The USDA Data Dissemination System

James E. Horsfield

Agricultural Economist

Economic Research Service

U.S. Department of Agriculture

Washington, DC

Outline

Figures

Keywords

Agricultural data system

Agricultural information system

Agricultural Marketing Service (AMS)

Client-server systems

Data access systems

Data base systems

Data bases

Data delivery systems

Data dissemination

Data dissemination systems

Data systems

Economic Research Service (ERS)

Electronic dissemination

Information systems

Institutional change

Integrated Services Digital Network (ISDN)

National Agricultural Statistics Service (NASS)

Research data bases

Structured Query Language (SQL)

Technology

THE USDA DATA DISSEMINATION SYSTEM

Introduction

The program on the opening day of this conference highlighted the timeliness of a fresh look at the agricultural and rural data system. The ever-changing dimensions of agricultural and rural problems, exciting new methods and technologies for social science data collection and research, and changing needs for economic, social, and natural resource data call for a dynamic system and "new directions".[1] However, as we attempt to reach a consensus on these new directions, we also need to recognize that the existing system is highly institutionalized and that institutional change will be needed to effect a permanent change in the direction of the information system.

The mere mention of an institutional process doubtless creates a feeling of helplessness. One might envision a steamroller moving inexorably down the street, fixed in its direction, and capable of crushing any impediment to its movement, let alone a mere committee or task force that might get in the way. Shouting at the driver to steer this way . . . or that way . . . is unlikely to change its direction and fighting the driver for control is a job for professionals in the sport. Nonetheless, change is possible. We now know from chaos theory what Poincaire knew at the turn of the century: ". . . that slight variances in the initial conditions produce very great differences in the final phenomena."[2] This suggests that we need only identify where to place a few pebbles, a few good ideas, in the path of the steamroller to realize a gradual, but permanent, change in direction. Many of the "drivers" of important institutions are at this conference. Perhaps they will sense a bump or two in the road.

---

[1] See Chapters 2-5.

[2] Quoted by Hugh Kenner in "Predicting Chaos", *BYTE* (June 1989):342.

Given the importance of the U. S. Department of Agriculture (USDA) and its constituent agencies in determining the content and direction of the rural information system, it is appropriate that we examine carefully the USDA data dissemination system. This key link between researchers and the Department's data collection system is, in the words of a popular television commercial, everything you want in a data dissemination system--and less. However, it is part of such a durable institution that a strategic pebble or two today could result in a change in direction with benefits extending well into the next century.

The USDA Data Dissemination System

Data Dissemination in the Agricultural Information System

James Bonnen (Bonnen 1975) developed the schematic in Figure 6.1, which represents an idealized view of an agricultural information system. The

Figure 6.1 here

interface between "data output" and "interpretation and analysis" in this framework is expanded in Figure 6.2 to include a data dissemination system and its component subsystems. These subsystems are:

1. A data base to store data from the underlying data system for dissemination. This data base can be a printed report or a sophisticated computer data base.

2. An access subsystem. This component is crucial for effective data dissemination. Data users need to find what is available and data developers need to know the target audience(s) for their output to design effective and timely delivery systems.

3. A delivery subsystem. This is often seen as the totality of a data dissemination system because it is often the only visible or active system.

Figure 6.2 here

USDA Data Dissemination Systems

In USDA, this ideal data dissemination system is many systems, where each reflects some combination of data output, user group, and dissemination technology. However, grouping them in three general types based on the relative importance of the three subsystems noted above gives some insight into the Department's overall system.

Data Delivery Systems: Perhaps the most familiar USDA system is the type shown in Figure 6.3, where the data delivery component is especially prominent. This is characteristic of the Department's dissemination systems

Figure 6.3 here

for agricultural statistics and market reporting where very timely delivery of data to very large user groups is essential. The Market News system of the Agricultural Marketing Service (AMS), for example, provides nearly real-time reports of supply, demand, and prices to market participants around the country. Similarly, the estimates released by the Agricultural Statistics Board reach thousands of users in minutes.

