

UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS

ORGANISATION FOR ECONOMIC  
COOPERATION AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE

**Joint UNECE/OECD Work Session on Statistical Dissemination and Communication**  
(14-15 February 2005, Henley-on-Thames, United Kingdom)

Topic (ii): Sketching out the future of statistical publications

## WHAT IS THE FUTURE OF STATISTICAL COMPENDIA IN THE 21<sup>ST</sup> CENTURY?

### Supporting Paper

Submitted by Bureau of Justice Statistics, United States<sup>1</sup>

#### I. INTRODUCTION

1. The statistical compendium has been widely produced throughout the world for many years. With the advent of the Internet, print products are on the decline and online accessibility to data from original sources is increasing. The question now is whether bringing together that data into one package is still important? In addition, the changing information landscape requires that statistical compendia evolve but what functionalities are necessary to support this evolution? Given all the changes that have and are occurring in information dissemination, what technological, political, organizational or socio-cultural requirements are there for supporting a statistical compendium? Findings from recent research supported by the U.S. Bureau of Justice Statistics (BJS) address these questions for statistical compendia in general and for the BJS *Sourcebook of Criminal Justice Statistics (Sourcebook)* in particular.

2. For the past 30 years the Bureau of Justice Statistics has published the *Sourcebook*, a compendium of over 600 statistical tables from a wide variety of sources. This document, produced by the Hindelang Criminal Justice Research Center at University of Albany, has been viewed as a critical resource in criminal justice statistics. With the advent of new electronic media, *Sourcebook* was also presented as a website and on CD-ROM.

3. After 30 years of publishing this compendium, BJS decided to take an integrated look at *Sourcebook*: its users, mission, and its future in light of current and emergent technologies and the information environment. No reviews of this project had ever been done prior to this initiative. Impetus for this project came from -

- Maturing of the web to enable widespread and robust access to data
- Mandates for increased access to government information and increased use of electronic technologies in support of the work of government
- Questions about the role of a statistical abstract in the new web environment where original source data is readily available.

4. Dr. Carol Hert, an information scientist from Syracuse University, conducted the research project from the spring of 2003 through the summer of 2004. A multi-method approach involving three different research activities was selected to meet specific objectives of the project. The first step was to determine current knowledge about users and to identify useful directions for new user research through document analysis, site visits, and focus groups. To find consensus on the future and on requirements for achieving

---

<sup>1</sup> Prepared by Marianne W. Zawitz, U.S. Bureau of Justice Statistics (Marianne.Zawitz@usdoj.gov)

the mission of *Sourcebook*, a Delphi study was conducted. Finally, a series of usability studies was conducted to look at a compendium in the information space on the Internet.

## II. BACKGROUND RESEARCH

5. *Sourcebook* is a set of about 600 pre-aggregated tables from a wide variety of government and non-government sources. Published annually in print, *Sourcebook* is also provided on CD-ROM and on the web. Many tables are initially published by BJS or others and republished in *Sourcebook*. Extensive source notes and metadata are provided for each table. The organization and index of the print version are used for all other formats. The CD-ROM contains several editions of the printed document in Portable Document Format (PDF) format. The website, also based on the print version, provides sections of the report in PDF for download as well as each table in separate PDF and spreadsheet files. The website is updated throughout the year as new data become available. BJS sells both the printed version and the CD-ROM for the cost of shipping and handling. It is also sold by the Government Printing Office at a much higher cost than that charged by BJS. Resellers also sell the document on Amazon.com. Many copies of the print version have also been made available for free to criminal justice officials. In recent years, the number of orders for the print version and the CD-ROM has been declining.

