**Official Statistics Use Case**
**Summary Description**

Government organizations often produce official statistics by conducting surveys and other forms of data collection. The data collected is typically tabular or relational and measures characteristics of individuals and organizations. The data is collected either directly, or through the use of authorized subcontractors, and is then managed by the producer through all lifecycle stages.

These data are disseminated in various ways [4]. First, derived "official statistics" – such as the unemployment index – are used for policy analysis and often have legal and regulatory weight in their own right. More complex derived tables and (less often) geographically aggregated data and/or cleaned microdata are made public, to support policy, public transparency (democratic governance), and for research. Finally, "data enclaves" are sometimes provided. In these enclaves one can, with permission, access the raw data within a secure environment controlled by the producer, generate tables and more complex model coefficients, and have these results vetted prior to publication.

Producers of official statistics are concerned with a range of disclosure threats, and tend to be highly conservative in disclosing data. The main threats include: identification of an individual (which typically is a violation of law, and threatens public confidence); inappropriate integration of data across multiple government organizations (which is legally constrained and raises concerns about government surveillance); and disclosure of competitive information about firms.

Producers of official statistics use a number of disclosure limitation mechanisms: legal mechanisms include data use agreements, which often carry significant legal (or even criminal) penalties; static statistical disclosure controls (perturbation, aggregation, suppression and recoding of data prior to publication); table-specific suppression and perturbation methods, and virtual data enclaves. Public analysis servers (which allow access to dynamically derived tables and maps) and synthetic data are also used, though less frequently.

The references identify detailed documentation of current methods used to address privacy in the U.S. Census [5], European national statistical offices [1], and U.S. federal agencies [2]. Also see [3] for coverage of emerging methods.

The emerging challenges in this area [6,7,8] are related to the velocity, data

integration, and increasing analytic sophistication. Agencies are pressured to release data faster and more cheaply, and to do it in a way that allows a greater range of analysis (including visualizations and data mining) and that provides estimates for finer time scales and geographic areas. Promising areas of research include: adaptive data collection, which continuously integrates data collection, management, and analysis to minimize sampling error and cost; integration of "organic" big-data resources to provide continuously updated small-area estimates; and management of versioning, provenance and authenticity from collection through publication.

*References:*


[1] ESSNET, Handbook on Statistical Disclosure Control
http://neon.vb.cbs.nl/casc/SDC_Handbook.pdf
[2] FCSM, Statistical Policy Working Paper 22
http://www.fcsm.gov/working-papers/SPWP22_rev.pdf
[3] Journal of Official Statistics: www.jos.nu
[4] Willenborg & Waal, *Elements of Statistical Disclosure Control*
http://www.amazon.com/Elements-Statistical-Disclosure-Control-Statistics/dp/0387951210/
[5] Census Confidentiality and Privacy, 1790-2002,
http://www.census.gov/prod/2003pubs/conmono2.pdf
[6] Workshop on the Future of Social Science Data Collection, National Academies
http://www.nap.edu/webcast/webcast_detail.php?webcast_id=456
[7] Robert Groves, 2011, "Future of Producing Social and Economic Statistical Information" (3 part series), Director's Blog.
http://directorsblog.blogs.census.gov/2011/10/18/the-future-of-producing-social-and-economic-statistical-information-part-iii/ ,
http://directorsblog.blogs.census.gov/2011/09/20/the-future-of-producing-social-and-economic-statistical-information-part-ii/ ,
http://directorsblog.blogs.census.gov/2011/09/08/the-future-of-producing-social-and-economic-statistical-information-part-i/
[8] Novak, K., Altman, M., Broch, E., Carroll, J. M., Clemins, P. J., Fournier, D., Laevart, C., et al. (2011). Communicating Science and Engineering Data in the Information Age. Computer Science and Telecommunications. National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13282