Census Big Data Analytics Use: International Cross Case Analysis

Akemi Takeoka Chatfield
University of Wollongong
Wollongong, NSW
Australia
akemi@uow.edu.au

Adegboyega Ojo
Insight Centre for Data Analytics
National University of Ireland, Galway
Ireland
adegboyega.ojo@nuigalway.ie

Gabriel Puron-Cid
Centro de Investigación y Docencia Económicas
Aguascalientes
Mexico
gabriel.puron@cide.edu

Christopher G. Reddick
The University of Texas at San Antonio
San Antonio, Texas
U.S.A.
Chris.Reddick@utsa.edu

ABSTRACT

Despite the growing practices in big data and big data analytics use, there is still the paucity of research on links between government big data analytics use and public value creation. This multi-case study of Australia, Ireland, Mexico, and U.S.A. examines the state of big data and big data analytics use in the national census context. The census agencies are at varying stages in digitally transforming their national census process, products and services through assimilating and using big data and big data analytics. The cross-case analysis of government websites and documents identified emerging agency challenges in creating public value in the national census context: (1) big data analytics capability development, (2) cross agency data access and data integration, and (3) data security, privacy & trust. Based on the insights gained, a research model aims to postulate the possible links among challenges, big data/big data analytics use, and public value creation.

KEYWORDS

Big data analytics, census big data, use, big data challenges, cross case analysis, public value creation, electronic census.

dg.o '18, May 30-June 1, 2018, Delft, Netherlands

ACM ISBN 978-1-4503-6526-0/18/05…$15.00
https://doi.org/10.1145/3209281.3209372

1 INTRODUCTION

Open and big data analytics use in government aims at discovering patterns, insights and trends that could impact government's policies, operations, public service delivery, and relationships with citizens. The government data and analytics landscape is rapidly changing across the world. Some of the contributing factors to the government's changing data and analytics landscape include the increased digitization of government and society, exploitation of big, open and linked data for smarter policymaking and co-creation of new digital services, and the new technological and organizational dimensions of mobility that are accelerated through open government data, social media, Internet of things, and cloud computing in government. To date, data analytics in government have been deployed to fight fraud and abuse in tax and welfare benefits services, develop safer and more efficient public transportation, fight crimes and cut down internal operational costs as well as to predict disaster impacts in the local government areas and epidemic spreading across national borders. Despite the growing anecdotal evidences and practices in this domain, there is still the paucity of digital government research on links between government use of big data analytics and public value creation. This paper is motivated to address this research-practice gap by raising the following two research questions:

RQ1: *What is the state of big data and big data analytics use in the national census context?*
RQ2: *What are key challenges for government in using big data analytics and creating public value?*

This paper aims to address these research questions in the census open data research context. We have conducted website analysis of the census big data analytics use in four countries: Australia, Ireland, Mexico, and U.S.A. We hold that the cross-case analysis findings provide a timely and critical examination of pathways from insights to public value creation with some important public policy implications to address key challenges and enhance public administration evidence-driven decision making through big data analytics use.

The structure of the reminder of this paper is as follows: The second section presents a review of the literature on the role of big data analytics, big data analytics use in government, and challenges. The third section describes our research methodology including a generic strategy for data and document collection in the national census context. The fourth section presents our within case analysis results of the four countries: Australia, Ireland, Mexico, and the U.S.A. The fifth section presents key cross-case

analysis results. Finally, the sixth section presents our discussion of a new research model and conclusions including our research limitations and future research directions.

2 LITERATURE REVIEW

2.1 The Role of Big Data Analytics

While volume is often a discernible dimension of big data, prior research on big data analytics-enabled customer agility [3] holds that big data is increasingly characterized by four V's: volume (scale of data), variety (different data formats), and velocity (streaming data) [19], [20], and veracity (uncertain data quality) [11]. In this paper big data analytics refers to:

1) Data analysis being undertaken that uses high volume of data from a variety of sources including structured, semi structured, unstructured or even incomplete data; and
2) The phenomenon whereby the size (volume) of the data sets within the data analysis and velocity with which they need to be analyzed has outpaced the current abilities of standard business intelligence tools and methods of analysis.
3) The complexity of the relationships within complex structures embedded in the data has reached a level that cannot be handled by the current tools and models of statistics and analysis [9].

At all levels of government, data-driven decision-making is considered one of the vital challenges of working with big data. There is no value in big data unless it is analyzed to produce critical information which can be effectively used to make informed decisions [4]. It is clear that poor data quality can impact the process of decision making negatively, leading to unacceptable and untrustworthy inferences or conclusions [13] and [12]. Because big data can be considered a trusted basis for decision-making in organizations, to achieve this trust, government decision makers need to embrace data-driven decision making [20], [29], [24]. In addition, data-driven decision-making can be a critical challenge if large amounts of data are not effectively processed through efficient analysis and effective implementation [31]. For this reason, it is essential for decision-makers to comprehensively understand big data-related technology so as to make optimal decisions.

2.2 Big Data Analytics Use in Government

The United Nations (UN) Global Working Group (GWG) on the Use of Big Data for Official Statistics was created in 2014 at the 45th session of the UN Statistical Commission [27]. The UN Statistical Commission recognized big data as a new source of information which government statistical agencies need to exploit and harness its challenges. The UN GWG was charged to take an inventory of activities and tangible examples of the use of big data for official statistics, propose methodological frameworks, and examine those countries in advanced IT use and in the use of open data platforms in the context.

