JTC Resource Bulletin

Big Data: What State Courts Should Know

Version 1.0
Adopted December 5, 2014

Abstract

While court data does not meet the criteria to qualify as Big Data, this data can and should be used to learn a great deal more than is currently known about the management of cases and the parties involved. By better utilizing their small but highly structured data, courts can gain insight into current business problems, and manage cases more efficiently and effectively.

Acknowledgments

Joint Technology Committee:

COSCA Appointments
David Slayton (Co-Chair)
Texas Office of Court Administration

David K. Byers
Arizona Supreme Court

Laurie Dudgeon
Kentucky Administrative Office of the Courts

Gerald A. Marroney
Colorado Administrative Office of the Courts

Robin Sweet
Nevada Administrative Office of the Courts

NCSC Appointments
The Honorable O. John Kuenhold
State of Colorado

The Honorable Michael Trickey
Washington Court of Appeals, Division 1

Ex-officio Appointments
John Greacen
Forum on the Advancement of Court Technology

NACM Appointments
Kevin Bowling (Co-Chair)
Michigan 20th Judicial Circuit Court

Paul DeLosh
Supreme Court of Virginia

Yolanda Lewis
Superior Court of Fulton County, Georgia

Kelly C. Steele
Florida Ninth Judicial Circuit Court

Jeffrey Tsunekawa
Seattle Municipal Court

CITOC Appointments
Jorge Basto
Judicial Council of Georgia

Craig Burlingame
Massachusetts Trial Court

NCSC Staff
Paul Embley
Jim Harris
Ilonka Dazevedo

Document History and Version Control

Version	Date Approved	Approved by	Brief Description
1.0	12/5/2014	JTC	Release document

Contents

Introduction

Big Data has become a household word, and for proof we need look no further than the publication of *Big Data for Dummies[1]* in 2013. Since so much of the focus in popular discussion and marketing messages is on claims about the revolutionary power of Big Data, a sense of near panic is being created: the data gold rush is on, and whoever is not joining the stampede will be left behind. The explicit message of these marketing message is that any organization not actively pursuing a Big Data initiative is foolishly missing out on the Next Big Thing: only with Big Data, we are told, can we compete, triumph, and arrive at higher intelligence about all that we do.

Despite the frequency with which the term Big Data is now used, many in the state court community remain confused about its meaning for their work. State courts are not alone in their confusion about the relevance of Big Data to their work. A recent Wall Street Journal blog described the hype of Big Data and reported results of a recent survey of executives at large- and medium-sized retail, financial services, and consumer goods firms: one-third of the respondents found the notion of Big Data "very confusing."[2] How do we separate the hype from the helpful in this discourse? Fortunately, for state courts, the buzz around Big Data is a distraction—here is why.

What is Big Data?

The term is used without much definitional precision and that lack of clarity is the basis for confusion among those managing public and private sector organizations.

Wikipedia offers this definition of the term big data: "a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications." Other sources point to the three Vs as a way to differentiate Big Data from other kinds of data: volume (the amount of data), velocity (the speed with which data is accumulated) and variety (the many kinds of data that become linked together). The intent of these descriptors is to underscore the fact that Big Data is not intended to simply refer to one or more very large databases, but to indicate that the differences along these three dimensions of volume, velocity, and variety create a qualitative difference in the data being described.

---

[1] Hurwitz, Judith, Alan Nugent, Fern Halper, and Marcia Kaufman. *Big Data for Dummies*. Hoboken, NJ: Wiley, 2013. Print.

[2] Dwoskin, Elizabeth. "Beware 'Big Data' Hype, Reports Warn." Web log post. *Digits: Tech News & Analysis from WSJ*. Wall Street Journal, 27 May 2014. Web. 17 Nov. 2014.

The investment in restructuring massive amounts of dissimilar data on an ongoing, continuous basis to create Big Data might be worthwhile for an organization that has abundant IT resources and is going to derive future profit from the investment. The firms engaged in this enterprise have a firm grasp on the basis metrics of their core business, and are seeking to extract additional value by exploring hidden relationships among the huge volume of disparate data.

But for every firm with a Big Data purpose clearly defined and operational, there are many more who remain confused. Failures to understand and create a shared understanding of the purpose of creating management information from Big Data results in confusion at best and demise at worst. Poorly guided data dumps result in what one analyst writing in Forbes described graphically as a "big data hairball,"[4] a tangle of uncoordinated information and business processes on which a firm can choke. More data is not always the answer, and volume and variety alone can obscure insights and generate spurious correlations that are not meaningful (the crime rate dropped while rainfall declined and the market share of the Chrome browser increased—all unrelated).

Do State Courts Have Big Data?

State courts do not have Big Data. State courts do have small amounts of highly structured data, which can be organized to support specific business processes and services so that court cases can be resolved in a fair, impartial, and timely manner.

