Managing Big Data with privacy in mind

Bert Bos

W3C Track at WWW 2017, Perth (Australia), 6 April 2017
Two related projects

Two research projects:

- **BDE**, 2015-2017: general platform for big data
- **SPECIAL**, 2017-2019: privacy for big data

By W3C and ERCIM (European host of W3C), in cooperation with a number of partners.
Big Data Europe: Consortium
Big Data Europe: Funding

Horizon 2020
European Union funding for Research & Innovation
What is Big Data?

‘Data too big to process with a single computer.’
The three V’s

Velocity

Volume

Variety

… but Variety is often neglected.
Handling ‘Volume’ & ‘Velocity’

- Use specialised software
  - HFDS, Spark, Flume, etc.
- Use parallel computing

BDE provides these and other software as Docker containers → easy to deploy on a ‘swarm’ of machines → scalable to many nodes.
Handling ‘Variety’

How to handle different kinds of data, and even correlate them.

Answer: metadata

(a.k.a., Semantic Web, RDF, LD, LOD… Did you expect a different answer in the W3C Track? 😊)
RDF and SPARQL

- Metadata tells how to process the data
- Allows to view the data as RDF
  → allows **SPARQL** queries

But:

- Most data isn’t actually converted to RDF
- Instead, the query is translated
Who owns the data?

Addition under consideration:

- Metadata about permissible actions on data

Could be **ODRL**.

(ODRL should be a W3C Recommendation by the end of the year.)
On to SPECIAL

And that brings us to privacy…

… and SPECIAL
SPECIAL: Consortium
SPECIAL: Funding

Horizon 2020
European Union funding for Research & Innovation
Metadata (again)

- Builds on Big Data Europe
- Privacy added by means of metadata
- In particular: a model of the EU General Data Protection Regulation (consent, anonymisation, erasure, etc.)
  - (It’s a European project, after all)
  - which should come into force in 2018
  - But likely useful outside EU
Additional modules

Also:

- Tools to import privacy policies in different formats
- Tools to anonymise data
- Tools to visualise & explain policies
- Versioning of policies
SPECIAL https://www.specialprivacy.eu/

Big Data Europe https://www.big-data-europe.eu/

Bert Bos <bert@w3.org>

GPG fingerprint: 7744 0204 52A5 14D9 147D 2A13 2D7A E420 184B 5BA4

https://www.w3.org/Talks/2017/0406-BDE-SPECIAL-WWW2017