Introduction to Enterprise MediaWiki

Yaron Koren
EMWCon Spring 2019
April 3, 2019
Daly City, CA

My companies:
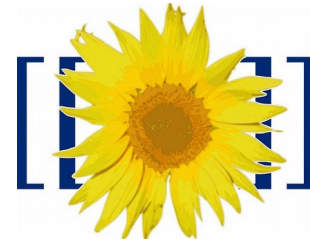

Also check out my book (Working with MediaWiki) and my podcast (Between the Brackets)!

2001: Wikipedia created

2002/2003: MediaWiki created

...for use on Wikipedia.

Key distinguishing features of MediaWiki (in my opinion):

- templates

- parser functions

- namespaces

2004: Wikicities launches

Later renamed to "Wikia", then "Fandom".

Soon becomes an extremely popular, MediaWiki-based wiki farm.

2005: Semantic MediaWiki (SMW) released
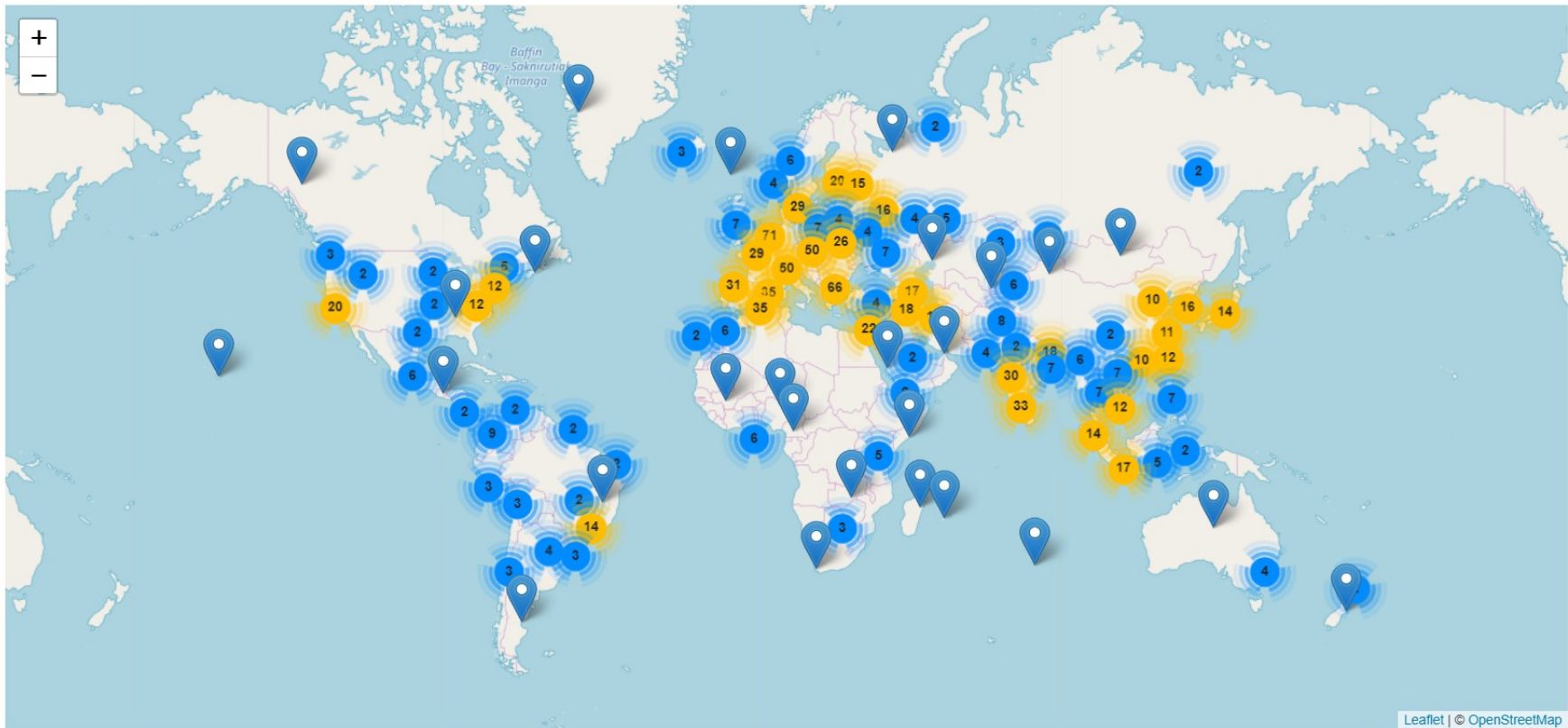
SMW syntax