In October 2015, the first meeting of the UN GWG agreed to address the following eight issues on: (1) sharing lessons learned from big data projects; (2) big data quality; (3) coordination of the group with the work undertaken in Europe; (4) existing big data technology gap of developing countries; (5) legal frameworks for access to the private sector big data resources and privacy concerns; (6) big data capacity building especially data science and IT; (7) limitations of big data for official statistics, and (8) building partnerships on the use of big data with the private sector, research institutes, and academia [25]. The UN Global Working Group proposed to examine not only mobile phone data, social media data, and satellite imagery data but also transaction data through e-commerce and transboundary big data.

More recently in 2017, the GWG agreed on the need to develop the business case for the global platform to formulate a policy framework for data governance and information management including issues of trust, privacy, confidentiality and security of data [26].

2.3 Public Value Challenges

With the growing adoption and use of big data in both private sector and public sector organizations, there is the emergence of the academic literature on the sector-specific big data benefits [23], [8] as well as cross-sector organizational, technological, political, and social challenges in adopting and using big data in organizations [4]. However, despite the emergence of big data analytics use in government, there remains a clear lack of the academic literature on various challenges facing different levels of government in using big data analytics and creating public value.

3 METHODOLOGY

3.1 Census Data & Census Data Provider

What is a census? A census is "a count of all people and households" which is undertaken every ten years in the U.K. [22]. Census data as open big data is a valuable asset for government as well as for the public for a number of reasons. First, a census is the only survey that provides the most complete source of information about the entire population [22], [1]. Second, census data provides all levels of government, health organizations, and other organizations to allocate their resources more effectively and more strategically for the planning of future public services [22], [1]. Third, census data are used to set electoral boundaries for all levels of government [2].

Table 1 shows country, the agency responsible for openly publishing the census data on the Internet for the public, and its website of the four countries we have studied in this paper.

Table 1: Country, Census Data Provider & Website

Country	Census Data Provider & Website
Australia	Australian Bureau of Statistics, http://www.abs.gov.au/websitedbs/censushome.nsf/4a256353001af3ed4b2562bb00121564/data
Ireland	The Northern Ireland Statistics and Research Agency (NISTRA), an agency of the Department of Finance, https://www.nisra.gov.uk/
Mexico	The Instituto Nacional de Estadística y Geografía (INEGI) (National Institute of Statistics and Geography), www.inegi.org.mx
U.S.A.	United States Census Bureau, https://www.census.gov/data.html

The challenge for census data analytics research is that the census data provider and the agency responsible for census data analytics may be same or different. On the one hand, Australian Bureau of Statistics (ABS) provides census data online as well as is in charge of census data analytics. On the other hand, in the case of Mexico, while the National Institute of Statistics and Geography (INEGI) provides census data online for the public, the National Population Council (CONAPO), under the Secretary of the Interior, is responsible for the analysis and research of population dynamics.

3.2 Secondary Source Data Collection & Analysis

We have conducted within case analysis of each of the four countries in the first phase of our research, which is followed by cross case analysis of the four countries at the second phase.

For each within case analysis of the four countries we have collected secondary source data from the government agency responsible for census data collection and/or analysis (see Table 1 for the website URL). In addressing the RQ1 we used the website search engine to find evidence of big data use and big data analytics use including the provision of geospatial data and big data analytics tools for the public. In general, we employed the following generic search strategy: (document) CONTAINS ("census data") as the primary key all years which was constrained by the use of secondary keyword ("big data") OR (“big data analytics”) OR (“data analytics”). The search strategy aimed to identify all relevant government documents publicly released on the Internet, which contain the primary and secondary keywords. We also replicated the same search strategy within the census agency website (see Table 1) using its website search engine to understand use cases in the census context.

In addressing the RQ2 we used the website search engine to perform a within case analysis of policy frameworks, agency annual reports, agency organizational chart, and other documents that may be relevant for us to identify organizational, technological, political and social challenges in creating public value through big data analytics use in process of achieving the agency’s mission.

For the second phase cross-case analysis we have compared and contrasted major findings of each within case analysis to identify common challenges faced by the four census agencies that would be required to answer the RQ2.

4 WITHIN CASE ANALYSIS RESULTS

4.1 Australia

4.1.1 Census Data and Big Data Analytics Use. Australian Bureau of Statistics (ABS) conducts a census every five years, for instance, most recently in 2016. The mission of ABS as Australian government’s central statistics and data provider is to seek “*to deliver the most public value*” by informing “Australia's important decisions through partnering and innovating to deliver relevant, trusted, objective *data*, *statistics* and *insights*” [2]. ABS has recently restructured to adapt to the dynamically changing data environment with three groups: Statistical Services Group, Transformation Group, and Census & Enabling Services Group. While the Statistical Services Group is responsible for the provision of official statistics such as macroeconomic statistics, industry statistics, and population & social statistics, the Transformation Group leads the agency’s digital transformation and methodological transformation required to adapt to the dynamically changing digital data environment. Finally, the Census & Enabling Services Group, where Chief Information Officer is located, provides technology & security services, census & statistical network services, and finance, risk & planning services.

For the 17th national census in 2016 ABS has pursued the “digital-first” strategy objective to obtain more than 65% of the 2016 national census response online. The result shows 63.3% of the Australians completed the census using the electronic form, which fell short of the ABS objective. The ABS claims that this was a significant increase from the 37.9% of the census completed online in 2011 [2].

