



Harnessing Big Data for Social Good: A Grand Challenge for Social Work

Harnessing Big Data for Social Good: A Grand Challenge for Social Work

Claudia J. Coulton

Case Western Reserve University

Robert Goerge

University of Chicago

Emily Putnam-Hornstein

University of Southern California

Benjamin de Haan

University of Washington

GRAND CHALLENGES FOR SOCIAL WORK INITIATIVE

Working Paper No. 11

July 2015

Grand Challenge: *Harness Technology for Social Good*

GRAND CHALLENGES FOR SOCIAL WORK INITIATIVE

The Grand Challenges for Social Work are designed to focus a world of thought and action on the most compelling and critical social issues of our day. Each grand challenge is a broad but discrete concept where social work expertise and leadership can be brought to bear on bold new ideas, scientific exploration and surprising innovations.

We invite you to review the following challenges with the goal of providing greater clarity, utility and meaning to this roadmap for lifting up the lives of individuals, families and communities struggling with the most fundamental requirements for social justice and human existence.

The Grand Challenges for Social Work include the following:

- Ensure healthy development of all youth
- Close the health gap
- Stop family violence
- Eradicate social isolation
- End homelessness
- Promote smart decarceration
- Reduce extreme economic inequality
- Build financial capability for all
- Harness technology for social good
- Create social responses to a changing environment
- Achieve equal opportunity and justice
- Advance long and productive lives

EXECUTIVE COMMITTEE

Co-Chairs

John Brekke
University of Southern California

Rowena Fong
University of Texas at Austin

Claudia Coulton
Case Western Reserve University

Michael Sherraden
Washington University in St. Louis

Diana DiNitto
University of Texas at Austin

Eddie Uehara
University of Washington

Marilyn Flynn
University of Southern California

Karina Walters
University of Washington

J. David Hawkins
University of Washington

James Herbert Williams
University of Denver

James Lubben
Boston College

Richard Barth (ex officio)
*American Academy of Social Work and
Social Welfare and University of
Maryland*

Ronald W. Manderscheid
*National Association of County
Behavioral Health & Developmental
Disability Directors*

Sarah Christa Butts (staff)
*American Academy of Social Work and
Social Welfare and University of
Maryland*

Yolanda C. Padilla
University of Texas at Austin

GRAND CHALLENGES FOR SOCIAL WORK INITIATIVE

Working Paper



Harnessing Big Data for Social Good: A Grand Challenge for Social Work

Claudia J. Coulton, Robert Goerge, Emily Putnam-Hornstein, and Benjamin de Haan

The digital revolution is producing vast quantities of social, psychological, and organizational data that social workers can harness to address society's most difficult problems. Data from computerized social service, education, and health records; open data portals; social media posts; web searches; and mobile GPS devices and sensors can help to illuminate social problems and propel effective solutions. Technological innovations have made it possible to manage and analyze such digital assets in real time.

Despite the technological progress, the social sector has been slow to incorporate a continuous flow of data analytics to inform policy and practice. Because most social program data remain in silos, it is difficult to demonstrate the efficacy and long-term cost benefits of programs and practices that make use of big data, which precludes social investments in such programs and practices. The fundamental knowledge of big data management is currently too limited among social work professionals to allow for effective collaboration with other disciplines and specialists to accelerate data-driven social innovation. The field also lacks an effective framework to address the legal, ethical, and privacy concerns inherent in the use of personal data. A grand challenge for social work is to build the capacity to deploy these powerful digital resources to discover and apply social solutions to benefit society.

Key words: Big data, digital technology, integrated data systems, data mining, computational social science

HARNESSING BIG DATA FOR SOCIAL GOOD IS A COMPELLING CHALLENGE

We live in a digital society in which vast quantities of data are produced, much of which remains untapped for social work practice, social policy, and social action. Instead of using this data, social work professionals rely on the empirical base of survey and other primary data collection methods to develop and test social interventions and policies. Meanwhile, a great deal of useful information remains trapped in the silos of legacy information systems or in commercial enterprises' digital footprints. Despite these existing troves of data, the social sector lags behind

GRAND CHALLENGES FOR SOCIAL WORK INITIATIVE

Working Paper



in using data driven strategies. Efforts to make social spending more evidence-based or to make social systems more effective often falter because of the high cost of gathering timely and complete data. Responses to emerging social problems often occur subsequently rather than preemptively because the social sector lacks the wherewithal to analyze informative data threads as they occur. The technological advances of the digital revolution can facilitate overcoming these roadblocks and accelerate the pace and scope of social discovery and development, but the potential of big data to inform action has yet to be fully recognized or acted upon within social work (Coulton, 2014; Getz, 2014; Grogan-Kaylor & Dunkle, 2014).