On a page named "California":

```
[[Has capital::Sacramento]]
```

On the page "Property:Has capital":

```
[[Has type::String]]
```

SMW has cool visualizations – like maps!

(Source: translatewiki.net)

…and calendars!

(source: iccl.inf.tu-dresden.de)

SMW is part of the greater Semantic Web initiative.

Other "semantic wiki" engines created at the time:

AceWiki, BOWiki, Braindump, COW, IkeWiki, Kaokolu, KendraBase, KiWi, KnowWE, Makna, OntoWiki, Paux, PlatypusWiki, Rhizome, SemperWiki, SweetWiki, SWiM, SWOOKI, TaOPis, Wikidsmart (for Confluence), WikSAR

(All of these became unmaintained by 2012 or so.)

2007: Semantic Forms released

- The first "Semantic ..." extension

- I'm the main author

- Initially: let users define forms for editing pages that call SMW-based infobox templates.

- Later renamed to "Page Forms"

Benefits of forms

Editing without a form:

```
{{Film
|Year=
|Genre=
|Director=
}}
```

Editing with a form:

Year: [   ]

Genre: [▼]

Director: [   ]

Page Forms syntax

On the page "Form:Film":

```
{{{for template|Film}}}
Year: {{{field|Year}}}
Genre: {{{field|Genre}}}
Director: {{{field|Director}}}
{{{end template}}}
```
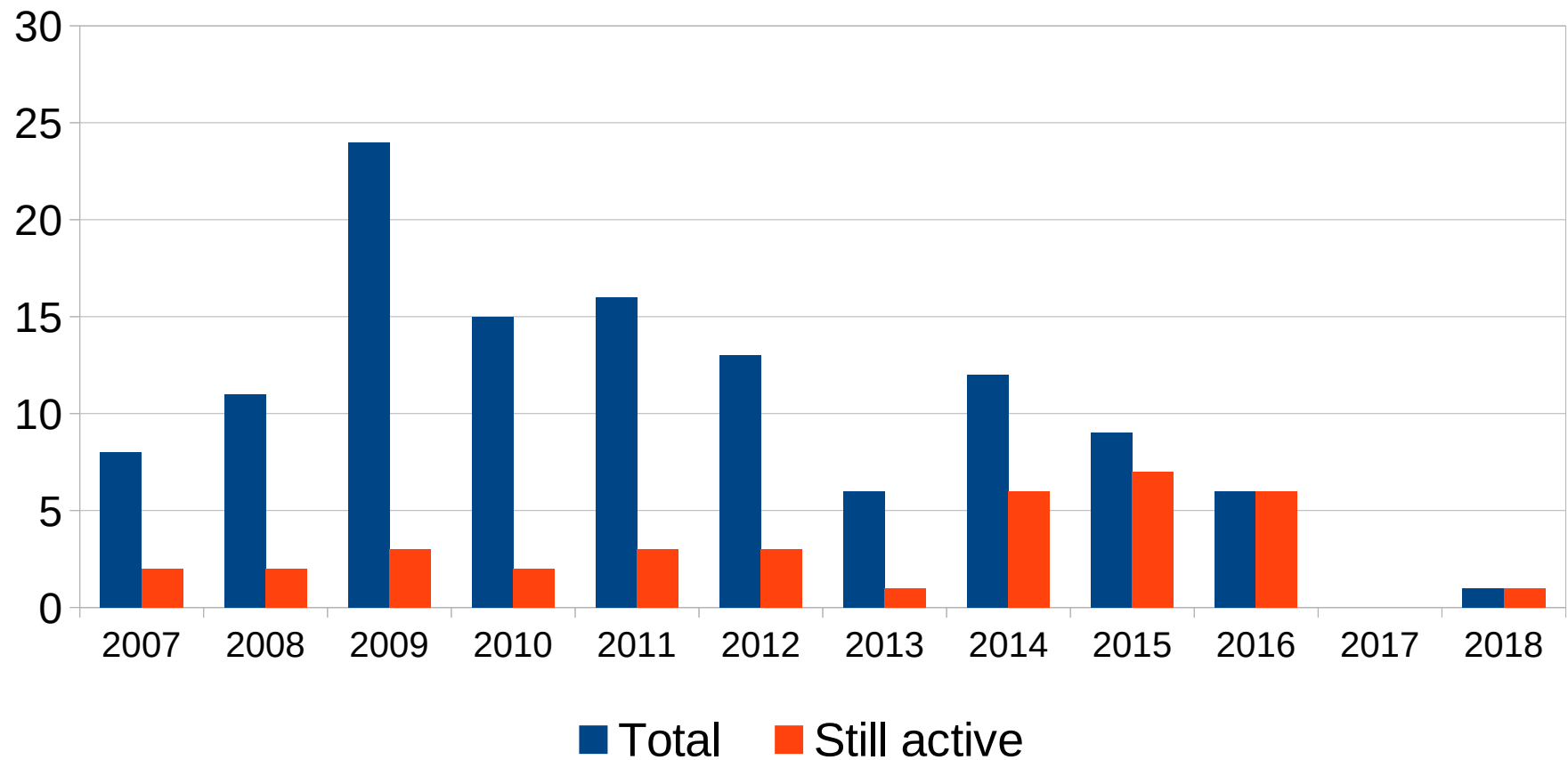
Then, there is **an explosion of SMW-based** extensions, some form-based and some not.