6. The document analysis, site visits, and focus groups provided insights into the existing usage as well as potential usage in the future. Before *Sourcebook*, the data on crime and justice were not easily accessible. Historically, *Sourcebook* has played an important role in providing data on crime and justice particularly trend and public opinion data. There was agreement that *Sourcebook* provides the core of criminal justice statistics and that the processes used to select the data for inclusion ensure that only methodologically sound data are included. According to the focus groups, *Sourcebook* was used for accessing quick facts for a variety of purposes, determining the source for data on a particular subject, getting trend information that is usually published one year at a time, finding public opinion data, and getting an overview of the field and its statistics. Users found the CD-ROM hard to use although they liked having multiple years available and suggested that this be a part of the online version. The patterns of use of the online version show frequent use of a limited number of tables. In general, the index was found to be hard to use.

7. This research also indicated that users were interested in new enhancements to *Sourcebook* including data from other related fields or on new topics; more State and local data; more current data; alternative presentations such as graphs, charts, and maps; and the ability to integrate and manipulate data. Some said that they thought *Sourcebook* should become a repository of links to the original tables. There was a tension between the value of having a set of static tables and tools that give users direct access to micro-level data so that they can conduct their own analyses.

## III. THE DELPHI STUDY

8. Because the intent was to look to the future, a forecasting technique - a Delphi study - was employed to gain a picture of *Sourcebook* three to five years in the future. The Delphi method was developed by the Rand Corporation as a method for developing consensus among a group of individuals while preserving their anonymity to each other. For the *Sourcebook* research, this approach was used to determine points of consensus on the future mission of *Sourcebook* and related requirements for achieving that mission. BJS invited 123 experts to participate of which 43 actually participated. They represented a number of fields, many agencies from throughout the United States and other countries, and included criminal justice experts (practitioners and academics), statistical agency personnel, information scientists, and computer scientists. Many were not familiar with *Sourcebook* but were familiar with other statistical compendia.

9. The Delphi technique provides for communication in "rounds" with individuals contributing within a set time period, followed by researcher analysis and synthesis, and then provision of this synthesized view for comment during the next round. The Delphi was conducted in four rounds. The first round consisted of three open-ended questions soliciting perspectives on the future mission of

*Sourcebook* and associated requirements to attain that mission. The researchers then coded and grouped the responses to produce statements based on the participants' actual words. Round Two provided the statements pertaining to the mission of *Sourcebook* and statistical abstracts and requested participants to indicate their level of agreement with individual statements (e.g., "*Sourcebook* should provide access to state and local data") on a five-point Likert scale. Round Three provided a researcher-synthesized statement of the mission and asked for feedback on the requirements to attain that mission (again using Likert scales). Round Four addressed issues that had not achieved a consensus in Round Three. In all Rounds, all participants were sent the entire questionnaire. The result was the following mission statement:

#### Delphi Generated *Sourcebook* Mission Statement

- *Sourcebook* serves, and is marketed, as the definitive source of criminal justice statistical information at the federal, state, and local level for criminal justice practitioners, researchers, policy makers, and statisticians.
- *Sourcebook* is a tool that promotes and champions the statistics and their producers. It provides data in a variety of ways (i.e., tables, chart, graphics) with associated metadata and explanations, and links to source data.
- *Sourcebook* links to relevant criminal justice sites, commentary, and analysis. It also provides access to downloadable data sets. It is able to present both the detailed view of criminal justice statistical information as well as a big picture of the range and nature of that information (partially by linking to new data sources as they become available). It provides a snapshot of criminal justice statistical information at a moment in time, as well as remaining current. Its information enables historical and other comparative analyses (i.e., geographic, etc.).
- If possible, *Sourcebook* is provided in English and other languages, has online tutorial functions, and provides a suite of data analysis tools.
- *Sourcebook* should not be privatized.

10. The Delphi study also identified an associated set of requirements considered essential or complementary to attaining that mission. Concerning content, there was agreement that attention needed to be paid to terminology and vocabulary issues, indexing, and understanding the key information requirements in the field. The Delphi participants also identified several critical requirements for the delivery of content in the future including focusing on usability, ensuring the information is accessible on the Internet, providing a robust search engine, and building archiving capabilities for the electronic world.

