

Journal of Technology Law & Policy

Volume XVIII – 2017-2018

ISSN 2164-800X (online)

DOI 10.5195/tlp.2018.212

<http://tlp.law.pitt.edu>

Pioneers in Computerized Legal Research: The Story of the Pittsburgh System

Tina Batra Hershey and Donald Burke

Abstract

The potential effects of law are far-reaching and research is ongoing regarding the intersection of law and technology. Given the widespread availability of online legal documents today, the laws of various jurisdictions can be reviewed and researched in their full text form. However, in the not-so-distant past, this task was overwhelmingly more difficult. Many jurisdictions, unable to keep pace with the increased volume of statutes, regulations, and judicial decisions, compiled indexes of legal information rather than catalogs of full documents. These indexes made comparisons between jurisdictions difficult and left researchers unsure of whether they had captured all relevant information. However, in the middle of the 20th century, researchers began to tap into the potential of computers in relation to information retrieval. Much of the early pioneering work in the legal field was conducted by researchers at the University of Pittsburgh, who developed the “Pittsburgh System” that was a precursor to the computerized legal research tools that are ubiquitous today.



This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License.



This site is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.

Pioneers in Computerized Legal Research: The Story of the Pittsburgh System

Tina Batra Hershey* and Donald Burke**

I. INTRODUCTION

Today, the ability to collect and research a plethora of information, including legal data, is taken for granted; however, this was not the case in the middle of the 20th century. During that time, the rate in which legal documents were published increased dramatically.¹ Information was collected in paper form and not updated frequently. In order to search this material, indexes were created, but found subjective, as they depended on the world-view and skill of the indexer and often did not capture the information being sought.² In the 1950s, attention turned to the use of computers to capture, compile, and retrieve data. Three intrepid pioneers from the University of Pittsburgh, John Harty, Nathan Hershey, and Eric Springer, were early innovators in this new arena. They were joined by fellow pioneers in the University's Computation and Data Processing Center. Together, they explored new methods of searching documents applicable to legal problems that allowed researchers to access the original documents. As Harty stated in 1960, "[w]hat is needed is a system which will store this mass of legal material, analyze it, and on demand, retrieve the information relevant to a given inquiry."³

These Pitt researchers were not alone in their exploration of the use of computers to retrieve legal information.⁴ However, they were at the forefront of this developing technology. As technology advances in the arena of information retrieval,

* Tina Batra Hershey is an Assistant Professor in the Department of Health Policy and Management at the Graduate School of Public Health at the University of Pittsburgh.

** Donald S. Burke is Dean of the Graduate School of Public Health at the University of Pittsburgh.

¹ William G. Harrington, *A Brief History of Computer-Assisted Legal Research*, 77 LAW LIBR. J. 543, 544 (1984-1985); see also Lee Loevinger, *Jurimetrics: Science and Prediction in the Field of Law*, 3 M.U.L.L. MOD. USES LOG. L. 187, 188 (1962).

² John F. Harty, *The "Key Words in Combination" Approach*, 3 MOD. USES LOG. L. 54, 56 (1962); see also Lee Loevinger, *Jurimetrics: the Methodology of Legal Inquiry*, 28 LAW & CONTEMP. PROBS. 1 (1963).

³ JOHN F. HARTY, APPLICATION OF INFORMATION RETRIEVAL TECHNIQUES TO LEGAL RESEARCH (Univ. of Pittsburgh 1960) (on file with author).

⁴ See, e.g., Harrington, *supra* note 1.

and we consider the accompanying legal ramifications, looking back at the history of computerized legal research provides an opportunity to understand and appreciate the innovations of the past.

