Faster, easier and safer access to microdata

Don McIntosh
Space-Time Research

Introduction
Query-based access is an alternative approach to dissemination-based access for making statistics available. It takes advantage of high performance computing and new information privacy protection methods to reduce the amount of up-front work required from the provider and increase the level of access to data for end users. It is particularly useful when your statistics are in demand from researchers.

Fast
• Faster to produce. You don’t have to design and build a set of cubes up front.
• Eliminate wait times resulting from manual consulting
• Using fast database engines, queries on microdata can be just as instantaneous as queries on pre-aggregated data
• Zero waste - create what is needed on demand

Easy
• Allow users to work directly from microdata to create their own outputs
• Data API to enable custom-built applications, integration with other systems, mashable data
• Link back to DDI-based metadata repositories to ensure data provenance is accessible
• Custom grouping of values to generate reliable aggregates for large cubes dimensions
• Analyze detailed query log to evaluate most popular queries for feedback into next cycle

Safe
• Links to metadata ensure context is understood
• End-to-end metadata flow
• Apply customizable disclosure control algorithms or base on confidentialized unit record files (CURFs)
• Maintain audit log of all queries
• Reduced error through automation

The challenge

Provider
• Provide sufficient access within budget constraints
• Reduce manual servicing of data requests
• High data quality to ensure customer trust
• Provide detailed statistics while protecting privacy of respondents
• “Duty of Care”: minimize risk of misinterpretation
• Maintain the process for many collections over time
• Minimize collection → release time to maintain relevance and usefulness

Customer
• Find the right statistics
• Meet tight research project deadlines
• Find most suitable data within tight project deadlines
• Ensure decisions are based on accurate interpretations
• Keep up-to-date when new data is released

From DDI to relational schema

Data standards
• Relational view of DDI “logical product” via JDBC
• DDI version 3.1 plus converters for older versions
• Drives automated data load and output production
• Access to detailed study metadata through use of complementary DDI tools, such as Colectica™ (eg: trace a variable back to the original question asked in a survey)
• Complex multiple response variables, hierarchical code schemes, weighed survey data
• Export to SDMX from browser-based application
• Machine to machine access via RESTful API (HTTP GET metadata queries, and POST new SDMX queries)
• Retain rich metadata from the DDI, such as study details, variable descriptions, multilingual labels
• RSEs and annotations included in SDMX output

Acknowledgments
Our sincere thanks to many people from the following organizations for their assistance in development of our DDI query-based access solution:
• Algeria
• Australian Bureau of Statistics
• Metadata Technology USA
• Minnesota Population Center

For further information
Please contact don.mcintosh@spacetimeresearch.com.
More information on this and related projects can be obtained at www.spacetimeresearch.com.