

Methods for preventing disclosure

If you have assessed that data poses an unacceptable risk in its current format, you should consider the following strategies:

- Grouping categories
- Aggregate to a higher level geography
- Aggregate to a larger population group i.e. Broad age bands instead of 5 year bands
- Aggregate tables across a number of years/months/quarters
- Rounding (rounding values to nearest 3 or 5 where values are small).
- Cell suppression – replace small values with a “X” or “...”

It is important to be open and transparent about the methods of suppression used in the metadata to allow the reader to understand the data.

If you have performed a risk assessment and are still unsure you can obtain further advice from the Information Management Team in Information Services.

Gemma Stephenson, Information Services

Tel: 01872 322763

Email: gstephenson@cornwall.gov.uk

Sources:

- Office for National Statistics
Disclosure Control Policy for Tables

Prepared by:

Abi Smith
Community Intelligence Team
Chief Executives Department

08 October 2012

If you would like this information in another format please contact:

Cornwall Council
County Hall
Treyew Road
Truro TR1 3AY

Telephone: 0300 1234 100
Email: enquiries@cornwall.gov.uk
www.cornwall.gov.uk

**Before you
publish
data...**

A bitesize guide to
preventing disclosure of
personal information



Introduction

This leaflet aims to help you understand how to understand the risks and implement strategies to avoid disclosure.

Statistics support a wide range of uses and provide essential information for government, business, academia and the community. Individuals have a general right of access to information held by public authorities as long as this does not contravene confidentiality constraints.

Types of disclosure

General attribute disclosure

This occurs when someone who has some information about people being able to discover details they didn't previously know, for example if someone was able to recognise their neighbour was claiming benefits by published details around age and gender.

'Motivated Intruder'

This occurs when data is combined with information from local sources to disclose further information. For example if someone was to look at conception data and identify a small number of women in a local area, and then uses another source to locate individuals.

Identification and self identification

This occurs where numbers are small and a person can identify themselves in a population and discover rareness or uniqueness that may cause harm or distress to the individual or cause them to feel 'exposed'. For example if an area had one crime broken down by characteristic of the victim.

Steps to understanding disclosure risk.....

Determine User Requirements

Understand main users and why they need figures. What is the demand and how will it be used?

Understand the key characteristic of the data and outputs

Think about which characteristics of the data may affect disclosure risks.

Understand circumstances where disclosure is likely to occur

Does the table contain small values, e.g. Under 5? Could the data be joined with other datasets to identify individuals?

Would disclosure represent a breach of public trust, the law or policy?

Consider Data Protection, Freedom of Information and other relevant legislation, information and management policies. The production and use of statistics depends on the cooperation and trust of citizens – would the release of data breach the privacy of individuals?

Risk Assessment

Once you have considered the risks of disclosure you need to make an assessment of the level of risk.

Low Risk: geography used and numbers mean that likelihood of identification is low.

Medium risk: if data has small populations or small geographies and contains small values that may identify if joined with other data or broken down by characteristic.

High risk: likelihood of identification is high and impact of disclosure would be great, e.g. Abortion statistics.

When considering risk it is also necessary to consider whether it is an acceptable risk or an unacceptable risk.

The legal guidelines state that statistical information can be widely and freely used provided confidentiality protection has been applied such that it is no longer likely that the information can be related to specific identifiable individuals.

The Office for National Statistics advise that disclosure risk is acceptable if the population is large enough that it would require a disproportionate amount of time, effort, and expertise to identify an individual. Disclosure risk is unacceptable if it allows a member of the public to claim to readily identify themselves or another data subject.

Cells identified as posing an unacceptable risk of disclosure are called 'unsafe'. Disclosure control methods should be used to reduce the risk by modifying unsafe cells.