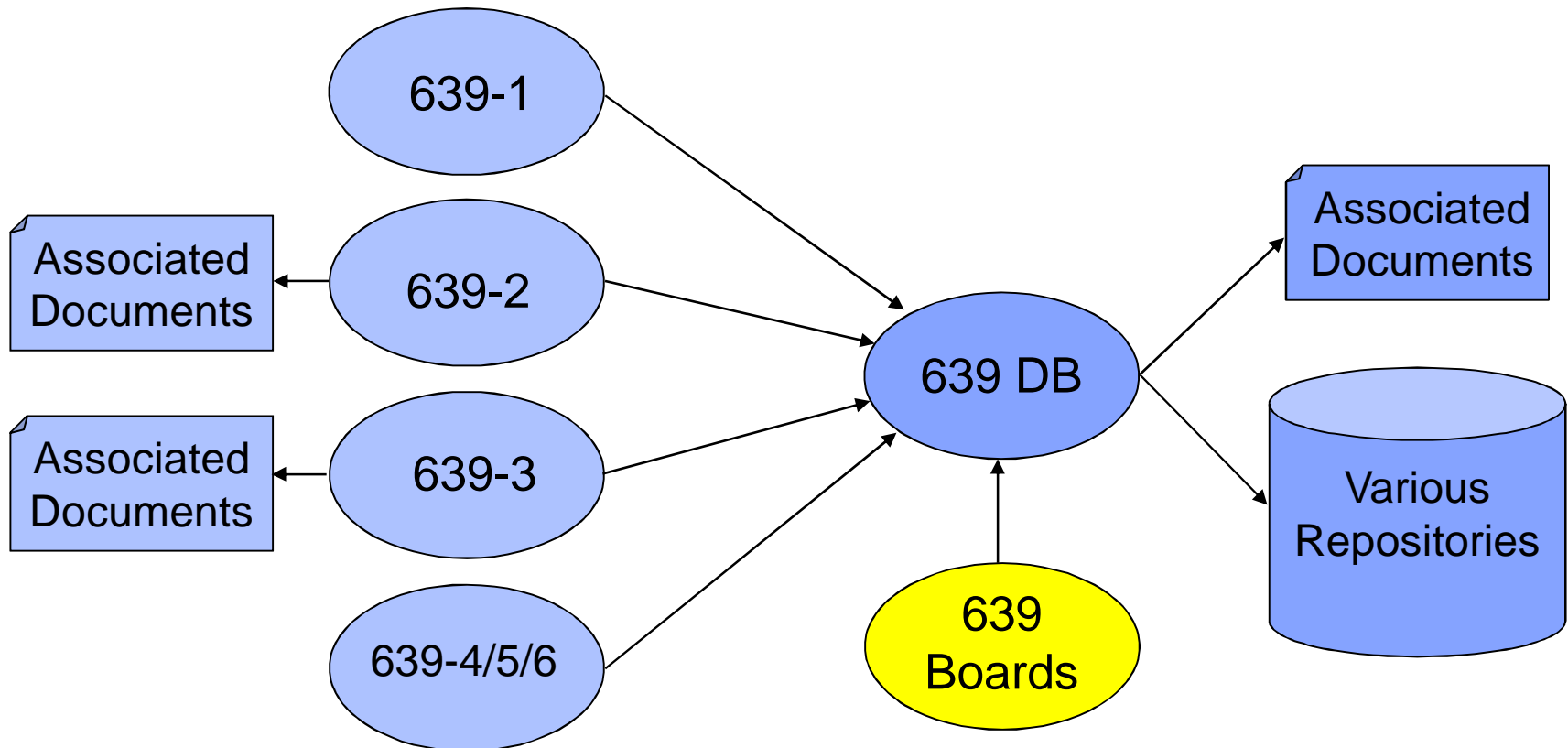
- ISO intends to
 - define general criteria for language codes
 - define language families
 - integrate dialects etc
 - bring all language code tables under one schema and process
 - maintain one authorized database for all codes that also implements the process
 - organize a widely respected governance structure based on informed communities
- thus the ISO approach can be compared with the UNICODE one
- ISO standards are the principles behind codes, the database content is not, i.e. the process can be faster
- has something already been done?