The Department of Finance of the Australian Commonwealth Government has published a national framework, *Australian Public Service Better Practice Guide for Big Data* [9]. Basically, it argues that the potential impact of big data analytics on government policy development, innovation, and public service delivery can increase the value of national data asset to government and the public. *The Big Data Strategy Working Group*, a multi-agency working group, was established in February 2013, while *the Whole of Government Data Analytics Centre of Excellence* (DACoE) Leadership Group was established in August 2013 to help the adoption of the national framework and the development of big data analytics capability and text analytics for government.

Against this national big data analytics backdrop, the ABS has developed *Graphically-Linked Information Discovery Environment* (GLIDE) [2]. GLIDE, an ABS developed big data analytics platform, enables to link complex big data in a variety of formats and from multiple sources, visualize it in different ways through interactive maps, and analyze it interactively. Using the GLIDE, ABS and other government agency collaborated on a Freight Performance Measurement Project. It used extant freight vehicle telematics data to provide regular road freight statistics and identify congested road freight network locations and calculate average travel times on major freight routes. This cross-agency collaborative project demonstrated the value of integrating the existing data to create new information and insights to inform important public policy decisions.

ABS provided nearly 500 statistical products had more than 16 million unique visitors to its website and enabled 2.5 million data downloads in one year ending 2017. Recently, ABS developed ABS DataLab to provide government and the public secure remote access to detailed microdata (e.g. census microdata), which refers to the specific variables recorded for each individual respondent. This DataLab provides 491 registered users from 40 different government and academic organizations with increased capability to handle large datasets with faster

processing through new servers for research, policy, and statistical research purposes [1].

The NationalMap platform, hosted by the data.gov.au website, offers a platform for discovering and visualizing a broad range of open government datasets across various geographies of Australia. In 2016–17, the ABS collaborated with Data61 (a business unit of Commonwealth Scientific & Industrial Research Organisation - CSIRO) to develop a new better way to provide ABS statistics directly to NationalMap with SDMX-JSON, a new developer-friendly data format which enables external users to directly access ABS statistics. This new functionality increased the amount of ABS latest data available to NationalMap without the need for manual update.

Finally, ABS also developed *the Statistical Spatial Framework* (SSF) for Australia in response to the global challenge, which the Secretary General of the UN Economic and Social Council 2012 refers to as "a better integration of geospatial and statistical information a basis for sound and evidence-based decision making" according to [21]. The SSF provides a national integrative approach for connecting people-centric (socio-economic) data to a geographic location to improve the accessibility and usability of this spatially-enabled data and metadata interoperability. Peter Harper, Deputy Australian Statistician, ABS stated:

> *The SSF will provide a bridge between the statistical and spatial communities, allowing information to move more effectively and consistently between these two groups.*

4.1.2 Barriers to Big Data Analytics and Public Value. Based on the within case analysis of the ABS, there are at least four key barriers that may affect public value of increased data use/reuse in the increasingly big data environment: challenges related to (1) *big data analytics*, (2) *data integration*, (3) *digital transformation*, and (4) *data security, privacy & trust*.

First, as discussed above in 4.1.1, ABS has developed capabilities in developing and using big data analytics technology platforms such as GLIDE and DataLab, enhanced the NationalMap platform functionality as well as developed the Statistical Spatial Framework (SSF). While the ABS used increasingly media and social media (Twitter and Facebook) to raise awareness of the public about the important role played by ABS statistics for informed decision making, the ABS still faces the challenge in big data analytics use across its stakeholders. Second, as discussed above, data integration adds value in creating new information and insights. However, data integration between traditional official statistics and big data such as geospatial data still challenges other government agencies. ABS needs to play a leadership role in addressing this data integration challenge. In addition, cross agency data integration challenges remain, indicating the need for collaborative network governance.

Third, the ABS has restructured to create a new Transformation Group which is charged to transform the ABS capability for "delivering official statistics in an 'information age' where there is increasing pressure to deliver more statistical information, more quickly and in more creative ways" [1] when the ABS faces "an increasingly complex environment with significantly less resources than previous years or comparable national statistical organisations" [1].

Finally, the ABS faced the criticality of data security, privacy and trust issues during the 2016 census when the ABS website for online census was maliciously hacked and disrupted by distributed denial of service attacks. In the aftermath of the chaos, various online media reported the incidence, for example, "Census hacked: Website for Australian census 2016 attacked, ABS says" [15]. The ABS annual report 2016-2017 mentioned that the Senate referred the 2016 Census to the Senate Economics References Committee for inquiry and report by 24 November. A public hearing was held on 25 October. While the "digital-first" strategy will be upheld for the 2021 eCensus, the ABS established a new independent census panel for data quality and data security audit review.

4.2 Ireland

4.2.1 Census Data and Big Data Analytics Use. The Irish Government has been conducting censuses of the population since 1821 and has carried out two censuses within the last decade (in 2011 and 2016) as basis for government planning and decision making. With census data generally considered as dataset of high value, the Central Statistics Office (CSO) of the Irish government has been investing on how to harness census data as open public data since 2011 [7]. The CSO in collaboration with one of its national research institutes published the 2011 census data as linked open data access (available at http://data.cso.ie/), which could be integrated with other public data across sectors such as education and health. The CSO has also pursued initiatives to demonstrate the potential public value of published 2011 linked open census data for private enterprises through the flagship European Commission Framework Programme 7 project, OpenCube[1]. Through the OpenCube project, the CSO sought to work with research institutions and SMEs in Ireland and Europe to develop the capability for publishing and enriching Linked Open Statistical Data for the development of data analytics and enhanced visualization services. This effort culminated in the publishing of a large number of linked open statistical data (LOSD) as open public data.