The term *big data* refers to the digital byproducts of human activity (e.g., carrying out government functions, delivering services, administering programs, conducting business transactions, communicating through social media, using digital devices). The scale and scope of these data represent a disruptive shift in relation to a given object of interest (Schroeder, 2014). “Big” in this case does not signify just volume, but also the velocity and variety of data elements and sources and the special analytic processes required to turn the raw data into useful information (Desouza & Smith, 2014). Proportionality is another distinguishing characteristic of big data because it typically encompasses all the data from a particular source and the dynamic capture of events in real time. Handling big data is known to be complicated because it is highly variable and prone to misinterpretation and error if not subjected to meticulous review and curation. Big data often involves data linkage or some type of “mash up” that brings together digital records or processes that cross sectors and points in time. It includes unstructured text, images, and recordings in addition to the more commonly used structured data. For these reasons, the usual methods of data management, processing, storage, and analysis are often inadequate for getting the full value from big data.

The Push for Big Data Application

The social sector does not benefit fully from the digital revolution. If this trend continues, the constituencies the sector serves could end up on the wrong side of the digital divide. For example, businesses glean a great deal about consumers’ behaviors and tastes from big data and customize products to match, whereas the social sector has yet to implement these types of practices to improve the effectiveness of social programs. Community groups have also been slow to access the data that could individually or collectively advance the quality of their members’ lives. If social work does not invest in building the capacity to fully make use of its existing big data, technologically savvy entities without a full appreciation of the data’s social context could take up the task for the wrong reasons.

A number of forces call for social work to increase its big data capacity. One is the increasing demand for evidence-based policy and practice. In the United States, much of the social spending is tied up in programs that have yet to be fully evaluated (Liebman, 2013a). Though there is a growing demand to prove what works, human service systems often lack the data management and analysis infrastructure necessary to efficiently evaluate programs. Moreover, such systems lack interoperability and standardization, making it challenging to integrate data across sectors to examine the long-term costs and benefits of social programs.

Another driving force comes from the private sector, where the reliance on conventional wisdom to make sound decisions has been replaced by reams of data and predictive analytics. These expectations are migrating from the corporate sector to the boards of non-profit organizations and influential foundation funders. Agency executives and managers are now expected to have data at the ready to support their operational decisions and long-term strategies. They are increasingly required to provide metrics to show what is working, and to assure that quality standards and outcomes are meeting the mark (Morino, 2011).

Additionally, the general public is beginning to call for more civic engagement with big data. Open government and civic hacking movements provide vehicles for mining data and generating information for communal use. The term *open data* describes the idea that certain kinds of data should be disseminated freely so that they can be reused, analyzed, published and transformed into new and useful products (Bertot, Gorham, Jaeger, Sarin, & Choi, 2014). Additionally, community coalitions that work to improve outcomes for target populations (e.g., vulnerable youth, homeless veterans, children growing up in deep poverty) require access to cross-system data as a way to keep all of the partners on track and working towards common goals (Kania & Kramer, 2011).

The successful mapping of the human genome has also fueled the hope that big data can yield a deeper understanding of how to improve social well-being. Individual behavior within social environments is complex, and there are potentially innumerable elements that are responsible for human variation (Kum, Krishnamurthy, Machanavajjhala, & Ahalt, 2014). Applying predictive analytics from numerous digital touch points to detailed data on individuals and their social settings could eventually lead to policy or practice interventions that could modify elements of behavior (e.g., Moore, Sacks, Manlove, & Sawhill, 2014). Such types of analyses consider the sequential nature of the elements and the potential for nonlinear and reciprocal relationships and the complex system dynamics that may evolve over various layers of social organization. Therefore, they hold great promise for eventually customizing social interventions with precision. Such discoveries could ultimately be the basis for social interventions that are finely tuned to the person in the situation, ones that will work with relatively higher levels of certainty. Like human genome mapping, big data allows many more variables to be taken into account in predicting what interventions will work for individuals with a unique social profile.

Obstacles to Overcome

Despite the push towards big data applications, the social sector needs to address numerous obstacles to benefit from the big data promise. Data security breaches are a central concern, especially with respect to human service, educational, and health records (Hoffman & Podgurski, 2013). Even though the law permits many of these records to be used for research, evaluation, and quality improvement purposes, agencies are often reluctant to share the records for analysis. Moreover, the social sector must consider the ethical matter of using digital material for purposes of which individuals may not have been aware when providing personal information. A related difficulty concerns data ownership and control, especially when data from various sources are combined to create applications beyond the scope of the original intent. Though such “mash ups” often yield information of very high value, exemplifying the power of big data, they require

agencies to cede individual control to achieve a larger societal benefit. Such agreements often encounter legal and practical complications.

Curation of big data is another enormous challenge because the data are a byproduct of numerous processes, not generated specifically for the purpose to which they are eventually applied. Understanding how these processes shape the data is crucial to producing valid information and correct interpretation. The data require careful cleaning and validation by specialists with a deep understanding of the data generation processes. A related issue is determining how to allow other analysts to replicate the results from big data. Unlike the fixed data sets associated with traditional research projects, big data platforms are typically refreshed by a continuous flow of new information. It may be difficult to exactly reproduce results if the data has shifted between the original and validation analysis. Before fully adopting big data applications, social work professionals need to develop, vet, and document rigorous quality assurance, versioning, and archiving techniques.

