

WP.39
ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)**

EUROPEAN COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

**STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

Joint UNECE/Eurostat work session on statistical data confidentiality
(Tarragona, Spain, 26-28 October 2011)

Topic (viii): International projects, groups and forum dealing with data access, release and related methodologies

Advancing cross-border access to microdata: Work of the “Paris Group”

Prepared by Heather Dryburgh, Statistics Canada

Advancing cross-border access to microdata: Work of the ‘Paris Group’

Heather Dryburgh, Ph.D.*

Microdata Access Division, Statistics Canada

Prepared for the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, October 26-28, Tarragona, Spain

1. Background of the Paris/OECD group

In May 2009 a few countries got together to discuss some options for cross-border access to microdata produced by National Statistical Offices (NSO). This workshop, jointly organised by the Australian Bureau of Statistics and OECD, and hosted by the OECD, aimed to explore the possibilities for collaborative development with regard to remote access to, or remote processing of, microdata. Participants discussed business requirements for remote microdata access/processing, as well as exchanging information on possible technical options and the approaches taken by different NSOs and agencies to ensure the de-identification and/or confidentialising of data. Participating countries presented the legislative frameworks underpinning microdata access in their countries.

Participating countries included ABS, Eurostat, OECD, ONS, Statistics Canada, Statistics Netherlands, Statistics New Zealand, Statistics Sweden and the US Census Bureau. From the discussions of that initial meeting, a number of action items were identified for the following year, including the preparation of a thematic issue on microdata access in the Official Journal of the International Association for Official Statistics and a paper on a common microdata access interface to be discussed at the meeting in 2010. Participants agreed that metadata describing microdata should be compliant to a common standard. While acknowledging that compliance would take some time, it was agreed to look at the Data Documentation Initiative (DDI) as a possible standard to adopt.

At the second meeting, the aims were: to broaden group membership and increase the group’s understanding of microdata access in new member countries; to develop a purpose / vision of the working group and identify key steps to achieve that purpose; to agree on a work plan for the following year; and to agree to a set of values and guiding principles for the working group. Initial discussions began on whether the group should request official status of the OECD as an expert group, or retain an informal status.

Some key deliverables from the 2010 meeting were the group’s mandate, values and ethics (Appendix A) and a discussion document about transitioning to an OECD expert group. During the year before the 2011 meeting, participants agreed to proceed with becoming an OECD expert group formally recognized by the OECD Committee on Statistics (CSTAT).

In 2011 the group met again. This time the work of the meeting had two main foci. The first was to come up with a model, or a series of models, for international access to microdata that could be agreed upon in principle (see Section 4). The second focus was on international standards of metadata and what could be accomplished through a first step of agreeing to adopt an international standard. One of the deliverables for the year following the 2011 meeting will be a document on a metadata standard. The paper will discuss the need for standards to describe microdata, what possible standards there are and their relationship to other standards and available tools for migrating to a new metadata standard.

2. Mandate and Work¹

As agreed by participants, the purpose of the working group is:

Statistical Institutes working together on practical steps to advance cross-border access to, and analysis of, microdata by leading the way and taking into account the needs of researchers and policy makers.

This will be achieved by

- Increasing coordination and communication between institutes and other expert groups to adopt best practices, promote a common understanding and minimise duplication of work in the area of cross-border access to microdata.
- Establishing procedures for efficient integrated cross-border access to microdata held by statistical institutes for statistical purposes while respecting confidentiality constraints.
- Conducting practical country experiments with actual data files.
- Advising and making recommendations to Chief Statisticians based on the Group's work and advice from other expert groups and practitioners on technical and non-technical issues.

CSTAT will provide the co-operative international leadership needed for strategic progress development. This Group will work closely with other relevant international groups discussing microdata access or related topics (*e.g.* the Workshop on Data Access – WDA). Some NSOs have already developed in-house methods for sharing microdata with clients, either nationally or internationally. These methods include a variety of on-site and remote access systems and practices. Other NSOs are commencing work on developing microdata access methods.

