Legacy Data

Legacy data

- Is this *all* data or do we expect gov's to shift their infrastructures to Linked Data?
- Various formats
  - Scans, PDFs, text, proprietary, Excel, CSV, Relational DBs
- Mostly
  - Geo
    - Maps
    - Lat/Long columns
  - Temporal
    - Timestamps
  - Record-oriented/relational
    - E.g. citizen data, statistical data

City opening data-concerns

Business and legal level

- What are the privacy considerations in publishing data? On one hand, city will like to respect the privacy of citizens and businesses, and on the other, it will like the data to be valuable enough to lead to positive change.

- How to pay for the cost of opening up data? Cities may or may not have legal obligations to open up data. Accordingly, they will look for guidance on how to account for the costs. Further, can they levy a license fee if they are not obligated to open data

- Which data should be opened and when? Should it be by phases? What data should not be shared?

- What policies / laws are needed from the city so that businesses can collaborate on open data, while preserving their IP?

City opening data-concerns

Technical level

- What should be architecture to share large-scale public data? How do we ensure performance and security?

- What visualization should be supported for different types of data?

- How can we facilitate non-experts in converting/annotating their data?

- Can we provide a template implementation for the reference architecture?

Some news

- Good news: Legacy models are typically less expressive than RDF.

- Bad news: Some of them are not practical in RDF (e.g. maps, images).

  – Where do we stop?