~ 120 SMW-based extensions created, so far

SMW-based extensions created per year

All SMW-based extensions

2007: *Halo*, *MOCA*, Page Forms, *Question API*, Semantic Drilldown, *Semantic Google Maps*, *Semantic Layers*, *SMW InlineQueryParserFunction* **2008:** *Evaluation WikiFlow*, *MetaVidWiki*, *Semantic Calendar*, Semantic Compound Queries, *Semantic Gallery*, *Semantic graph extensions*, Semantic Result Formats, *Semantic Tasks*, *SemanticMappoint*, *SemanticQueryFormTool*, *Tetherless Map* **2009:** *Access Control List*, *Collaboration extension*, *Data Import*, *DSMW*, *Enhanced Retrieval*, External Data, *FolkTagCloud*, *Freebase*, Maps, *RDFa*, *Rich Media*, *Rule Knowledge*, *Semantic Forms Inputs*, *Semantic Gardening*, *Semantic History*, Semantic Internal Objects, *Semantic Maps*, *Semantic Multi Edit Extension*, *Semantic Notifications*, *Semantic NotifyMe*, *Semantic Signup*, *TreeView*, *Triple Store Connector*, *WYSIWYG* **2010:** *CreateVirtualPages*, *ITSM-Wiki*, LinkedWiki, RDFIO, *Semantic Incoming Properties*, *Semantic Need*, *Semantic Project Management*, *SemanticAnnotations*, *SemanticCSV*, *SemanticQueryRDFS++*, *SemanticUpdateOnPurge*, *ShowIncomingRelations*, *SMW OntologyEditor*, *SMWAskAPI*, *SparqlExtension* **2011:** *AskQ*, *Frequent Pattern Tag Cloud*, *ImageRefer*, Semantic Extra Special Properties, Semantic Glossary, *Semantic Image Input*, *Semantic Social Profile*, Semantic Watchlist, *Semantic Web Browser*, *SemanticACL*, *SemanticFormsToolbar*, *SemanticPropertyTools*, *SemanticQueryComposer*, *SemanticSVGPrinter*, *SMWAutoRefresh*, *SolrStore* **2012:** *AkAPI*, *JSON*, *LightweightRDFa*, *OfflineImportLexicon*, *Semantic Access Control*, *Semantic Expressiveness*, Semantic Forms Select, *Semantic Genealogy*, *Semantic Image Annotator*, *Semantic Page Maker*, Semantic Page Series, *Semantic Tagbox*, Semantic Title **2013:** *ChapTimeline*, *Semantic Sifter*, *SemanticComments*, *SemanticMailMerge*, *SemanticMustacheFormat*, Title Icon **2014:** *AnalysisTool*, *AssembleFormLink*, HierarchyBuilder, PhpTags SMW, Semantic Dependency, Semantic Text Annotator, SemanticDummyEditor, *SemanticHighcharts*, Surl, *TaskManagement*, *VIKI*, *VikiSemanticTitle* **2015:** AutoFillFormField, Semantic Breadcrumb Links, Semantic Cite, Semantic Interlanguage Links, Semantic Meta Tags, *SMW Ontologies*, Technology Portfolio, Technology Radar, *VikiTitleIcon* **2016:** Cognitive Process Designer, SemanticDependencyUpdater, Semantic MediaWiki Graph, Semantic OLAP, Semantic Scribunto, SMWConnector **2018:** Semantic Organization