New ISO standards



- ISO 639-4 *Codes for the representation of names of languages* -
- *Part 4: [Implementation guidelines and general principles for language coding](#)*
 - state: FDIS
- ISO 639-5 *Codes for the representation of names of languages* -
- *Part 5: [Alpha-3 code for language families and groups](#)*
 - state: version 2008
 - RA: LoC
- ISO 639-6 *Codes for the representation of names of languages* -
- *Part 6: [Alpha-4 representation for comprehensive coverage of language variation](#)*
 - state: FDIS
 - RA: GeoLang
- [RA to be changed soon](#)

Integration intention



finally one language code table with sub-sections and a unified board structure based on ISO principles

Concept Database



ISO Concept Database

guest@iso.org
2009-09-26

Process Search Settings Help Logout

Terms Graphical symbols Codes All

Search by Term criteria

English

Term: management Any word

Definition: _____

Term Type:

Normative status:

Search by Standard / Project criteria

Search

Reset

Results per page 10

306 items found

1 - 10

Page 1 of 31. Go to: 1

View selected

Identifier	Stage	Name	Reference	Entry No.	Definition
<input type="checkbox"/> CDB-0000531-001	90.60	(N)-layer management protocol	ISO/IEC 10040:1998	3.6.24	An (N)-layer protocol for the exchange of (N)-layer management information supported solely by protocols of the layers (N-1) and below.
<input type="checkbox"/> CDB-0000532-001	90.60	(systems management) operation	ISO/IEC 10040:1998	3.6.25	An operation on a managed object to effect systems management.
<input type="checkbox"/> CDB-00002742-001	90.92	operational binding management	ISO/IEC 9594-2:2005	25.1.7	The process of establishing, terminating or modifying an instance of an operational binding. This management may be achieved via information exchanges defined by Directory Specifications, via exchanges defined in other Specifications, or by other means.
<input type="checkbox"/> CDB-00002836-001	90.93	project management	ISO 10006:2003	3.6	planning, organizing, monitoring, controlling and reporting of all aspects of a project (3.5) and the motivation of all those involved in it to achieve the project objectives
<input type="checkbox"/> CDB-00002837-001	90.93	project management plan	ISO 10006:2003	3.7	document specifying what is necessary to meet the objective(s) of the project (3.5)
<input type="checkbox"/> CDB-00002845-001	90.93	configuration management	ISO 10007:2003	3.6	coordinated activities to direct and control configuration
<input type="checkbox"/> CDB-00002872-001	90.60	measurement management system	ISO 10012:2003	3.1	set of interrelated or interacting elements necessary to achieve metrological confirmation and continual control of measurement processes
<input type="checkbox"/> CDB-00003498-001	90.60	management domain	ISO/IEC 10040:1998	3.6.11	A specification of a collection of managed objects of interest.

Current comparison work



Microsoft Access - [Main table : Table]