Overreliance on big data could be dangerous. Relying on only big data could lead to problems such as important issues going unnoticed because they only affect a segment of the population that does not leave a significant digital footprint. For example, older adults and individuals with low levels of income and education are underrepresented in social media data (Hargittai, 2015), whereas young adults and more affluent individuals are overrepresented in data on consumers of public services. At the same time, digital multiplier effects from data duplications or redundancies may divert attention to concerns that appear larger than they are. Unlike controlled studies and probability samples, well developed techniques for assessing the representativeness of big data do not yet exist. To avoid big mistakes, it is important to remember that big data must be interpreted within the context of other information and based on a good understanding of the factors that affect the data generation process (Lazer, Kennedy, King, & Vespignani, 2014).

Another challenge is avoiding the intentional or inadvertent statistical discrimination that arises from big data. For example, commercial data aggregators already mine big data for market segmentation, classifying some neighborhoods in unflattering ways that companies use to make decisions about whether to business in the area cite operations, often to the disadvantage of residents (Parker, Uprichard, & Burrows, 2007). Though “red lining” is not new, big data makes it easier for vulnerable consumers to be targeted for push marketing of harmful products or for individuals to be labeled in ways that diminish their life chances. Moreover, errors that creep into big data might be magnified in ways that incorrectly classify individuals as high risk or make them targets of investigation by various authorities. Community advocates need their own big data analytic capacity to counter and correct discriminatory and other harmful practices.

Technological advances allow data to be generated and stored at unprecedented scales and rates, but the principles and practices to unleash their power for social benefit while limiting unintended harm have not yet been developed. Education for institutional review boards about the use of such data resources is necessary.

Harnessed correctly, existing big data can inform and empower the social sector in its work to improve social well-being. To hasten progress, the social sector must incorporate digital

technology and data analytics into practice and policy. Data silos and digital divides across sectors need to be removed to foster strategic investments and system reforms that demonstrate long-term cost benefit. Social workers with knowledge of digital applications and data science will be needed to communicate with computer scientists to drive innovation and applications in the field. Finally, a rigorous framework is required to balance privacy protection and the public interest along with equity considerations. Harnessing big data for social good is a compelling challenge for social work.

THE FEASIBILITY OF HARNESSING BIG DATA FOR SOCIAL GOOD

The feasibility of using big data to drive innovation and accelerate discoveries is now being recognized across numerous fields and disciplines (Burrows & Savage, 2014; Graham & Shelton, 2013; Japac et al., 2015; King, 2011; Rudin et al., 2014; Taylor, Schroeder & Meyer, 2014). Federal investments in raising the nation's capacity for big data and open data further reinforce the achievability of this goal. Such investments will allow big data applications at the highest levels of government, in the business sector, and across many scientific communities (Mervis, 2012). Additionally, commercial vendors and the open source community are offering a number of tools (e.g., distributed storage and processing, open data portals) that have the possibility of being adapted to the big data needs of the field.

Since the late 1990s, some social work and policy researchers have analyzed administrative records from social service agencies and applied the results for scientific and practical purposes. Based on these experiences, the field has already started building a core of knowledge and skill in data base management, record linkage methods, data cleaning and harmonization, and the statistical analysis of large data sets. A number of studies related to welfare reform have demonstrated the use administrative records in research and policy development. Cancian, Han, and Noyes (2014) matched and merged monthly public assistance and social service records to understand how specific elements of state policy, such as welfare time limits and sanctions, were affecting program participants. Other efforts (e.g. Allard & Danziger, 2002; Beimers & Coulton, 2011; Davis, Lim, & Livermore, 2011; Larson, Singh, & Lewis, 2011; Needell, Cuccaro-Alamin, Brookhart, & Lee, 1999; Wu, Cancian, & Wallace, 2014) also included administrative records from other systems, such as employment and child welfare, to examine how the policies implemented by one agency could affect outcomes observed in another.

The social sector can draw upon the considerable experience of the child welfare field in using large administrative databases. Wulczyn, Chen, and Hislop (2007) aggregated and analyzed child welfare administrative records from multiple states to derive comparative service patterns, outcome metrics, and policy recommendations. Large numbers of child welfare records have been linked to birth and death certificates to examine a range of epidemiological issue such as rates of teen childbearing among foster children (Putnam-Hornstein & King, 2014) and injury death following a report of maltreatment (Putnam-Hornstein, Cleves, Licht, & Needell, 2013). Recent efforts have also examined whether integrated administrative records can be used to identify children at high risk of subsequent maltreatment in an effort to more strategically target prevention services (Vaithianathan, Maloney, Putnam-Hornstein, & Jiang, 2013).

The homeless services field has also benefitted greatly from discoveries from analyzing large caches of administrative records. Studies have combined homelessness system records from several cities and linked them to other agency records to develop a typology for the field that has been widely applied (Culhane, Metraux, Park, Schretzman, & Valente, 2007). Additionally, systemic analysis of outcomes and costs found through mining administrative records have led to significant modifications of public policy regarding homelessness (Culhane, Metraux, & Hadly, 2002).