Electronic dissemination of time-sensitive USDA data has a long history beginning with retransmission of data through the private broadcast and wire services. Today, time-sensitive data flow directly to the private dissemination systems and users through the Computerized Information Delivery System (CIDS) and the satellite communication system for market news.[3] These delivery systems also demonstrate the enhanced access and data base capabilities provided by automation. Subscribers to CIDS, for example, can use a series of menus to locate and download all current estimates on the system.

Data Access Systems: The second type of USDA data dissemination system is characterized by its emphasis on an access system to serve user needs for

_____

[3]CIDS was originally called the USDA Electronic Dissemination of Information (EDI) system.

timely use of a very broad data base of current information. These access-driven systems are perhaps the most common, and least understood, part of the overall data dissemination system. They are easily identified, however, by the passive role of the Department in the dissemination process. Many USDA agencies, for example, maintain published and unpublished data that are available on request, and many agencies promote this service. Actual data delivery, however, must be initiated by a specific user request. These systems often supplement more active data delivery systems by providing access to data for users not in the target group for delivery. They also provide user access to data collected for program administration that would not ordinarily be included in a data delivery system.

Access systems can be improved substantially with automation and electronic dissemination of "metadata"--the catalog and directory information needed by users to identify what data are available and where. The FINDERS advisory system for ERS information and data is one attempt to automate this type of information. Experience with FINDERS will be used to evaluate the distribution of knowledge bases combining data bases with intelligent access tools. Compact disk, read only memory (CD-ROM) technology also shows considerable promise as an automated delivery system with powerful user tools for search and retrieval. ERS is working with the National Agricultural Library to assess the potential of this technology for dissemination of its information and data.

Data Base Systems: The third type of USDA data dissemination system is particularly interesting to researchers. This type is characterized by investment in comprehensive data bases to serve user needs for analysis and research--often as a spin-off from data base development for USDA program administration. While users need functional access and delivery systems, preferably automated, for these data bases, the overriding user need is for data bases with depth, breadth, and quality. A few USDA data dissemination systems address these needs. The National Agricultural Statistics Service (NASS) and ERS offer machine-readable data bases on a variety of topics, including many of interest to agricultural economists and rural sociologists.

Other comprehensive data bases are distributed as statistical bulletins or, in a more piecemeal way, as appendices to USDA reports.

Changing Directions

This typology of data dissemination systems highlights the form and timing of USDA-data user interfaces, but it is essentially neutral in its description of either the conceptual content of the larger data system or the information and data needs of decisionmakers. This suggests that changes in USDA data formats, in the timing of dissemination, and in the costs and benefits of alternative dissemination systems must be addressed independently of data collection issues. Thus, pebbles must be placed carefully in front of the right institutional steamrollers.

Status and Trends

Current USDA Systems

A fair evaluation of the USDA data dissemination systems described above would likely assign grades of "A" to the time-sensitive systems, "C+" to the access-driven systems, and "C-" to the data base systems. The Department's delivery systems for time-sensitive data deserve a high grade. The value of timely data to market participants has stimulated much public and private investment in the delivery system for these data and the CIDS and Market News services reflect state-of-the-art technologies. Areas for improvement include adding more of the Department's time-sensitive data to CIDS and planning for the next wave of computer and communication technologies. The shortcomings of CIDS for research data base maintenance will remain, however, including relatively high data transmission costs, the dissemination of text files not adapted to use with analytical software, and the need for continuous analysis of current data series to incorporate revised data or conceptual changes in the data base.

The lack of comprehensive directories and indexes for USDA data is the major impediment to better user access in the second type of dissemination system.