II. A PROBLEM EMERGES

As with many innovations, the idea was stumbled upon while working on a different issue. Researchers at the Health Law Center at the University of Pittsburgh (Health Law Center) embarked on a two-year study in 1956 to conduct research on legal issues related to hospitals in jurisdictions across the country.⁵ The research team included four lawyers: Horthy, Hershey, Springer, and Kennard Hirsch, as well as other non-legal personnel.⁶ At the outset of the project, Horthy and his fellow attorneys believed that the law would not vary significantly from one jurisdiction to the next.⁷ They also felt that the research primarily would be focused on case law rather than statutes.⁸ However, as the project progressed over the next two years, they found great variation among the relevant statutes that they manually reviewed in all 50 states. The research was challenging, as the researchers had to rely on indexes, which were dependent on the world view of the indexer.⁹ Moreover, the indexes, while perhaps useful in other domains, were not conducive to research related to hospitals.¹⁰

The project culminated in publication of the Hospital Law Manual, a loose-leaf publication with quarterly supplementation.¹¹ The Hospital Law Manual was a unique publication at the time, as two different manuals were written: one for hospital attorneys with complex legal language, and one with simpler legal language for use by hospital administrators.¹² The Hospital Law Manual was a success when

⁵ John F. Horthy, *Experience with the Application of Electronic Data Processing Systems in General Law*, 2 M.U.L.L. MOD. USES LOG. L. 158, 158 (1960).

⁶ Interview with John Horthy, Managing Partner, Horthy, Springer & Mattern, in Pittsburgh, Pa. (Jan. 27, 2014). Hirsch left after one year.

⁷ Horthy, *supra* note 5.

⁸ *Id.*

⁹ Horthy, *supra* note 6.

¹⁰ Horthy, *supra* note 5.

¹¹ NATHAN HERSHEY, HOSPITAL LAW MANUAL (Health L. Ctr., Univ. of Pittsburgh 1959).

¹² Nathan Hershey, *Hospital Law Manual and Quarterly Service*, (Alicia Ni'Tracy ed., Jan. 2011) (on file with author).

published¹³ and remains a significant resource to this day. Furthermore, the development of the Hospital Law Manual was critical for another reason: it caused Horthy and his fellow researchers to think of other approaches for conducting large-scale research of legal data.¹⁴

III. GENESIS OF AN IDEA: AUTOMATED LEGAL SEARCHING

Through brainstorming sessions in 1959, Horthy, Hershey, and Springer quickly realized that computers were the answer to a more efficient research process that involved legal text sources.¹⁵ They developed a legal information retrieval methodology called the “Pittsburgh System” which was different from other research approaches being developed at that time, because it did not rely on existing indexes. Rather, the Pittsburgh System involved inputting the full text of the legal material in question (statute, regulation or case law) and then searching the entire text for research terms of interest.¹⁶ Moreover, the Pittsburgh System was unique in that “[i]t [had], almost from the beginning, been angled at the solution of specific research problems in statute law.”¹⁷

By Spring 1960, the University’s Computation and Data Processing Center and the Health Law Center, in conjunction with IBM’s Advanced Systems Development Division, had developed a set of information retrieval programs for the IBM-provided University computer, the IBM 650.¹⁸ The IBM 650 was based on vacuum tubes and could store up to 2,000 words via drum storage.¹⁹ Considering the degree of computer sophistication at that point in time, turning to computers was a bold approach.²⁰

¹³ *Id.* (stating that 1,100 sets were sold in the first year at a price of \$150 for a three-year subscription to both volumes).

¹⁴ Horthy, *supra* note 6.

¹⁵ *Id.*

¹⁶ William B. Eldridge & Sally F. Dennis, *The Computer as a Tool for Legal Research*, 28 LAW & CONTEMP. PROBS. 1, 88 (1963) (“It is quickly apparent that the Pittsburgh system is the first to bring to bear on legal research problems computer capabilities other than speed. . . . Professor Horthy and his associates have proven that a completely unindexed body of literature can be searched effectively by a computer.”).

¹⁷ *Id.* at 87.

¹⁸ JOHN F. HORTHY, SEARCHING STATUTORY LAW BY COMPUTER (Univ. of Pittsburgh 1962) (on file with author).

¹⁹ Jon Bing, *Let there be Lite: A Brief History of Legal Information Retrieval*, 1 EUR J. L. TECH. 3 (2010).