Against this background and the strategic interest of Irish government in the big data sector in the context of its '*Disruptive Reforms' in Action Plan for Jobs 2013*[2] initiative, the Central Statistics Office and its partners have been looking at how to harness big data technologies in census and statistical data. The increasing realization that big data offers the opportunity to produce more timely official statistics, significantly reducing their processing, and to do so in an on-going basis is another major pull for the CSO towards big data adoption(Eurostat 2013) [17]. In fact, in its latest strategy (2016-219) document, one of CSO's

[1] http://opencube-project.eu/

[2] https://www.siliconrepublic.com/enterprise/ibm-and-dcu-create-masters-degree-in-big-data-analytics-and-smart-cities

strategic priorities is *increasing the use of secondary data sources including administrative and big data* [7]. To realize this, the CSO also wishes to modernize its statistical processing and systems. This is expected to include electronic sensors and developing big data processing capabilities.

As a step in this direction, the CSO produced and large volumes of open census data from the results of 2016 Census which took place on April 24, 2016. Specifically, in addition to the traditional electronic census report, the CSO provided interactive web reports and interactive maps from census report[3]. In addition, CSO has also in collaboration with Ordinance Survey Ireland (OSi) launched a new census website based on private-sector owned (Esri Ireland) ArcGIS digital mapping platform. This platform allows the Census 2016 data to be visualized in map form and at the same available for exploitation by other organizations and the public for different ends. These interactive maps cover covering 15 different themes (see http://census2016.geohive.ie/).

However, like other National Statistical Institutions in Europe, CSO faces a number of challenges articulated in the Scheveningen Memorandum (Eurostat 2013) and [18] which are discussed below.

4.2.2 Challenges to Big Data Analytics and Public Value.

Against the backdrop of the 2016 Census, we highlight here some of the key challenges that have been identified regarding the attainment of CSO's goals regarding the use of big data as secondary statistical data sources.

Capability and skills – Big data skills are still relatively in low supply in Ireland. It will take a few years for some of the forward looking initiatives such as collaborative postgraduate degrees programs in Big Data Analytics between universities and industry[4] and setting up national big data research centre to have real impact in terms of big data skill supply [10]. The CSO will have to compete with industry players to attract the available big data talents to help realize their goals of human resource

Legislative - The second challenge is the lack of necessary legislative framework to deal with the implications of Big Data with regard to data protection and personal rights; for instance when considering access to Big Data sources held by third parties [10].

Access to the data – The third challenge identified in [17] includes gaining access required secondary big data such as weather data, administrative data from other government agencies, transportation, social media and retail data. Given that these data are valuable assets to the organizations that own them, CSO will have to develop strategic multidisciplinary partnerships with these organizations. Once the data is accessed, there is a secondary challenge that it will remain available.

Coverage and Suitability of data – A fourth challenge relating is ensuring that sourced data is suitable for producing official statistics and representative of the population. Given the CSO like other National Statistical Institutions set their sampling parameters carefully to represent the population, big data although exhaustive generally not representative [17].

Reputational Risk – A fifth challenge is the reputational risk that CSO could be exposed to by working with commercial entities driven primarily by profit in obtaining the big data that would be used as a basis of official public statistics [17]. Scandals resulting from underlying data will be considered as too costly for the reputation of a national statistical institution such as the CSO.

4.3 Mexico

4.3.1 Census Data and Big Data Analytics Use. The census data use in Mexico as a Big Data project has its origins in two mainstream initiatives. First, the Statistical Data and Metadata eXchange (SDMX) initiative was launched in 2001 by seven partnership organizations: Bank for International Settlements, European Central Bank, Eurostat, International Monetary Fund, Organization for Economic Cooperation and Development, the Statistics Divisions of United Nations and World Bank. They were the key project sponsors of SDMX to develop organizational capability for more efficient use of census data and metadata among international member countries. A model for SDMX was created to standardize census and survey data/metadata.

There are three main elements of SDMX: (1) an information model that includes statistical data (dimensions, attributes and measurements) and statistical metadata (structural metadata and reference metadata), (2) a group of content-oriented guidelines to define cross-domain concepts (cross-cutting), thematic statistical domains, and common metadata vocabulary, and (3) an IT architecture for data exchange that establishes standard formats for the exchange of data and metadata, architectures for the exchange of data. The model uses several tools that allow processing data, convert, relate, repositories, etc. that work under the SDMX standards. This model involves four levels: (1) multiple datasets and data formats, (2) extraction and transformation of data and metadata according to SDMX standards (open standard), (3) body of standardized data (open data), and (4) open source tools to use standardized data, such as: apps, graphics, maps, dashboards and data warehouse (DWH) integration.

