

Positioning *Sourcebook of Criminal Justice Statistics*
for the 21st Century

Final Report
for U.S. Department of Justice cooperative agreement # 2003-BJ-CX-K004

Carol A. Hert
Syracuse University
October 28, 2004

EXECUTIVE SUMMARY

Over the last several years, the maturing of Web technologies, Federal mandates for increased efficiency and effectiveness in information dissemination, and the public's changing perceptions of the role of information in their lives have influenced the dissemination activities of Federal agencies. It was in this climate that the current project was developed and undertaken. The project reported here had the goal of providing an integrated look at *Sourcebook of Criminal Justice Statistics* (hereafter referred to as *Sourcebook*): its users, mission, its future in light of current and emergent technologies and the information environment.

The project consisted of three research activities. These were:

- A preliminary assessment of the state of knowledge about users of *Sourcebook* developed using document analysis, site visits, and focus groups (Section two of the report).
- Usability studies to examine user behavior with a range of criminal justice statistical websites (Section three).
- A Delphi study to determine points of consensus on the future mission of *Sourcebook* and on the related requirements for achieving that mission (Section four).

These methodologies, findings, and specific recommendations of the three activities are summarized in separate sections of this report. The rationales for the three activities and their relationships are provided in an introductory section. Finally, a synthesis of findings across the three activities and resultant recommendations appears in the concluding section (Section five).

KEY FINDINGS AND RECOMMENDATIONS FROM EACH PROJECT ACTIVITY

Each project activity resulted in a set of findings and recommendations. These are summarized here.

The User Assessment Activity

The goal of the user assessment was to provide a baseline picture of what was currently known about users of *Sourcebook* in order to determine what further user studies should be done. Findings are not detailed here (but are in Section two), given their role in supporting further work. The knowledge of existing usage of *Sourcebook* in its current manifestation and the Delphi preliminary results led the researcher to recommend the following user research strategies. Their intent was not to further document existing usage but to look towards how users and usage might change as *Sourcebook* moves into the future. Two views of *Sourcebook* were expressed. The first supports *Sourcebook's* role as a set of pre-aggregated tables. The second considers *Sourcebook* within the larger information environment.

Sourcebook as set of pre-aggregated tables

Critical to its function as the access point for pre-aggregated tables is assuring that the "right" pre-aggregated tables are available. The current *Sourcebook* has an established rigorous procedure for determining which tables should be included. This procedure, combined with an analysis of usage of the tables in the online version and an exploration of possible redundancy of tables (because they exist in

other sources) can indicate which tables are important to provide in pre-aggregated format, as well as how to identify new ones and monitor that list in the future.

The researcher recommends that SUNY-Albany continue monthly listings of the top 100 tables accessed online (with titles and section listings) so that the comparisons made in this document can be extended. In addition, a list of pages not accessed during the month should be tabulated.

Sourcebook in the Criminal Justice Statistical Information Landscape

Sourcebook is one component of the system of criminal justice statistical information that includes the creators and users of the information, the information and its presentation containers, the technologies available to create, disseminate and use it, and the surrounding socio-political landscape. To understand users and uses of *Sourcebook*, particularly as they might be changing in the future, therefore involves investigations beyond *Sourcebook*.

One challenge to understanding user behavior in relationship to *Sourcebook*, and criminal justice statistical information more generally, is that users change and adapt as rapidly as the tools they use. Thus, looking at user behavior in association with one tool (such as *Sourcebook*) in one instantiation at a particular moment in time provides limited guidance in developing an understanding of users that can be utilized for a long period of time. A strategy that can mitigate this situation is to shift the focus to understanding user behavior on particular tasks without unduly constraining the information strategies and sources users employ. This approach provides a more generalized picture of those strategies and the desired characteristics of sources enabling a researcher to make recommendations about the types of features and services a source could provide to facilitate user information seeking. This approach provided information that complements that of the Delphi study, by providing future-oriented information grounded in current behaviors. The researcher designed a study to meet the requirements above. The study addressed the following questions: When engaged in specific criminal justice statistical information tasks, how do people use the primary criminal justice statistical sources? What features and services of these sources facilitate or hinder the resolution of those tasks? The study and its results are described in Section three of the report.

