

# In Praise of Data Archives: Finding and Recovering the 1963 Kennedy Assassination Study

## Preface

On the morning of September 11<sup>th</sup> America had the tragic misfortune of being the target of the largest terrorist attack in history. Four hijacked airlines were crashed into the twin towers of the World Trade Center, the Pentagon, and a field in rural Pennsylvania. In the span of a few minutes more than 3,000 people were killed. Upon learning about these terrorist attacks, I recalled the study that the National Opinion Research Center (NORC) had carried out in 1963 in the days following President Kennedy's assassination. For the first time in the intervening four decades, I felt that the Kennedy Assassination Study (KAS) should be replicated. I consulted with my NORC colleague, Ken Rasinski, and we decided to launch the National Tragedy Study (NTS). In the course of less than two days, we secured NORC's commitment to conduct the NTS; raised funds from the National Science Foundation, the Robert Wood Johnson Foundation, and the Russell Sage Foundation; and settled on the sample design and questionnaire content. By Thursday September 13<sup>th</sup> we were conducting interviews. The NTS drew heavily from KAS, allowing comparisons of these two great national tragedies, and from NORC's General Social Surveys (GSSs), providing for more proximate pre- and post-September 11<sup>th</sup> comparisons.

## Hunting for the Kennedy Assassination Study (KAS) Data

To select items from KAS to be included in NTS, we turned to NORC's librarian, Patty Cloud and NORC's archivist, Michael Forstrom. They immediately pulled the extensive hardcopy files on SRS-350, as KAS was formally known.<sup>1</sup>

These files had been prepared by NORC's long-time librarian (1961-1997), Patrick Bova. They consisted of 1) codebooks that contained copies of the original questionnaire, coding/data processing instructions and deck/column location information, marginals, and certain select data tabulations; 2) drafts and final versions of papers and articles based on KAS; 3) data processing memos, especially about creating work decks for cross tabulating data stored on different decks; 4) correspondence about the study after collection; and 5) miscellaneous related material. With this complete documentation in hand the NTS content was quickly finalized.

*by Tom Smith &  
Michael Forstrom\**

With the NTS launched, we now turned to finding the machine-readable, case-level records for KAS. Cloud and Forstrom searched the Access master list of NORC tapes and found no reference to SRS-350, Kennedy, or any other term that connected a tape with KAS.

Tom W. Smith, co-PI of the GSS and NTS, recalled that the GSS had in the mid-1970s collected copies of various NORC and non-NORC studies as part of the Social Change Project (SCP). He searched the GSS archives and found a tape and related printouts from 1975 with data from KAS. With assistance from the Fay Booker, data librarian from Social Science Research Computing at the University of Chicago, the data were recovered from this old tape. Unfortunately, as was typical when SCP acquired data from earlier NORC studies, it only contained a sub-set, consisting only of most demographics, the 10 affect-balance scale items, and a scattering of miscellaneous items.<sup>2</sup> None of the assassination-specific items had been extracted and saved.

Smith also checked with the major data archives where NORC typically archived data - the Roper Center, University of Connecticut, and the Interuniversity Consortium for Political and Social Research, University of Michigan, and confirmed that NORC had not sent them the data. He also followed up a lead in an early NORC bibliography (Allswang and Bova, 1964) that the International Data Library at the Survey Research Center, University of California - Berkeley had the KAS codebook and questionnaire. They too did not have the data.

So the search turned to finding the punched cards, the original machine-readable medium for the data. Cloud and Forstrom consulted records of both on- and off-site records storage. There are about 4,000 cubic feet of records stored at NORC's headquarters and upwards of 20,000 cubic feet at a remote site, O'Hare Records and Retention Center.<sup>3</sup> Based both on the storage records and frequent past access to the on-site stored material, there was no indication that any punched cards were stored at NORC's headquarters.<sup>4</sup>

Cloud and Forstrom also did not find any punched cards indicated by the Access database listing of 3,669 boxes in the off-site records storage. Based on this Cloud therefore

reported on September 14th that “the likelihood of our finding cards is virtually nil.”

Smith then contacted Bova who indicated the next day that the cards should still be in storage. So Forstrom next consulted hard-copy inventories of the material of bulk storage.<sup>5</sup> These inventories from 1997-1999 identified 37 skids of boxes (about 15-20 boxes per skid) as well as bays of filing cabinets. The inventories included no references to KAS, but the contents of some skids and cabinets were either cursory or unidentified, so one couldn't rule out the punched cards being there.