The plethora of agencies, reports, subject areas, and timetables for the Department's data outputs overwhelms potential users, even those well-informed about USDA. Use of these resources for research is similarly difficult and is complicated further by the need to get a variety of reports from multiple sources in USDA, each with their own subscription systems and pricing policies. Improvements could include both the development of a Departmental data directory or locator system and more coordination on the development of USDA-wide data bases for electronic dissemination. As noted earlier, CD-ROM technology shows considerable promise for delivering enormous quantities of information in a compact medium with powerful user access tools. An expanded version of the annual *Agricultural Statistics* report is an example of a potential product.

Data base dissemination systems are the weakest component of the USDA system. In part, this reflects the Department's important role in providing data as a flow resource and the concentration of resources on this activity in the data-producing agencies. It also reflects the lingering influence of the User Fee Statue,[4] which, despite its name, effectively prohibited the dissemination of data bases on a cost recovery basis until the Food Security Act of 1985 authorized the Secretary to prepare materials for electronic dissemination and to charge user fees to cover the costs of their preparation. The ERS data products program is an example of the beneficial effects of the 1985 Act.

The Future Data Dissemination System

Trends in computer hardware, software, and communication technologies point toward an ever-expanding role for data bases as the core component of data dissemination systems. These trends also imply a gradual merger of the three types of dissemination systems discussed earlier into one system built around data base technology. Figure 6.4 illustrates the data dissemination system of the future--a system where a nationwide integrated digital services network (ISDN), which will replace the current telephone network, links personal

---

[4] 31 U.S Code 9701.

computers to data bases down the hall or across the country.

<center>Figure 6.4 here</center>

The client-server data base software technology implies a generic data access and delivery system *for all data*. Thus, a request from a personal computer analysis system for economic data is functionally equivalent to a similar request for demographic data or nutrition data. The actual physical location of the underlying data base(s) may be unknown to a user.

These emerging technologies have many implications for data dissemination policy. The boundaries between public and private roles in the dissemination of Federal statistics, for example, are likely to become even more indistinct than they are now. Also, the ability to obtain and use data from many public sources about individuals and firms will heighten concerns about privacy. The costs of the new technologies, the training needed to use them, and access to the communication network are additional concerns. These factors may segregate the user community into the "data-rich" and the "data-poor" with adverse consequences for incomes and earning potential. One may expect, however, that the USDA data dissemination systems of the future will retain the equal-access principle that has guided development of current systems.

<center>**The Data Base for Research**</center>

<u>A Pessimistic Assessment</u>

The availability of USDA data bases for agricultural and rural social science research will likely improve as data bases become more important for data dissemination. However, the immediate availability of a comprehensive agricultural economic data base is unlikely. There are three reasons for this pessimistic assessment. First, USDA data dissemination systems are probably close to an equilibrium with respect to current assessments of user needs. This means that USDA contributions to a comprehensive data base will be forthcoming only as data base technologies are incorporated in existing dissemination systems. Second, no single agency such as ERS or NASS has either the responsibility or the resources to develop a unified Departmental

data base. Third, only 96 of the 225 datasets identified in the survey conducted for this conference are datasets provided by USDA.[5] Therefore, a research data base built only on USDA data bases would include *less than half* of the data identified as important for rural social science research. Such a data base would not include, for example, the datasets of the Census or of the Bureau of Labor Statistics.

Shifting the timetable forward for the use of data bases in USDA data dissemination systems is possible, however, and the rural social science community can play a pivotal role in changing the Department's perception of user needs for electronic data, especially in data base form. Unfortunately, the inexpensive, straightforward approach of lobbying USDA agencies to direct more resources toward meeting the needs of researchers in this area is unlikely to succeed because, first, the research data users are a small minority of all data users and, second, building and maintaining a economic data bases is both expensive and demanding of professional expertise--data sources must be monitored continuously, historical data must be vigilantly updated, and data base management systems must be adapted to rapidly changing technology. Given this balance of benefits and costs, the agencies, including ERS, will make only marginal adjustments in spending for data base development.