²⁰ *Id.*

IV. DEVELOPMENT OF THE PITTSBURGH SYSTEM: THE IBM 650 COMPUTER SEARCH SYSTEM

The first step in establishing the system was to select material to form the “library.” In this instance, approximately 2,200 statutory sections of Pennsylvania hospital law constituted the “library.”²¹ The text of each statutory section was punched on IBM cards.²² To punch a card, the operator typed a letter and the machine punched holes in the card to represent the letter.²³ Each punch card could hold 68 letters; therefore, 20 or more cards might be necessary in order to enter one statute.²⁴ When the cards were fed into the IBM 650 computer, the holes would trigger electronic impulses that would transfer individual characters of text into magnetized spots on the tape reel.²⁵ Thus, the “library” was stored on reels of magnetic tape.

As each Pennsylvania statute was placed on the computer, it was given a document number.²⁶ The first statute placed on tape was number one, the next statute was number two, and so forth.²⁷ After the statutes were placed on tape, the IBM 650 computer analyzed the “library” and constructed an alphabetical list of the total vocabulary contained therein (after eliminating certain common words).²⁸ This list could be printed for use in the search.

The Pittsburgh System was demonstrated for the first time at the 1960 American Bar Association (ABA) Annual Meeting in Washington, D.C.²⁹ The event marked the first demonstration of a working system of legal information retrieval anywhere in the country.³⁰ The weeklong demonstration was co-sponsored by the University of Pittsburgh, IBM, and the Electronic Data Retrieval Committee of the Bar Activities Section of the ABA.³¹

²¹ HORTY, *supra* note 3.

²² *Id.*

²³ *Id.*

²⁴ *Id.*

²⁵ HORTY, *supra* note 18.

²⁶ *Id.*

²⁷ *Id.*

²⁸ *Id.*

²⁹ HANDBOOK OF LEGAL INFORMATION RETRIEVAL 261 (Jon Bing ed., 1984); CHARLES P. BOURNE & TRUDI BELLARDO HAHN, A HISTORY OF ONLINE INFORMATION SERVICES, 1963–1969, at 230 (2003).

³⁰ HORTY, *supra* note 3.

³¹ *Id.*

Horty and Hershey conducted demonstrations for 6–7 hours each day at the IBM facility in Northwest Washington, D.C.³² IBM provided hourly bus service from the ABA conference to the facility.³³ The demonstrations were a success, with over 1,100 people in attendance over 5 days,³⁴ and were somewhat of a sensation, as the computer analyzed 400 statutes and typed out all the citations in less than 10 minutes.³⁵

V. REFINEMENT OF THE PITTSBURGH SYSTEM: THE IBM 7070 COMPUTER SEARCH SYSTEM AND DEVELOPMENT OF A LEGAL THESAURUS

The IBM 650 search process was limited to document numbers only.³⁶ If a word appeared anywhere in the statute, that document was considered relevant and printed.³⁷ However, even this simple search technique strained the capabilities of the IBM 650 computer.³⁸ Commentary during the 1960 ABA demonstration from the information retrieval field revealed the fear that the IBM 650 program was not sophisticated enough to function efficiently when confronted with a mass number of documents.³⁹

Moreover, the researchers in Pittsburgh realized that they needed to develop a program that could locate each *occurrence* of a word more precisely than merely its existence somewhere in a document so that searches could be based on more specific word relationships.⁴⁰ Financial assistance for this next phase of the project was provided by several entities, including the National Institutes of Health, the Ford

³² Interview with Nathan Hershey, Professor Emeritus, Univ. of Pittsburgh Graduate School of Public Health, in Pittsburgh, Pa. (Feb. 12, 2014).

³³ *Id.*

³⁴ Horty, *supra* note 3; Horty, *supra* note 6. In fact, then-Vice President Nixon was scheduled to attend one day of the demonstration but had a scheduling change that prevented him from attending. In the end, Nixon met with Horty for over an hour at a later date to discuss a search regarding the powers of the presidency.

³⁵ *Automated Legal Research*, AUTOMATIC DATA PROCESSING NEWSL., May 11, 1964 (Diebold Group, Inc., New York, N.Y.) (on file with author).

³⁶ HORTY, *supra* note 18.

