

The Semantic Web and why Wikipedia should bother

Jakob Voß

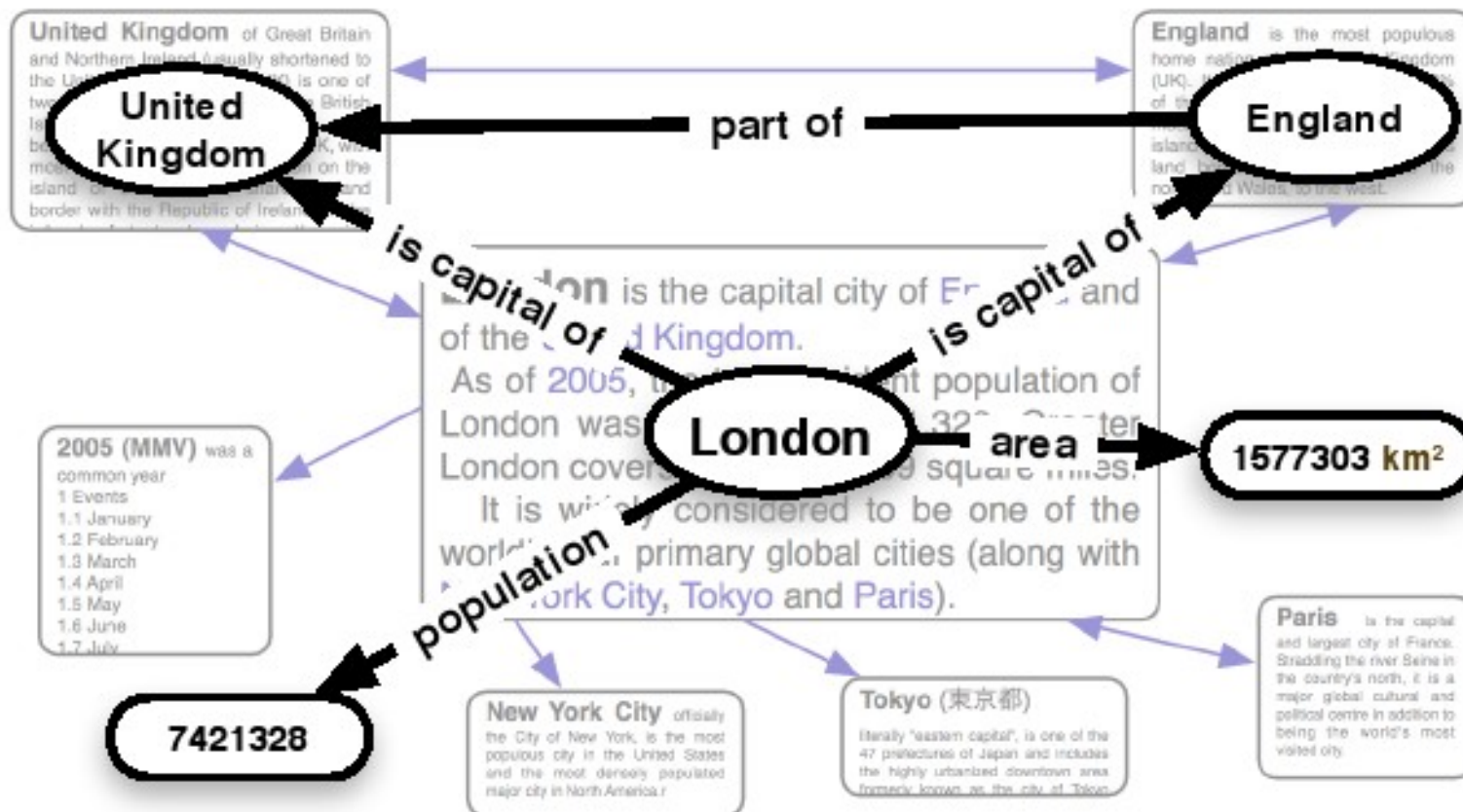
**Wikimania 2007
Taipei, Taiwan, 2007-08-03**

Agenda

- (1) The Semantic Web
- (2) Wikipedia's contribution
- (3) Examples and problems
- (4) Possible solutions

The Semantic Web

- **Everything** can be linked via its **URI**
- Every data in **triples** with typed links



The Semantic Web

- **Ontologies** define common structures and rules
- More data is generated by **aggregation** and **reasoning** on distributed data from several sources
- **Software agents** understand your commands, aggregate, reason, decide and act independently (at least in theory)

Wikipedia's contribution

- Largest source of freely available non-specialized data
- Templates and categories contain structured data
 - Persondata
 - DBpedia.org
 - Geodata
 - ...
- Semantic MediaWiki adds typed links and attributes

Aggregating and Reasoning

Aggregating and Reasoning

Which polish authors are currently most published in Germany?