#### IV. THE USABILITY STUDIES

11. The usability studies' goals were to determine:

- how people make choices among sources of information;
- what tools, presentations, information, etc. help and hinder them as they search for and use criminal justice statistical information;
- how *Sourcebook* fits into the larger landscape of criminal justice information.

12. To gain insights on these issues the usability studies tested a number of expert and non-expert subjects as they completed statistical information retrieval tasks on the Internet. Five tasks were developed where data were available from multiple sources including *Sourcebook*. (See table at the end of this section for a listing of the tasks and sources.) These tasks and sources were identified by experts including staff at BJS and *Sourcebook*. All of the tasks were considered routine and reflected frequently requested information. The correct answer was available on the source sites provided for each task. By using more than one source, comparisons could be made across sources including the original source. Users were asked to select 2 or 3 of the 5 tasks and then to use at least two of sources for each task. A total of 27 users at three sites throughout the country were tested.

13. The usability tests provided specific findings concerning compendia on the Internet. First, general behavior on the web dominated the results rather than aspects related to specific use of criminal justice statistical information. How users sought information required by the tasks was consistent with the research on general information seeking behavior. Specifically, users perform word matching activities, looking for the keywords that describe the information that they seek. As the number of words on a page increases, users have more trouble finding the information. Replicas of book pages with many words are not helpful to users in matching words. In particular, users had difficulty using PDF files because they replicate printed pages, are often at resolutions hard to see on a screen, lack navigation, and require knowledge of a different program besides the web browser to manipulate the content.

14. The usability study also found that users do not modify their search engine behavior to the search engine or collection being used. They generally enter one or two words and rarely use advanced features or online help for search. What was or was not covered by the search created problems for users. Several users expected the search to cover only the compendium when it often covered the entire parent site.

15. Some of the findings were more specifically related to compendia and statistical information. While different sites with different functionalities such as database interfaces were preferred for different tasks by different users, there were no strong expressions of “what functionalities should be in a statistical compendium website.” Users also expressed frustration at the lack of currency with the *Sourcebook* data and commented that they would always go to the source site to see if more recent data were available. There was also confusion when the data from the original source was not the same as the data from *Sourcebook*. Users expected the data to be harmonized.

16. In addition, the study found:

- Predictability of presentation of statistical information was used to facilitate retrieval. If the same table were in each edition, users would seek that table in earlier editions to get data for the time period specified in the task.
- Searching for information about geographic entities was challenging, as people want to search by the name of the entity, such as California, rather than type of entity, such as by State, which was generally all that was provided in table titles.

17. Some findings specific to criminal justice information, to *Sourcebook*, and to each of the sources tested also resulted. A full listing of these findings is included in the final project report, which is available upon request from BJS.

