Insider's Guide to Using Information in Government

Case

Moving from paper to the Web:

how the Council on Children and Families is transforming a static information resource into a dynamic one

Moving from paper to the Web: how the Council on Children and Families is transforming a static information resource into a dynamic one

Transforming an annual statistical publication into a continuously available information resource on the Web means changes in policies, data, and processes, as well as new technology.

Introduction

Services to benefit children and families cost billions of dollars in New York State each year. These expenditures provide an extensive array of services, including early childhood education and child care, services for the disabled, mental health and chemical dependency treatment, employment and skills training, and health maintenance and wellness.

The New York State Council on Children and Families and its member agencies have developed a set of interagency goals, objectives, and outcome measures for children's health and well-being known as "Touchstones" indicators. The Council coordinates the compilation of data from the member agencies in support of the indicators and makes them available for policy development, planning, and accountability at the state, county, and community levels. This data has traditionally been collected and distributed in print form in a publication known as the *Touchstones Data Book*.

Over time, both member agencies and other users of the book expressed interest in having the data be more accessible, possibly over the Web. As a result, the Council sought and received funding from the US Department of Health and Human Services, as well as a partnership with the Center for Technology in Government, to develop a publicly accessible Web-based information repository to complement the *Touchstones Data Book* by making data more available and more useful to policy-makers, service providers, and communities.

Project environment

The membership of the Council on Children and Families consists of 13 state agencies that provide programs and services to children and families. A small staff carries out the programs and priorities of the Council. One of the Council's many programs is the Touchstones Indicators Initiative. The Indicators Initiative grew out of recognition that in order to improve outcomes in the area of health and human services, it was necessary to shift to a new paradigm characterized by prevention, early intervention and youth/family involvement. Further, to increase the effectiveness of the various systems, the agencies embarked on an effort to develop a common set of measurable goals and objectives that lead to improved outcomes for children and families. New York State touchstones emerged from these efforts. New York State is a leader in the development of state-level indicators of the health and well-being of children and families. Such vital indicators as the numbers of low birth weight babies, children living in poverty, high school dropouts, and teen pregnancies that occur in the state every year can be derived from these data sets. Significant efforts have been invested to develop the indicators for New York and to identify the state agency data sets that would provide the necessary information for each indicator. The Council coordinates the compilation of data from the member agencies and makes it available as a tool for policy development, planning, and accountability at the state, county, and community levels.

The indicator effort in New York State is linked to a national effort supported by the Annie E. Casey Foundation which publishes a national *KIDS COUNT Data Book*. Policy-makers and citizens use *KIDS COUNT* as a major source of information about the educational, social, economic, and physical well-being of children. The Foundation funds a network of state-level *KIDS COUNT* projects that provide a more detailed community-by-community picture of the condition of children. The Council produces the New York State *Touchstones Data Book* as part of this network.

Each release of the *Touchstones Data Book* comprises 5000 copies for distribution in NYS. The compilation of data proved very useful to the children's service community and people began to seek more and more copies of the book. They also began to advocate for more data and for more capability to manipulate the data. Users wanted more timely distribution and greater ability to conduct cross-indicator comparisons and

longitudinal analysis. The Council recognized the limitations of the print medium and began to consider alternative distribution vehicles that would complement the *Data Book*.

Working with the Center for Technology in Government (CTG), the Council developed the Kids Well-being Indicator Clearinghouse (KWIC) to overcome the limitations of access and the constraints on analysis inherent in the printed document. The book provided the results of many different analyses of the data, but these analyses were predefined and, of course, the medium did not allow for additional real-time analysis. To overcome these limitations, a Web-based repository was created to provide users access to the data available in the book, as well as to other data that was not included in the book due to space constraints. An Internet-based Clearinghouse approach was used to provide users access to related materials about children's indicators, to related efforts in NYS and elsewhere, and to tools that allow them to go beyond the analyses provided in the book to conduct the cross-indicator and longitudinal analyses they desired.

Developing a governing structure for KWIC

The Touchstones book and the planned Web site required data in two very different formats, demanding different decisions about how the data would be stored and presented. From the beginning of the effort, important questions emerged. For example, which process (book publication or Web update) would take precedence for data compilation? Would the Web site be updated prior to the book's next publication? The Web site also raised considerations about how text would be developed or changed, since print format and Web format can and should be different. Then there was the issue of how information would be stored. One available version was stored in a Microsoft Access database and multiple queries were used to change the views of the data. While this method was productive for the short-term, long-term strategic decisions needed to be made as new data elements were defined and added to the database.