The Usability Study

The analysis yielded three types of findings from the usability study:

- Generalized patterns of behavior for each site for each task performed on that site. Findings generalized across all tasks, tools, users.
- Usability problems identified for *Sourcebook* with recommendations for their resolution.

The generalized patterns for each site are not reported here but are in Section three.

Generalized findings

The findings generalized across tasks, tools, and users do not indicate that users tailor their behavior to the specific website they are using. They:

- Perform word matching activities.
- Don't modify their search engine behavior to the search engine or collection being used.
- Have difficulty using pdf files.

A fourth general finding is that the mismatch between search engine boundaries and site boundaries caused problems for users. In addition to the findings that reflect general Web behavior, some of the findings are more specifically related to compendia and statistical information. These were:

- Among the study participants, the tasks did not present situations where their domain knowledge was lacking.
- The organization of information by criminal justice system (most noticeable on *Sourcebook*) was rarely used or caused difficulties in some of the tasks.
- Predictability of presentation of criminal justice statistical information was used to facilitate retrieval.
- Searching for information about geographic entities is challenging, as people want to search by the name of the entity rather than type of entity.

These findings incorporated with others from the usability study led to a set of summary recommendations from the usability study and follow the section on specific recommendations related to usability for *Sourcebook*.

Usability problems identified for *Sourcebook* with recommendations for their resolution

Because of this project's focus on *Sourcebook of Criminal Justice Statistics*, the researcher reframed the usability findings specific to *Sourcebook* from the perspective of usability problems and their resolution. The usability problems identified for *Sourcebook* Online were:

- The introductory text isn't helpful in working on the tasks. All the information on content is below the fold.
- Users felt that *Sourcebook* had the feel of "just taking the print version and putting it online".
- There was an unclear organization of contents and tables within sections. People realized both listings weren't alphabetical and couldn't find another structure it represented to help them use the organization.
- The search syntax doesn't map to other search engines and every user of search engine got it wrong at least once (and many, multiple times).
- When you have invalid search syntax, you get a message saying to check link for search rules but it is not linked to that page. A user has to go back to search first and find the rules on that page.
- Some people thought the pdf and excel files for each table actually represented different tables and would look at both.
- No way to get to earlier years (mentioned when someone wanted to use the strategy of looking for the same table in a different year).

- The multiple search options (what component of *Sourcebook* was being searched) didn't seem to be clear to people. People input a search without considering where the search would be executed.
- Options for searching across sections (via search of index or table titles) were not obvious leading people to attempt to navigate by section (which was challenging for the juvenile task).
- All available options for accessing data are not clear on a section's introductory screen (contents and tables listings re below fold, and sometimes the search option is below the fold as well).
- There were comments on the data lag from UCR and BJS.
- URL includes "1995" appeared to some as representing the 1995 data.
- Many tables in pdf versions have footnotes to Appendices which people wanted to be directly linked.
- The indexing doesn't always follow standard practice reducing the indexing's ability to get people to the correct content.
 - For example, using the term "death penalty" leads to three terms that have "see" references rather than "see also" references. Even better indexing practice would be to use an entry such as "death penalty—public opinion" (indicating a subset of the available information on the death penalty) with a link directly to the set of relevant tables.
- Searching the index resulted in very long listings of entries of a different format from search listings in the table section. Search option on index page should be at the top of the screen.
- The pdf files of the full sections were rarely used.

The following recommendations address these problems in two tiers of recommendations with tier one representing problems that are most easily resolved.

