Cargo and the future of Semantic MediaWiki

Yaron Koren
SMWCon Spring 2015
St. Louis, Missouri
May 7, 2015
The original idea

SMW started as a Semantic Web tool.

But SMW + templates = structured data = more like a database than like the Semantic Web.

How can we make the system simpler, given that?
- An extension I released in January 2015
- Stores data in “true” DB tables, instead of SMW's DB tables that hold triples
- One table per template, more or less
- Syntax is a wrapper around SQL, instead of a custom query language
Table data = triples

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
</table>
In a page called “St. Louis”:

```markdown
{{City|Country=USA|Population=12000000}}
{{Mayor|Name=William Carr Lane|Start date=...}}
{{Mayor|Name=Daniel Page|Start date=...}}
```

This could store one row in a table called “Cities”, and two rows in a table called “Mayors”. 
Cargo's main parser functions

• `#cargo_declare` – instead of property pages
• `#cargo_store` – instead of property tags
  (\[
  \[A::B]\]\]), #set, #subobject and #set_internal
• `#cargo_query` – instead of #ask
#cargo_declare and #cargo_store

In the “Mayor” template's <noinclude> section:

```
{{#cargo_declare:_table=Mayors|Name=Text|Start_date=Date|End_date=Date|Important_bills=List (, ) of Page}}
```

In the same template's <includeonly> section:

```
{{#cargo_store:_table=Mayors|Name=|Start_date=|End_date=|Important_bills=}}
```

...and you're done!
Cargo DB storage

- A Cargo table called “People” is stored as a DB table called “cargo__People”
- This creates a “sandbox”, so Cargo can't query non-Cargo DB tables
- Cargo DB tables can also be stored in a separate database
• One example:

```json
{{#cargo_query:
  tables=Mayors
  |fields=_pageName=Mayor, Start_date, End_date
  |format=ul
}}
```
#cargo_query - syntax

- **tables** - the Cargo tables to query on
- **join on** - JOIN conditions, if more than one table
- **fields** - the table fields to display
- **where** - the WHERE clause
- **group by** - the GROUP BY clause
- **order by** - the ORDER BY clause
- **format** - the display format
- **plus** limit, intro, outro, default
Cargo display formats

- list
- ul
- ol
- category
- template
- embedded
- outline
- tree
- table
- dynamic table
- gallery
- calendar
- timeline
- googlemaps
- openlayers
- csv, json, excel
Cargo's other parser functions

- `#cargo_attach` – instead of property pages
- `#cargo_compound_query` – instead of `#compound_query` (from Semantic Compound Queries)
- `#cargo_display_map` – instead of `#display_map`, `#display_point` (from Maps)
- `#recurring_event` – instead of `#set_recurring_event` (from SMW)
Extensions that Cargo tries to replace

User-facing:
- Semantic MediaWiki
- Semantic Result Formats
- Semantic Maps
- Maps
- Semantic Drilldown
- Semantic Compound Queries
- Semantic Internal Objects

“Library” extensions:
- Validator
- DataValues
- DataValues Common
- DataValues Geo
- DataValues Interfaces
- DataValues Validators
Extensions used by both SMW and Cargo

- Semantic Forms, Semantic Forms Inputs, etc.
- External Data
- Page Schemas
- Data Transfer, Replace Text, Admin Links, Header Tabs, Widgets, ...etc.
Extensions with no equivalent in the Cargo system

- Semantic Watchlist
- Semantic Glossary
- Semantic Title
- ...etc.
Cargo: simpler to install

- One extension replaces around 15
- Around 10% of the code size (not counting 3rd party JS libraries)
- No need for: Composer, Semantic Bundle
Cargo: simpler to create data structures

- No property pages – there can be thousands of these!
- Drill-down available automatically – no need to define filters
Cargo: SQL-like querying

- SQL is more widely-known than #ask syntax
- SQL is more powerful than #ask syntax
Queries possible in Cargo but not SMW

- Display of joined fields (like “?A.B”)
- A table showing the distribution of different values for a field/property
  - Distribution is possible for charts only
- Showing only pages that contain a blank value for a field
- Text-matching queries, like getting all values that start with a certain substring
Cargo: simple DB storage

- Much simpler for other extensions, and outside software, to query the data – they can just use SQL! (No need to even talk to Cargo.)
- “Semantic MediaWiki turns your wiki into a database; Cargo turns your wiki into an SQL database.”
Disadvantages of Cargo

1) No support for RDF export (yet?)
2) No equivalent yet for Semantic Watchlist, etc.
3) Free-form tagging not allowed
   • ...though can be simulated with a template
4) No querying of categories
5) No automatic metadata storage (page creation date, etc.)
6) Less mature software
The Semantic Web has lost some “buzz” in the last five years.

Is it still important? I don't know.

Regardless, the vast majority of SMW users do not use it!
What does Cargo mean for SMW, SMWCon, etc.?

I plan to keep supporting Semantic MediaWiki to the extent that I did before.

Beyond that – let's talk about it.