## Tasks and sources for usability testing

Task	Sources presented
<p><b>Death penalty public opinion</b></p> <p>Polling organizations frequently ask the American public about their views on the death penalty. In recent years, what percentage of the public favors the death penalty? Generally, how does this compare to surveys done in past decades? Finally, what percentage of people currently thinks the death penalty has a deterrent effect?</p>	<p>Harris Polls <a href="http://www.harrisinteractive.com">www.harrisinteractive.com</a></p> <p>POLLINGREPORT.COM <a href="http://www.pollingreport.com/">http://www.pollingreport.com/</a></p> <p><i>Sourcebook of Criminal Justice Statistics</i> <a href="http://www.albany.edu/Sourcebook/">http://www.albany.edu/Sourcebook/</a></p>
<p><b>Violent crime rates</b></p> <p>Find the latest statistics for the number and the rate of violent crime in your state and in the United States. How does your state compare to the United States?</p>	<p>U.S. FBI Uniform Crime Reports <a href="http://www.fbi.gov/ucr/ucr.htm">http://www.fbi.gov/ucr/ucr.htm</a></p> <p>U.S. Bureau of Justice Statistics Data Online Website <a href="http://bjsdata.ojp.usdoj.gov/dataonline/">http://bjsdata.ojp.usdoj.gov/dataonline/</a></p> <p><i>Sourcebook of Criminal Justice Statistics</i> <a href="http://www.albany.edu/Sourcebook/">http://www.albany.edu/Sourcebook/</a></p>
<p><b>Juvenile crime</b></p> <p>Juveniles and related crime trends are of increasing concern. For example, in 2000, there were over 1.6 million juveniles processed in juvenile courts in the U.S. What percent of those were female; what percent male? After being processed, many juveniles are placed in residential custody. For the most current year available, what percentage is placed in private vs. public facilities?</p>	<p>U.S. Office of Juvenile Justice and Delinquency Prevention, Statistical Briefing Book, Easy Access Analysis Tools <a href="http://ojjdp.ncjrs.org/ojstatbb/html/EZACCESS.html">http://ojjdp.ncjrs.org/ojstatbb/html/EZACCESS.html</a></p> <p><i>Sourcebook of Criminal Justice Statistics</i> <a href="http://www.albany.edu/Sourcebook/">http://www.albany.edu/Sourcebook/</a></p>
<p><b>Number of prisoners executed or on death row</b></p> <p>Find the latest year's statistics on the number of prisoners under sentence of death and the number actually executed in the United States. How do the numbers differ from the numbers five years ago?</p>	<p><i>Sourcebook of Criminal Justice Statistics</i> <a href="http://www.albany.edu/Sourcebook/">http://www.albany.edu/Sourcebook/</a></p> <p>Statistical Abstract of the United States <a href="http://www.census.gov/statab/www/">http://www.census.gov/statab/www/</a></p> <p>Bureau of Justice Statistics Website <a href="http://www.ojp.usdoj.gov/bjs/">http://www.ojp.usdoj.gov/bjs/</a></p>
<p><b>Law enforcement officer deaths</b></p> <p>Find the latest year's statistics on the number of law enforcement officers killed in the United States. How many of those deaths were accidents (i.e., not felonious killings)?</p>	<p><i>Statistical Abstract of the United States</i> <a href="http://www.census.gov/prod/www/statistical-abstract-03.html">http://www.census.gov/prod/www/statistical-abstract-03.html</a></p> <p><i>Sourcebook of Criminal Justice Statistics</i> <a href="http://www.albany.edu/Sourcebook/">http://www.albany.edu/Sourcebook/</a></p> <p>FBI Uniform Crime Reports <a href="http://www.fbi.gov/ucr/ucr.htm">http://www.fbi.gov/ucr/ucr.htm</a></p>

## Integrated Findings

18. From all of the research activities, the researcher constructed a set of integrated findings. To some extent, the findings of the usability study stand alone. Since it was the only project activity that explicitly investigated user behavior in real time (rather than relying on perceptions of those behaviors on the part of those users or experts), its findings do not overlap with those of the other two study activities. The importance of the web in current information dissemination highlights the importance of the findings of the usability study even though its findings are not confirmed in the other two. The Delphi and earlier user assessments both provide perceptions of the use and value of *Sourcebook*. The results of these two activities were generally in agreement, with the Delphi study able to provide a more detailed expression of the requirements for providing the information available in *Sourcebook* in the future as well as the context in which that dissemination would occur.

### 19. **Finding 1. There is still a recognized need for the functions of a statistical compendium.**

20. The early user assessment work pointed strongly to the recognized historical value of *Sourcebook* in the pantheon of criminal justice statistical information sources. The Delphi confirmed the ongoing importance of its role with the consensus-generated mission statement that points to the role of *Sourcebook* as that of a definitive source of information covering all areas of the criminal justice statistical information spectrum.

21. Traditionally, statistical abstracts have been used to:

- Access commonly used statistics;
- Identify quality information due to editorial processes;
- Point to original sources enabling users to get a broad view of statistics within a domain as well as identify sources and strategies for delving deeper;
- Enable historical comparisons as they tend to provide the same information year after year in a consistent fashion.