Changing USDA Priorities

A more fruitful approach (and a pebble not so easily crushed) is to articulate the growing importance of electronic information and data to farmers, ranchers, and rural communities. This could result in accelerated development of USDA data bases through a shift in the Department's appraisal of benefits and costs. If, for example, the Department was convinced that information and data are needed to improve farm productivity, to assess worldwide market opportunities, and to incorporate rural areas fully in a national computer-communications network, then the needs of USDA and data users would converge

---

[5]See Chapter 5.

on data dissemination using data base technologies. Moreover, this convergence would likely occur sooner than anticipated in current plans for improvements in USDA data dissemination systems--giving additional impetus to faster data base development.

Regardless of improvements in the USDA data dissemination system, however, the development of a truly comprehensive research data base is unlikely given existing institutional arrangements. There are many "institutional" data bases including economic and social science data bases offered by many government agencies, commercial firms, and universities. None of these were developed to serve the collective needs of rural social scientists and, given the public goods nature of data, there is no incentive for any one system to address these needs.

A Cooperative Approach

A possible solution to this dilemma is to coordinate existing data base development and maintenance activities in a cooperative effort involving researchers, USDA agencies, and commercial concerns. This would take advantage of the emerging data base technology described above to develop and maintain very large, decentralized time-series and cross-sectional data bases on a national computer network. One approach is to establish a pool of cooperators around a subject area to share data maintenance responsibilities in exchange for access to the pooled data resources through the network. Multiple, interrelated pools could follow with each pool having its own data maintenance arrangements, but with all groups using a common data management system. Development and operating costs for these cooperative data bases could be covered by both in-kind contributions of data, analyst time for quality control, and computer time but also by subscription fees based on access and use. One might also expect that the value of the resource would attract non-member users who would pay unsubsidized prices for data services.

Conclusions

Technological change presents new problems and new opportunities for data dissemination. In USDA, this will likely result in continuing evolution of existing data dissemination systems toward ever faster and cheaper modes of access and delivery. However, these systems may not advance the technology of agricultural and rural social science research unless the research community becomes actively and continuously involved in shaping the public institutions for data dissemination. This involvement can take the form suggested above of articulating user needs with respect to the form and timing of data dissemination, especially when the interests of broad groups of users are affected. It can also take the form of developing new institutional arrangements that better serve research needs. An appraisal of the feasibility of cooperative data bases, for example, would be a good first step toward planning the research data utility of the future.