File Edit View Insert Format Records Tools Window Help Adobe PDF

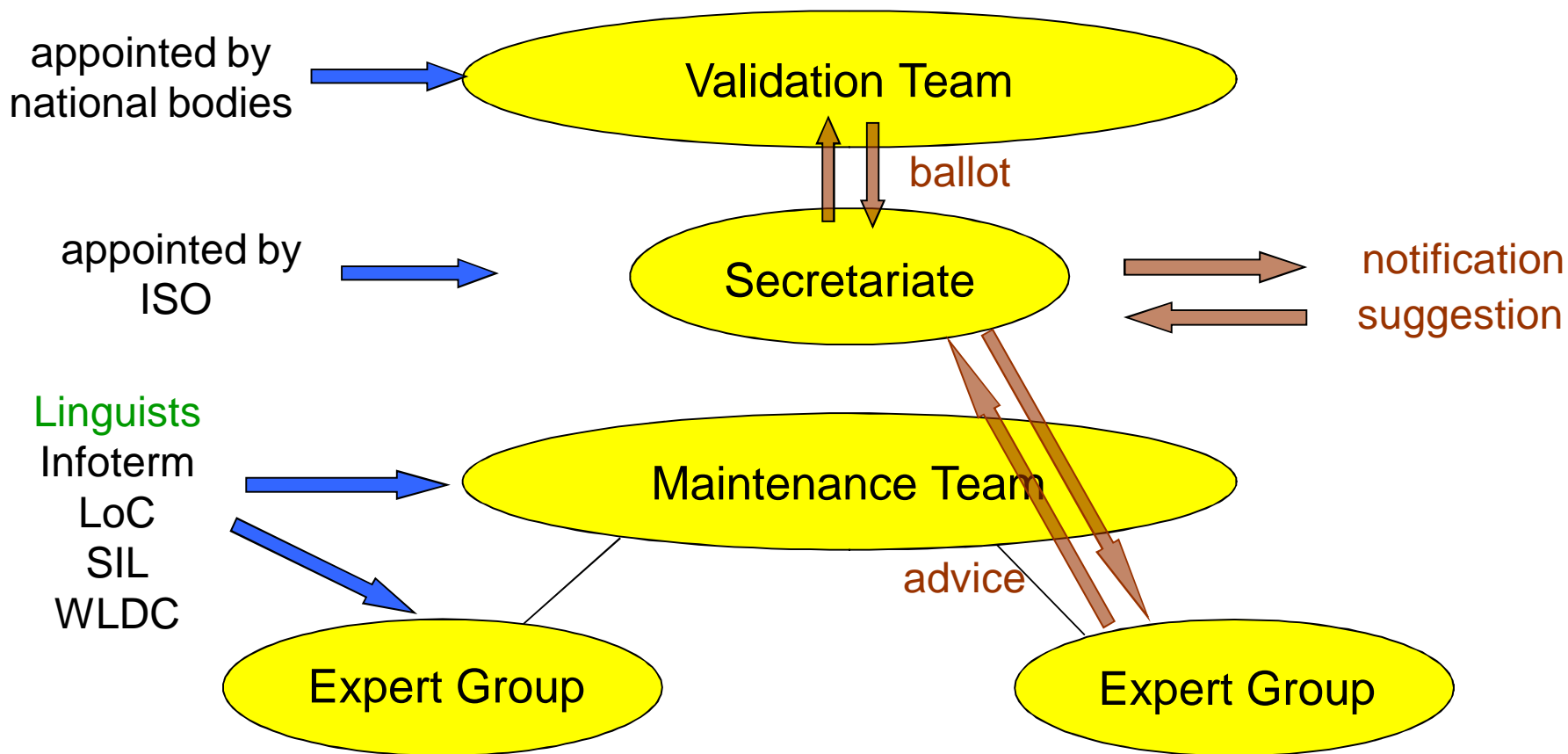
Type a question for help

ID	alpha3T	alpha3B	diff3T3B	alpha2	status 639-1	status 639-2	part	scope	type	English	English note	French	French n
2047	fos	fos					3	I	E	Siraya			
8572	fox	fox					5	G	L	Formosan languages		formosanes, langues	
2048	fpe	fpe					3	I	L	Fernando Po Creole English			
2049	fqs	fqs					3	I	L	Fas			
2050	fra	fre	+	fr			1 2 3	I	L	French		français	
2051	frc	frc					3	I	L	Cajun French			
2052	frd	frd					3	I	L	Fordata			
2053	frk	frk					3	I	E	Frankish			
2054	frm	frm					2 3	I	H	Middle French	ca. 1400-1600 CE	français moyen	ca 1400-16
2055	fro	fro					2 3	I	H	Old French	842-ca. 1400 CE	français ancien	842-ca 1400

Record: 1 of 7803

Datasheet View

Future decision structure



Nominations and decision process

Message



Please, start using ISO 639-3 now!!!!

Include it as pick list in your tools!!!!

IMDI Editor as an example.

Is picking up the ISO 639-3 and offers it as vocabulary.

