Overview of Micro Data Services to Researchers with focus on European Countries

Approvals (Process 1 – 3):

> Process 1: **Approval of the research environment**
>
> Process 2: **Approval of researchers**
>
> Process 3: **Project approvals (**Overall description of the project)

Data delivery process (Process 4):

- Selection of variable (Counseling, from metadata description)
- Selection of access facility (on-site, remote, remote execution)
- Selection of type of data ('secure-use files' (RR); 'scientific-use files' (MUC), (PUF)
- Extraction of data
- Statistical disclosure control
- Production/extraction of metadata (Customized/from metadata)

Contract (Process 4):

- Costs of service
- Contract
- Conditions

Output control (Process 5):

Overview of Micro-Data Services to Researchers with focus on European Countries

ACCESS MICRO-DATA IS GRANTED AFTER THE 'NEED TO KNOW PRINCIPLE' AND ONLY FOR SCIENTIFIC PURPOSES

- But great variations among NSO's on how this statement is interpreted

Overview of Micro Data Services to Researchers with focus on European Countries

Generally speaking organizational structure, procedures and processes for micro-data access depend on:

Legislation:

- The Data Protection Directive (European Parliament and Council Directive 95/46/EC) is the foundation of data protection rules in the EU and in the individual Member States.
- National regulations: E.g. Personal Data Act, Statistics Act, Public Administration

Maturity:

- Maturity in the NSO toward a common understanding, harmonization and description of processes involved in providing micro-data services to researchers

Overview of Micro Data Services to Researchers with focus on European Countries

Maturity challenges:

- Guidelines for all processes (Published internal and external)
- A standardized and transparent catalogue of available data
- Meta-information (survey/register level as well as variable level)
- Classifications
- A common data library or a data warehouse

Process 1: Approval of research environments (in most countries expectations occur)

Common sub-processes in most countries

A. Check whether the research entity complies with the criteria for access to microdata according to national laws

B. Approval of research entity (Published)
- The authorization manager – usually the head of the environment
- One central person in the NSS - usually the National Statistician

C. An authorization contract is signed by
- The authorization manager – usually the head of the environment
- One central person in the NSS - usually the National Statistician

Process 1: Approval of research environments

Example of information requested in the approval application are:

- Legal name of the research entity
- Legal address of the research entity
- Type of entity

    e.g. universities, research institutions, a research department within public   administrations, banks,

    a non-profit foundation, organization, private sector

- Web-page
- Organizational diagram
- Research performed / research profile
- Experiences in handling confidential data

Process 1: Approval of research environments

The core essence in the authorization contract could be:

- State the laws and regulations applying to microdata access
- State the confidentially rules applying to the access
- Legal and administrative sanctions
- Length of the authorization – some countries require the authorization renewed at regularly basis e.g. every five years other only in case of changes in the organization
- Name and title of the authorization manager
- The responsibility of the authorization manager

Process 1: Approval of research environments

Significant differences

- In some countries, the person signing the authorization agreement on behalf of the research environment undertakes personally to supervise associated researchers that the confidentiality of data is always kept

- Size of an entity (e.g. University or Department level)

- Whom is eligible to gain access (e.g. private companies, foreign research Institutions etc.)

- How often the authorization need to be renewed

- Option of filling out the request on-line

- Template can be downloaded from the internet

- Whether or not a list of authorized institutions are published on-line

Process 2: Approval of researchers

Common Sub-processes

A. Check whether the researcher complies with the criteria for access to microdata according to the NSO (e.g. educational level etc.)

B. Check whether the researcher is associated with an authorized institution

C. Signing of a Confidentiality Agreement

  • The authorization manager – usually the head of the entity

  • One central person in the NSS – e.g. the National Statistician, the NSO lawyer, the Head of the Research Service unit

Process 2: Approval of researchers

Example of information requested in the application are:

- Name

- Country of birth

- Country of residence

- ID number

- Document of institutional affiliation

- Curriculum Vitae

The core essence in the contract are:

- State the laws applying to microdata access

- State terms and conditions of the confidentiality rules applying to the researcher

- Legal and administrative sanctions in case of confidentiality breaches

Process 2: Approval of researchers

Significant differences

- Whom is eligible to gain access (e.g. Students, foreign researcher)
- Option of filling out the request on-line
- Template can be downloaded from the internet
- Some countries have the following additional requirements:
  - Document making researcher a volunteer employee
  - An on-line test on confidentiality rules has to be passed
  - Meetings in order to ensure the researchers understand the confidentiality rules

Process 3: Project approvals
Overall description of the project

Purpose

- To check whether the project complies with the criteria for access to microdata according to the NSO

- To select a contact person(s) in the NSO
  - Central with or without team division
  - From subject units