In the first meeting to develop a vision for the project, the team members recognized that both policies and structures would be needed to guide the creation and ongoing management of KWIC. The team brainstormed the characteristics of a suitable governance structure, identified resources that would be required to put the structure in place, and the barriers they might expect to encounter. These and other similar issues prompted the group to identify the need for a governance body, reporting to the Council's leadership, to work in concert with the Data Book staff to ensure the integrity and consistency of the data in both media.

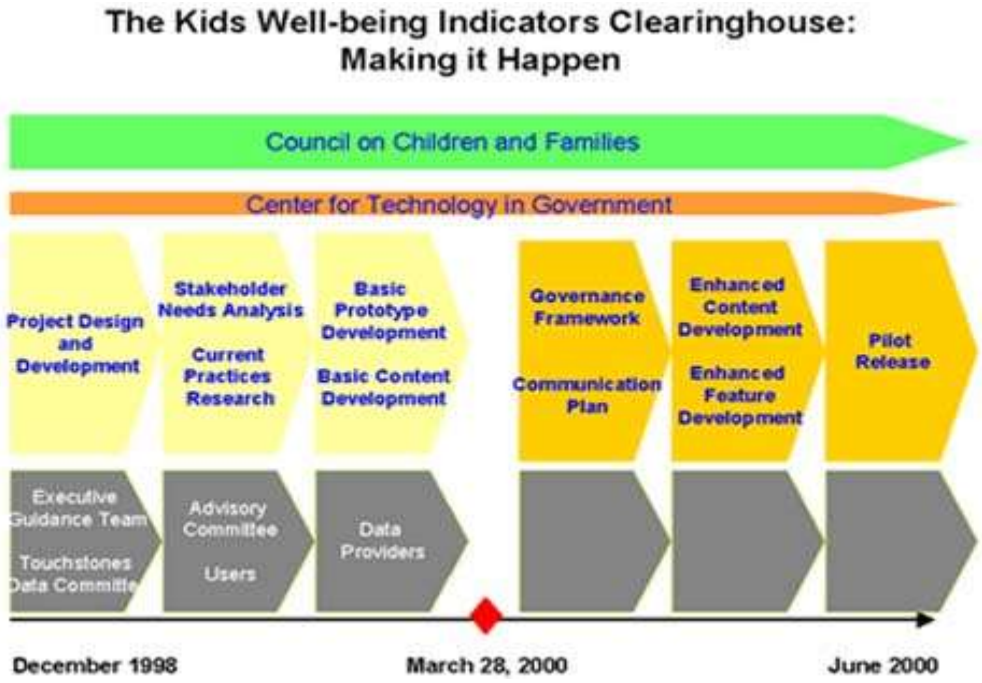
The group proposed an action plan for a governance body that would work within the existing management structure of the Council, and focus on two main responsibilities:

1. Build a framework for the development, direction, upkeep, and evaluation of the Web site and its content, policy, and management
2. Oversee and manage the major interagency issues related to the governance and operation of KWIC

Within these two broad themes, the governance body would address such topics as:

- clarification of data ownership
- establishment of minimum standards about data quality and meta data
- establishment of business rules regarding operations, data inclusion, and fitness for use
- establishment of a communication policy for interaction with agency partners as well as users
- establishment of a communication plan

The group then prepared a draft governance document and presentation to the Council's member agencies. The presentation described the planned KWIC prototype and outlined what needed to occur in order to reach pilot status. The figure below was used in this discussion. It outlines the tasks and the main players to be involved, the need for a governance framework, and plans for ongoing development. Three sub groups discussed and developed recommendations and lead roles for completing the three main areas of work: instituting the governance structure, continuing development of the repository, and releasing the pilot Web site.



User needs and information gathering stages

To better understand user needs, the project team established an advisory committee drawing from the 13 member state agencies, local governments, community-based organizations, research institutions, and others with an interest in children's health and well-being. This advisory group played a key role in helping to define the Web site's contents and features.

The advisory committee, along with the members of the Council's standing Executive Guidance and Data Committees, engaged in a series of facilitated group decision meetings to elicit the full range of needs and uses for indicator data. These one- and two-day meetings were designed to identify, explain, and prioritize uses for indicator data, desired types of data, and associated information and other resources that would help increase the usability of the data.

The same group of users and data suppliers identified the supporting resources that should be available on the Clearinghouse. These resources fell into several categories: search tools, data manipulation tools, data presentation tools, research and evaluation reports, federal statistical data, education and technical assistance programs, conference announcements, grant opportunities, directories, and electronic discussion lists.