Extension names in italics are no longer maintained.

2009: First SMWCon

In Cambridge, MA (MIT).

2011: VisualEditor released

2012: Scribunto extension released

Allows for defining "modules" that do text processing (and other stuff) via a scripting language, Lua.

Why Scribunto is great: a case study

Displaying chess boards on Wikipedia.

The "Chess diagram" template on the English-language Wikipedia, until 2013:

```
<div class="{{{1}}}" style="clear: right; width: 260px; text-align:center">{{{2}}}
<div style="border:1px solid #b0b0b0; background:#f9f9f9; padding-top:6px">
<table cellpadding="0" cellspacing="0" style="border: 1px solid #b0b0b0; margin:auto">
<!--
-->{{#switch:{{{letters|both}}}|top|both={{Chess diagram/letters|numbers=
{{{numbers|both}}}}}}}{{Chess diagram/row|row=8|size={{{size|26px}}}|{{{3|}}}|
{{{4|}}}|{{{5|}}}|{{{6|}}}|{{{7|}}}|{{{8|}}}|{{{9|}}}|{{{10|}}}}}<!--
-->{{Chess diagram/row|row=7|size={{{size|26px}}}|numbers={{{numbers|both}}}|
{{{11|}}}|{{{12|}}}|{{{13|}}}|{{{14|}}}|{{{15|}}}|{{{16|}}}|{{{17|}}}|{{{18|}}}}}<!--
-->{{Chess diagram/row|row=6|size={{{size|26px}}}|numbers={{{numbers|both}}}|
{{{19|}}}|{{{20|}}}|{{{21|}}}|{{{22|}}}|{{{23|}}}|{{{24|}}}|{{{25|}}}|{{{26|}}}}}<!--
-->{{Chess diagram/row|row=5|size={{{size|26px}}}|numbers={{{numbers|both}}}|
{{{27|}}}|{{{28|}}}|{{{29|}}}|{{{30|}}}|{{{31|}}}|{{{32|}}}|{{{33|}}}|{{{34|}}}}}<!--
-->{{Chess diagram/row|row=4|size={{{size|26px}}}|numbers={{{numbers|both}}}|
{{{35|}}}|{{{36|}}}|{{{37|}}}|{{{38|}}}|{{{39|}}}|{{{40|}}}|{{{41|}}}|{{{42|}}}}}<!--
-->{{Chess diagram/row|row=3|size={{{size|26px}}}|numbers={{{numbers|both}}}|
{{{43|}}}|{{{44|}}}|{{{45|}}}|{{{46|}}}|{{{47|}}}|{{{48|}}}|{{{49|}}}|{{{50|}}}}}<!--
```
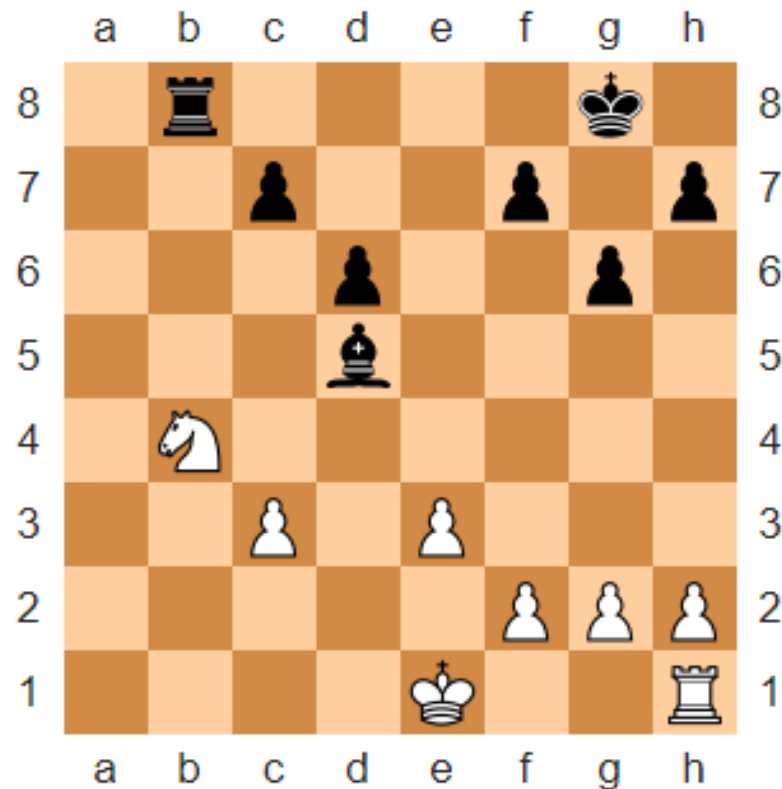
The "Chess diagram/row" template until 2013:

```
<tr style="vertical-align:middle;">
{{#switch:{{{numbers|both}}}
 |left|both=<td style="background-color:white; text-align:center">{{{row}}}</td>
}}{{#switch:{{{row}}}
|2|4|6|8=<!--
-->{{Chess diagram/square|size={{{size|}}}|color=light|piece={{{1|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=dark|piece={{{2|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=light|piece={{{3|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=dark|piece={{{4|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=light|piece={{{5|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=dark|piece={{{6|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=light|piece={{{7|}}}}}<!--
-->{{Chess diagram/square|size={{{size|}}}|color=dark|piece={{{8|}}}}}
|1|3|5|7=<!--
-->{{Chess diagram/square|size={{{size|}}}|color=dark|piece={{{1|}}}}}<!--
```