IMDI Editor as one example



The screenshot shows the IMDI Metadata Editor application window. The title bar reads 'IMDI Metadata Editor'. The menu bar includes 'File', 'View', 'Options', and 'Help'. Below the menu bar is a toolbar with icons for saving, opening, and help. The main interface is divided into several sections:

- Left Panel:** A tree view showing the metadata structure. The top level is 'Standard (Session) Resource B'. Underneath, there are 'Project', 'Content', and 'References'. The 'Content' folder is expanded, showing 'Unspecified', 'Actors', 'Resources', and 'References'. Below this is a 'Local Repository' section with icons for 'Projects', 'Content', 'Actors', 'Languages', and 'Access'.
- Editor Area:** Contains a 'Language' section with several fields:
 - Name:** Unspecified
 - ID:** A list of language IDs: Enets, Forest; English; Ewe; Fanagalo; Farsi, Western; Finnish; French; Fulfulde Fuuta Jalon; Unspecified. 'Finnish' is currently selected.
 - Dominant Language:** A dropdown menu with 'CV' next to it.
 - Source Language:** A dropdown menu with 'CV' next to it.
 - Target Language:** A dropdown menu with 'ISO639-3:fin' selected and 'CV' next to it.
- Bottom Section:** Two tabs are visible: 'Resource References' and 'Descriptions'. The 'Descriptions' tab is active, showing a large empty text area.

End



Thanks for your attention.

- ISO 639 Registration Authorities' Joint Advisory Committee (ISO 639 RAs-JAC)
 - Representatives of ISO/TC 37 and ISO/TC 46
 - Votes on ISO 639-1, -2, and -5 items
- ISO 639-3 has separate procedures
 - Before an item can be included in ISO 639-2 (or ISO 639-1), it has to pass ISO 639-3 procedures

Will contain:

- ✓ Terms and definitions (of all ISO standards)
- ✓ Symbols (of a large number of ISO standards)
- ✓ Codes (language, country, currency, script)

Timeframe:

- ✓ October 2009: publicly available read-only version
- ✓ January 2010: decentralized data maintenance functions

- ISO 639 (-1, -2, -3, -5) data have been submitted to ISO Central Secretariat for inclusion in the first version of the CDB. This is a preliminary version, including:
 - all alpha-2 and alpha-3 identifiers
 - English names; “alternative names” only to a limited degree
 - French names for -1, -2, and -5 items
 - part to which item belongs (-1, -2, -3, -5)
 - some notes

