Metadata standards to support controlled access to microdata

Wendy Thomas – MPC, University of Minnesota
Arofan Gregory – Open Data Foundation
Alistair Hamilton – Australian Bureau of Statistics

Focus of Presentation

- Technical and semantic approaches to dealing with legal and organisational considerations of trans-national microdata access
  - Enforcing clear communication
  - Leveraging the use of DDI and SDMX
- Support for the capture of metadata along the statistical process – industrialization
  - Applying industrialisation to enable and manage microdata access
- Combined DDI-SDMX model

Technical and Semantic Approaches

- Areas of interaction with the access system
  - Depositing metadata covering data content, access protocols, and confidentiality rules
  - User interface for discovery and comparability, access authorisation, and analysis
  - System operationalisation of access constraints, data retrieval and processing, confidentiality rules, and output generation
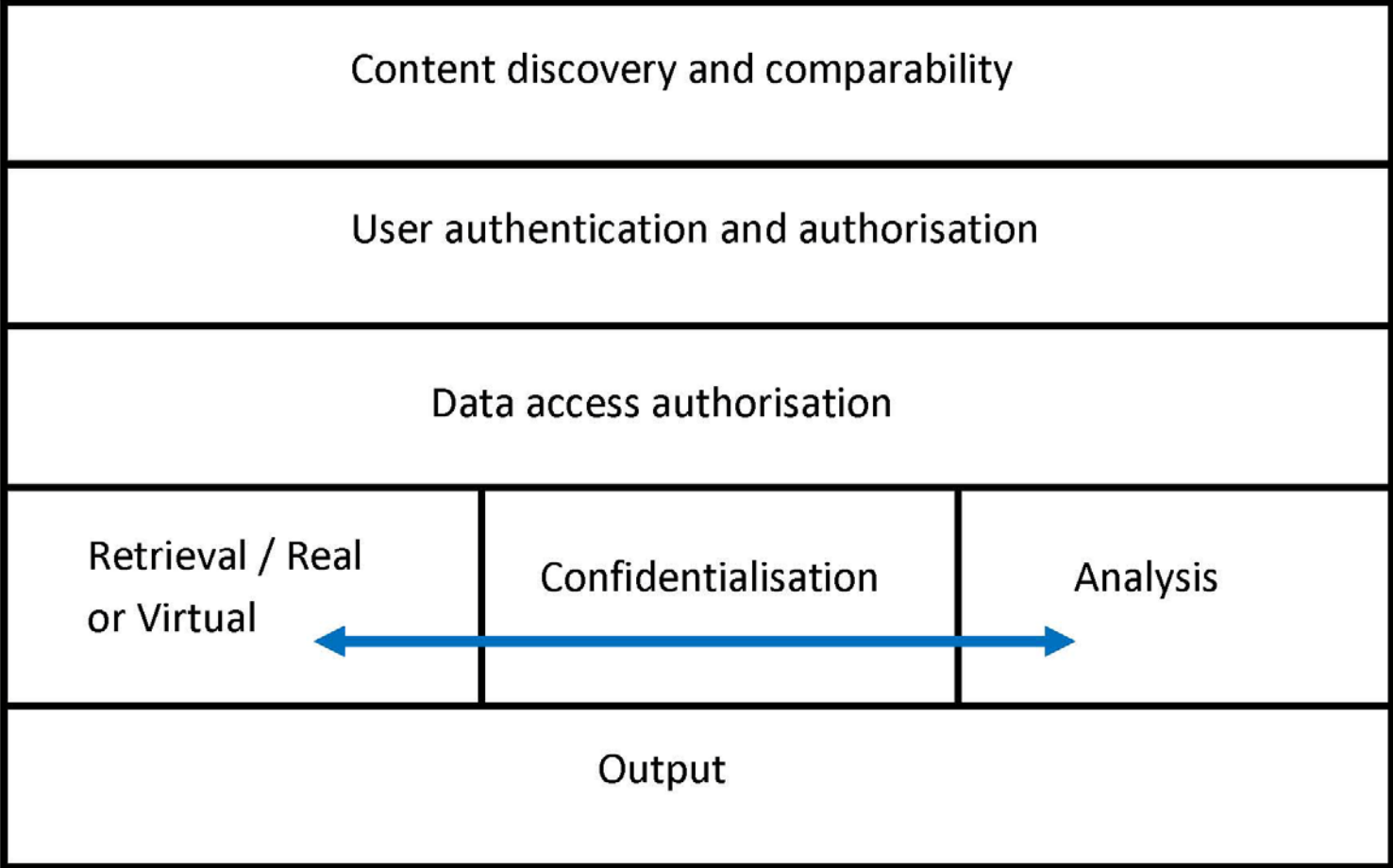
Single standard

- A single standard that is designed to interact in a consistent and reliable manner with current commonly supported standards

- Metadata driven processes within the system to extent feasible

- Delivery of output in the desired structure to support the work of a diverse research community

DDI Lifecycle Model

- Designed to interact with ISO/IEC 11179, SDMX, DDI Codebook, ISO/IEC 19115, Dublin Core, and more

- Used in both official statistical and academic research communities in a range of disciplines

- Core focus is on microdata and its conception, collection, management, analysis, distribution, and preservation

- Designed to support metadata driven systems throughout the lifecycle of the data and metadata, it represents the top row on the GSBPM chart

Support for capture of metadata along the statistical process

- Metadata to support an *external* access system will only be created if it can be used to drive an *internal* statistical process
- The industrialisation and use of industry standards to manageme the statistical business process enables:
  - Shared process and tools development
  - The ability to share metadata and provide a broader range of access to data where desirable

IHSN Microdata Toolkit

- Expanded the use of metadata standards (DDI-Codebook) through a toolkit supporting metadata management, best practices, and dissemination options

- Used in 81 countries in the developing world, producing approximately 5,000+ file descriptions

- IHSN microdata catalog

The DDI-SDMX model

- DDI-L describes the full process of research or statistical production including the aggregation process
- The strong overlap with a cross-walk between DDI-L and SDMX standards supports SDMX output from DDI-L metadata describing the source microdata
- The DDI-SDMX model covers the majority of the GSBPM stages
- NSOs pursuing this model have noted its value in supporting both metadata/data management and quality control systems

Supporting a trans-national access system

- DDI can be used to support statistical production as well as used within controlled-access environments

- Defining a harmonised profile of the DDI metadata to support both functions would provide data producers and researchers with a consistent and useful set of metadata and documentation

- DDI is widely used within secure data centres around the world notably in those providing remote access

Conclusion

- In summary, DDI is positioned to become an increasingly popular standard for managing secure microdata in all types of environments, both within the domain of official statistics and for researchers, supplemented by SDMX in some official statistical organisations.

- By providing a common language between statistical organisations, shared access systems, and researchers, the application of DDI can provide benefits for users at every stage of data lifecycle