References

Bonnen, James T. "Improving Information on Agriculture and Rural Life."
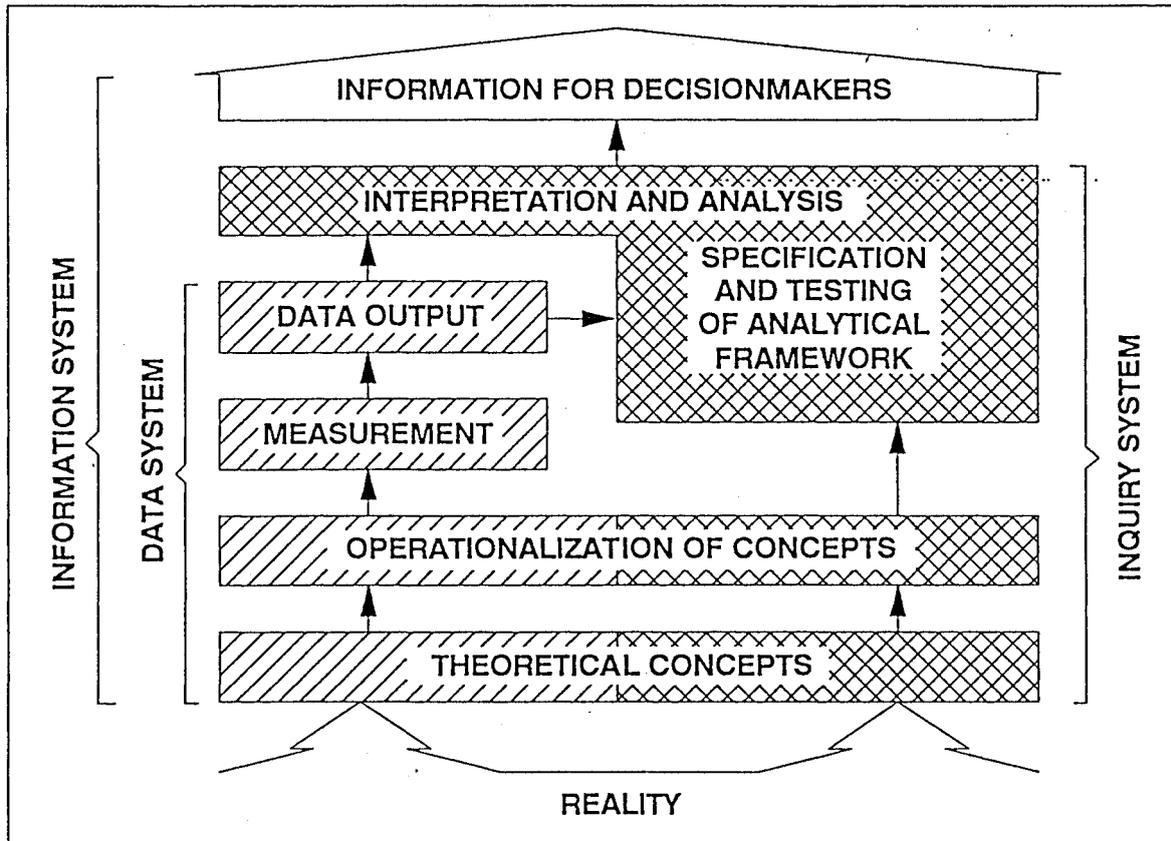*American