Randomized control trials are also beginning to find applications for big data. If interconnecting information from a variety of available databases can track outcomes for experimental and control subjects, findings can be produced quicker and at lower cost than conventional studies that require expensive follow-up studies (Coalition for Evidence Based Policy, 2012). For example, an experimental test of a new case management model with disabled Medicaid beneficiaries relied on linked individual records from several social service agencies to show positive impact on emergency hospital readmissions, homelessness, and receipt of substance abuse treatment, yet still failed to produce cost savings (Bell et al., 2014). This type of information was directly pertinent to government decisions. The for-profit sector has fruitfully used big data in a form of low-cost continuous experimentation to test the impact of small changes in practice or production. If social sector data systems were brought up to speed, there is reason to believe that similar experiment-driven innovation could be useful in refining programs on a regular basis (Manzi, 2012).

Analyses of administrative records data have yielded a great deal of useful information, but that work was typically carried out in the context of discrete research projects and, thus, did not increase institutionalized big data capacity. However, there is a growing movement to build and maintain multiagency integrated data systems (IDSs) as a permanent utility for the social sector. In most IDSs, administrative records from many agencies are retrieved on an ongoing basis, linked at the individual level, cleaned and organized, and made available for analysis. Though these systems are under development, they have great potential to deliver high quality big data with almost unlimited possibilities to yield vital information to transform social policy and practice. For example, such systems are already making it possible to estimate the scope of multiple system use and the associated costs, suggesting starting points for overall improvements in the social service sector (Goerge, Smithgall, Seshadri, & Ballard, 2010). They are also foundational for data driven social innovations such as community-wide, collaborative impact projects (London & McLaughlin, 2014) and social impact financing (Stoesz, 2014; Butler, Bloom, & Rudd, 2013). A series of case studies of selected IDSs affirm the technical and practical feasibility of building these systems and generating actionable information for policy¹.

Social services IDSs are not the only innovative type of integrated data platform emerging in the social sector. Community revitalization work is benefiting from “mash ups” of property,

¹ See the work of the *Actionable Intelligence for Social Policy* network at <http://www.aisp.upenn.edu/resources/network-site-case-studies/>

housing, and neighborhood data using geographic information system technology (Kingsley, Coulton, & Pettit, 2014). For example, community partners in some cities have been enabled to fight blight and disinvestment using open data portals that link numerous transactional records (e.g., foreclosure filings, deed transfers, evictions, complaints) in real time and make them available through a user friendly interface for action on the ground (Nelson, 2014). Among other things, the data have been instrumental in alerting foreclosure prevention counselors and mitigation specialists to reach out to specific housing units and their occupants based on predictive models of risk found from mining the data and to fight predatory lending with data-based evidence (Coulton, Schramm, & Hirsh, 2010).

In addition to administrative records, other types of big data are proving useful in the social sector. Internet searches are an important source that could be analyzed to track trends and patterns in social conditions. For example, a study of Google searches related to child maltreatment during the last recession showed trends that tracked closely to rates computed from child mortality records (Stephen-Davidowitz, 2013). Interestingly, the trends found in Google searches were opposite from the declining trends based on official child maltreatment reports during the same period. These discrepancies among sources demonstrate a particular value of the capacity to mine big data. Because almost any data source is subject to some type of bias, there is high value in comparing alternative data sources to gain deeper insight into social processes.

Similarly, social media posts also promise to be fruitful data sources for informing the social sector. One application has been to identify emerging social movements and issues. Using Twitter's own application program interface (API), Tinati, Halford, Carr, and Pope (2014) used tweets during a student movement to monitor the expansion and scope of political activism in a dynamic fashion. By applying these tools, it is also feasible for community activists to observe trending social network activities in real time to target their efforts, prevent emerging problems, or promote positive change.

To get the most out of big data, it is necessary to move beyond the reliance on structured data fields and standard statistical models, an area in which the social sector has made some progress. In particular, there is a great deal that can be learned from detailed case notes, assessment reports, or other digital text that are part of agency records. Text mining methods are evolving rapidly and are now being applied to unstructured notes to gain a more complete picture of behaviors. For example, unstructured text fields in child welfare records have been analyzed using knowledge engineering and text mining tools to discover patterns of substance abuse risk that could not be ascertained by simply tabulating structured data fields (Goerge, Ozik, & Collier, 2015). Similarly, the application of a hybrid text mining approach to massive numbers of text records in the Veterans Administration's data warehouse yielded an innovative method to calibrate suicide risk in that population (Hammond & Laundry, 2014). Moving beyond standard statistical approaches, machine learning algorithms using hold-out samples and cross validation are promising for the analysis of big data (Hindman, 2015). Microsimulation models are also proving useful for evaluating how policy or practice changes might play out over entire human service systems (Goldhaber-Fiebert et al., 2012).

MEANINGFUL AND MEASURABLE PROGRESS CAN BE MADE IN A DECADE

In the foreseeable future, it is unlikely that social workers will be considering data on the scale of the Large Hadron Collider at CERN, which produces more than 10 petabytes of data on an annual basis and is connected to the largest computer network on earth (Brumfiel, 2011). Though the scale is modest compared to the analytical challenges physicists and nuclear engineers face, many of the complex data sets now available in the social sector have the potential to inform policy and practice responses to persistent social issues such as poverty, crime, child maltreatment, and health disparities. However, these data are not simply ready for the taking. Getting there will require significant advances in the field's capacity to access, manage, and curate complicated data sets. The social sector will also need to tackle legal and structural barriers to data sharing, enable agencies and communities to incorporate data analytics into their work, apply innovative methods for data analysis, and develop a new generation of personnel with the expertise and skills to meet the technical and methodological demands of big data. Nevertheless, measurable progress on all of these fronts can be made in a decade.

