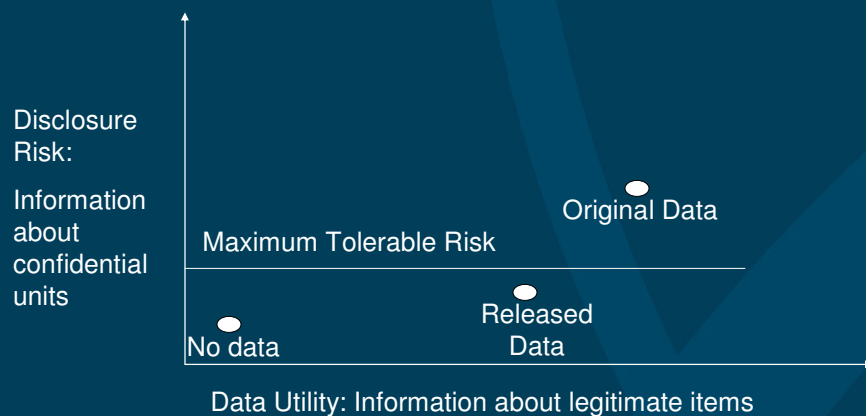


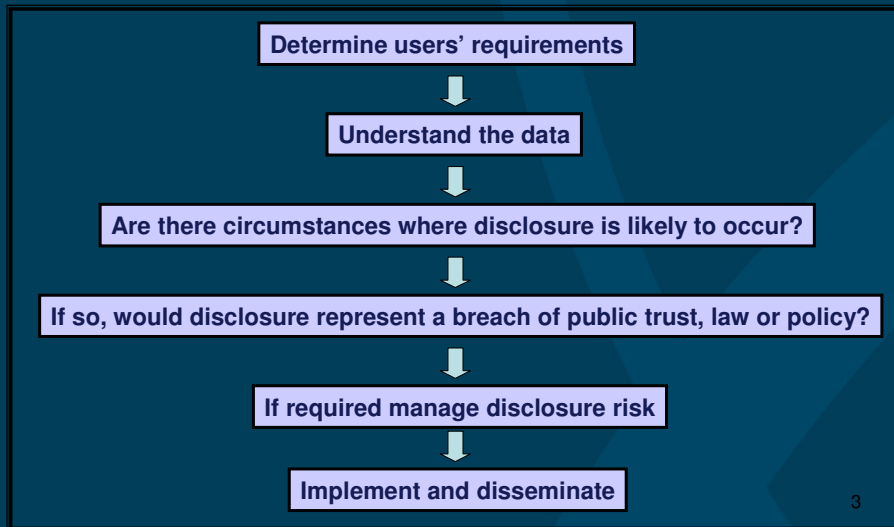
Balancing Risk and Utility – experiences in Official Statistics

Marie Cruddas
Methodology Directorate, ONS

The Issue



Framework for confidentiality protection



Determine users' requirements

- Identify main users
- Understand why they need the statistics
- How they will use them

Understanding the Data

- Data source
- Sensitivity
- Age and quality
- Identify statistical units
- Coverage
- Characteristics of the output

5



Risk Assessment

- What disclosure risks need to be protected against?
- What parts of the output pose a risk of disclosure?
- Consider a range of different examples
- If situation is likely to occur then protection may be required for the data

7

Risk Assessment

- Challenge – scenarios are theoretical
- Need for more empirical evidence:
 - Assess what information might be available to an intruder
 - 'Penetration' testing
 - Consideration of 'jigsaw' identification

8

Need for Statistical Disclosure Control

- Legal
 - National and International standards for statistics
 - Statistics and Registration Service Act 2007
- Ethical
- Practical