The "Chess diagram" template since 2013:

```
{{#invoke:Chessboard|board}}<noinclude>
{{documentation}}
</noinclude>
```

The "Chessboard" Lua module

```lua
local function innerboard(args, size, rev)
    local root = mw.html.create('div')
    root:addClass('chess-board')
        :css('position', 'relative')
        :wikitext(string.format( '[[File:Chessboard480.svg|%dx%dpx|link=]]', 8
* size, 8 * size ))

    for trow = 1,8 do
        local row = rev and trow or ( 9 - trow )
        for tcol = 1,8 do
            local col = rev and ( 9 - tcol ) or tcol
            local piece = args[8 * ( 8 - row ) + col + 2] or ''
            if piece:match( '%w%w' ) then
                local img = image_square(piece:match('%w%w'), row, col, size )
                root:tag('div')
                    :css('position', 'absolute')
                    :css('z-index', '3')
                    :css('top', tostring(( trow - 1 ) * size)    'px')
```

2015: Cargo released

Intended as an alternative/competitor to Semantic MediaWiki.

Stores data in *table* format, instead of *triple* format.

Cargo syntax

In the page "California":

```
{{State or province
|Capital=Sacramento
}}
```

In "Template:State or province":

```
{{#cargo_declare:_table=States_and_
provinces|Capital=String}}
{{#cargo_store:_table=States_and_
provinces|Capital={{{Capital|}}} }}
```

2016: North American SMWCon becomes EMWCon

A generally popular name change – proved the viability of the "Enterprise MediaWiki" term.

This is now the 4th EMWCon – first on the West Coast.

Would be nice to see the European events (in the fall) be renamed to EMWCon too.

2016: Semantic Forms becomes Page Forms

- No longer requires the presence of SMW.