Expand Data Access and Use

Data sharing and access is a complicated issue whenever personal information is involved or data are viewed as proprietary. Many federal agencies are beginning to modify statutes or issue directives permitting data sharing for research or program improvement purposes as long as rigorous privacy protection is observed. Yet resistance to data sharing is widespread and there is tremendous unevenness in data governance practices at state and local levels and in the non-profit sector. A thorough review of the status of data access in the social sector at all levels is necessary. Investing in systematizing and standardizing these policies and practices would be a first step toward progress in this area.

One-time access to big data for specific projects is inefficient. It is also necessary to significantly expand the development of IDSs that link records across social service agencies, add value through validating data elements, and create meaningful variables and make deidentified data available for analysis and use. These systems rely on a regular stream of data, flowing periodically or in real time. Once established, these systems will be able to feed relevant data to various applications, ranging from “dashboards” for agency staff to policy research to public-facing websites that provide insightful data visualizations. They will also supply empirically based information to support effective action in the social sector.

Some social researchers have access to big data and others do not, which needs to be addressed. Relatively few archives, other than open data portals, readily provide big data to anyone interested in using it. To access big data, social work professionals must form and maintain relationships with numerous data providers. Increasing the stake of data providers in the research and applications of the findings would benefit researchers by elevating these relationships to partnerships and collaborations. Such communication channels would also increase the usability of the information and its value for addressing societal concerns.

Implement Advanced Data Management, Security, and Analysis

Gaining access to many data sources is just the beginning. Big data is notoriously messy and a lot of work goes into cleaning and organizing it to make it meaningful. These processes (often referred to as *data wrangling*) can be overwhelming if done manually, so it is essential to develop standardization tools that allow data users and analysts to efficiently extract and prepare data that come from these distributed data systems. In the social sector, progress is needed in applying data capture mechanisms that allow crowdsourcing and greater automation. A method for sharing and refining these methods throughout the sector must be established.

Turning administrative records, images, transactions, and text into useable information requires significant investment in validating the data. This begins with a very deep and careful assessment and documentation of the processes that generate the data in collaboration with the agencies from which it is sourced. To be practical, these understandings need to be converted to algorithms and code that allow the raw data to be efficiently standardized, harmonized, and evaluated for quality. Developing robust methods of evaluating and addressing missing data and error structures is necessary.

Privacy and confidentiality have always been core values of social workers in their practice. But with big data, the usual methods of informed consent may not be applicable, and rigorous methods of protecting personal data are clearly required (Strandburg, 2014). As universities and other research units move to use big data, they will need to take data security and data stewardship to a higher level. To ensure the confidentiality of identified data with a limited number of staff who work with sensitive data, universities and other research units must also enact specific policies and procedures to protect against disclosure and its negative effects.

Big data requires the social sector to move beyond standard analysis techniques. This involves scaling up and adopting statistical methods consistent with the nature of the data (National Research Council, 2013). The social sector's existing focus on hypothesis testing, statistical significance, and power analysis—necessary when sample size is modest—becomes nearly irrelevant with big data. The data lend themselves to techniques such as data mining, predictive modeling, and simulation, all of which are quickly evolving. In this regard, social work should become active participants in the growing movement for a computational social science (Lazer et al., 2009). Adding expertise specifically in methods applicable for big data will help the field avoid erroneous statistical conclusions that could arise through the application of traditional sample based methods.

Both a deep understanding of the represented population and also the eventual application of the generated knowledge must inform the analysis of big data. For example, big data analysts must be aware that data trails that result from a human choice (e.g., making a social media post, running a search) are limited to populations that engage in such activities. Similarly, public agency records represent only those who interact with or are known to government agencies. With big data, the usual assumptions of probability based sampling seldom apply, and methods of compensating for the unrepresentativeness of samples, such as weighting, are not

straightforward and need innovative solutions. It is crucial that big data analysts not mistakenly communicate certainty based on large sample sizes alone; rather, they must carefully evaluate the limitations before issuing estimates of prevalence rates, effect sizes, and trends (Kaplan, Chambers, & Glasgow, 2014).

Prepare Data Savvy Social Workers

The field needs to attract, train, and retain a generation of social workers who are passionate about data and able to work in teams to manage, curate, analyze, interpret, and apply big data for social good. Some social workers will choose to specialize, but it will also be important to prepare all social workers to appreciate their role in data generation and become adept applying data in their practice (Naccarato, 2010; Shaw, Lee, & Wulczyn, 2012).

A significant problem is that most social work and social welfare degree programs do not explicitly train students in computational skills beyond standard statistics. Few offer education on large and complex data sets. Social work masters and doctoral programs will need to incorporate special training in data science. Schools of social work should consider developing joint programs with computer science departments to train some social workers in both fields. This is already occurring in selected public policy and social science departments, suggesting that social work can also move in this direction.