In Mexico, the INEGI adopted the version 1.0 of the SDMX standards in April 2005 through an ISO technical specification for government (ISO / TS 17369: 2005). A version 2.0 was adopted in November 2005. Several consolidated versions were adopted in March 2006, February 2008, and January 2009. The model has improved the quality of structural census data and metadata conducted by the INEGI. Today, the INEGI promotes the use of consistent statistical concepts and code lists across domains and between international organizations and other National Statistics Institutes (NSIs) to share and facilitate data files automatically from their DWH. In March 2007, the sponsoring institutions signed a memorandum of understanding (MoU), which aims to establish the necessary arrangements for a lasting collaboration by the sponsors in all aspects of the initiative. During the 39th meeting of the UN Statistical Commission in New York (February

[3] http://census.ie/how-we-do-it/publishing-the-data/

[4] https://www.siliconrepublic.com/enterprise/ibm-and-dcu-create-masters-degree-in-big-data-analytics-and-smart-cities

2008), SDMX was recognized as an invaluable tool for INEGI and other National Statistics Institutes around the world.

Second, in 2011 the Federal Government signed an agreement to become a member of the Open Government Partnership (OGP) which has generated until now a set of 63 compromises that became eventually initiatives for open data initiatives in different areas of government, including census and national survey data. Among these initiatives, the INEGI has launched more than 380 projects using census or survey data. The introduction of each of these projects exceeds the purpose of this section, but here are some examples: the microdata, the bulk of urban geostatistics cartography, and the bulk of business directory (Directorio Estadístico Nacional de Unidades Económicas, DENUE).

Microdata provides two forms of access to microdata according to the Calendar of Dissemination of Statistical and Geographic Information: Direct through free download of microdata files through the website; and indirect, through the Microdata Laboratory assisted by government analysts and representatives of international organizations, researchers and postgraduate students. Another option is through the regular processing service from the INEGI information centers. The bulk of urban geostatistics cartography has reinforced the construction of a unique and national geostatistical framework to reference reliably statistical information of the censuses and surveys with geographic locations. The bulk of business directory (DENUE) offers identification, location, economic activity and size of active businesses in the national territory.

In sum, the SDMX and Open Government Partnership has transformed the way traditional census, survey data, and metadata are offered, updated, and used by adopting new models, frameworks, standards and techniques based on the ideas of Big Data, data analytics, and DWH.

4.3.2 Challenges to Big Data Analytics and Public Value. Despite the advance of developing big data analytics innovations on census and survey data, the use of the new sophisticated tools is poor and limited. Four main challenges are considered critical for the development of Big Data analytics in statistical and census: (1) *data policies*, (2) *large investments - low returns*, (3) *organizational transformation*, and (4) *digital divide*.

First, the census and survey data is today available on digital versions and travels across different organizational boundaries inside and outside of federal government. Several policy issues arose during the implementation of the SDMX and Open Government initiatives, which are still emerging, for example, with state and local governments, and private sectors. The challenge in the coming years is to create a robust legal framework and a set of policies to sustain Big Data initiatives in a context of organizational cross-boundary settings.

Second, the adoption of Big Data analytics at the INEGI demands large amounts of investment at the initial stages of the implementation with poor or unclear returns of this investment. This type of projects requires the acquisition of advanced technology and contracting of specialized staff that increased costs during the first years of implementation. Recent budget cuts and the lack of clarity about the use of Big Data analytics applications and tools by different users have created an environment where these initiatives are more a fashion or a luxury project than a real "priority" for governments.

Third, Big Data analytics require comprehensive organizational transformation based on leadership that understands the value of this emerging technology. The existing organizational structures and staff had neither the background, knowledge and expertise to support the initiative, nor the new talented staff with the knowledge of internal processes and expertise to operate within INEGI. So, the need for collaboration was critical and still is, when the original leader who supported the initiatives was changed by the new administration.

Finally, many expected greater value from the new technological and computational applications and the creation of a reliable framework of data and metadata standards and guidelines. However, most citizens and businesses have not been benefited from these advances largely due to the serious digital divide in the country: deficient capacity and technical expertise of data users to benefit from the open sources; high cost of adequate Internet services, ICT equipment and software, and still many regions, most of them rural areas, have no Internet access. These challenges explain why these tools do not create the expected greater public value.

4.4 U.S.A.

4.4.1 Census Data and Big Data Analytics Use. The U.S. Census Bureau has a long history of innovation; the punch card for the 1890 Census, the first civilian computer for the 1950 Census, and the first official survey sample was used by the Census Bureau to measure unemployment in 1937 [5].

The Census Bureau conducts a census every 10 years that primarily provides counts of people and counts of all habitable dwellings in the U.S. Every five years there also is a census of the economy, counting different businesses, and a census of governments. In addition, survey samples are conducted more frequently (i.e., monthly, quarterly, and annually) of people, housing, businesses, and government. Data collected cover a broad range of economic, social, and demographic data [5].

For the Census Bureau the largest undertaking is the 10-year census. For the next 2020 Census, the U.S. Census Bureau is adopting the latest big data technology such as a large Hadoop cluster, real-time stream data processing, and advanced mapping and visualization [30]. For the 2020 Census, there will be about 326 million Americans that the Census Bureau will collect and store data on. The Census has contracted with Hortonworks Data Platform (HDP) Hadoop distribution which will provide a main data lake, or the Census Data Lake, that will store most of the census data. It will be able to join data from other databases, including data from other agencies and store structured data such as names, addresses, and individual answers to demographic questions as well as unstructured data such as pictures taken from Google Maps.