The results of these discussions, plus research on similar projects elsewhere, guided the development of KWIC. The project team reviewed user needs and suggestions by categorizing them according to specific features or functions of a Web site. For example, users requested "meta data and information about the data collection methodology" and "data descriptions and caveats to prevent misuse and miscorrelation." These requests were translated into three features of the Web site. One is the requirement for comprehensive meta data. The second is a separate section of the Web site that outlined how to properly analyze the data. The third is more of an overarching principle than a feature, calling for the design team to prevent users from making invalid comparisons of non-comparable data sets. To do this, the design team had to "technologically" block comparisons that would produce invalid results due to different population sources, time frames or other dimensions.

These features and functions were then placed into a framework which rated each one as modest, moderate, or elaborate. The rankings considered how technologically complex or advanced the feature was from both the user's standpoint and the developer's standpoint. For example, from a user's perspective, query capabilities were an elaborate enhancement of capability over the current practice of using the book. However, from a developer's standpoint, query capabilities fell into two categories - pre-

defined queries were in the moderate category, whereas ad-hoc query capabilities were considered elaborate due to the technological challenges they presented.

The project team then conducted current practice research in existing Web-enabled government clearinghouses and indicator Web sites. They categorized the features of each reviewed site into a similar framework, and found that a majority fell into the modest category as defined by this framework. The team also reviewed advocacy sites to identify additional functionality and added this to the overall framework.

Based on the user needs assessment, user capabilities, current practice research, and review of available data and associated files, the project team decided on a moderately featured Web site for the initial pilot, with elaborate functionality planned for future versions. The CTG technical staff designed an underlying structure for the Web site and presented it to the Council. The structure included both information housed on the Web site itself and links to data and resources that resided elsewhere on the WWW. Development of the prototype Clearinghouse then proceeded.

Meta data helps make data usable

The Council collects data annually from each of its member agencies to produce the *Touchstones Data Book*. Meta data in the context of the printed book was limited to footnotes. As the Web site design took shape, different meta data requirements emerged. The user needs assessment had identified two major areas of need ("meta data and information about the data collection methodology" and "data descriptions and caveats to prevent misuse and miscorrelation"). As the team developed the new database these needs rose in importance.

Meta data is defined as "data that describes data." The user community asked for better meta data to establish the validity and limits of analysis for the data. One way to meet this need was to provide descriptive information such as population source, agency source, data source, date compiled, and any other contextual information that might assist in the analysis of the data in a note. For example: *"These rates represents the three-year average number of self-inflicted injury hospital discharges per 100,000 youth ages 10-19 years. Mid-year estimates (1994 base year and 1997 current year) are used as the denominator in calculating rates for 1993-95 and 1996-1998, respectively. Regional totals may not equal sum of counties due to rounding. Rates are not stable when the number is less than 20."*

The Council recognized the importance of this additional information not only in the data analysis feature of the Web site, but also for the *Touchstones Data Book.* It therefore added a focus on improving meta data to its agenda for developing the *2000 Touchstones Data Book* by asking all data-providing member agencies to update the

meta data associated with their data set(s), following the format of the new Web site design.

Managing differences in data

The transition from a static paper-based data resource to a dynamic Web-based one required a new understanding of the implications of the differences among the data sets and ways to manage those differences.

The *Touchstone's Data Book* presents information derived from related but separate data sets provided to the Council by its member agencies. The Council accepts data in a variety of forms and formats to create the book. In most instances, the member agencies provide data in its final format for the production of the book. In some case, the Council is responsible for manipulation of the data to produce the desired information in the format required for the book.

One of the primary advantages of the Clearinghouse for users is the data manipulation capability. Creating this capability required the project team to review all the data with this new capability in mind. In particular, they had to translate what they knew about limits to cross indicator comparisons from their experience in producing the book, into rules for using the data on the Web. This process also needed to identify data differences that were either not considered or were not relevant in the production of the book, but might be critical in the Web-based repository.

As each data provider sent data to the project team, critical differences among the data sets became better known. These differences began to raise questions about what would need to be communicated to users about comparability across indicators. The Council and their advisers recognized the criticality of communicating the limits to cross-indicator comparability that arise when data is drawn from different populations, at different points in time, or by incompatible data collection methods. The team recognized that a significant part of the effort in developing the Clearinghouse would be in identifying differences among the data sets and developing business rules to guide their use. The differences had to be fully understood, used to inform the development of the Clearinghouse and, ultimately, communicated to potential users.

Differences among sources of population data is a good example of the subtle non-comparability problems of data sets. Some agencies used the US Census as their

source for population numbers, others used alternative sources, while still others used an internal calculation prepared by one of their analysts. Differences in data collection cycles were one characteristic that needed to be considered. The Education (SED) and Health Departments (DOH) provide a good example of this data-oriented work. Much of the information available about children is naturally based on data collected by SED and DOH. SED collects data on the basis of the school year, by school district; DOH collects data based on the calendar year, by county. SED's data, although available soon after the end of the school year, could not be available to the Council for use in the prototype until the SED staff had a chance to rework it into a county format - a time-consuming and labor-intensive process.