Such programs would attract a new type of student to the field and give those already committed to social work an opportunity to employ technology and data science to their practice, analysis and research. The availability of such opportunities, when widely promoted, will bring in a number of talented individuals who want to be part of the digital revolution, but also have an interest in having an impact in their communities.

Promote Data Driven Social Policy and Practice

The real benefit of big data for the social sector will become evident as the new information translates to improved decision making and social action at all levels. Reaching that point will require significant changes in organizational culture, along with advanced technology to convert data into actionable information. Ideally, policymakers, administrators, and practitioners will increasingly call for data to inform decision making until it becomes the norm. At that point, the field will need the capacity to deliver data in forms that support thinking and action. Data querying and display capability has to be crafted to support the ways administrators, practitioners, and clients approach problems and choices. Visualization, simulation, and crowdsourcing capabilities are interfaces whose implementation could support the social sector's uses of data in action. Results from predictive analytics can be programmed into clinical and policy decision support systems so that the most promising approach can be chosen for specific individuals or situations. With a high degree of empirically derived customization, the social sector could maximize the positive impact of social spending on the well-being of society.

HARNESSING BIG DATA REQUIRES CROSS-SECTOR, INTERDISCIPLINARY COLLABORATION

Turning big data into socially useful information requires a combination of substantive knowledge and a newly developing mix of skills referred to as data or computational science. This work is inherently interdisciplinary and would be carried out by teams of specialists across social work, social science, computer science, statistics, and related domains. Educational and training programs are being launched to deliberately cross and even blur these boundaries. Social work specialists, with the training and skills to work across the boundaries, will be essential to moving the field forward.

Government Collaboration

There are some additional boundaries to cross if big data is to be applied successfully in the social sector. Both the bulk of social program funding and a plethora of useful data come from government agencies. Though some government officials are pushing for open data and data informed policy, there are also barriers at various levels of the bureaucracy. Through cooperative efforts and coalition-building across professions and levels of government, such challenges can be overcome. Some government agencies are already out in front in this regard. For example, in the early childhood sector there is a growing emphasis on the development of shared data systems (Demma, 2010). Similarly, the federal government has made cross agency data sharing and analysis a vital component of several of its place-based initiatives such as Promise Neighborhoods and Choice Neighborhoods (Rubin & MacAfee, 2014).

Business Sector Collaboration

Given big data's value in commercial enterprises, the business sector is another strong potential partner. Through philanthropies and private investment, companies are taking a strong interest in improving performance in the social sector. The growing interest in social impact bonds, also known as "pay for success," demonstrates the synergy of big data analytics and the desires of investors to get involved with social innovations that have a quantifiable possibility of achieving social benefit and return on investment. These projects rely on projections of cost savings and impact that can only be obtained through the rigorous analysis of data from multiple administrative data systems (Leibman, 2013b; Overholser, 2014). The strong business interests in this arena, along with the pressing need for improved access to data and analytic models, are forces that can move the field forward in the future.

Social Movement Collaboration

Important social movements are evolving around advocacy for open data and citizens engaged in community coding projects. The power of these movements can be harnessed to develop novel applications for data to tackle social problems. The resulting civic engagement would strengthen communities and spread the profession's influence.

THE SOLUTION TO THE CHALLENGE OF HARNESSING BIG DATA REQUIRES SIGNIFICANT INNOVATION

Big data has the power to transform the social sector. It promises to change the way government agencies and non-profit organizations make decisions and can markedly shift the curve on the rate of knowledge development and impact. There is no question that technical innovations borrowed from computer and data science will be required, and that social organizations will have to adopt new ways of managing and using data. Moreover, if the social sector is successful, it can increase the pace of new discoveries regarding social processes and problems that will contribute to novel solutions. Practice and policy will be informed by a continuous feedback loop, allowing resources to be deployed toward their highest value. The sector will become more transparent, accountable, and effective as a result.

REFERENCES

- Allard, S. W., & Danziger, S. (2002). Proximity and opportunity: How residence and race affect the employment of welfare recipients. *Housing Policy Debate*, 13(4), 675–700.
- Beimers, D., & Coulton, C. J. (2011). Do employment and type of exit influence child maltreatment among families leaving Temporary Assistance for Needy Families? *Children and Youth Services Review*, 33(7), 1112–1119.
- Bell, J. F., Krupski, A., Joesch, J. M., West, I. I., Atkins, D. C., Court, B., & Roy-Byrne, P. (2014). A randomized controlled trial of intensive care management for disabled Medicaid beneficiaries with high health care costs. *Health services research*. doi:10.1111/1475-6773.12258
- Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, 19(1), 5–16.
- Brumfiel, G. (2011). Down the petabyte highway. *Nature*, 469(20), 282–283.
- Burrows, R., & Savage, M. (2014). After the crisis? Big data and the methodological challenges of empirical sociology. *Big Data & Society*, April-June. 1–6.
- Butler, D., Bloom, D., & Rudd, T. (2013). Using social impact bonds to spur innovation, knowledge-building, and accountability. *Community Development Investment Review*. San Francisco, CA: Federal Reserve Bank of San Francisco.
- Cancian, M., Han, E., & Noyes, J. L. (2014). From multiple program participation to disconnection: Changing trajectories of TANF and SNAP beneficiaries in Wisconsin. *Children and Youth Services Review*, 42, 91–102.
- Coalition for Evidence-Based Policy. (2012). *Rigorous program evaluations on a budget: How low-cost randomized controlled trials are possible in many areas of social policy*. Washington, DC: Coalition for Evidence-Based Policy.
- Coulton, C. J. (2014, October). *Stepping up to harness big data for social good*, Presented at American Academy of Social Work and Social Welfare Induction Ceremony, Tampa, FL.
- Coulton, C. J., Schramm, M. & Hirsh, A. (2010). REO and beyond: The aftermath of the foreclosure crisis in Cuyahoga County, Ohio. In E. Rosengren & S. Pianalto (Eds.), *REO & vacant properties. Strategies for neighborhood stabilization* (pp. 47–54). Boston/Cleveland, Federal Reserve Banks of Boston and Cleveland.
- Culhane, D. P., Metraux, S., & Hadley, T. (2002). Public service reductions associated with placement of homeless persons with severe mental illness in supportive housing. *Housing Policy Debate*, 13(1), 107–163.
- Culhane, D. P., Metraux, S., Park, J. M., Schretzman, M., & Valente, J. (2007). Testing a