Aggregating and Reasoning

Which polish authors are currently most published in Germany?

- Currently published in Germany
 - List of published books by book vendors or by the German National Library

Aggregating and Reasoning

Which polish authors are currently most published in Germany?

- Currently published in Germany
- Authors
 - National Library catalogue contains author and uniquely identifies author by PND-ID

Aggregating and Reasoning

Which polish authors are currently most published in Germany?

- Currently published in Germany
- Authors
- Polish authors
 - German Wikipedia contains PND => article
 - Article linked via Interwiki => more articles
 - Biographical articles contain place of birth
 - Place of birth linked to country via category

Aggregating and Reasoning

subject	predicate	object
Publication	published-in	Germany
Publication	has-author	Person
Person	born-in	Town
Town	place-in	Poland

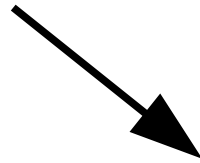
Where is Poland?

Where is Poland?

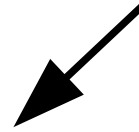


Where is Poland?

Or five times here in
Maine, Ohio, or NY

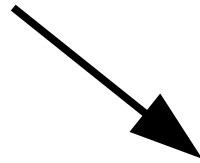


Somewhere here

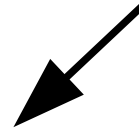


Where is Poland?

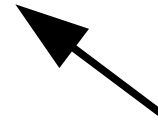
Or five times here in
Maine, Ohio, or NY



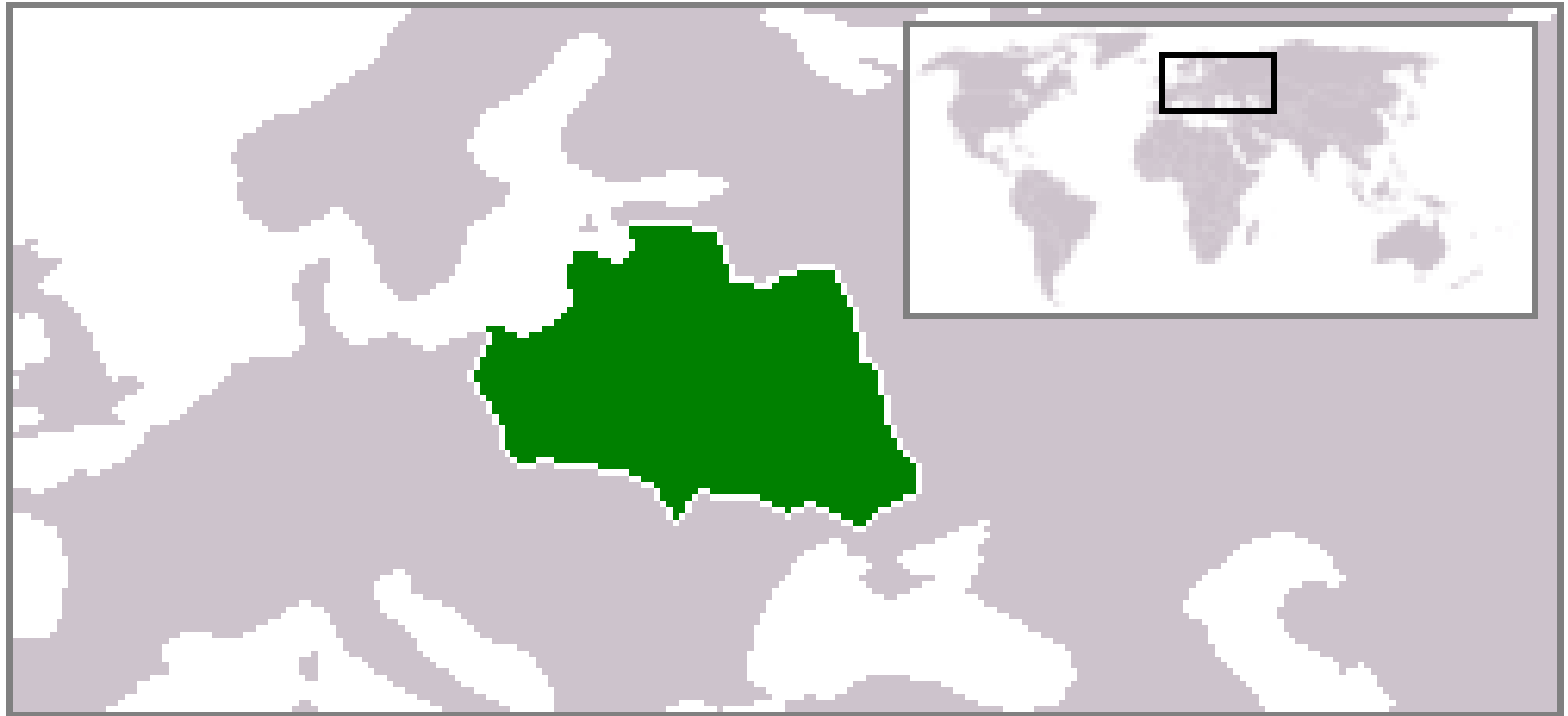
Somewhere here



Or did you mean
Poland, Kiribati?