Tier One Recommendations

1. Provide navigational options at top or left side of screens. These would include search option, access to index, and section links. Usability testing is suggested to see if users locate links and understand functionality to be found via the links.
2. Information on how to use the search engine should be available on the search page and also when the user gets error messages. A no-click option to get the information would be best.
3. Current text appearing on homepage as introductory information should be relocated—some may be appropriate to add to the existing "about *Sourcebook*" section, other information might go to a "how to use *Sourcebook*" section which provides information on downloading, searching, etc. Usability testing is suggested to determine what content should appear under what link and what the link label should be. Card sorting techniques may be a useful approach for both components of usability testing.
4. Instead of using file names for files (e.g., table3.46 wk1), provide links that say Table 3.46 pdf file format, Table 3.46 spreadsheet format.
5. Remove "1995" from main *Sourcebook* URL.
6. Provide links to appendices and sources from within a table.
7. Reorganize listings of table titles and section headings.
8. Consider ongoing provision of pdfs of file sections. Monitor their usage.
9. Provide additional years of data online.
10. Conduct usability tests to determine structure of search result entries.

Tier Two Recommendations

These recommendations may require more extensive usability testing/research than Tier One recommendations or might need to be considered in the context of a shift away from pdf formats to HTML or XML markup strategies. In addition, many of the recommendations, while arising from the usability studies, are not recommendations for specific usability changes.

1. Examine indexing practice. Monitor usage of the index and search to determine the extent to which an index is used.
2. Reorganize listings of table titles and section headings.
3. Enhance access to earlier years of data by direct linking from within a table (to its earlier iterations).
4. Enhance access to related information by inclusion of links to related tables from within a given table.
5. Work with BJS and other data providers to address issues of harmonization and currency of data.

Summary recommendations from the usability study

The findings of the usability studies point to the extent to which the Web and the standard approaches to web design currently shape user behavior. The dominant results of the usability study indicate that users don't use criminal justice statistical websites differently than they might use any other website: they perform word recognition tasks and don't tailor their behavior to the specific content or features of a given website. Pdf files represent special problems and equivalent boundaries for content and search engines help. This implies the baseline for any redesign of a website. It is recommended that:

- The extent to which user terminology matches agency/expert terminology be investigated. The extent to which users can find the terms they are looking for in the context of few non-relevant terms, the better they are at finding information. User terminology should be identified and strategies for mapping it to the more specific expert terminology developed.
- Pdf file usage needs to be further investigated. For those who are printing, pdf's provide a useful printing format. However for activities involving finding information within a document or using a document online, they represent significant challenges.
- Where possible, search engines should be indexing the content of the site they are associated with. Larger sites may benefit from allowing users to choose to search specific components of the site but should not over-partition the contents (as is the case at the moment on *Sourcebook*). In all cases, there should be an explanatory note indicating the contents the engine is searching and the depth of indexing of entities (such as to the table title level or document title level).

In terms of the findings related to the usage of Web-based criminal justice statistical information, it is recommended that:

- The effort involved in organizing content by the criminal justice system should be investigated. If it is not helpful online, it may not be a reasonable use of resources. In any event, access will be enhanced by a topical index.
- The topical index might include names of geographic entities with pointers to appropriate data. If this is too cumbersome, entries could be added for geographic names with a note indicating

how to access relevant data (e.g., by topic with subdivision by geographic entity). If possible, the associated search engine should incorporate the names of geographic entities.

- Since predictability of presentation was useful to some study participants, one strategy to consider would be the making of links among the documents for all years of a publication.

The Delphi Study

The final result of the Delphi study was a mission statement for *Sourcebook of Criminal Justice Statistics* and an associated set of requirements considered essential or complementary to attaining that mission. The final mission statement follows:

Final Mission Statement for the Sourcebook of Criminal Justice Statistics as Generated by the Delphi

Sourcebook serves, and is marketed, as the definitive source of criminal justice statistical (CJS) 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 CJS 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 CJS 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.

Due to its length, the set of requirements is not presented here but can be found in Table 4.3.5.

The findings lead to a set of recommendations concerning the positioning of *Sourcebook of Criminal Justice Statistics* in the coming years. These are:

- *Sourcebook* should continue to provide vetted, quality information. *Sourcebook* is playing a key role is “setting the bar” for quality information in the domain of criminal justice statistics, particularly for information that is produced outside of the Federal statistical agencies. Sophisticated procedures are currently in place at the University at Albany to assure the quality; these should be maintained.
- *Sourcebook* should continue to identify new sources of information and incorporate them when they meet appropriate quality standards. Numerous participants commented on new fields of endeavor that are important to be represented in *Sourcebook*. The field of criminal justice will continue to change, and *Sourcebook* will need to continually reflect those changes. This might mean that the current organization of *Sourcebook* will change to reflect changes in the organization of the field.