For the 2020 Census will use extensive aerial and street-level imagery for data collection and analysis. The visualization will

enable the Census Bureau to streamline address identification, which will minimize the numbers of workers going door-to-door. The Census Bureau expects that the HDL will help to funnel the flow of data gathering from census fieldworkers to the Census Data Lake (see Figure 1). In order for the Census Bureau to provide the public with more data with greater refinement it will use SQL-based query engines run on Hadoop with the results shared with the data.census.gov website.

4.4.2 Challenges to Big Data Analytics and Public Value. There are three important challenges that the Census Bureau as identified for big data implementation [14]. First, and foremost, it is *privacy and confidentiality of data*. With the growing public concern over privacy issues of online information this is noted as paramount importance for the Census Bureau to address this issue. The accuracy and validity of the data depends upon addressing confidentiality and integrity of Census Bureau data. To address this challenge Congress in 2016 approved funding for the Census Bureau to fortify its platform that will provide federated infrastructure that permits data from sources to be curated, integrated, and provisioned in ways that improve data integrity and privacy.

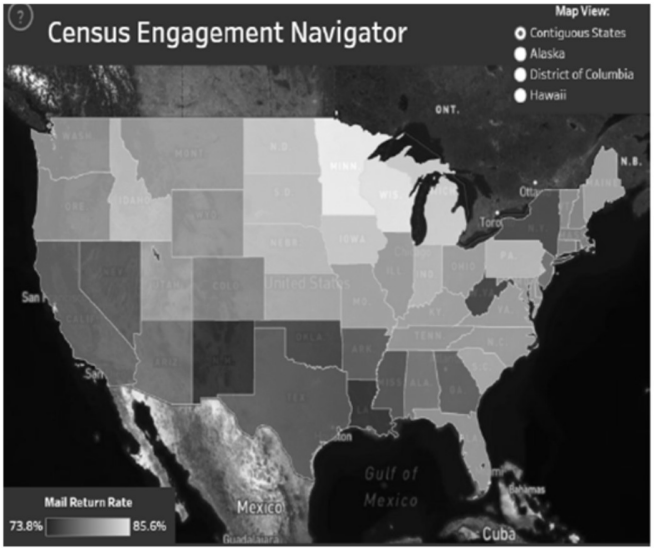

Figure 1: Decennial census collection map (Woodie, 2018).

Another challenge for the implementation of big data is *small-area estimation*. Most of the monthly statistics and economic indicators are only released at the national level. However, many analysts for big data usage need timely estimates at the sub-county level, data outside of the major cities. Poverty, crime, education, transportation, and economic development data are different for the cities than the suburbs and within the same county. According to the U.S. Census Bureau, small area estimates are increasingly important and this challenge needs to be addressed for greater impacts of big data implementation.

A third challenge is *the need to support data quality and validation work*. Crowdsourcing is essential to take advantage of emerging data sources and exploring the changing data landscape of moving from structured to unstructured data. This will involve greater interaction and collaboration across the different federal government agencies, with the public, business community, and researchers in the emerging areas of data science.

A final challenge is *the human capital aspect* of big data implementation. There needs to be a data science focus to help hire, train, and retain evidence builders which would require investment in internships, fellowships, and agency rotations. There needs to be a focus on training data scientists who are able to analyze and collect big data. This is a major challenge for the Census Bureau to be able to recruit the best and brightest, given the many opportunities in this emerging field.

5 CROSS CASE ANALYSIS RESULTS

We have performed cross case analysis of the four countries' big data and big data analytics use in the national census context. Table 2 shows the summary of the cross case analysis findings of these agencies. Overall, our findings show that Australia is the leader in big data implementation among the four countries. U.S.A. does not currently have big data analytics capability, but has been planning to roll out this capability for its next 2020 Census through the Big Data Working Group located within the Census Advisory Committees [28]. Of the four countries, Australia is the only country with a formal big data framework for the ABS. Mexico uses big data in the national census context. In comparison, while the U.S.A and Ireland do have big data use within the context of broader initiatives of government, they do not have a specific big data framework for the national census. As Table 2 indicates, in terms of a publicly proclaimed big data policy framework for the national census, Australia is slightly ahead of other three countries. In terms of big data analytics use for the national census, Australia is the only country in this study.

Table 2: Cross Case Analysis of Big Data Analytics Use in Census

	Australia	Ireland	Mexico	U.S.A.
Agency	Australian Bureau of Statistics	Central Statistics Office	The Census Inegi in Mexico	US Census Bureau
Census undertaken every years	5	5	10	10
Electronic census	2016	2021	2010	2010
Electronic census response	63.3%	No	N/A	55.4%
Current big data use/reuse	Yes	No	Yes	No
Current big data analytics use	Yes	Partially, geospatial data analytics	Partially	No but the plan for the 2020 census

In terms of the common challenges facing the four countries, both developing big data analytics capability and facilitating cross agency data access and data integration in the context of the national census have emerged as the key challenges commonly shared across the four countries. All four counties mentioned the issue of training human capital, especially big data and big data analytics capabilities. Although large government agencies have invested in enterprise data warehouse technologies, business intelligence, and data analytics tools/systems/platforms [6], exponential growth in geospatial big data, multi-media data, sensor data, and social media text data in government has overwhelmed extant big data analytics capability in other government contexts such as the Internal Revenue Service [16]. This challenge is also true in the national census context. Next, all four countries mentioned issues regarding data, in terms of quality, quantity, and integration. Interestingly, only Ireland mentioned the challenge of creating a legislative agenda, while Mexico acknowledged the need for a legal framework for greater public value creation.