- typology of family homelessness based on patterns of public shelter utilization in four US jurisdictions: Implications for policy and program planning. *Housing Policy Debate*, 18(1), 1–28.
- Davis, B. C., Lim, Y., & Livermore, M. (2011). A fresh look at an old debate: Assigned work activities, employment, and post-program earnings in TANF work programs. *Journal of Policy Practice*, 10(2), 108–127.
- Desouza, K. C., & Smith, K. L. (2014). Big data for social innovation. *Stanford Social Innovation Review*, 9.
- Getz, L. (2014). Big data's impact on social services. *Social Work Today*, 14(2), 28.
- Goerge, R., Ozik, J., & Collier, N. (2015). *Bringing big data into public policy research: Text mining to acquire richer data on program participants, their behavior and services*. manuscript submitted for publication.
- Goerge, R., Smithgall, C., Seshadri, R., & Ballard, P. (2010). Illinois families and their use of multiple service systems. Chicago: *Chapin Hall at the University of Chicago*.
- Goldhaber-Fiebert, J. D., Bailey, S. L., Hurlburt, M. S., Zhang, J., Snowden, L. R., Wulczyn, F., & Horwitz, S. M. (2012). Evaluating child welfare policies with decision-analytic simulation models. *Administration and Policy in Mental Health and Mental Health Services Research*, 39(6), 466–477.
- Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261.
- Grogan-Kaylor, A., & Dunkle, R. (2014, January). *Promises and perils of big data for building the science of social work*, Presented at the Society for Social Work and Research Annual Conference, San Antonio, TX.
- Hammond, K. W., & Laundry, R. J. (2014, January). *Application of a hybrid text mining approach to the study of suicidal behavior in a large population*. Presented at System Sciences (HICSS) 2014 47th Hawaii International Conference, Waikoloa, HI.
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76.
- Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 48–62.
- Hoffman, S., & Podgurski, A. (2013). Big bad data: Law, public health, and biomedical databases. *The Journal of Law, Medicine & Ethics*, 41(s1), 56–60.
- Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., & Usher, A. (2015). *AAPOR*

- report on big data*. Retrieved from https://www.aapor.org/AAPORKentico/AAPOR_Main/media/Task-Force-Reports/BigDataTaskForceReport_FINAL_2_12_15_b.pdf
- Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*, 9(1), 36–41.
- Kaplan, R. M., Chambers, D. A., & Glasgow, R. E. (2014). Big data and large sample size: A cautionary note on the potential for bias. *Clinical and Translational Science*, 7(4), 342–346.
- King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331, 719–721.
- Kingsley, G. T., Coulton, C. J., & Pettit, K. L. (2014). Strengthening communities with neighborhood Data. Washington, DC: Urban Institute.
- Kum, H. C., Krishnamurthy, A., Machanavajjhala, A., & Ahalt, S. C. (2014). Social genome: Putting big data to work for population informatics. *Computer*, 47(1), 56–63.
- Larson, A. M., Singh, S., & Lewis, C. (2011). Sanctions and education outcomes for children in TANF families. *Child & Youth Services*, 32(3), 180–199.
- Lazer, D. M., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, 343, 1203–1205.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. & Van Alstyne, M. (2009). Computational Social Science, *Science*, 323(5915), 721–723.
- Leibman, J. (2013). Advancing evidence-based policymaking to solve social problems. *Issues in Science and Technology*, 30(1), 47–55.
- London, R. A. & McLaughlin, M. (2014). *The youth sector: Supporting cross-institutional community collaboration through shared data*. San Francisco, CA & Washington, DC: Federal Reserve Bank of San Francisco & Urban Institute.
- Manzi, J. (2012). *Uncontrolled: The surprising payoff of trial and error for business, politics, and society*. New York: Basic Books.
- Mervis, J. (2012). Agencies rally to tackle big data. *Science*, 336(6077), 22.
- Moore, K. A., Sacks, V. H., Manlove, J., & Sawhill, I. (2014). “What if” you earned a diploma and delayed parenthood? Research Brief (Publication #2014-27). Bethesda, MD: Child Trends.
- Naccarato, T. (2010). Child welfare informatics: A proposed subspecialty for social work. *Children and Youth Services Review*, 32(12), 1729–1734.
- National Research Council. (2013). *Frontiers in massive data analysis*, Washington, DC: The National Academies Press.