- *Sourcebook* should expand its efforts at attaining feedback from users and soliciting their input. If *Sourcebook* is to retain its premier status more attention needs to be given to understanding its users and their needs for criminal justice statistics.
- *Sourcebook* should exploit the technological opportunities of the Web. In the short term, *Sourcebook* should start providing links to source materials, linking technical information directly to appropriate tables, and link to relevant criminal justice sites. It might be appropriate for *Sourcebook* to include a “reference shelf” section that would provide links to sources, commentary, indexes, etc. As a second tier of activities, it should explore the technical options available for transforming statistical presentations (such as tables to graphics) or provide such additional presentations “manually”. Having a robust search engine is also a concern. Also at the second tier, it should assess technologies for enabling users direct access to data to perform unique manipulations. This effort would need to occur in the context of *Sourcebook*’s important role as a vetted source of the most important information. In the longer term, ongoing efforts need to be in place to stay abreast of, and incorporate new technological advances.
- Technological advances need to occur with attention to usability and other user concerns. Delphi participants indicated attention to usability, indexing, and terminology issues as critical requirements for attaining the mission of *Sourcebook*.
- *Sourcebook* should develop a mechanism for providing an archive of past *Sourcebooks* and enabling reoccurring tables to be linked from one *Sourcebook* to the next.

INTEGRATED FINDINGS AND FINAL GENERAL RECOMMENDATIONS

The final section of the report brings together the findings of the study into an integrated picture and presents recommendations for action. Integration of the findings from the multiple project activities involved 1) identifying agreements in findings among all study activities, 2) assessing the meaning of findings which were the result of one study, and 3) considering conflicting findings. The findings are:

- There is still a recognized need for the functions of a statistical compendium.
 - However, only a small number of tables in online *Sourcebook* receive large usage.
 - There is a desire on the part of users for new features that take advantage of technological advances.
- 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.
- Information media have different “affordances” resulting in a lack of direct transferability of function, content and form among the media.

Finding: There is still a recognized need for the functions of a statistical compendium.

Recommendations: It is 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.
 - The existing *Sourcebook* product should form the launch pad for these efforts.
 - Reallocation of resources from table presentation to efforts to identify and provide a uniform approach to linking to sources should be considered.
 - Establishing a metadata standard that can support identification, transmission, and presentation of information from those linked sources should be explored.

Finding: 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.

Recommendations: It is recommended that:

- Existing usability problems with *Sourcebook* should be addressed. In section three, these were provided in two tiers of feasibility.
- Efforts to support word recognition tasks should be increased. These include:
 - Enhancement of indexing structures to incorporate user terminology
 - Assessment of web page look and feel to enable more rapid recognition of words on the part of users or as one person suggested one site provided “all the right words without other words”. Haas and Hert (2000) provide the details of a strategy to identify and map user terminology to expert terminology. Additional usability studies (including strategies such as card sorting) would be helpful in identifying terminology issues associated with information organization and categorization.
 - Providing a new search engine on *Sourcebook*, the functionality of which is compatible to other search engines.
 - Ongoing monitoring of technological advancements in search engines.
- Strategies to minimize time lag and assure harmonization across multiple sources of the same data need to be considered. These will include:
 - Development of standardized metadata across sources.
 - Development of Web services that could import data in real-time thus assuring that compendium data are identical to source data.

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

Recommendations: It is 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.

The coming years will be exciting and challenging for those charged with disseminating criminal justice statistical information. Users will still need the value provided by the traditional statistical compendium but technologies will continue to shape the ways in which that value can be provided or new value can be added. This project has presented a picture of how *Sourcebook of Criminal Justice Statistics* is currently providing value to users and how it can be enhanced to meet the changing needs of users within the changing technological environment so as to continue to provide the values it has provided for over thirty years.