Languages

Als u dit begrijpt, dan is Nederlands niet vreemd voor u ...

... if not, it is Dutch to you

What is a language?

- Languages are what we use to communicate
 - when it is a shared language
 - when the language is supported
- Many "rules" make up a language
 - the terminology
 - the grammar
 - a (formal) orthography
 - the pronunciation ...
- A "linguistic entity" is recognised
 - informally when people understand it
 - formally when it is ratified by a standards body

Standard Languages

- With formal recognition for a language
 - it can be referred to in an unambiguous way
 - nl is both Dutch, Nederlands, ھولندي , オランダ語
 - when it is part of a document's meta-data, you can change the behaviour for the document
 - spell checking for instance
 - data mining
- ISO 639 deals with linguistic entities
 - there are several versions ratified
 - ratified versions do not deal with language families, dialects or orthographies
 - this will be different in future versions

Languages II

- IANA language subtags are used to indicate languages on the Internet
 - ratified standards are part of its make up
 - ISO 639, ISO 15924, ISO 3166
 - aims to provide stable references
 - is more for computers then for humans to use
- WMF uses ISO 639 codes for its project names and IANA tags internally in the meta data

WMF Language Committee

- The acceptance of new language was always a fight
 - some were politically "unacceptable"
 - some were not what they advertised to be
 - some were not even formally recognised
 - some do not have a community / activity
- Incubator
 - first create some content and user interface
 - allows for the verification of the language
 - allows for the process of first getting a code
 - indicates the existence of a community

WMF Language Committee II

- It deals primarily with new languages in WMF
 - languages need to be recognised
 - languages need to be what they advertise
 - is intentionally not democratic no votes
 - the rules are fuzzy
- It deals on request with an existing language
 - it would prefer a "global arbitration committee"
 - would prefer projects to revert to Incubator and not be blocked

Sign languages

- We would like to welcome projects in sign languages
 - the script, SignWriting, is not included in UTF-8
 - the positioning is top down
 - only specialised software supports it as a consequence
- To remedy this, a big investment is needed
 - to include SignWriting in UTF-8
 - to have browsers support top to bottom content
 - to have MediaWiki support SignWriting
 - to popularise SignWriting for the sign languages where it needs more exposure

Fund raising

- Organisations create WMF content
 - would like to participate in WMF projects
 - already do it on an ad hoc basis
 - there is no partner framework in the WMF
- Organisations have content that need a safe harbour
 - WMF is not always the right partner, but it might have partners that would do better
 - data often needs curation to be optimally usable
- Sharing objectives means sharing resources
 - technology, content, effort and money

Fund raising II

- Collaboration / shared objectives are key
 - combining what partners have together
 - technology, data, community
- Single issue fund raising
 - donations, putting your money where your mouth is
 - Kamusi available under a Free license
- Fund raising together is more successful
 - it indicates a bigger commitment
 - it strengthens our position to build a Free Culture
- Present of Open Progress
 - OmegaT with Wiki read functionality for now

Assisted Reading & Semantic Support

- People read articles and want to understand more about the concepts and relations
 - Assisted Reading defines and translated
 - PositanoNews adopted it first
 - to support the Neapolitan language
 - further development by University Bamberg
 - Semantic Support provides additional services
 - analyses text for its concept cloud
 - we want to bring this to Wikipedia in any language