³⁷ *Id.*

³⁸ *Id.*

³⁹ *Id.*

⁴⁰ *Id.*

Foundation, and the Council of Library Resources, Inc.⁴¹ IBM provided a more advanced computer, the IBM 7070, that could better achieve these goals.⁴² The IBM 7070 computer, which was a transistorized version of the IBM 650, possessed magnetic core storage for 9,990 numbers of 10 digits each, was installed in September 1960 and the research team spent the remainder of 1960 and the beginning of 1961 developing the search program.⁴³ The program was fully operational by July 1961.⁴⁴

The IBM 7070 search system was similar to the IBM 650 search system, with one critical difference: the precise location of words in a statute could be identified.⁴⁵ This time, the “library” consisted of the full Pennsylvania statutes, which had 31,113 sections (documents) filling four tape reels.⁴⁶ These four tape reels contained 6,230,529 words and 112 common words (which occurred 2,815,340 times).⁴⁷ An alphabetical list of every non-common word of the “library” was generated by the computer.⁴⁸ This alphabetical listing identified the exact location by document, line, sentence, and position within the sentence of each non-common word, and became the basic tool for searching the statutes.⁴⁹

The IBM 7070 search system included several improvements. First was the ability to search for phrases with 3–4 words by using the command “within three words.”⁵⁰ Another improvement was the ability to eliminate words from a search through the “but not” command.⁵¹ These two improvements allowed for a more focused search of the text.⁵²

While the IBM 650 program limited the number of words that could be processed, the IBM 7070 system allowed up to 500 inquiry words to be processed at

⁴¹ HORTY, *supra* note 18; *see also* BOURNE & HAHN, *supra* note 29.

⁴² HORTY, *supra* note 18.

⁴³ *Id.*

⁴⁴ *Id.*

⁴⁵ William B. Kiehl, John F. Horthy, Charles T. Bacon & David S. Mitchell, *An Information Retrieval Language for Legal Studies*, 4 COMMC’N. ASSOC. COMPUTING MACH. 9, 380 (1961).

⁴⁶ HORTY, *supra* note 2.

⁴⁷ *Id.*; *see also* HORTY, *supra* note 18.

⁴⁸ HORTY, *supra* note 18.

⁴⁹ *Id.*

⁵⁰ *Id.*

⁵¹ HORTY, *supra* note 18.

⁵² *Id.*

once, enabling the use of as many synonyms and word variants as desired.⁵³ Any number of searches could be conducted so long as the group of searches contained no more than 500 inquiry words (versus single searching only with the IBM 650 program).⁵⁴ This ability to combine searches led to cost savings in computer rental as total search time was reduced.⁵⁵

Finally, the IBM 7070 program printed, as part of the search results, any inquiry word that did not occur as “no list” so that the searcher could see which searches words were not present in the statutes.⁵⁶ This was an important feature; previously, a researcher would not know whether certain words had not been included in the search or whether they were not present in the statutes.

Thus, the IBM 7070 search system was “the first operational computer retrieval system based on the concept of placing on tape the full text of a large body of materials without manual indexing of any sort.”⁵⁷ Experiments conducted by the Pittsburgh team revealed that the search methodology developed under the IBM 7070 search system would be particularly beneficial to the revision of laws, which require precise searching for specific word combinations.⁵⁸ Other types of searches that benefitted from this approach were searches for all laws related to a subject in one jurisdiction or searches for similar laws across jurisdictions.⁵⁹

VI. THE PITTSBURGH SYSTEM IN ACTION

By 1961, the complete statutes of Pennsylvania were available for full text searching.⁶⁰ All 77 volumes fit onto four reels of magnetic tape.⁶¹ The Pennsylvania Department of Justice requested assistance from the Health Law Center to prepare a collation of statutes relating to the administration of welfare services due to a

⁵³ *Id.*

⁵⁴ *Id.*

⁵⁵ *Id.*

⁵⁶ *Id.*

⁵⁷ *Id.*

⁵⁸ *Id.*

⁵⁹ HORTY, *supra* note 18.

⁶⁰ AUTOMATED LAW SEARCHING, AUTOMATED LAW SEARCHING (n.d.) (on file with author).

⁶¹ *Computer Tapes State Laws, Scans 77 Volumes in Hour*, PITT. POST-GAZETTE, Aug. 16, 1962, at 25.

mandate from the Pennsylvania Legislature.⁶² One example of the type of “clean up” envisioned by the Legislature was the removal of any references to shillings and pence.⁶³ The full text of the Pennsylvania statutes were searched for the words “shilling” and “pence” and those statutes identified as containing the obsolete words were flagged for revision.⁶⁴ Another area that the Legislature wanted to focus its modernization efforts on was Pennsylvania’s welfare laws.⁶⁵ Use of the Pittsburgh System made modernizing the Pennsylvania statutes a much simpler task.