3. Challenges for Cross-border access to microdata

¹ Taken from the OECD document: *Sub-Committee/Sous-Comité: Expert Group for International Collaboration on Microdata Access/Groupe d'experts de l'OCDE pour la collaboration internationale autour de l'accès aux microdonnées*. Reporting to/Faisant rapport à: Committee on Statistics / Comité des Statistiques. Date of Creation/Date de création: 17 June 2011 / 17 juin 2011

The biggest challenges for NSOs when it comes to providing access to data beyond their own borders centre around legislation, costs, technical expertise and other risks associated with all microdata access, such as ensuring data confidentiality and maintaining good respondent relations.

The Group will, in particular, review the current situation and work towards developing and implementing new solutions regarding:

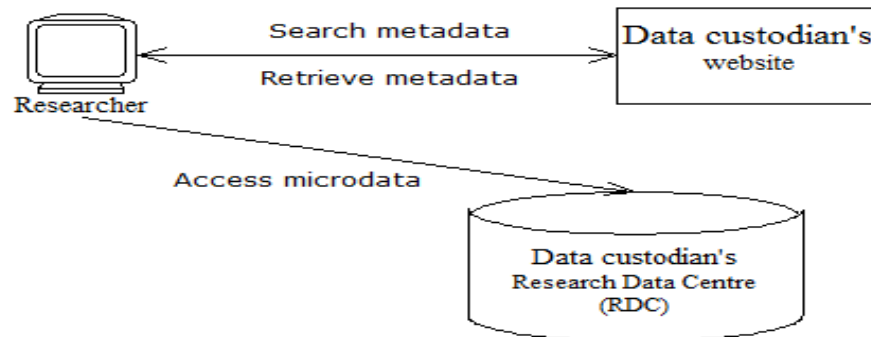
- Conceptual issues, such as principles of risk assessment
- Standards and procedures for accessing microdata, both nationally and internationally
- Technical solutions to facilitate international microdata access
- Administrative requirements for accessing microdata, including metadata.

The Group will determine agreed approaches for accessing microdata across countries.

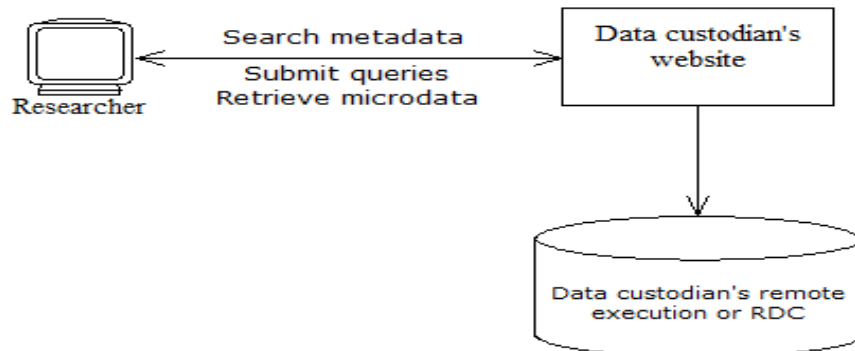
4. Models for microdata access²

Model 1: Current model for microdata access

Example 1



Example 2



² *Access Continuum Models Presentation*, for discussion at Paris Microdata Group meeting 4- 6 July 2011.

Model 2: Syndicated discovery model

Description of syndicated discovery model:

- Data custodians contribute metadata to a single basic metadata registry accessible from a web gateway. The registry would include some broadly comparable metadata.
- Contributed metadata not necessarily directly comparable as may use different metadata standards and concepts. Needs a standard set of data items to power the Search engine.
- Gateway to include a search engine to search metadata registry.
- Gateway will also include hotlinks back to data custodians' web pages about microdata