- Works well with either SMW or Cargo (or neither of the two.)

What is Enterprise MediaWiki?

- Literally: any use of MediaWiki outside of Wikipedia and other large-scale Wikimedia projects.

- However, there's also frequently an interest in structured data and "the supremacy of the infobox".

- Plus a large group of (non-Wikimedia) extensions.

Thoughts on infoboxes

establish. Later the same year, the Franciscan missionary Francisco Palóu founded the Mission San Francisco de Asís (Mission Dolores).[16] As part of the founding, the priests claimed the land south of the mission for sixteen miles for raising crops and for fodder for cattle and sheep.[17] In 1778, the priests and soldiers marked out a trail to connect San Francisco to the rest of California.[17] At the top of Mission Hill, the priests named the gap between San Bruno Mountain and the hills on the coast *La Portezuela* ("The Little Door").[17] La Portezuela was later referred to as Daly's Hill, the Center of Daly City, and is now called Top of the Hill.[17]

During Spanish rule, the area between San Bruno Mountain and the Pacific remained uninhabited.[18] Upon independence from Spain, prominent Mexican citizens were granted land parcels to establish large ranches, three of which covered areas now in Daly City and Colma.[18] Rancho Buri Buri was granted to Jose Sanchez in 1835 and covered 14,639 acres (59.24 km^2) including parts of modern-day Colma, Burlingame, San Bruno, South San Francisco, and Millbrae.[18][19] Rancho Laguna de la Merced was 2,219 acres (8.98 km^2) acres and covered the area around a lake of the same name.[18][19] The third ranch covering parts of the Daly City–Colma area was named Rancho Cañada de Guadalupe la Visitación y Rodeo Viejo and stretched from the Visitacion Valley area in San Francisco, to the city of South San Francisco covering 5,473 acres (22.15 km^2).[18][19]

Following the Mexican Cession of California at the end of the Mexican–American War the owners of Rancho Laguna de La Merced tried to claim land between San Bruno Mountain and Lake Merced. An 1853 US government survey declared that the contested area was in fact government property and could be acquired by private citizens. There was a brief land rush as settlers, mainly Irish established ranches and farms in parts of what is now the neighborhoods of Westlake, Serramonte, and the cities of Colma and Pacifica.[20] A decade later, several families left as increase in the fog density killed grain and potato crops. The few remaining families switched to dairy and cattle farming as a more profitable enterprise.[20] In the late 19th century as San Francisco grew and San Mateo County was established, Daly City also gradually grew including homes and schools along the lines for the Southern Pacific railroad.[21] Daly City served as a location where San Franciscans would cross over county lines to gamble and fight.[22] As tensions built in approach to the American Civil War, California was divided between pro-slavery, and Free Soil advocates. Two of the main figures in the debate were US Senator David C. Broderick, a Free Soil advocate, and David S. Terry, who was in

Location in the United States

Coordinates: 37°41′11″N 122°28′06″W

Country	United States
State	California
County	San Mateo
Incorporated	March 22, 1911[1]
Named for	John Daly
Government	
• **Type**	Council-Manager
• **Mayor**	Juslyn C. Manalo[2]
• **City council**[2]	**Council Members** [show]
• **State Assembly**	Phil Ting (D)[3]
• **State Senator**	Scott Wiener (D)[3]
• **U. S. Rep.**	Jackie Speier (D)[4]
Area[5]	
• **Total**	7.64 sq mi (19.79 km^2)
• **Land**	7.64 sq mi (19.79 km^2)
• **Water**	0.00 sq mi (0.00 km^2) 0%
Elevation[6]	344 ft (105 m)

Enterprise users tend to make much greater use of infoboxes than Wikipedia does.

In many cases, pages consist entirely of an infobox − no "free text".

"Smart infoboxes"

Cargo, SMW and Page Forms can turn templates into "smart infoboxes" - data is easily editable and queriable.

Best of both worlds: structure of a database + flexibility & undo capability of a wiki.

No data is ever lost – not even if you delete a field!