Journal of Agricultural Economics.* 57(1975):753-63.

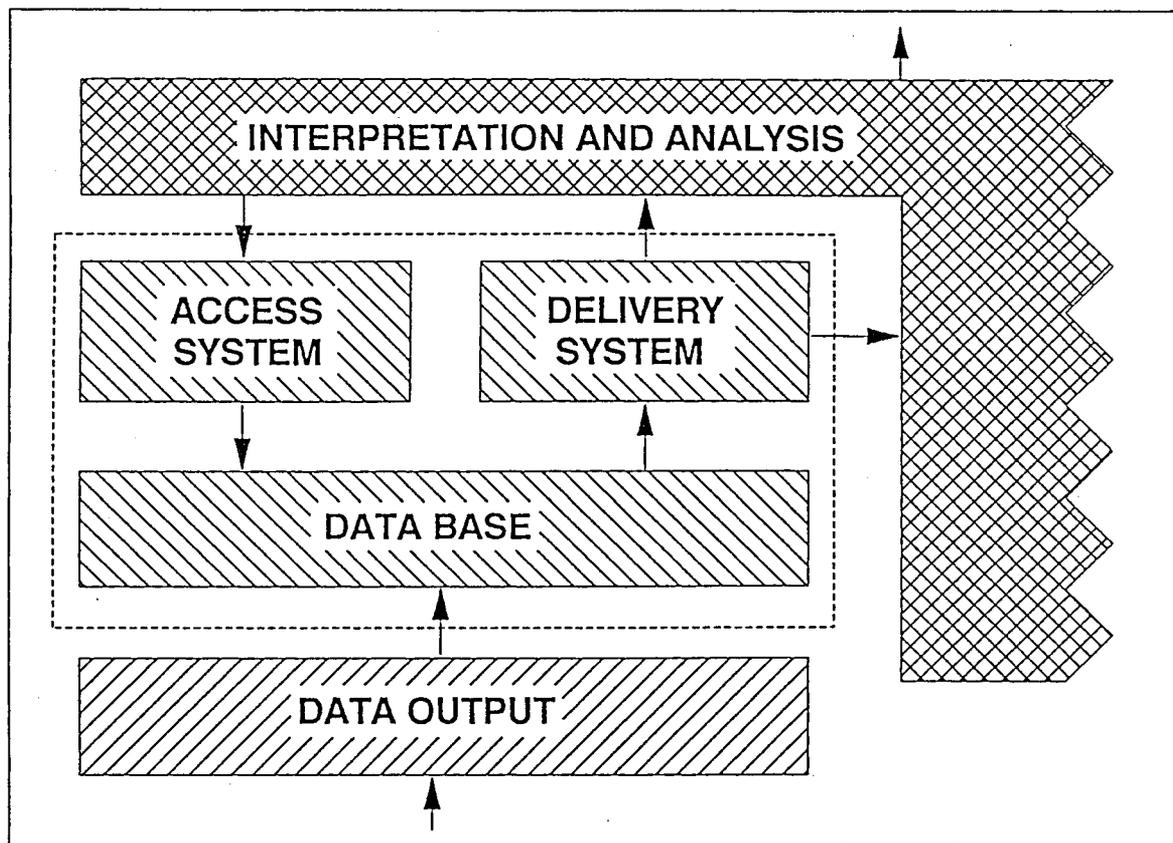Figure 1. An agricultural information system