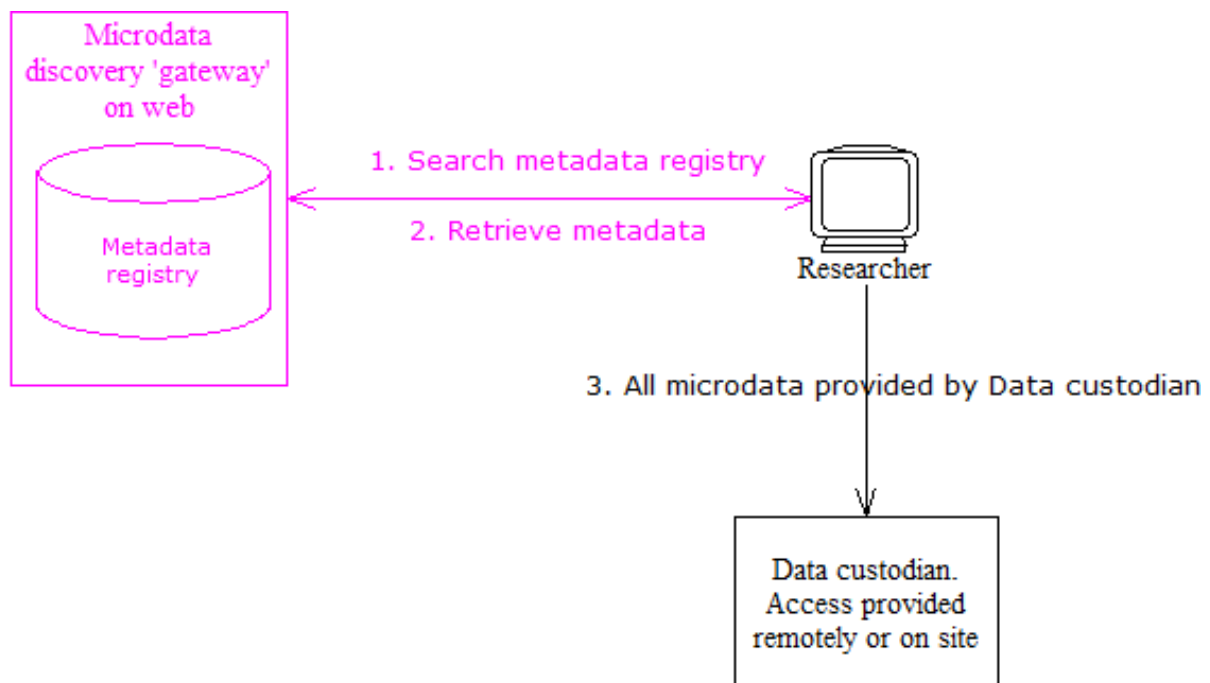
Services for researchers:

- Researchers search metadata registry to discover what microdata is available, its characteristics (as defined by the data custodian), and its data custodian.
- Researchers apply for access directly to the data custodian.
- Data custodian assesses researcher's application for access and either approves access or rejects application.
- Approved researchers use microdata either remotely or in RDCs.

Prerequisites/Assumptions:

- That most metadata is not confidential and can be shared publicly.

Model 2



Model 3: Syndicated discovery and interrogation model

Description of syndicated discovery and interrogation model:

- Basic metadata registry and search engine from Model 2.
- Plus, a 'harmonised' metadata registry. Are a range of options for harmonisation. Could start with structural harmonisation (e.g. harmonised variable definitions), then provide conceptual harmonisation and later data integration. Harmonisation would be mainly within collections (e.g. LFS, Health) unless classifications cross collections
- Plus, a query generator for tabular outputs for microdata held by data custodians,

