Open Data in Electoral Administration

Digitization of the Electoral Process

Peter Wolf
International IDEA
E-Vote-ID
Bregenz, 26 October 2017

Definition(s) of open data

1. Complete: all data, except for valid privacy, security, legal limitations
2. Primary: data is available at the source and in the finest level of granularity
3. Timely: data is made available as quickly as needed to preserve its relevance
4. Accessible: data is accessible to the widest range of users for the widest range of purposes
5. Machine-processable: in format and structures that allow processing
6. Non-discriminatory: available to everybody without a need for registration
7. Non-proprietary: data formats should not be under exclusive control of any entity
8. Licencse free: not subject to copyright, trade-secret, patents. Reasonable restrictions may be allowed
9. Permanent: it should remain online, be properly archived
10. Free of usage costs: no usage fees should be applied for public use

What election data should be open?

- Not only results data
- Voter registration
- Campaign and party financing
- Electoral districts, boundaries
- Polling station assignment
- Legal framework, regulations, etc.

Current application

- Several success stories:
  - Afghanistan (2009 & 2014)
  - Burkina Faso (2015)
  - Indonesia (2014)
  - UK (2015)
- But overall, slow progress

Typical Limitations

- Type of data
- Granularity of data
- Time and duration of publication
- Machine readability
- Registration requirements for accessing detailed data

Reasons for the limited uptake

- Purely technical, institutional
- Unavailability of data – especially in paper elections
- Worries about related risks of too much openness
- Too much openness also not desirable for electoral stakeholders
  → **What are reasonable restrictions?**

What e-voting data can be opened?

- Specifications
- Procurement details
- Security measures
- Source codes
- Detailed results
- Log files
- Any other data produced

Transparency of e-voting

Less transparency as technology is opaque, hard to understand

VS

More transparency as everything can be logged, traced, recorded … and potentially published

What e-voting data should be opened?

- Can everything be published without endangering the security of the process and the secrecy of the vote?
- What are reasonable limitations?
- What about log files, encrypted data …?
- When should data be published (e.g. source codes)?
- Is restricted access to some users acceptable?
- Can too many details be misused, have opposite effect and create doubt?

Thank You!