Data Provenance for GDPR Compliance

Benjamin E. Ujcich and William H. Sanders
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
GDPR Background

- European Union General Data Protection Regulation (GDPR) took effect on May 25, 2018

- Scope is global
  - US companies who have EU customers
  - Data stored by EU companies around the world

- Penalties for non-compliance can be up to 4% of global annual revenue
Data Protection Is Highly Relevant!

A big part of GDPR is about data breaches and there’s a punishment if companies don’t notify regulators about a data breach within 72 hours of it happening. Firms can also be hit with fines if they are found to have not done enough to prevent a data breach or went against any of the principles around the processing of information laid out in GDPR legislation.

How Should We Think About Data?

Under what conditions was consent or justification given?

For what specific purposes can the data be used?

For what period of time is such use valid?

Personal Data

Without this additional metadata, no lawful basis for processing!
A Data Provenance Approach

- Captures how data were generated and used, what processes acted on data, and the responsibility for such processes
- Allows for domain-specific modeling, automatic capturing, and automatic querying

Why Is Provenance Relevant to GDPR?

- Right to consent: get subjects’ consent (opt-in)
- Right to withdrawal: stop future data processing
- Right to explanation: can ask how data has been processed, particularly for automated (machine learning) decisions
- Right to removal: erase data if no longer needed
- Right to portability: collect data and transfer it
- Obligation of data minimality: only use data necessary for task
GDPR Data Model

Future Work and Thoughts

- Data provenance and compliance for future data protection regulations
- Implementation of automated metadata/provenance collection in systems that process personal data
- Technical assurances about security of provenance collection systems and security of provenance data
Thanks for Listening

- Ben Ujcich
  E-mail: ujcich2@illinois.edu
  Web: http://ujcich2.web.engr.illinois.edu/

- Bill Sanders
  E-mail: whs@illinois.edu
  Web: http://whs.crhc.illinois.edu/