As word began to spread regarding this novel text retrieval methodology, the Pittsburgh System attracted attention beyond Pennsylvania.⁶⁶ In 1962, the New Jersey Supreme Court, seeing the need to modernize its court rules, engaged the Health Law Center for the preparation of its Rules of Court, Rules of Evidence, and Constitution.⁶⁷ The New Jersey Governor’s office, upon hearing of this effort, requested that all the statutes of New Jersey be included.⁶⁸

The federal government was also interested in the Pittsburgh System, but wanted further testing in order to validate the methodology. In 1963, the Air Force Accounting and Finance Center commissioned a test of the Pittsburgh System for a six-month trial using the full text of the United States Code and half of the published decisions of the Comptroller General as test materials.⁶⁹ To conduct the test, the materials were searched manually and by computer regarding questions submitted by Department of Defense users from around the country.⁷⁰ The Air Force then compared the search results.⁷¹ The comparison revealed that the Pittsburgh System searches were almost twice as effective as the manual searches.⁷² Thus, the Pittsburgh System was used to build the first computerized service, Legal

⁶² ERIC W. SPRINGER & JOHN F. HORTY, *SEARCHING AND COLLATING THE WELFARE LAWS OF PENNSYLVANIA BY COMPUTER* (Univ. of Pittsburgh 1962) (on file with author).

⁶³ Interview with Eric Springer, Of Counsel, Horty, Springer & Mattern, in Pittsburgh, Pa. (Jan. 24, 2014).

⁶⁴ *Id.*

⁶⁵ SPRINGER & HORTY, *supra* note 62; Bing, *supra* note 19, at 11.

⁶⁶ *AUTOMATED LAW SEARCHING*, *supra* note 60.

⁶⁷ *Id.*

⁶⁸ *Id.*

⁶⁹ *Id.*; *see also* BOURNE & HAHN, *supra* note 29.

⁷⁰ *AUTOMATED LAW SEARCHING*, *supra* note 60.

⁷¹ *Id.*; BOURNE & HAHN, *supra* note 29.

⁷² *Id.*

Information Through Electronics (LITE), that provided batch searching of full-text legal statutes for the Air Force Staff Judge Advocate in Colorado.⁷³

In 1964, the Pittsburgh System was installed in New York for the use of the New York State Legislature.⁷⁴ Tapes with the New York State statutes, all training of legal and technical personnel, and the programs necessary for searching were provided by the Health Law Center, allowing the New York Legislature to operate its own legal retrieval center.⁷⁵ The system installed in New York searched for obsolete law, rearranged statutes into logical arrangements, and provided background information for continuing legislative modernization.⁷⁶

As part of ongoing efforts to promote the Pittsburgh System and to highlight the further search refinements that occurred between 1962 and 1964, a demonstration was held for state and federal officials in Spring 1964 at IBM's Education Center in Washington, D.C.⁷⁷ The audience included attorneys general, heads of joint legislative committees, leaders of state congress, and officials in the federal government.⁷⁸ The IBM 1401/1301 System, an even more advanced computer, was being utilized at this point. The purpose of the demonstration was simple: to display how the Pittsburgh System was revolutionizing legal research in its ability to search the entire text of statutes and other relevant legal material.⁷⁹ At the demonstration, the computer reviewed approximately 30,000 statutes and produced the requested citations in less than 20 minutes.⁸⁰ Attendees were impressed with the technology.

Sold on the value of automated computer searching, other federal government agencies, including the United States Department of Defense, adopted the Pittsburgh System.⁸¹ In 1965, the General Counsel's Office of the Internal Revenue Service (IRS) began using the Pittsburgh System to search the IRS Code and Regulations.⁸²

⁷³ *Id.*

⁷⁴ AUTOMATED LAW SEARCHING, *supra* note 60.

⁷⁵ *Id.*

⁷⁶ AUTOMATED LAW SEARCHING, *supra* note 60.