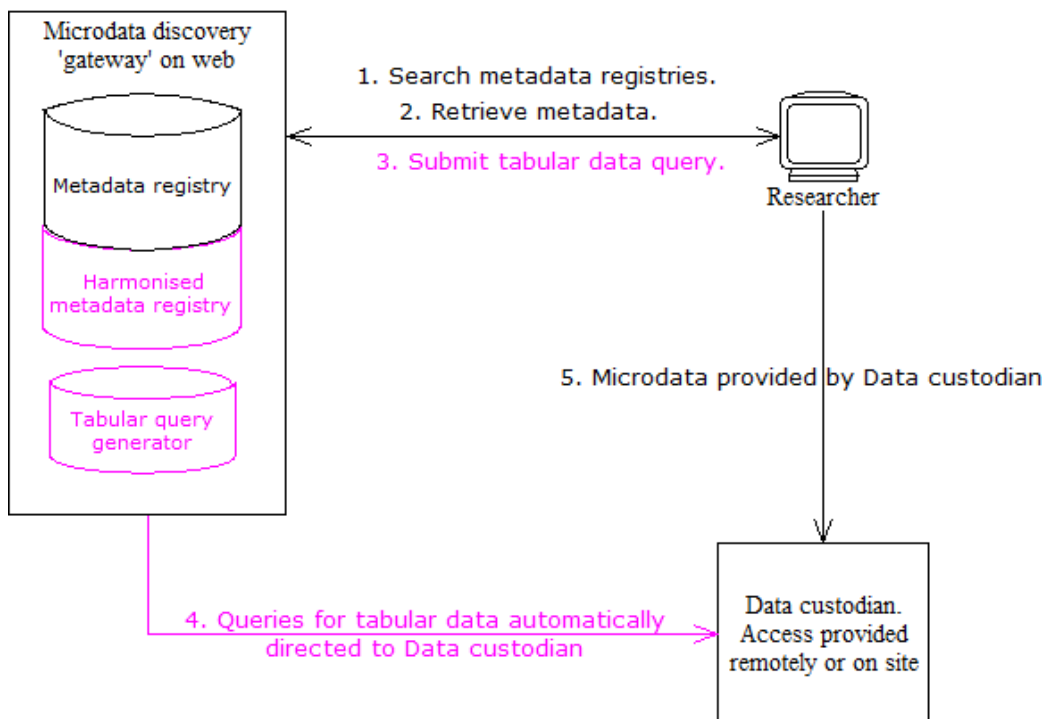
Services for researchers:

- Researchers search metadata registries and retrieve metadata.
- Researchers generate queries for tabular outputs for data held by data custodians.
- Researchers apply to the data custodian for access to the microdata.
- Data custodian assesses researcher's application and either approves access or rejects access.
- Approved researcher analyses microdata either remotely or in a RDC
- Researchers can make international comparisons

Assumptions and prerequisites:

- As for Model 2.
- Each data custodians' metadata described using an internationally accepted metadata standard (e.g. SDMX or DDI) to facilitate harmonisation of metadata
- Query generator to use an internationally accepted query language (e.g. SQL). Data custodians need to be able to read the same query language in order to receive queries directed to them from the query generator.

Model 3



Model 4: Integrated Access Model

What is provided:

- Metadata registries, search engine and query generator for tabular outputs from Model 3
- Plus, analysis query generator
- Plus, selected harmonised microdata and functionality to enable remote analysis of that microdata data store

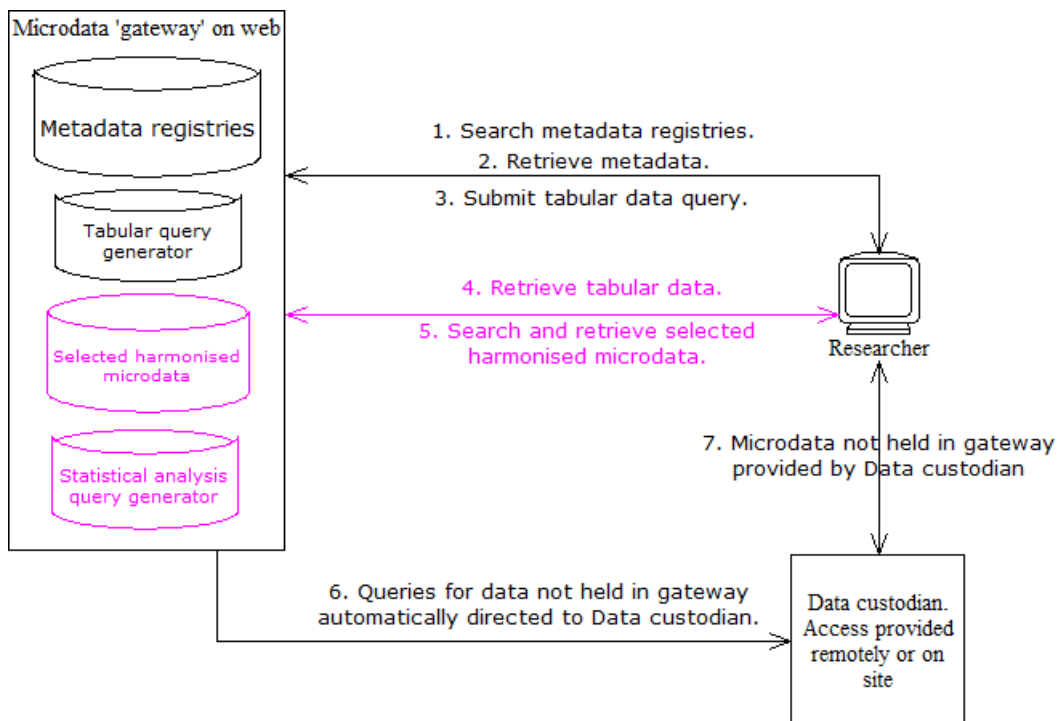
Services for researchers:

