

# Data Profiling

---

**Do "Where You Go" and "What You Do" Become "Who You Are"?"**

**Ralf Bendrath**  
University of Bremen

**Francis Tonyan**  
IBM Threat and Fraud Intelligence

**Joe Zenobio**  
EPC Global Canada

# Wikipedia

---

- **Profiling**, the extrapolation of information about something, based on known qualities, may refer specifically to:
    - Stereotypes
    - Offender profiling
    - Racial profiling
    - Gender bias
    - Cultural bias
    - Performance analysis in (software engineering)“
-

# Bruce Schneier

---

- ❑ To profile is to generalize. It's taking characteristics of a population and applying them to an individual.
-

# Roger Clarke

---

- Profiling is (...) a means of generating suspects or prospects from within a large population, and involves inferring a set of characteristics of a particular class of person from past experience, then searching data-holdings for individuals with a close fit to that set of characteristics.
-

# General Idea

---

- Model of population classes
  - Information about individual
  - Sorting individual into a class
  - Acting on these classes
-

# Origins of Profiling

---

- ❑ Middle Ages: Inquisition
  - ❑ 1800: criminal profiling
  - ❑ 1900: early biometrics
  - ❑ today: forensic psychology
  
  - ❑ requires intense manpower
-



# Technologies of Surveillance

---

Observing

Recording

Distributing

Deciding

---

# Surveillance as Decisions

---

- „Dataveillance“ (Clarke)
  - „Social sorting“ (Gandy)
  - „Digital discrimination“ (Lyon)
-

# Basis of Decisions: Individual Data

---

- Income
  - Shopping history
  - Search records
  - Movements
  - Relationships
  - ...
-

# Basis of Decisions: Mass Data

---

- Data from large populations
  - Correlation analysis
  - Modelling
-

# Data → Profiles / Models

---

- „urban hipster“
  - „rural conservative“
  - „silver surfer“
  - „chronic complainer“
  - „credit risk“
  - „terrorist suspect“
-

# Effectiveness Problems

---

- Individual data basis
    - wrong information?
  - Modelling error
    - millions of Amazon buyers
    - too few known terrorists!
  - Probabilistic error
    - profile  $\neq$  individual
    - false positives / false negatives
-

# Ethical Problems

---

- ❑ Treating humans as objects, as data sets
  - ❑ Decisions by machines that affect real humans
  - ❑ Discrimination reified in machinery
-

# EU Data Protection Directive 1995

---

Article 15:

Automated individual decisions

„Member states shall grant the right to every person to not be subject to a decision which (...) is solely based on automatic processing of data intended to evaluate certain personal aspects relating to him (...).“

---

# Web 2.0 Issues

---

- self-generated profiles
  - gossip and chatter about others
  - public → democratization of profiling?
  - future of reputation (Solove)
- 
- lots of data, but not standardized
  - semantic web?
-