In addition to these two challenges, the third challenge in data security, privacy and trust was shared only by Australia and the U.S. government agencies responsible for the national census. Finally, the fourth and the fifth challenges in operational efficiency and agency digital transformation in the big data era were only found in Australian Bureau of Statistics. Table 3 shows a summary of the common challenges. These cross-case analysis results show common challenges to creating greater public value through big data analytics capability in these Census bureaus.

Table 3: Challenges to Data Use/Reuse by the Public

	Australia	Ireland	Mexico	U.S.A.
Agency/ Challenges	Australian Bureau of Statistics	Central Statistics Office	The Census Inegi in Mexico	US Census Bureau
Big data analytics capability	☑	☑	☑	☑
Cross agency data access & integration	☑	☑	☑	☑
Data security, privacy & trust	☑			☑
Operational efficiency	☑			
Agency digital transformation	☑			

6 DISCUSSION & CONCLUSION

As we have shown in our review of the academic literature, despite the growing government practices in big data use, there is still the paucity of digital government research on links between government use of big data analytics and public value creation. This international cross case analysis paper is motivated to address this research-practice gap by raising the following two research questions:

RQ1: *What is the state of big data and big data analytics use in government?*

RQ2: *What are key challenges for government in using big data analytics and creating public value?*

As we have presented the within case analysis results in the previous section, the four census agencies are at varying stages in implementing digital transformation of the mission-critical national census data collection and analysis process. Against this backdrop, the within case analyses of the four countries and the cross-case analysis in the previous sections indicate that there are some important similar uses and challenges that have emerged in this multi-case study of an international scope.

As shown in Table 3, both the big data analytics capability challenge and the cross agency data access & data integration challenge were faced by all four census agencies analyzed in this study. On the other hand, the data security, privacy & trust challenge was mentioned only by the two census agencies: Australia and the U.S.A. These three challenges may inhibit national census agencies from using big data and big data analytics systemically and effectively in alignment with their mission-critical strategic goals. This in turn may inhibit them from creating public value through their big data and big data analytics investments. Finally, the awareness of two additional challenges in affecting greater digital transformation of the national census agency and achieving its greater operational efficiency gains was found only in Australia. Despite this lack of consensus, the lack of high-level digital transformation of national census agencies may preclude them from their assimilation and systemic use of big data and big data analytics in the process of national census data collection and analysis.

Given the paucity of empirical research on the links between big data/big data analytics use and public value creation in the context of national census, Figure 2 shows a proposed research model. The proposed working model aims to provide a better understanding of the key challenges across the four national census agencies that are at various stages in not only digital transformation but also the assimilation and use of open big data and big data analytics in the process of providing official census data, statistics, and insights in the national census context. We posit that the three common challenges, namely (1) data security, privacy and trust, (2) big data/big data analytics capability development, and (3) cross agency data access and data integration, will likely prevent the census agency's systemic use of big data/big data analytics in its mission-critical national census data collection and analysis services processes. Although Australia is the only country that mentioned the digital transformation challenge, the research model postulates that the maturity in digital transformation within the census agency will likely to facilitate the agency's systemic use of big data and big data analytics tools in increased agency efficiency gains in providing census services and products to other government agencies and the public. While this multi-case study did not focus on the public use of the census agency's census services and products, the research model for the future study suggests that the public demand in the form of big data and analytics use via the census agency website or portal will likely stimulate the agency's systemic use of big data and big data analytics to meet the agency's mission-critical services objectives. Finally, the model

postulates that public value creation through big data and big data analytics in the context of national census is predicated on the census agency's systemic use of big data and big data analytics which would involve the public use of the census big data and big data analytics tools that can generate new insights and mine new patterns among the hidden big data.

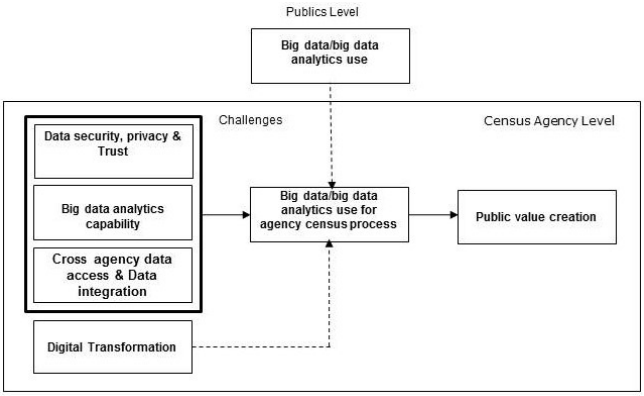

Figure 2: A research model for creating public value through census big data analytics use.

This proposed research model has important policy implications. A public policy framework for big data is required to either reduce or remove the existing challenges to maximize the national census agency's systemic and effective use of big data and big data analytics. The digital transformation of the agency needs to be advanced, since the maturity in digital transformation is thought to facilitate the agency's systemic and effective use. While these policy implications are primarily at the national census agency level, we also acknowledge the criticality of public policy that can promote the public's demand and use of census big data use/reuse. Increased use of big data/big data analytics at the agency level and at the public level is necessary to create greater public value. While not all countries found the challenge of increasing operational efficiency in the dynamically changing big data environments, the UN GWG has recognized this challenge. In this regard, the U.S. and Ireland in particular identified the need to develop human capacity in government to be able to use big data and big data analytics tools capably.