- Researchers search metadata registries and retrieve metadata.
- Researchers generate tabular queries and receive confidentialised tabular outputs
- Researchers undertake remote analysis of microdata and retrieve confidentialised
- For microdata not held in the store, researchers' queries automatically directed to data custodian

Assumptions and prerequisites:

- As for Model 3, except that selected microdata could be analysed remotely, with no RDC required.
- Requires technical infrastructure to provide remote analysis of microdata, though access to unit records is not an assumption of this model.
- Only selected microdata held in the store (e.g. microdata that requires no legally based registration process, or synthetic data) Other microdata would continue to be held by data custodians.

Model 4



5. Outstanding challenges and conclusions

The proposed series of models presented show the results of initial discussions. There remain key issues to resolve before implementation will be possible. These include:

1. Communicating the vision – to decision makers
2. Maximising the power of metadata – working toward standardisation - DDI
3. Linkages with other developing systems and projects
4. Meeting legal requirements of data custodians
5. Accreditation system - of researchers and of projects
6. Technical infrastructure
7. Costs modelling
8. Logistics of a gateway

Working together with projects such as Data without Boundaries and the DDI Alliance, the working group will continue to seek solutions for the remaining challenges.

Appendix A: Paris Microdata Workshop: Purpose, Objectives, Long Term Goal, Principles, Values, Success Criteria

PURPOSE

Statistical Institutes working together on practical steps to advance cross-border access to, and analysis of, microdata by leading the way and taking into account the needs of researchers and policy makers.

OBJECTIVES

Increase coordination and communication between institutes and other expert groups to adopt best practice, promote a common understanding and to minimise duplication of work.

Advise and make recommendations to Chief Statisticians based on our own work and advice from other expert groups and practitioners.

Undertake work commissioned by Chief Statisticians.

Serve as a forum for other countries or other expert groups to raise issues.

LONG TERM GOAL

To establish efficient integrated cross-border access to microdata held by statistical institutes for statistical purposes while respecting confidentiality constraints.

PRINCIPLES

1. Reuse, adapt and integrate into common framework for microdata access
2. Develop a project governance approach based on a variety of complementary projects
3. Communicate and share information proactively
4. Break long term goals into deliverable shorter term projects; focus on outcome orientation, and strive for a good balance between top down standardisation and bottom up integration of practical implementations
5. Conduct the business of the group in English

6. Acknowledge and respect the diversity of contexts, views, needs and approaches to microdata access
7. Network with other related groups to share knowledge and identify common goals and streamline work programs (focus work and avoid duplication)
8. Periodically check and if necessary realign, the goals, role and value of the group
9. Make decisions based on evidence.
10. Build projects with sharing and re-use across the whole community in mind.

VALUES

1. Active engagement by participants, while respecting any agreed upon differences in engagement.
2. Be open and respectful and explore the views of others
3. Group decisions should be based on consensus, where consensus is defined as "the absence of sustained opposition" arrived at after open debate and discussion
4. Develop a climate for innovation and for empowering individuals
5. Keep the group coordinator informed of progress and sent out early warning signals in case of arising problems.
6. Appoint single point of accountability for each task and only accept this role in full understanding of the desired leadership role accompanying it.

SUCCESS CRITERIA

Alignment of group goals with goals of individual NSOs and endorsement of group goals by Chief Statisticians.