A further example relates to the data DOH collects on in-patient discharges. This data is collected on a continuous basis rather than on a particular collection schedule. The data is shared with the Council at a point in time, rather than at the end of a data collection cycle. To accommodate this nuance in the book, the Council would report the data at a specified point in time and note that time in the narrative. This nuance needed to be accommodated differently on the Web.

The KWIC designers needed to understand the limitations on cross-indicator comparisons and develop a design for the Clearinghouse that was appropriate for each limitation. Each data set was reviewed in the context of every other potentially related set to identify characteristics that might result in non-comparability, or to identify some manipulation that would be required prior to use in cross-indicator comparisons. The implications of each difference was identified. Then Clearinghouse developers and the Council worked to determine when the differences among the data sources was sufficiently important to formally block comparisons across indicators, when they warranted a warning to the user, or just a statement in the "How to use the data" section of the Clearinghouse. A critical resource in this process was knowledge about each data set. In some cases the necessary knowledge came from the meta data. In other cases, the formal meta data was insufficient, but the team had access to the individuals who had knowledge about the data.

Technology choices

Given the fairly short term of the project and the need to eventually transfer it to the Council, it was important to select technology that was easily available and required little "ramp-up" time for the technical staff. The team selected software that either already existed at the Council (such as Microsoft Access) or was available on the Web as "open source" (PHP3 and MY SQL). The selection of "open source software" allowed the project team to acquire software quickly without going through a long procurement process. Open source software can also be of higher quality because it has been subjected to extensive peer review in the open source community. Open source

software offers the additional benefit of many existing applications that the project team could use and customize, further reducing development time. This choice will also allow the Council to freely choose "open source" software or move to commercial software for the future of KWIC.

Prototypes test the waters

The KWIC prototype allowed both users and technical staff to test the waters with a portion of the full system. Once developed in prototype form, KWIC was tested both to learn about the way it would be used and to ease the transfer of operation to the Council. This shared experience provided a platform for learning and discussion that was much better than a passive model, such as screen captures and work flow diagrams,that leave too much to each person's imagination.

Initially the prototype was developed to give substance to the many ideas that various user groups provided to the designers. The prototype allowed the design team to demonstrate to the Council staff and Executive Guidance Committee the specific features and functions that were recommended by users. The prototype also allowed the project team to demonstrate problem areas, such as the lack of sufficient meta data and its impact on users. It also provided the Council staff and Executive Committee a forum for further discussion of the governance framework and expanded features of the site, acting as an impetus for decisions that were essential to moving the site forward to the pilot stage.

Technically complex projects demand high-level technical skills

The KWIC project also presented a number of technical challenges, from Web design to data analysis and data management. These tasks required deep knowledge and experience in data base design and development as well as database administration and Web design and development including Web server configuration and implementation. In addition, there was a need for knowledge and experience in data manipulation and analysis tools and in the quantitative aspects of data use. These high-end skill requirements were needed for several reasons:

- the wide range of data sources
- high expectations for the analytical capabilities of the site
- maintenance once the site was established.

Like many small agencies, the Council has expert policy and program staff, but very few technical staff. As a result, one of CTG's roles in the DHHS grant was to facilitate the design and lead the development and first year operation of the Web site. CTG would

then transfer operation of the site to the Council. Over the 2-year project period, the Council would acquire the human resources necessary to maintain the site through direct hire, outsourcing, or other possibly more creative resource sharing relationships.

The system was designed using the Council's existing Microsoft Access database, which was developed by Council staff with suggested enhancements from CTG to allow for use in the Web work. CTG staff customized that database as a staging area for the Web site database. They also designed automated work flows, allowed for easy customization in the event of data changes, and created the Web application to operate with minimal human intervention.

Transition from CTG to the Council

The process of transitioning responsibility for the Clearinghouse out of CTG and into the Council is underway. This process will play out very much like the project up to this point. Multiple streams of activity will address the range of decisions and tasks that must be attended to to ensure a smooth release of the Clearinghouse. One stream of activity will focus on the physical maintenance of the site at the Center. One stream will involve the Council and the agency partners in implementing the governance structure necessary to provide guidance to decisions about the Clearinghouse. Another will require working with the necessary other partners to plan the public release of this new and highly valuable resource. Throughout this time the Council is continuing to work on the challenges of ensuring the Clearinghouse continues to meet the needs of users over time, of managing the differences in the data, of dealing with the technical complexity of the resource, and of providing the necessary technical, managerial, and policy infrastructure.