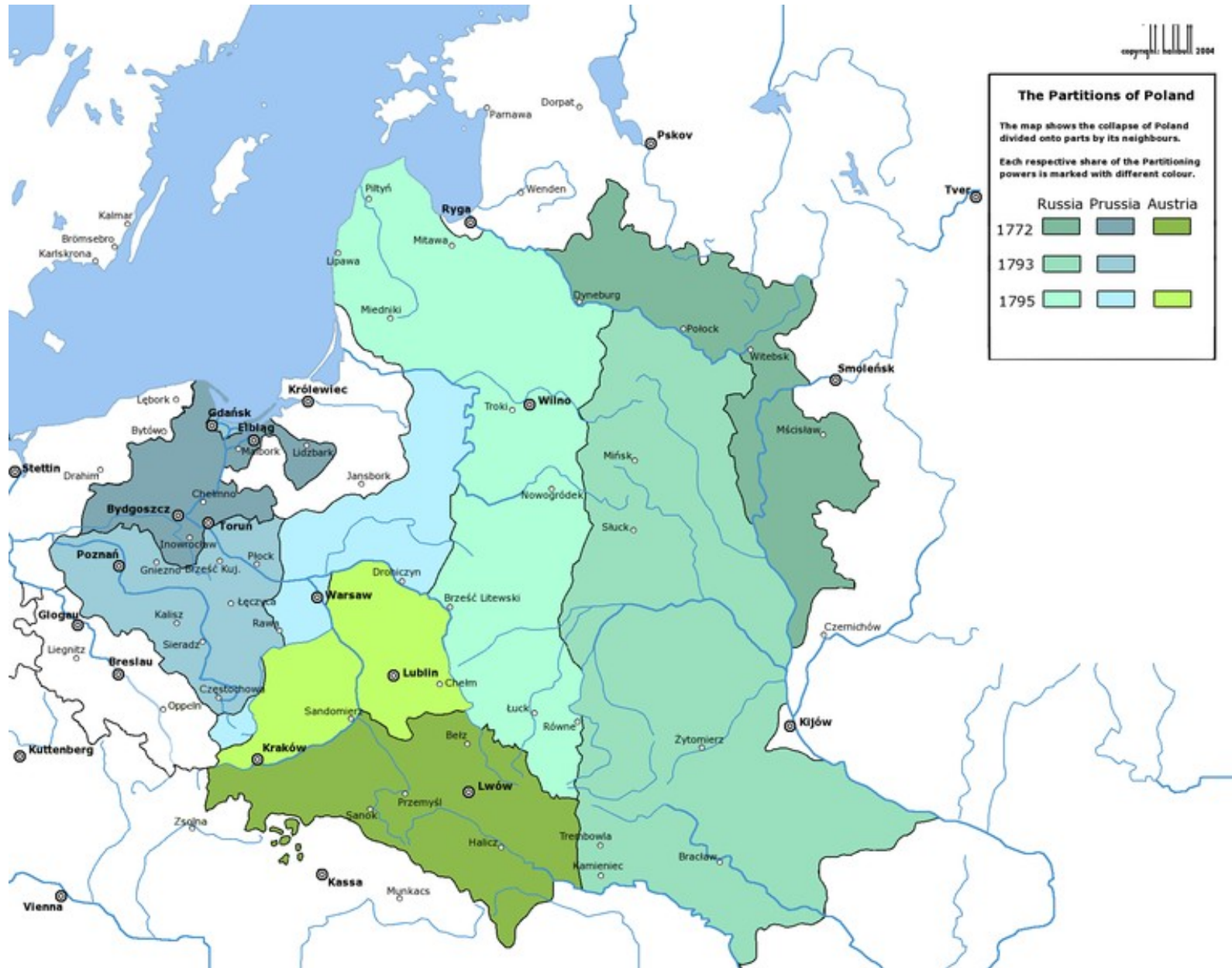


Poland around 1619



Polish-Lithuanian Commonwealth

Poland 1772...1793..1795



Poland 1945-



Where is Poland?