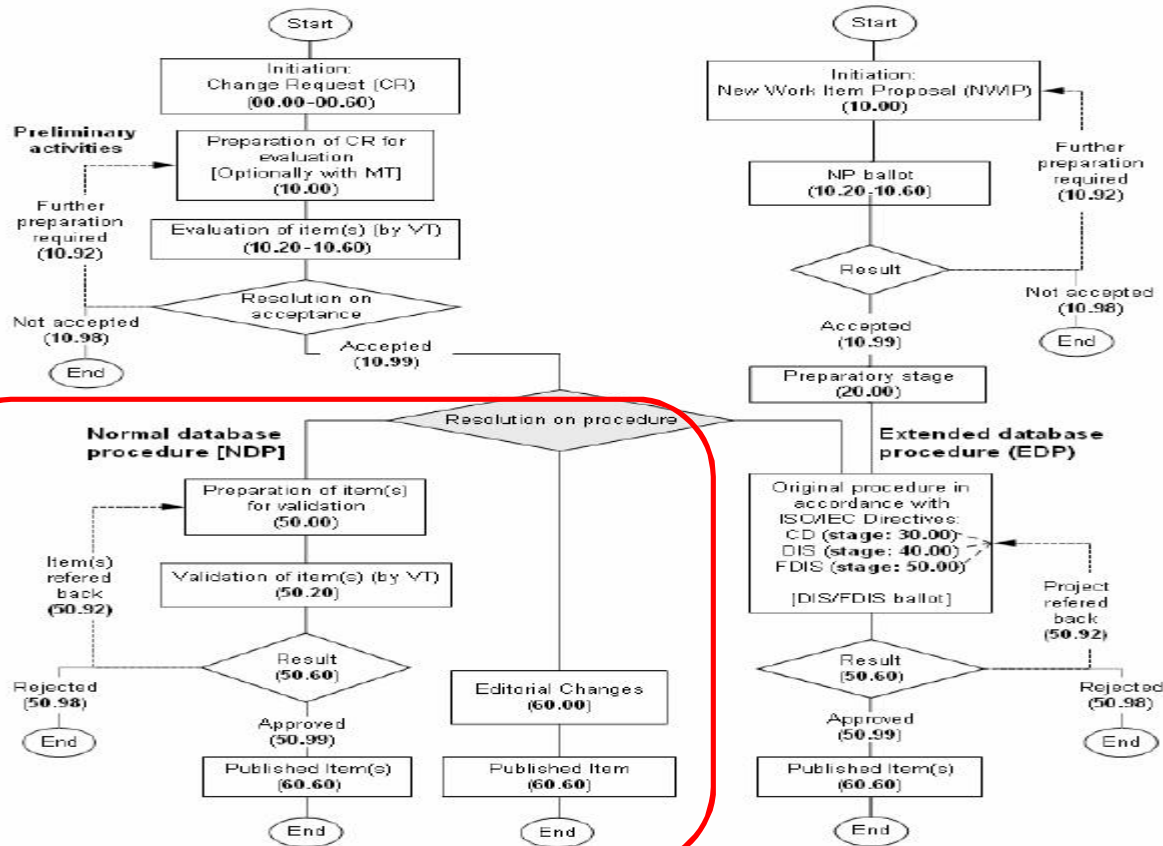
ISO standards as databases

- Annex ST of the ISO Supplement to the ISO/IEC Directives
 - from June 2008
 - <http://isotc.iso.org/livelink/livelink?func=ll&objId=4230452> (click “ISO Supplement – Annex ST ...”)
 - Procedures for maintenance of ISO standards in databases (ISO CDB or external databases)
 - Items in database may be process individually through formal stages (WD, CD, DIS, FDIS, IS)
 - Annex ST has some governance-related rules:
 - Maintenance Team
 - Validation Team

ISO standards as databases



Håvard Hjulstad,
2009-09



This part is the most relevant

CR: Change Request
 VT: Validation Team
 MT: Maintenance Team
 NDP: Normal Database Procedure
 EDP: Extended Database Procedure

NP: New Work Item Proposal
 CD: Committee Draft
 DIS: Draft International Standard
 FDIS: Final Draft International Standard

Figure ST.1 – Process for the development of new and the maintenance of existing standards

ISO 639 – new governance



- All “old” groups and structures are formally dissolved, including
 - Registration Authorities
 - Joint Advisory Committee
 - JAC Secretariat
- New bodies are established:
 - Validation Team (VT)
 - Maintenance Team (MT)
 - ISO 639 Secretariat

ISO 639 – new governance



- Validation Team
 - Individual members appointed by committee National Member Bodies
 - The VT makes the formal decisions
- Maintenance Team
 - A “large” body of experts that may be called upon (by the VT or the Secretariat)
 - Includes all current JAC members, and many more
 - Exact membership needs to be decided
 - Input?
- ISO 639 Secretariat
 - “permanent”, “outsourced”

- The standard itself will be restructured:
 - All code elements in one table (in the ISO CDB)
 - Current ISO 639-4 will be expanded and be the only ISO 639 linear document
 - Current ISO 639-2 will be a “defined subset” (maintained in the database itself)
 - Additional information will be included in the CDB:
 - Linguistics-based navigation in the database (e.g. by linguistic hierarchy)
 - Links to external information resources (Ethnologue, LinguistList, etc.)

ISO 639 – timeline



- Not yet decided in detail, except “as soon as possible”
- So; as soon as
 - ISO CDB is fully operational (early 2010)
 - ISO 639 Validation Team is in place (a few months)
 - ISO 639 Maintenance Team is in place
 - ISO 639 Secretariat is in place