Importantly, we hold in this research model that the government's ability to increase big data use/reuse may be affected by four factors: (1) effective use of big data analytics; (2) to the extent which the agency's data is integrated with data held by other agencies for added value; (3) assurance of government data security, privacy & public trust in data quality; and (4) the agency's increased operational efficiency through a strategy of digital transformation such as Australia's "digital-first" strategy towards national census (eCensus by default) to do away without paper-based census.

Similarly, we argue that for the agency to increase operational efficiency to create greater public value the agency needs to be aware of three enabling or inhibiting factors: (1) effective and systemic use of big data analytics, (2) data integration, and (3) digital transformation. In other words, if these factors are not effectively governed and managed in the big data census environment, these will pose serious challenges to the agency in creating greater public value.

Based on the four within case analysis and the cross-case analysis, we hold that the research model offers utility in guiding policy making and future research in this new and emerging field of open big data and big data analytics use in the national census context. However, we acknowledge some research limitations, including the use of secondary source data for this multi-case study and the selection of the four countries which are advanced in the use of ICT in advancing e-government. Our findings and insights gained may or may not be directly generalizable to other countries. Our future research directions include a survey of other countries both in the developed and developing countries.

REFERENCES

[1] Australian Bureau of statistics, "Annual report 2016 - 17 2017a," 2017.

[2] Australian Bureau of statistics, "Annual report 2016-17 2017b," 2017.

[3] A.T. Chatfield and C.G. Reddick, "Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services.," 2017.

[4] A.T. Chatfield, C.G. Reddick, and W.H.A. Al-Zubaidi, "Capability challenges in transforming government through open and big data: Tales of two cities.," in *Exploring the Information Frontier, ICIS 2015*, 2015.

[5] C. Capps and T. Wright, "Toward a vision: Official statistics and big data. Amstatnews," 2013.

[6] C.L.P. Chen and C-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information Sciences*, vol. 275, pp. 314-347.

[7] CSO, "Central Statistics Office Statement of Strategy 2016-2019," 2016.

[8] A. Darrel, M. Hume, T. Hardie, and J. Soar, "The benefits of big data analytics in the healthcare sector: What are they and who benefits? Big Data Analytics in Bioinformatics and Healthcare," 2014.

[9] Department of Finance, "Australian Public Service Better Practice Guide for Big Data," 2015.

[10] Eurostat, "Scheveningen Memorandum Big Data and Official Statistics," no. 99, pp. 12–29, 2013.

[11] P. B. Goes, "Editor's comments big data and IS research," *MIS Q.*, vol. 38, no. 3, 2014.

[12] W. Groves, J. Collins, M. Gini, and W. Ketter, "Agent-assisted supply chain management: Analysis and lessons learned," *Decision Support Systems*, vol. 57, pp. 274-284, 2013.

[13] M. Hu. "Biometric ID cybersurveillance," *Indiana Law Journal* vol. 88, no. 4, pp. 1475-1558, 2013.

[14] R. S. Jarmin and A. B. O'Hara, "Big data and the transformation of public policy," *J. Pol. Anal. and.*

Manag., vol. 35, pp. 715–721, 2016.

[15] D. Killalea, "Census hacked: Website for Australian census 2016 attacked, ABS says," Aug-2016.

[16] G-H. Kim, S. Trimi, and J-H. Chung, "Big-data applications in government sector," *Communications of the ACM*, vol. 57, no. 3, pp. 78-85.

[17] R. Kitchin, "Big data and official statistics: Opportunities, challenges and risks (a)," 2015.

[18] R. Kitchin, "What does big data mean for official statistics (b)," 2015.

[19] D. Laney, "3D data management: Controlling data volume, velocity, and variety," *META Gr. Res. Note*, 2001.

[20] A. McAfee and E. Brynjolfsson, *Big data: The Management Revolution," Harvard Business Review (90:10)*. 2012.

[21] National Statistical Service., "Statistical Spatial Framework for Australia," 2018.

[22] NISTRA, "The U.K. Office for National Statistics. 2017. What is a census?," 2017.

[23] J. Roski, G. W. Bo-Linn, and T. A. Andrews, "Creating value in healthcare through big data: Opportunities and policy implications," *Health Aff.*, vol. 33, no. 7, pp. 115–1122, 2014.

[24] M. Schermann, H. Hemsen, C. Buchmüller, T. Bitter, H. Krcmar, V. Markl, and T. Hoeren, "Big Data," *Bus. Inf. Syst. Eng.*, vol. 6, no. 5, pp. 261–266, 2014.

[25] United Nations, "Statistical Commission Report on the Forty-Sixth Session," 2015.

[26] United Nations, "Statistical Commission Report on the Forty-Sixth Session," 2017.

[27] United Nations, "Using big data for the sustainable development goals," UN Big Data GWG Task Teams, 2014, https://unstats.un.org/bigdata/taskteams/sdgs/.

[28] U.S. Census Bureau. Census Advisory Committees (CAC). Big Data Working Group. https://www.census.gov/about/cac/sac/wg-big-data.html.

[29] M. White, "Digital workplaces: Vision and reality," *Bus. Inf. Rev.*, vol. 4, pp. 205–214, 2012.

[30] A. Woodie, "The big data tech inside the 2020 Census," 2018.

[31] D. Zhang, "Granularities and inconsistencies in big data analysis," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 23, no. 6, pp. 887–893, 2013.