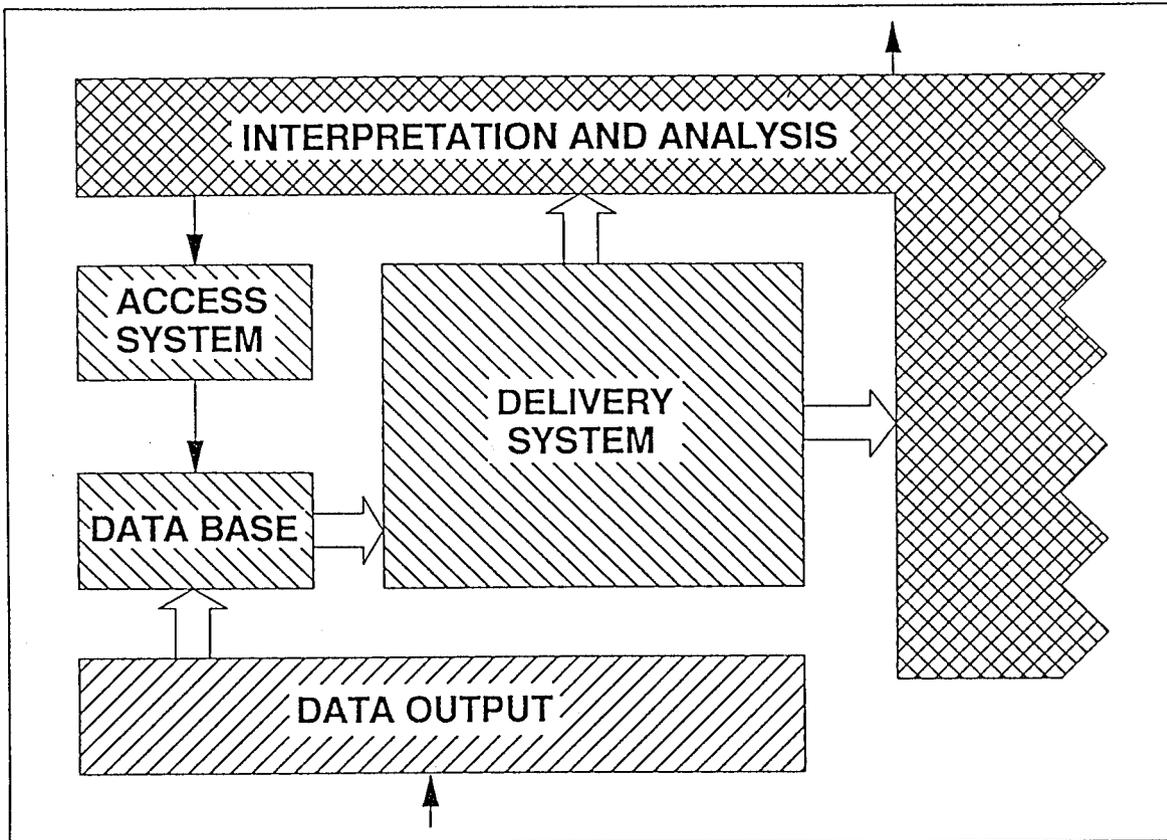Figure 2. Data dissemination in the agricultural data system
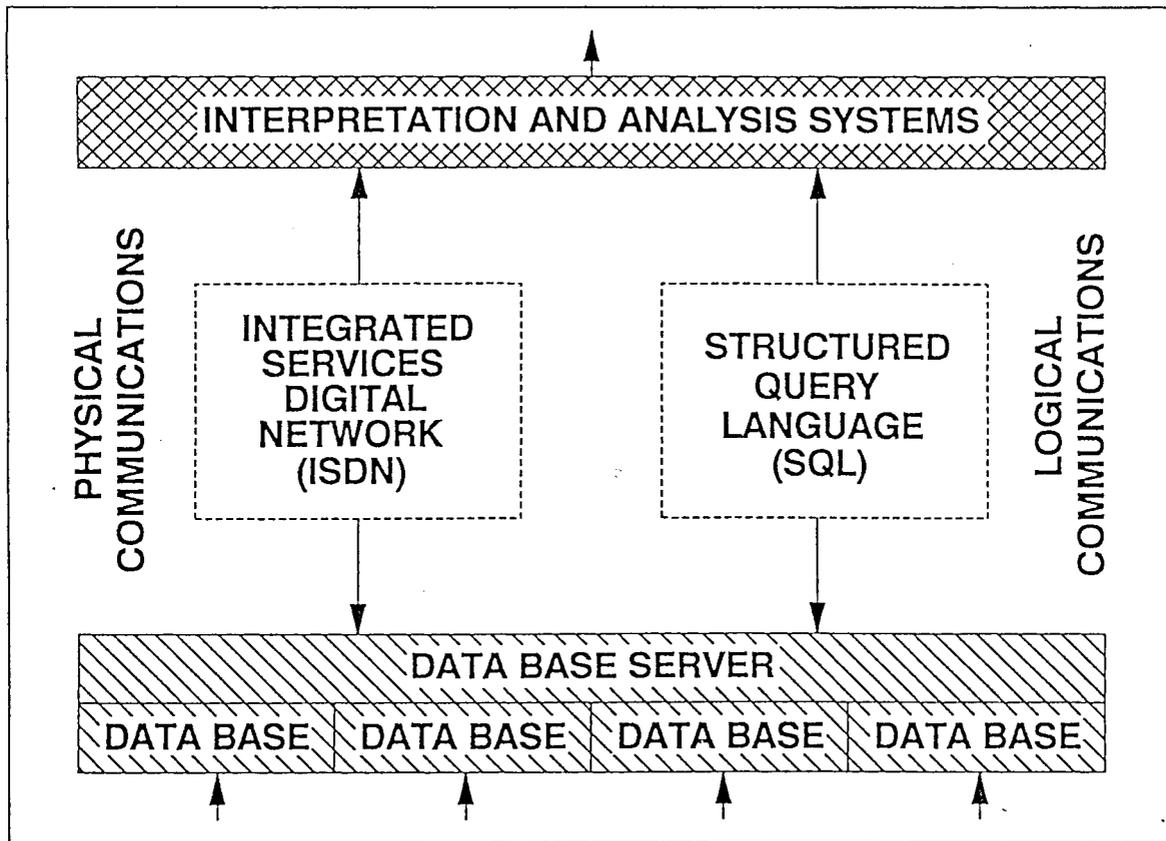
Figure 3. A delivery-driven data dissemination system

Figure 4. The data dissemination system of the future