The findings of the user assessments and Delphi point to the ongoing need for these services. What is in question is the strategy for providing these services. Both the user assessments and the Delphi identified areas of new topical coverage for the statistical compendium and new functionality to support compendia goals. The Delphi generated a consensus statement that a *Sourcebook* for the twenty-first century “provides data in a variety of ways (i.e., tables, chart, graphics) with associated metadata and explanations, and links to source data. *Sourcebook* links to relevant criminal justice sites, commentary, and analysis. It also provides access to downloadable data sets.” These functions would serve to enhance the tool’s role in providing access to key statistics and the ability for users to find more detail.

22. The role of a compendium in enabling historical analysis also needs to be addressed. The online *Sourcebook* does not currently facilitate this although the paper volumes (and the CD-ROMs) do. Two strategies for enabling historical analysis would be to provide links from a table to its earlier versions and to provide the full contents of each *Sourcebook* volume at the site.

23. Seemingly in contradiction to the findings of the importance of the statistical compendium is the finding that a relatively small number of tables are accessed frequently online. However, it was found that the tables accessed were highly consistent over time. This suggests that *Sourcebook* might be able to reduce efforts to present the number of tables currently presented and increase efforts to find innovative ways to provide links to those less often used.

24. The research suggested that a future online compendium might consist of a set of locally mounted data presentations, remotely mounted sources (from collaborating agencies) which are connected to on-demand, with the data ported to the compendium for presentation in a standardized format, and additional links to other sources. There would be an associated hierarchy of “vetting” with

the locally mounted sources fully vetted, the remotely mounted source agencies vetted (and perhaps the data as well) and the links with the lowest level of editorial review.

25. The researcher recommended that strategies be identified to provide a source of vetted criminal justice statistical information that appears coherent and unified to the user and supports the traditional roles of a statistical compendium. She also suggested that BJS consider reallocation of resources from table presentation to efforts to identify and provide a uniform approach to linking to source. In addition, she recommended establishing a metadata standard that can support identification, transmission, and presentation of information from those linked sources should be explored.

26. The final project report also found the following:  
 “The existing *Sourcebook* product is the likely starting point for these efforts. It has a well-established editorial process designed to identify new sources of information as well as assess its quality. It is a coherent and unified package that is widely recognized in the field. The Delphi and usability studies identified a set of near-term enhancements to functionality (along with some longer term strategies) that continue to support the statistical compendium’s role in information dissemination.

In the longer term, the challenge will be to manage the diverse linkages to constantly changing sources while retaining the perception on the part of users that there is a “source” they can go to for the range of statistics. As *Sourcebook* exploits web technologies to enhance linking to sources and ongoing update, there will be a tendency for the look and feel of the product to become less uniform and for users to quickly lose the sense that they are using a stand-alone product. It will be important to investigate ways in which data and information could be ported into a standardized *Sourcebook* information presentation. This is likely to rely on 1) standardized metadata across sources and 2) web services that could import and present data in consistent fashions.”

27. **Finding 2. Usage of compendia in the web environment is impacted by the behavior of web users, general usability issues, and the lag of data from original sources.**

28. The usability study results clearly point to the extent to which users have a set of web behaviors that are not determined by the nature of the sites they are using. They perform word recognition tasks and do not tailor their behavior to the specific content or features of a given website. Another feature of the technological environment that impacts compendia usage is the time lag in providing data that results from the editorial and publishing processes associated with consolidating information into a compendium.

29. To continue to support the role of a statistical compendium in an online environment, the researcher recommended that existing usability problems with *Sourcebook* should be addressed including improvements to the search capability. In addition, she indicated that efforts to support word recognition tasks should be increased including the enhancement of indexing structures to incorporate user terminology, assessment of webpage look and feel to enable more rapid recognition of words on the part of users, and identification of terminology issues associated with information organization and categorization. Consistent with her recommendations under Finding 1, she also recommended that BJS consider strategies to minimize time lag and assure harmonization across multiple sources of the same data. These include development of standardized metadata across sources and development of web services that could import data in real-time thus assuring that compendium data are identical to source data.

