

Diversity and Wikidata



Leveraging data as information
for all languages, cultures and interests



Wikidata

- Stores semantic data with multi-language support
- Had a flying start by servicing the interwiki process
 - There are pros and cons for this approach
 - The link to WP allows bots to enrich WD
 - Most statements are derived from WP, any WP
- Terminology:
 - Item
 - Label
 - Link

Issues

- Links without labels
- Items not referring to “your” language (no labels)
- WP articles are not what they seem; lists for instance
- No statements for an item
- Statements with no label in “your” languages
- However, these issues are better than not having data to work with

Using WD for search



- WD needs to support the “fat head” of search in each language
- Not only for WD but also WP and Commons
- We do not know what people look for
- We do not know what labels WD does not serve




User participation

- Build upon the data that we have
- Indicate what is missing in “their” language
- Allow people to translate for the languages they know
- Compile a “concept cloud” for each subject
 - Ask help with missing labels
 - Suggest their use in WP articles
 - Suggest their use in WD
 - Present a basic info-box

Specific subjects



- Consider the use of ontologies
 - WP categories ARE an ontology in their own right
 - Linking red links to the “concept cloud”
 - Make use of global (ie WD) and external ontologies
 - Make the “concept cloud” part of the watchlist
 - Consider that another WP may be more advanced
 - Labels may be lacking in “your” language
- 

Some statistics

- Most items do not have statements
 - This affects precision, but not the trends
- Many items have links and no labels
 - They cannot be found
- Basic WD gender ratio is for all and no Wikipedias
 - Please apply your research data !!
 - Learn how it affects ratios in other languages
- PS What sex is an eunuch?

More statistics

- WMF statistics – not really relevant
- Magnus's and his existing query functionality
 - Been adding sex info and it affects the numbers
- We could do with statistics on failed searches
 - So far it has no priority

Visualisation

- It helps people understand what the data is about
- It motivates to add labels and statements
- When served to “WD/WP readers” they get structured data and it may motivate to write WP articles

Your agenda

- Gender ratios are not that relevant; quality information is.
 - Make sure your category of subjects is full of statements
 - Do the subjects you like best first
 - Make sure that you enjoy yourself
 - Soccer, hockey, swimming, gymnastics
- IGNORANCE is imho what we should concentrate on
 - Gapminder type data and visualisation is what we need



Questions ?

User:GerardM

UltimateGerardM.blogspot.com

