

Vertical Integration in the Statistical Knowledge Network

Deborah Barreau, University of North Carolina at Chapel Hill, barreau@ils.unc.edu

Chang Su, University of North Carolina at Chapel Hill, suc@ils.unc.edu

Kristina M. Spurgin, University of North Carolina at Chapel Hill, kristina@infomuse.net

What is the Statistical Knowledge Network?

Building the Statistical Knowledge Network (SKN) is a goal of the GovStat project. The SKN would make statistical information across agencies and levels of aggregation seamlessly available to people in a context that would help them understand what they found and relate it to their needs.

How does Vertical Integration fit into the SKN?

To provide access to statistical information, the SKN must account for the vertically integrated relationships of the organizations that produce it. Representation of vertical integration in the SKN would make clear the ownership of and context for information in the network and would allow users to search for the information they need across agencies at all levels.

Vertical Integration in Agricultural Statistics Information

National agricultural statistical information is a product of cooperative agreements and collaboration across agencies at the state and federal levels. This allows for sharing of responsibilities for data gathering, aggregating and disseminating information, and providing services. The agencies involved in cooperative arrangements to accomplish these tasks include: US Department of Agriculture (USDA), National Agriculture Statistical Service (NASS), state statistical offices (SSOs), and state departments of agriculture (SDAs) and/or state university agricultural extension services. The SSOs function as field offices for NASS in the states, and their funding and staffing are shared by the federal and state agencies.

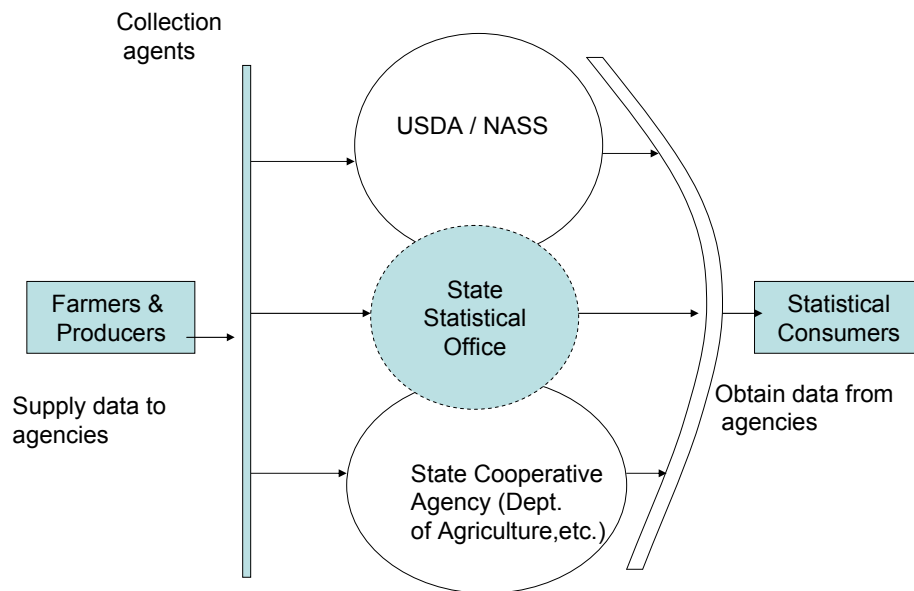


Figure 1. Process of Vertical Integration in the National Agricultural Statistics Service

This research examines the relationships between these Federal and state agricultural agencies as they are characterized on the Web in order to understand the vertical integration of information among them and the implications it may have for building the SKN.

How are vertical cooperative relationships among agencies characterized on the World Wide Web?

To answer this question, links connecting the websites of USDA, NASS, SSOs, and SDAs were examined to assess whether and how the agencies refer to each other and how they represent statistical information. A link from an SSO to the SDA (and vice versa) was counted as an acknowledgment of a cooperative relationship between them. Sometimes this relationship is described and the link contained within the textual description, but other times there are hyperlinked images void of context. Analysis of the USDA and NASS sites identified eight broad functional categories of content, including statistical information. SSO and SDA site content was then compared against these categories.

Findings:

- Vertical relationships between agencies are represented more consistently from the top down than from the bottom up. A path from the NASS home page through the SSO home pages to the SDA sites was found to 96% of the states, but only about 25% of the SDA sites linked upward to the SSO page.
- Inconsistent URL construction and varied content categorization among the states pose challenges for developers trying to design automated tools that identify and index statistical information comprehensively. It will be difficult to identify relevant directories and topics reliably without manual intervention.
- The terminology used to label links leading to statistical data varies greatly, as does the placement of this information within the site. The most meaningful information concerning vertical relationships and describing statistical content is usually found below the top level of a site's hierarchy, and it is organized and labeled inconsistently across sites.

Implications and questions for further research:

The diversity of vertical relationships among agencies is reflected in linking patterns and site structures. We suspect that navigation of these different agency sites for statistical information access may be frustrating and difficult for people. Future studies should focus on statistical consumers, both within and outside the agencies, to discover how interagency relationships and data provenance may influence information seeking and use.

One goal for the SKN is to facilitate information sharing among agencies. For the USDA, the SSOs are a mechanism for vertical data collection, reporting, and sharing between Federal and state agencies. This is reflected in the SSO websites as well, as they provide the most links (both up and down) to the statistical data and the cooperative agencies responsible for it. Studies and comparisons of linking mechanisms in other agencies may lead to notions of best practices for vertical integration.

Finally, we need to explore how the use of vertically integrated statistical data may influence the organizations and the processes that produce it.