By comparison with the Big Data examples above, the largest court in the country, the superior court in Los Angeles, averages 57,000 felony cases per year. That is not big data. Nor are its 1.6 million infraction cases each year. Nor are the 100 million cases filed annually in all the state courts combined. Once the volume of Big Data is understood, it becomes clear that courts simply do not have Big Data.

What Kind of Data do State Courts Have?

Courts have small data, data that has architecture and a purpose, optimized to produce particular views (e.g., a court's daily calendar, lists of the number of pending cases) and perform specific tasks (e.g., identify cases pending beyond a time standard, or orders awaiting signature). Courts sometimes struggle to combine appropriate kinds of data from different structured databases—for example, treatment data for drug court participants maintained by treatment providers with case management data maintained by the court to track progress through the program, or criminal history data maintained

---

[4] Arthus, Lisa. "Are You Tangled In A Big Data Hairball?" *Forbes.* Forbes Magazine, 1 Aug. 2013. Web. 17 Nov. 2014.

by law enforcement agencies and repositories with criminal case data to inform sentencing decisions by judges.

Any court that has tried the comparatively simple task of integrating structured data from different systems (e.g., legacy case management system with new case management system, interpreter assignment database with case management system, drug court treatment provider data with law enforcement and court data) has experienced the pain and expense that comes with combining data that differs in structure, relationships, and definitions—even when that data is describing the same kinds of things (criminal cases, defendants, sanctions). This is not a problem of Big Data. It is a problem of integrating small data effectively. Unlike Big Data, these data are of the same variety (records on individuals in structured databases), similar volume (in the tens, hundreds, or thousands of cases), and are created at a relatively slow and similar velocity (the number of cases that proceed in a month).

Case-related data is not Big Data, but it is big enough data for courts to develop business intelligence about these cases. Courts stand to learn a great deal more than is currently known about the management of these cases and the parties involved and to allow the court to manage those cases more efficiently and effectively, if these data are organized and analyzed correctly. But to do that, courts need to define what it is they need to know to adjudicate cases in a fair, timely, and effective manner.

What Should State Courts Do With Their Data?

What should courts do? Small data, done well. There is no substitute for clear analytical management and judicial thinking, and such thinking does not require large amounts of data. It does require an understanding of what it is that courts are trying to manage and what data is needed to inform that process.

Courts today need to utilize their existing small data, along with data from their justice system partners, to provide insight into current business problems. These include classic caseflow management problems (e.g., real time reporting on the status of cases, sources of delay, and forecasting of future workload), coordination of resources (e.g., effective scheduling to allow efficient and cost-effective assignment of interpreters to batches of cases), monitoring the effectiveness of court orders (e.g., the collection rates on fees and fines and the disbursement rate to victims and restitution funds), and a host of other everyday management issues that remain out of court control. For this purpose, well-designed data warehouses and data marts can provide the focused empirical basis for answering key management questions.

In many instances, court data remains poorly organized for answering core business needs of its leaders and managers. Judges need criminal justice data to allow for empirically-based decision making about pretrial release (Who does not constitute a threat to public safety?) and to support informed sentencing decisions (Who will benefit from a residential alcohol treatment program? Who is like to recidivate? When is incarceration appropriate?). Court managers need to understand filing trends and the deployment of resources (jurors, interpreters, self-help staff, family law facilitators, mediators, and more) to meet the needs of those using court services.

Today many courts are not able to go beyond simple calendars and lists of cases, and cannot produce even the most basic management reports that are regarded as accurate (e.g., time to disposition, clearance rate, jury trial rate, age of pending cases, juror yield and utilization, and more). The majority of policy and management decisions are based on faith in gut instinct or accepted wisdom rather than clear empirical facts. Many courts and AOCs report that they are seriously deficient when it comes to IT and analytical staff and resources, and that management information remains out of reach, always sacrificed to meet basic operational requirements of keeping the courts functioning.

Beyond managing current resources and cases more effectively based on well-organized data, courts could greatly benefit from the development of meaningful, predictive analytics. That is, the utilization of data contained in case files and past case histories that would allow courts to understand and anticipate future patterns of development in current cases, before they occur. For example, examination of adult guardianship cases that have ended in malfeasance by a court-appointed guardian might reveal a set of characteristics that could predict when that might occur in current cases. The court could evaluate current cases against this profile and identify those that it should monitor more closely now, before anything goes wrong. Similarly, a set of characteristics of cases involving self-represented litigants could be develop that would predict when those cases might stall out along the way, and allow the court to intervene proactively before that happens. The use of risk and needs assessment tools, which rely on a variety of data sources, are also key to managing offenders before, during, and after adjudication.

These tools do not require big data. Instead, they require small data to be well organized, consistent in quality, provided timely, and presented in a manner that is easily understandable. The limited IT and analytical resources of courts make it all the more important that these be focused on key business problems. The data that courts do have is grossly underutilized and much can be gained from organizing and deploying it to improve court services.