Trust and Usability

Participants from:

- Statistical sciences/bio-statistician/Statistics Canada

- Database/data quality researchers

- User applications: BI from The Cooperators

Nicely prepared by: Fei Chiang

Questions

1. What is your impression of the state-of-the-art in trustworthy data management?

2. Challenges inspired by 1.

3. How does data trust worthiness interact with other areas?
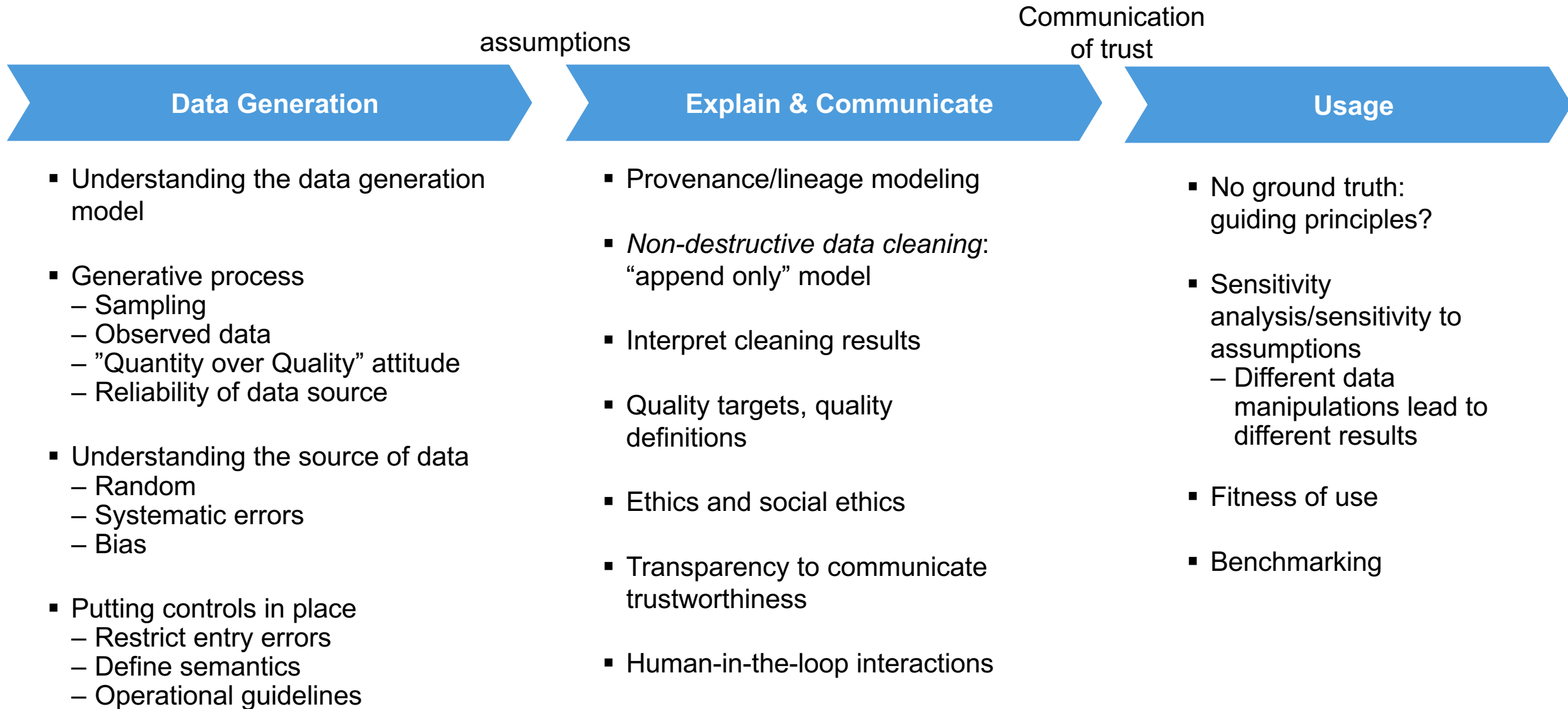
State-of-the-Art: Challenges

- Piece-meal approach towards data cleaning
- Unstructured vs. structured data
- Data quality Framework/dimensions: accuracy, accessibility, timeliness, etc.
- Data quality is based on the usage of data
  - Varying attitudes: quantity over quality
- Statistical modeling: data checking and verification of errors
- Propagation of errors

State-of-the-Art

- Bias
  - How can we measure the influence of bias in data analysis results?
- Master Data Management
  - Alignment of metadata and data
  - Data semantics (e.g., interpretation of null values)
- Open Data: source reliability, persistence, lack of documentation
- Timeliness and data currency
  - Temporal aspects of data quality

Trust & Usability Pipeline

assumptions

Communication
of trust

| Data Generation | Explain & Communicate | Usage |

Data Generation
- Understanding the data generation model
- Generative process
  – Sampling
  – Observed data
  – "Quantity over Quality" attitude
  – Reliability of data source
- Understanding the source of data
  – Random
  – Systematic errors
  – Bias
- Putting controls in place
  – Restrict entry errors
  – Define semantics
  – Operational guidelines

Explain & Communicate
- Provenance/lineage modeling
- *Non-destructive data cleaning*: "append only" model
- Interpret cleaning results
- Quality targets, quality definitions
- Ethics and social ethics
- Transparency to communicate trustworthiness
- Human-in-the-loop interactions

Usage
- No ground truth: guiding principles?
- Sensitivity analysis/sensitivity to assumptions
  – Different data manipulations lead to different results
- Fitness of use
- Benchmarking

Interactions

Big Data Management	Analytics	Security & Privacy
▪ How much data to ingest? ▪ Storage and scalability	▪ Specifies usage and inform quality assurance ▪ How to communicate trustworthiness in analytics	▪ Access methods ▪ Data publishing, analysis ▪ Data privacy vs. data utility

Trust Worthiness and Usability

Ethics, Policy and Social Impact	Dissemination & Visualization
▪ Transparency helps explainability ▪ Social responsibility, fairness	▪ Understand the data generation model ▪ Opportunities to communicate trust worthiness ▪ Capacity building ▪ Transparency of actions ▪ Understanding of errors