

Wikidata Tech

Wikimania 2012

Wikidata Tech Challenges and Solutions

Daniel Kinzler
Wikimedia Deutschland

Wikidata Mission

- Provide a database of the world's knowledge that anyone can edit
- Collect references and quotes for millions of data items
- Engage a sustainable community that collects data from everywhere in a machine-readable way
- Increase the quality and lower the maintenance costs of Wikipedia and related projects
- Deliver software and community best practices enabling others to engage in projects of data collection and provisioning

Multilingual Content

- Canonical, numeric ID used as Title internally
- Language-Specific label used in the UI
- Aliases
- Descriptions

Labels and descriptions can't be used to address pages.

Site Links

- Interlanguage links from the Wikidata perspective
- Are unique identifiers of the data item:
- A wikipedia page can only be about one Item (topic)
- Only one Wikipedia page per language can be about a specific topic.

Multilingual Content

Questions?

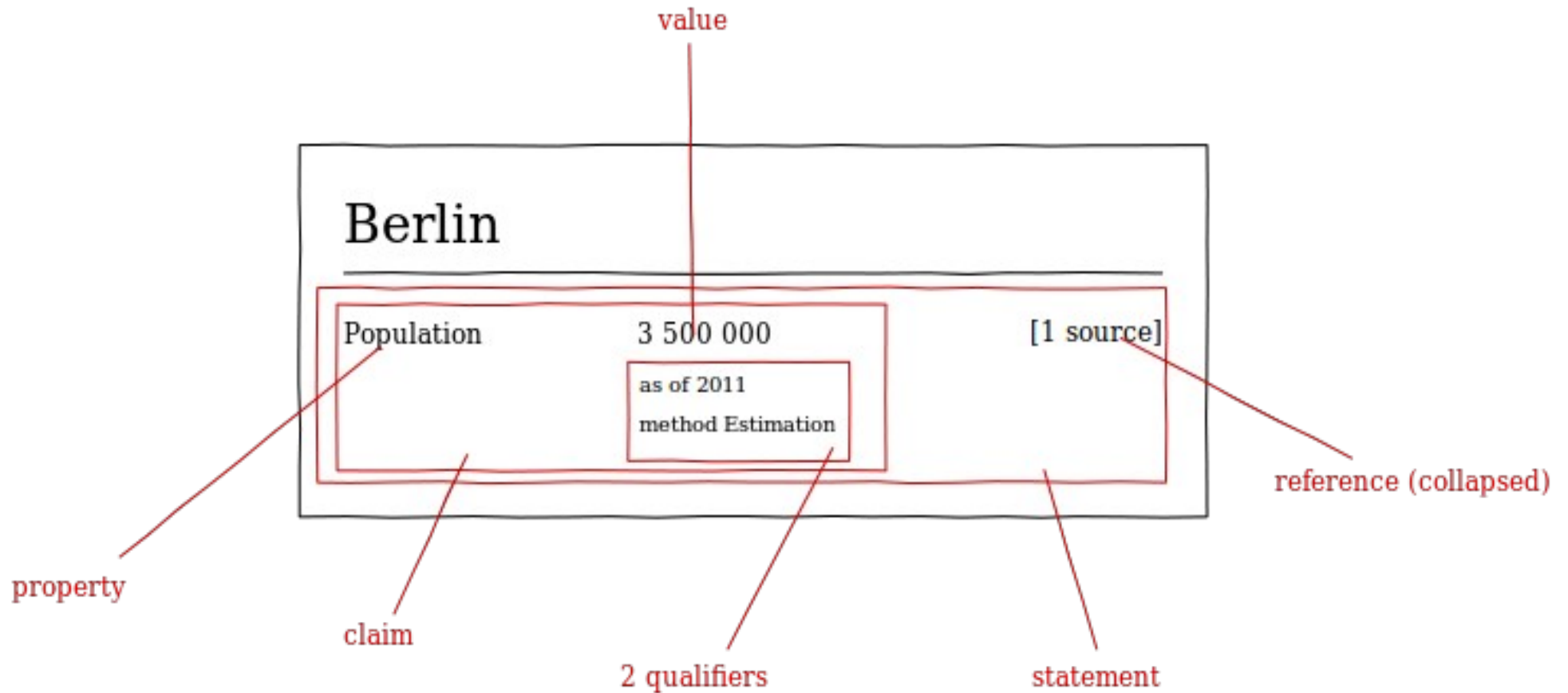
Structured Data

- JSON structure stored on wiki page
- *Instead* of wikitext

JSON

```
{
  entity:"q8"
  label:{
    "ar":"\u0647\u064a\u0644\u064a\u0648\u0645",
    "de":"Helium",
    "en":"Helium",
    "it":"Elio",
  },
  description:{
    "en":"A chemical element",
    "de":"Ein chemisches Element.",
  },
  aliases:{
    "en":["He"],
    "de":["He"],
  },
  links:{
    "arwiki":"\u0647\u064a\u0644\u064a\u0648\u0645",
    "dewiki":"Helium",
    "enwiki":"Helium",
    "itwiki":"Elio",
  },
}
```

Diversity and Disagreement



ContentHandler

- Introducing the ContentHandler facility
- Handlers for different *content models*.

- Serialization
- Rendering
- Diffs, etc

ContentHandler

Questions?

Inclusion in Wikipedia

- Parser functions for accessing property values.
- To be used in infobox templates, etc.
- Per default, property for this page's "own" data item
- Options for formatting, collation of multiple values, etc

Cacheing

- Local cache of data items
- Updated by a background process polling an update feed
- Currently MySQL
- future may be PubHub aka PuSH.

Inclusion on Wikipedia

Questions?

Watching

- Data changes written to recentchanges, shown in watchlist
- Based on cache update mechanism

History

- "Virtual" entries in the page history
- Diff view shows data when appropriate

Not fully spec'ed yet. Needs data revision ID in page history.

Inclusion on Wikipedia

Final questions?

Wikidata

Thank You!

<https://meta.wikimedia.org/wiki/Wikidata>

<http://wikidata-test-repo.wikimedia.de>

<https://meta.wikimedia.org/wiki/Wikidata/Development/Setup>

<https://lists.wikimedia.org/mailman/listinfo/wikidata-l>