⁷⁷ *Automatic Data Processing, supra* note 35.

⁷⁸ *Id.*

⁷⁹ *Id.*

⁸⁰ *Id.*

⁸¹ SEARCHING LAW BY COMPUTER: HOW IT WORKS, AUTOMATED LAW SEARCHING (n.d.) (on file with author).

⁸² *Young IRS Attorneys Get Computer Assist to Speed Experience*, WALL ST. J., Aug. 31, 1965, at 1.

VII. GOING PUBLIC: ASPEN SYSTEMS CORPORATION

The Pittsburgh System researchers continued to refine their search process throughout the 1960s, taking advantage of advances in computer technology. By the late 1960s, the Pittsburgh System researchers realized that they needed outside funding in order to continue with their efforts to maintain compilations of law.⁸³ They established Aspen Systems Corporation (Aspen) in 1967,⁸⁴ named after their experiences with the Air Force base in Aspen, Colorado.⁸⁵ The University of Pittsburgh was the largest shareholder.⁸⁶

By 1969, Aspen had input and stored the statutes of all 50 states, which contained over 200 million words.⁸⁷ To provide context, if 200 million words were printed in books, the books would be stacked higher than a five-story building.⁸⁸ This effort, which used nearly 400 typists, proofreaders and keypunch operators, took over one year and cost over one million dollars to achieve.⁸⁹ The resulting system, called the “System 50 for State Statutes,” was the largest computer information file in the United States.⁹⁰ Intended users were industrial corporations, federal and state agencies, law firms, state legislatures, trade associations, and insurance companies.⁹¹ Additionally, searches could be ordered of the statutes of a single state, a specific combination of states, or all 50 states.⁹²

VIII. SIGNIFICANCE OF THE PITTSBURGH SYSTEM

The innovative work by Horty, Hershey, and Springer has been proclaimed as “the first successful text retrieval system . . . [that] started the development of computerised legal information services . . . with major international examples as

⁸³ Horty, *supra* note 6.

⁸⁴ BOURNE & HAHN, *supra* note 29.

⁸⁵ Horty, *supra* note 6.

⁸⁶ *Id.*

⁸⁷ *Memory Bank Stores Laws of 50 States*, PITTSBURGH PRESS, Apr. 15, 1969; *see also New System of Legal Research by Computer Tapes Developed*, WASH. POST, Apr. 28, 1969.

⁸⁸ Horty, *supra* note 6.

⁸⁹ ASPEN SYSTEMS CORPORATION ANNOUNCES SYSTEM 50 FOR STATE STATUTES—THE NATION’S LARGEST FULL TEXT DATA BANK (Aspen Systems Corp. 1969) (on file with author) [hereinafter ASPEN SYSTEMS CORP.].

⁹⁰ BOURNE & HAHN, *supra* note 29, at 238.

⁹¹ ASPEN SYSTEMS CORP., *supra* note 89.

⁹² *Id.*

Reed Elsevier’s LEXISNEXIS service, or Westlaw and other services of the Thompson group.”⁹³ Indeed, the Pittsburgh System demonstrated the “feasibility of the concept of using the digital computer for legal research.”⁹⁴ Thus, as lawyers around the globe perform their everyday research tasks on computers, they are utilizing a technology that has its roots in Pittsburgh.⁹⁵ The work of the Pittsburgh researchers serves as an inspiration to think outside the box when tackling difficult tasks, to strain the boundaries of convention, and to continue refining methods until a goal is achieved. It is also beneficial to look back at this work to appreciate the tools at our disposal for easy retrieval of legal information that may be taken for granted today. The Pittsburgh System was an historical achievement that deserves recognition in the modern era.

⁹³ Bing, *supra* note 19, at 4.

⁹⁴ Harrington, *supra* note 1.

⁹⁵ Indeed, John Harty and the Health Law Center at the University of Pittsburgh are included in the LexisNexis Timeline. *LexisNexis Timeline*, LEXISNEXIS, http://www.lexisnexis.com/anniversary/30th_timeline_fulltxt.pdf. (last visited Nov. 23, 2017).

PIONEERS IN COMPUTERIZED LEGAL RESEARCH
