Data Security and Privacy

Mark Hubbard , Mike Branch, Sally Keller, Rai Safavi-Naeini, Supinder Pandher, Bridget Moloney, **Florian Kerschbaum**, Gerald Penn

Challenges to Security and Privacy in Operational Data Science

- Inference-proof release of data
    - Aggregate data
    - Micro data
  - Secure
    - Against linking with additional data sources
    - Under multiple release
- Distributed storage and processing
  - Compliance (GDPR, etc.)
- Data sharing agreements

Sources of challenges

- Distributed data sources and processing
  - IoT
  - Cloud
- Lack of technical solutions
  - Lack of knowledge about them
  - Lack of proven, successful deployment
  - Lack of guidelines (parameters)
- Legal compliance
  - Diverse across geographies
  - Purpose binding
  - Means of last resort

Technical Solutions

- Differential Privacy
  - Problems with
    - Micro data release
    - Security parameter
    - Unknown data correlation
- Cryptography (Homomorphic, Functional, ZKP, etc.)
  - Problems with
    - Scalability
    - Flexibility

Overcoming the Gap

- Need to test and deploy
  - Identify operational challenges
- Need to devise guidelines (parameters)
  - Security parameters
- Need to (security) test
  - Helps determining security parameters
- Initiate social debate

Non-Technical Solutions

- Data marketplaces
- Liability
- How important is user-verifiable enforcement?

Looking beyond

- Operational challenges around confidentiality/privacy
  - Access control
  - Inference control
  - Privacy and security go hand-in-hand
- Integrity is an open problem
  - What does integrity in this case mean?
  - Integrity across the data science chain?
  - Provenance/logging as a tool?
  - Does integrity/transparency contradict confidentiality?
  - Does blockchain play a role in all of this?

Selected problems across the data science chain

- Making data trustable and usable
  - Privacy-preserving record linkage
- Management of big data
- Modelling and Analysis
  - Secure storage and processing
- Dissemination and Visualization
  - Privacy-preserving data release