On September 17th Bova then supplemented his earlier comments that the cards should be in remote storage with a listing of storage records, including barcode numbers for the KAS material from NORC's previous off-site storage facility, before the holdings were transferred to the O'Hare site. This documentation encouraged us to search on, but unfortunately these earlier control numbers did not match anything in the Access database.

By September 24th it was decided that if the punched cards were in off-site storage, they had to be in the unidentified, bulk-storage material. From September 26th to October 11th, Cloud and Forstrom spent 4-5 days at the O'Hare facility going through 37 skids of boxes and the collection of around 50 filing cabinets. No KAS material was found.

At this point, Cloud and Forstrom wondered if the KAS punched cards could be in boxes in records storage, but not listed among the contents of boxes, if they had been discarded, or if they had been among a set of records lost during the a move to the O'Hare facility. Finding the cards under any of these situations seemed unlikely, so Smith pursued another line of inquiry. From the hard-copy KAS files the names of two individuals who had been sent data in the early 1970s were obtained and from a NORC bibliography (Bova and Worley, 1991) the names of another two non-NORC researchers who had published an article using the data in the 1960s were obtained. The two authors were tracked down and indicated that one had had the data, but abandoned them many years earlier. The two data orderers were being traced when this line was then abandoned after a breakthrough.

On October 20th, Forstrom received from accounting a question about a late fee attached to an O'Hare records storage invoice. This inquiry led Forstrom to examine the invoices from O'Hare more closely. They listed 8,348 boxes in general, records storage, but the Access database documented only 3,669 boxes. Thus, there were 4,679 undocumented boxes in NORC's general, records-storage holdings at O'Hare.

Forstrom contacted NORC's former records manager, Connie Schumacher, and she indicated that there was a drawer

in the NORC library's old card catalog that listed at least some of the O'Hare holdings and that Bova had a printed index to punched cards. Cloud also found a 1987 memo that referred to holdings of punched cards for about “60 studies in 700 boxes.”

These records allowed the old, record-control numbers to be matched to barcodes and locations at O'Hare. O'Hare was contacted on October 24-25th and asked to pull 11 boxes designated as KAS punched cards.<sup>6</sup> On October 26th Forstrom picked up ten boxes from O'Hare, but found that they had pulled one incorrect box. The final box was then retrieved by Forstrom on October 29. After six weeks of effort we had the KAS data.

### **Dealing with Multiple-Punched Cards**

We now had to convert the cards to a current, machine-readable medium. There were three challenges. First, we had to know how to interpret the data. Second, we had to read the cards. Third, we had to convert the multiple-punched data<sup>7</sup> to single-punched data. The first task was made possible by the detailed codebook that Bova had prepared in 1964. We knew what data were in each column and that there were 3 decks per case. Next, we had to find the right cards. Smith examined the 11 boxes and determined that they consisted of a single copy of cards in deck 2, multiple copies of cards in decks 1 and 3, and several boxes of work decks used in the early analysis. Thus, we quickly knew what to read and how to interpret the data. The second and third task depended on finding someone who could read cards and translate or spread multiple-punched data. NORC, like most organizations, had given up using cards about 25 years ago and no longer had a card reader. We searched for an organization that could read cards and convert multiple-punches. After considering several possibilities, we settled on November 9th on the National Data Conversion Institute (NDCI) in New York City, a firm that NORC was already using for old, seven and nine-track tape conversions. Since there was only a unique copy of deck 2, we were very concerned about shipping the cards to NDCI. We arranged for Isabel Guzman, a NORC administrative assistant, to fly two boxes of cards to New York as carry-ons on November 16th. On November 12th American Flight 587 crashed taking off from LaGuardia, but Guzman still was willing to fly to New York and delivered to NDCI on November 16th.