FRUIT & VEG

CHEF CLUB

37

BETHANY BROCCOLI

Subgrouping

ID

Image

Grouping

Name

We have an *instinctive* desire to categorize and structure the information we have.

The Enterprise MediaWiki ecosystem

- Used by companies, organizations, govt. agencies, personal use...

- 10,000s of active installations?

- Dozens of consultants and consulting companies

- Several wiki farms (Wikia/Gamepedia, Miraheze, Referata, etc.)

Competitors to Enterprise MediaWiki

- Microsoft SharePoint
- Confluence
- Drupal, Wordpress, etc.

Trend #1: MediaWiki-based packages

"Acme Consulting presents Acmoo, an one-stop solution for all your [CRM/ERP/BPM/GDPR/TOGAF/etc. etc.] needs! For the low price of $5,000/month, you get this software, plus all the support you need to customize it for your data requirements."

Sometimes, there's no mention of MediaWiki or even "wiki"!

This seems to be a successful formula, especially in Europe.

I have no problem with this − whatever increases usage is good.

Trend #2: Moving away from the Semantic Web

Outside of Wikidata/Wikibase, it's been a while since we have heard any of these terms:

- Reasoner
- RDF
- SPARQL

(Is Wikibase a component of Enterprise
MediaWiki? That's an open question.)

Trend #3: Mobile-friendly display

Responsive skins like Timeless, Chameleon, Foreground, Tweeki have become very popular.

Possibly people don't care as much anymore about "looking like Wikipedia".

MediaWiki and the Wikimedia Foundation

MediaWiki is in an unusual position as a software application

The organization that develops it (the Wikimedia Foundation (WMF)) does not benefit if more people install it.

How do open-source software makers usually benefit from greater usage?

- Paid hosting
- Consulting
- Donations, corporate and individual

The WMF doesn't do any of these.

(Of course, they take donations, but those overwhelmingly come from users of the *sites*, not users of the *software*.)

Often, when the software is not directly part of the company/org's "mission", it gets spun off.

Examples:

- **Hadoop** (Yahoo → Apache/Hortonworks)
- **Kubernetes** (Google → Cloud Native Computing Foundation)
- **Phabricator** (Facebook → Phacility)
- **PhoneGap** (Adobe → Apache)

Should the WMF spin off MediaWiki? Who knows.

In any case, it won't happen any time soon.

So, what does that mean for "Enterprise MediaWiki"?

- It may be helpful to have our voices heard on core MediaWiki development (see: MediaWiki Stakeholders' Group)

- Hosting, consulting and even marketing are all up to us!

Question: should "Enterprise MediaWiki" have its own name?

*"**Ubuntu** is based on **Debian**, which includes **Linux**"*.

So, how about:

*"Our CRM software package, **Acmoo**, is based on **???**, which includes **MediaWiki**."*

Benefits: more clarity in marketing, avoid the word "wiki" (?!), avoid copyright issues, …

Some thoughts on SMW vs. Cargo

The concept is the same: both let you **store data** in the wiki, then **drill down** on it, **display it** as lists, tables, calendars, maps, etc., and **export it**.

Both make your wiki much more powerful!

Cargo vs. SMW

Feature	Cargo	Semantic MediaWiki
Storage	Tables	Triples
Querying	SQL-based	Custom query language
Extensions required	Usually 2 (Cargo + Page Forms)	~ 8-15
Pages where data is defined	Templates	Templates, Properties, Concepts
Lines of code	~15,000	~150,000

Cargo vs. SMW: performance

Cargo's querying runs faster than SMW's:

- NASA EVA group found Cargo to be ~30% faster.

- Gamepedia switched from SMW to Cargo to avoid server crashes.

So, what does that mean?

Whether you should use one or the other is a matter of opinion.

There are successful SMW-based wikis, and successful Cargo-based wikis.

(And some use both!)

Other important "EMW" extensions

- Approved Revs – Mark one revision of a page as approved, removing much of the fear of bad edits

- External Data – Query outside data sources (APIs, Dbs, etc.) and display the data on your wiki

- Data Transfer – Import outside data into wikitext.

- Oauth, PluggableAuth, etc. – authentication and authorization, single sign-on

That's it!

Comments/questions/demo?