9

General Laws to Protect Confidentiality of Data about Living Individuals

Data Protection Act 1998

Common Law

Statistics and Registration Service Act 2007

Human Rights Act 1998

Freedom of Information Act 2000

10

Statistics and Registration Services Act, 2007

- Personal information held by the Board must not be disclosed
- Personal information means information which relates to and identifies a particular person (including a body corporate)
- That is if the identity of that person is specified in the information can be deduced from the information, or can be deduced from the information taken together with any other published information

11

Code of Practice for Official Statistics

- Published by UK Statistics Authority January 2009
- Principle 5: Confidentiality
Private information about individual persons (including bodies corporate) compiled in the production of official statistics is confidential, and should be used for statistical purposes only.
- National Statistician's Guidance: Confidentiality of Official Statistics

12

Ethical

- Moral viewpoint: it is the right thing to do
- Trust - *public perception*

13

Practical

- **TRUST** - if the people whom we ask to give us their data don't trust that we'll keep it safe.....
- They won't give us the data in the first place.
- No data = no statistics!

14

Managing disclosure risk

- Maximise utility while reducing risk to an acceptable level
- Statistical disclosure control (SDC) methods
- Control access
- Alternative methods for presenting data

15

SDC methods

- Perturbative:
 - Rounding
 - Perturbation
 - Record swapping
- Non-perturbative:
 - Recoding
 - Suppression

16

Controlling Access

- Virtual Microdata Laboratory – identified data
- Special Licence – access to disclosive (identifiable) data
- End User Licence – access to non-disclosive data
- Currently no public use microdata files

17

Microdata release panel

- Implements microdata release procedure at the ONS
- Centralised process for authorising access to all microdata
- Decision making body for requests to access data
- Decisions are recorded on a central database
- Involves senior officials with experience of releasing microdata

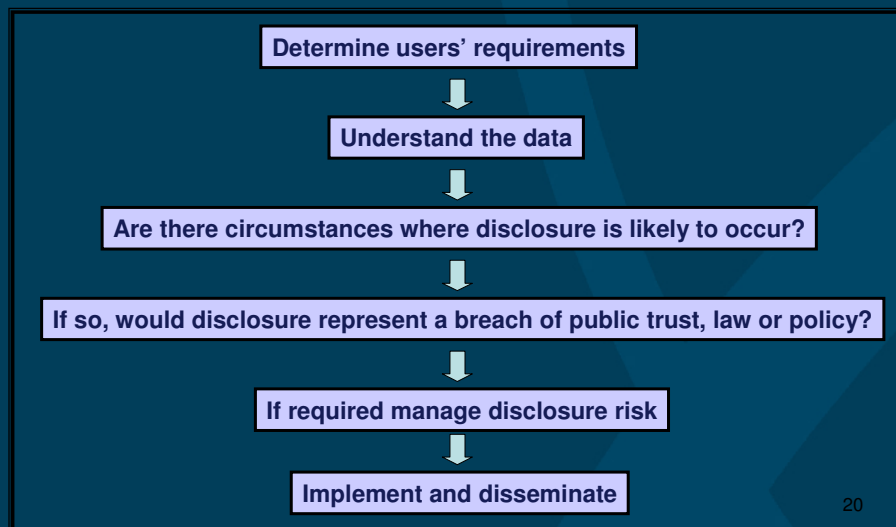
18

Implementation

- Solution (disclosure rules and protection method) should be clear, consistent and practical
- Limited information on disclosure control should be published, e.g. disclosure control method but not specific parameters
- Evaluation of utility

19

Framework for confidentiality protection



20

GSS Standards for SDC

- Developed to ensure consistent approach to SDC for different outputs
- GSS Standard for tables produced from administrative data sources
- GSS Standard for tables produced from surveys
 - Social surveys
 - Business surveys
 - Subsamples
- GSS Standards for microdata from social surveys

21

Future Challenges

- Increasing demand to release more data, in particular publicly
- Currently little experience of releasing public microdata
- Increasing concerns around privacy issues
- More data available to potential intruders
- Empirical evidence to back up decisions around managing confidentiality risks

22

Discussion and Questions

23