NDCI originally thought the conversion would take only a week or two, but complications soon developed. First, although the 38 year-old cards were in excellent condition, they had problems reading them because of the age of their card reader. They discovered that they “had to refurb our punched card equipment, it had been sitting around so long it got a little rusty.” It took them until the end of December to send a test file of deck three data. The file sent was corrupted, but this fortunately had nothing to do with the original data on the cards. An uncorrupted deck 3 test file

was sent on January 9th. This file revealed that they did not fully understand multiple-punched data and that such data needed to be spread. Smith was familiar with spreading multiple-punched data from the 1970s and was able to explain to NDCI what needed to be done. Over the next three weeks, NDCI sent various versions of the three decks separately which Smith checked against the original documentation and marginals. Various data interpretation issues were ironed out and a problem with new variable names being too long was resolved. On January 28th, a merged SPSS file with all three decks combined into one record per case was received that Smith verified as matching the original file.

At this point the KAS data had been fully recovered, but the SPSS file was very crude and not user friendly: 1) variables had names that referred to their original column locations rather than meaningful mnemonics, 2) there were no value or variable descriptors, missing values, or other data definition information, 3) some variables had long alpha values where short numerics were needed, and 4) other reformatting was needed. After close examination of the data and the original documentation, a final SPSS system file with detailed labels was finished and the data set was archived at the Roper Center.

### Lessons

NORC is to be praised for its thorough documentation of the 1963 KAS and for preserving that documentation and the data for 38 years. But NORC is not a data archive and it merely stored the information. As a result, the study itself did not appear on any readily accessible listing of NORC surveys. It was only because Smith had worked with the data on the SCP in the mid-1970s that KAS came to light after the September 11th terrorist attacks. Even within NORC many stored records are not well indexed and it took persistent efforts, the assistance of two ex-employees, and a bit of serendipity to unearth the data. Moreover, once recovered we had data on a medium that was so antiquated that it took four months of extensive efforts to convert it to a modern, user-friendly format.

The lessons from the KAS experience are simple, but important. Survey data must be sent to survey archives like the Roper Center and ICPSR where the documentation and data will be preserved, backed-up, periodically updated as technologies change, indexed, and made routinely and easily accessible to researchers. Failure to archive studies is poor science and a disservice to other contemporary researchers and those in the future.

### References

Allswang, John M. and Bova, Patrick, NORC Social Research, 1941-1964: An Inventory of Studies and Publications in Social Research. Chicago: NORC, 1964.

Bova, Patrick and Worley, Michael Preston, NORC Bibliography of Publications, 1941-1991: A Fifty Year Cumulation. Chicago: NORC, 1991.

\* Paper presented at the IASSIST Conference, Storrs, CT, June, 2002. Tom W. Smith, e-mail: smith@norcmail.uchicago.edu, Michael Forstrom, national Opinion Research Center, University of Chicago.

### Footnotes

<sup>1</sup>The Survey Research Service (SRS) was a division of NORC that mostly conducted extramural research during the period 1963-1966.

<sup>2</sup>SCP was mainly interested in identifying studies that included items that had been adopted by the GSS to study trends from these baseline studies to the GSS.

<sup>3</sup>The O'Hare storage is equivalent to the contents of about 17 three-bedroom houses.

<sup>4</sup>On-site storage consists mostly of more recent material.

<sup>5</sup>NORC has two types of storage: records storage contains bar-coded boxes with some indication of the contents and exact location of each box. Bulk storage consists of boxes and filing cabinets that are not bar-coded, whose content is less precisely known, and which are stored by skid (for boxes) or by bay (for filing cabinets), and not by individual box or filing cabinet. The Access database covered the former, but not the latter. Records storage was itself divided into general, records storage and two project specific holdings. Only the general holdings were relevant in this case.

<sup>6</sup>Punched cards are stored in corrugated cardboard boxes approximately 3.5 inches tall by 8 inches wide by 21.5 inches deep. They were stored as separate, stand-alone records and not placed in larger storage boxes.

<sup>7</sup>NORC, like many other organizations, multiple-punched data to reduce the number of columns that data took up in order to save space which was very important when data appeared on punched cards and computing storage and analysis were very limited. For example, the study number (350) was entered in the 80th column of each deck. That is, the 3, 5, and 0 punches were all punched in column 80. This had to be converted into three, three-column fields with 350 for all cases. Another example was a code-all-that-apply question in which codes 0-9, x, and y in column 28 in deck 2 each stood for a different reason for Oswald killing Kennedy and people could mention as many reasons as they thought applied. Data from this one column had to be spread into 12 columns, which represented each of the 12 reasons as separate variables.