30. **Finding 3. Information media have different “affordances” resulting in a lack of direct transferability of function, content and form among the media.**

31. Various information media (in this case, paper, CD-Rom, and websites) all have different ways in which they support information seeking and use. In the domain of human factors and human computer interaction, the general features of a medium that allow that medium to support particular functions are

referred to as “affordances.” For example, a paper book affordance is that a reader can judge how long the book is or where he or she is within the volume because of the physical nature of a bound set of paper pages. Each of the media currently used to disseminate *Sourcebook* has a set of different affordances which makes it difficult to directly translate function, content, and form among the media.

32. In the paper version of *Sourcebook*, the page layout supports ease of reading and copying and the table numbering supports finding a table within the printed volume. Some features are not helpful (or may be harmful) in the online environment. For example, table numbering is less important when one can link to the table directly from the pointer information of a table title or index entry though the numbering may help in historical searching. Individual tables presented as PDF files support printing tasks but the usability study results point out that for location and on-screen utilization they are less helpful.

33. Functionality, content, and form are dependent on the affordances of a given medium which suggests that each type of medium must be designed for individually. It does not work to design the paper product and then just present it electronically. Nor does it work to design for the electronic and then produce it in paper. The researcher recommended that *Sourcebook* and BJS should determine their priority for information dissemination in different media, develop individualized missions for those dissemination media, and do this in the context of the affordances of those media. These priorities need to be assessed in the context of resource, workflow, and marketing ramifications.

34. The usability studies suggest some of the affordances of the online and paper versions. It was found in the user assessments that the CD-ROM version was generally not used and likely should be discontinued. This need to individualize products is likely to have a variety of ramifications. Creating multiple products will require either reallocation of resources (such as reduction in the number of tables presented) or additional resources.

## **VI. FUTURE DIRECTIONS FOR *SOURCEBOOK* AS A RESULT OF THE RESEARCH**

36. Based on the findings and recommendations that came from the research, BJS has made several decisions about the future directions for *Sourcebook*. First, we will continue to support the *Sourcebook* as the research clearly showed a need for this type of resource.

37. Second, we decided to focus on *Sourcebook* dissemination on the Internet. While the Delphi study did not explicitly come to this conclusion, the requirements for successfully attaining the mission defined by the study participants clearly focused on the Internet. In addition, the research results showed that success on the Internet would require a decoupling of the print version from the web version. Given that we do not have additional resources to do both, we decided to cease printing after the next edition and focus all of our resources on the web version. We also decided to drop CD-ROM production.

38. Third, the *Sourcebook* staff began working on short-term improvements to usability suggested by the research. They are in the process of acquiring a new search engine and are planning to implement a new development process that integrates usability.

39. Fourth, we are committed to creating a more seamless *Sourcebook* while maintaining the source data. We do not want to have the *Sourcebook* staff repurposing static data from our reports. Rather, we want to move toward a distributed model and are exploring the use of web services to accomplish this. We are also discussing this direction with other major data providers. This initiative holds the promise of providing the most current data and avoiding data harmonization issues. We are also beginning an initiative to develop a metadata standard for all of our data and for that produced by our partners.



## VII. IMPLICATIONS FOR STATISTICAL PUBLISHING

40. Perhaps the most important finding for statistical publishing in general from this research is that we need to design materials for each type of media. As we moved from the printing paradigm to the Internet, we merely produced replicas of the printed document. Obviously, this was only an interim step but what is the next step? Our models for web dissemination continue to be grounded in how we produce print materials. For example, we frequently produce large tables that squeeze as many numbers as possible onto a printed page. Perhaps, we need to reconsider what a table should look like on the web. We need more information about what works for users in presenting statistical material on the web. Although we had hoped to discover if users preferred specific methods of presentation, no conclusions could be drawn from this research. We also need to get more information about what supports users of statistics on the web such as keyword placement, indices, and metadata. Clearly, more work needs to be done to determine what works for users on the web.

-----