- Reality is complex, confusing, and fuzzy
- What's the »*default*« Poland?
- Humans can look up context in Wikipedia
- Semantic Web only consists of statements

Example #2

Presidents of the United States

- Bill Clinton 1993-01-20 – 2001-01-20

Example #2

Presidents of the United States

- Bill Clinton 1993-01-20 – 2001-01-20
- George W. Bush 2001-01-20 – 2009-01-20

Example #2

Presidents of the United States

- Bill Clinton 1993-01-20 – 2001-01-20
- George W. Bush 2001-01-20 – 2009-01-20
- Barack Obama 2009-01-20 –

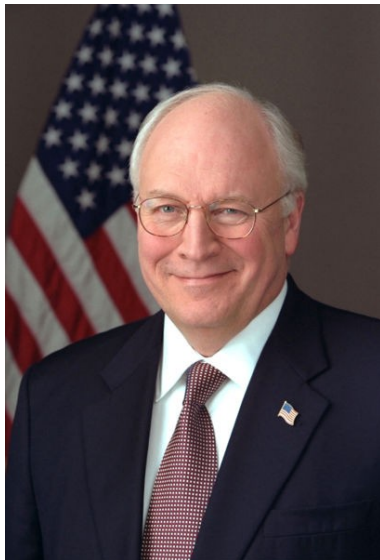
Example #2

Presidents of the United States

- Bill Clinton 1993-01-20 – 2001-01-20
- George W. Bush 2001-01-20 – 2009-01-20
- Barack Obama 2009-01-20 – 2013-01-20
- A. Schwarzenegger 2013-01-20 –

Presidents of the United States

- George W. Bush 2001-01-20 – 2002-06-29
- **Dick Cheney** 07:09 – 09:24 a.m.
- George W. Bush 2002-06-29 – 2007-07-21
- **Dick Cheney** 07:14 – 09:21 a.m.
- George W. Bush 2007-07-21 –



Twice president of the US
(see 25th amendment)

Presidents of the United States

- The devil is in the details ;-)
- Automatic reasoning will give you inconvenient results

Example #3

Finally a clear division

男性

XY



女性

XX



So let's formalize...

owl:disjointWith

“Classes may be stated to be disjoint from each other. For example, Man and Woman can be stated to be disjoint classes. [...] a reasoner can deduce that if A is an instance of Man, then A is not an instance of Woman.”

OWL Web Ontology Language Guide

<http://www.w3.org/TR/owl-guide/>

A clear division?

Other chromosomal sexes (karotype)

- Turner syndrome (X_{-}), Trisomy X...
- Klinefelter syndrome (XXY), XYY -Syndrome ...

A clear division?

Other chromosomal sexes (karotype)

- Turner syndrome (X_{-}), Trisomy X...
- Klinefelter syndrome (XXY), XYY -Syndrome ...

Intersexuality, Hermaphroditism

- Chromosomal sex inconsistent with phenotypic sex or phenotype is not just male or female

A clear division?

Other chromosomal sexes (karotype)

- Turner syndrome (X₋), Trisomy X...
- Klinefelter syndrome (XXY), XYY-Syndrome ...

Intersexuality, Hermaphroditism

- Chromosomal sex inconsistent with phenotypic sex or phenotype is not just male or female

Gender identity

- Gender with which a person identifies independent from biological sex.

A clear division?

- Reality is far more complicated
- Many kinds of exceptions

Problems

- Clear divisions discriminate
- Discussion and context gets lost
- **Example #4**
 - IF your name = X
 - AND X on a list of suspected terrorists
 - THEN you have a problem

Not our problem?

- Wikipedia is already used as source by millions of people
- People can think, judge and ask, computers cannot
- We create definitions that will be used in thousands of applications
- Statistics lie
Aggragation/resoning even lies better

Possible Solutions

- More of all (data, aggregation, reasoning)
- Less of all
- Statements about statements
- Fuzzy logic
- Data provenance / data lineage
- Allow exceptions
- Teach people to be careful
- Do not expect or believe simple answers
- It's just dirty data

Summary

- Semantic Web is great
- Reality is based on exceptions
- Simplification is useful but dangerous
- Data POV \neq NPOV
- We also bear responsibility for stupid use of Wikipedia data
- Never stop analyzing and thinking instead of relying on computers

More to read

- Shadbolt, Berners-Lee, and Hall: *The Semantic Web Revisited*. IEEE Intelligent Systems 21 (3) pp. 96-101. May/June 2006.
http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf
- Völkel, Krötzsch, Vrandečić, Haller, and Studer: Semantic Wikipedia. Proceedings of the WWW2006.
http://www.aifb.uni-karlsruhe.de/Publikationen/showPublikation_english?pub
- Doctorow: *Metacrap: Putting the torch to seven strawmen of the meta-utopia*. August 2001.
<http://www.well.com/~doctorow/metacrap.htm>
- Geoffrey and Star: *Sorting Things Out: Classification and Its Consequences*. MIT Press, 1999.