- Needell, B., Cuccaro-Alamin, S., Brookhart, A., & Lee, S. (1999). Transitions from AFDC to child welfare in California. *Children and Youth Services Review, 21*(9), 815–841.
- Nelson, L. (2014). Cutting through the fog. In G.T. Kingsley, C. J. Coulton, & K. L. Pettit (Eds.), *Strengthening communities with neighborhood data* (pp. 205–218). Washington, DC: Urban Institute.
- Overholser, G. (2014). Hearing on social impact bonds and families in need, Submitted to House Committee on Ways and Means, Subcommittee on Human Resources (September 9).
- Parker, S., Uprichard, E., & Burrows, R. (2007). Class places and place classes geodemographics and the spatialization of class. *Information, Communication & Society, 10*(6), 902–921.
- Putnam-Hornstein, E., Cleves, M.A., Licht, R., & Needell, B. (2013). Risk of fatal injury in young children following abuse allegations: evidence from a prospective, population-based study. *American Journal of Public Health, 103*(10), e39–e44.
- Putnam-Hornstein, E. & King, B. (2014). Cumulative teen birth rates among girls in foster care at age 17: An analysis of linked birth and child protection records from California. *Child Abuse & Neglect, 38*(4), 698–705.
- Rudin, C., Dunson, D., Irizarry, R., Ji, H., Laber, E., Leek, J., & Xue, L. (2014). *Discovery with data: Leveraging statistics with computer science to transform science and society*. A Working Group of the American Statistical Association.
- Schroeder, R. (2014). Big data and the brave new world of social media research. *Big Data & Society, 1*(2). doi: 10.1177/2053951714563194
- Shaw, T. V., Lee, B. R., & Wulczyn, F. (2012). “I thought I hated data”: Preparing MSW students for data-driven practice. *Journal of Teaching in Social Work, 32*(1), 78–89.
- Stephens-Davidowitz, S. (2013). *Unreported victims of an economic downtown*. Harvard University. Retrieved from <http://static1.squarespace.com/static/51d894bee4b01caf88ccb4f3/t/51e22f38e4b0502fe211fab7/1373777720363/childabusepaper13.pdf>
- Stoesz, D. (2014). Evidence-based policy reorganizing social services through accountable care organizations and social impact bonds. *Research on Social Work Practice, 24*(2), 181–185.
- Strandburg, K. J. (2014). Monitoring, datafication and consent: Legal approaches to privacy in a big data context. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum, (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (pp. 5–43). New York, NY: Cambridge University Press.
- Taylor, L., Schroeder, R., & Meyer, E. (2014). Emerging practices and perspectives on big data analysis in economics: Bigger and better or more of the same? *Big Data & Society, 1*(2). 1–10.

- Tinati, R., Halford, S., Carr, L., & Pope, C. (2014). Big data: Methodological challenges and approaches for sociological analysis. *Sociology*, *48*(4), 663–681.
- Vaithianathan, R., Maloney, T., Putnam-Hornstein, E., & Jiang, N. (2013). Children in the public benefit system at risk of maltreatment: Identification via predictive modeling. *American Journal of Preventive Medicine*, *45*(3), 354–359.
- Wu, C. F., Cancian, M., & Wallace, G. (2014). The effect of welfare sanctions on TANF exits and employment. *Children and Youth Services Review*, *36*, 1–14.
- Wulczyn, F., Chen, L., & Hislop, K. B. (2007). *Foster care dynamics, 2000-2005: A report from the multistate foster care data archive*. Chicago: Chapin Hall Center for Children at the University of Chicago.

ABOUT THE AUTHORS

CLAUDIA J. COULTON² is the Lillian F. Harris Professor of Urban Research and Social Change at the Jack, Joseph and Morton Mandel School of Applied Social Sciences at Case Western Reserve University.

ROBERT GOERGE is a Chapin Hall Senior Research Fellow at the University of Chicago.

EMILY PUTNAM-HORNSTEIN is an Assistant Professor at the University of Southern California School of Social Work.

BENJAMIN DE HAAN is the Executive Director of Partners for Our Children.

ACKNOWLEDGMENTS

Sandra Audia Little at the University of Maryland School of Social Work designed the cover. John Gabbert at Washington University's Center for Social Development provided editorial support.

SUGGESTED CITATION

Coulton, C. J., Goerge, R., Putnam-Hornstein, E., & de Haan, B. (2015). *Harnessing big data for social good: A grand challenge for social work* (Grand Challenges for Social Work Initiative Working Paper No. 11). Cleveland, OH: American Academy of Social Work and Social Welfare.

CONTACT

American Academy of Social Work and Social Welfare
Sarah Christa Butts, Assistant to the President

academy@aswsw.org

² Corresponding author. Email: claudia.coulton@case.edu; office phone: 216-368-2304.