- To check that the need to know principle is followed - justify need

- To check the availability of data

- To check if additional approvals are needed in case of own data or external data

Process 3: Project approvals Overall description of the project

Example of information requested in the application are:

- Objective of the research project

- Description of population

- Data set required including external data

- Justify need to use microdata

- Research methodology

- Expected results

- Expected duration of project

- Forms of dissemination

Process 3: Project approvals
Overall description of the project

Signature requested on the project description:

- The authorization manager – usually the head of the entity (but not needed in Denmark)

- The principal researchers (but not needed in Denmark)

- One central person in the NSS – e.g. the National Statistician, a Director of the NSO, The Head of the RS

Process 3: Project approval

Significant differences

- Signature needed for approval
- Whom is eligible for approving the project in the NSO
- Core fields to be filled out in the application
- Option of update
- Option of filling out the request on-line
- Template can be downloaded from the internet
- Whether or not access mode and data-type has to be selected
- Whether or not exact variable list need to specified

Process 4: The data delivering process

Common Sub-processes

- Selection of variable (Counseling, from metadata description)

- Selection of access facility (on-site, remote, remote execution)

- Selection of type of data ('secure-use files' (Research room files); 'scientific-use files' (MUC), Public use files (PUF)

- Extraction of data

- Statistical disclosure control

- Production/extraction of metadata (Customized/from metadata)

Process 4: The data delivering process

Selection of variable

- +/- Consulting

- +/- Catalog of available data (including standard suppression/hierarchy such as e.g. age group etc)

- Metadata available (Survey/register level and variable level as well as standard suppression/hierarchy such as e.g. age group etc)

Process 4: The data delivering process

Selection of type of data

- 'Secure-use files' (Research room files)
- 'Scientific-use files' (MUC)
- Public use files (PUF)

May depend on sensitivity of data – some countries e.g. has classified their data into degree of sensitivity.

Selection of access facility

- **On-site** (Very different settings from big rooms to small individual cubes with fingerprinting recognition, camera etc)
- **Remote** (from work or home; fingerprinting or password protected)
- **Remote execution** (manual or automated - Researcher does not see data but get the output)

The decision of accesses facility the decision often depend on type of data (RR, MUC), category of data (e.g. Business data)

Process 4: The data delivering process

Extraction of data

- From a common data warehouse (with or without direct identifier)

- Entirely decentralized to the subject units (with or without direct identifier)

- Partly from a common data warehouse but enriched with data from subject units (Approach taken by DST in the early period of centralization – IDA database)

- One dataset or multiple datasets per project

- Extraction performed by programmer or academic staff

Process 4: The data delivering process

Statistical disclosure control (SDC)

- i) non-perturbative techniques, such as recoding and local suppression, which suppress or reduce the detail without altering the original data;

- ii) Perturbative techniques, such as adding noise, Post-Randomization Method (PRAM), micro-aggregation and shuffling, which distort the original micro-dataset before release;

- iii) Techniques that generate a synthetic microdata file that preserves certain statistics or relationships of the original files.

- +/- Approval of NSO e.g. a confidentiality committee

Process 4: The data delivering process

Production/extraction of metadata (Customized/from metadata)

- From a common source of metadata (manual or automated)
- Tailor made for each project by the subject units - if no common metadata structure is available

Process 4: The data delivering process

Price calculation and contract

Basically expenses are divided into the following components:

- Start up fee (Administration)

- Labor cost for advise, extraction of data, production of metadata (based on a standardized model or estimation of time used)

- Usage of IT tools (Programs, disk space, advise). Daily, monthly or yearly fee, fee based on usage (time in the RR, time logged on to the system), usage of specific programs, use of disk space)

- Output control

MONITOR AND FOLLOW UP!!!

Process 5: Output control

What is checked?

- All output checked / sample checked

- Automated or manual

- +/- Application form

Restrictions

All files

Restriction of output format e.g. SAS files etc.

Restrictions on output type e.g. +/- graphs

Restrictions individual type (e.g. number in a cell. The Nederland's require minimum 10 units in a cell Denmark only 3, max min values)

Number of output files that can be send home per day/project

Solution often depend on responsibility by the NSO

Process 5: Output control

Time spend by NSO on output checking

- From a few minutes to one hour per input

Time for output checking by the NSO

- Before or after the researcher receive the output
- From a few minutes to one hour per input

Time for recival of output check by the Researcher

- Instantly
- After a few days
- Before or after the researcher receive the output
- From a few minutes to one hour per input

Process 5: Output control

How does the researcher get the output?

- Mailed on placed on a server

Actions if breaches of confidently occur

- Administrative actions

- Legal action

Are rules that apply clearly described and published on the internet?