Development Informatics

Working Paper Series

The Development Informatics working paper series discusses the broad issues surrounding digital data, information, knowledge, information systems, and information and communication technologies in the process of socio-economic development

Paper No. 62

Measuring the Barriers to Big Data for Development: *Design-Reality Gap Analysis in Colombia's Public Sector*

LILIANA FERNÁNDEZ GÓMEZ & RICHARD HEEKS

2016

ISBN: 978-1-905469-56-7

Published by: ***Centre for Development Informatics***
Global Development Institute, SEED
University of Manchester, Arthur Lewis Building, Manchester, M13 9PL, UK
Email: cdi@manchester.ac.uk Web: http://www.cdi.manchester.ac.uk

View/Download from:
http://www.gdi.manchester.ac.uk/research/publications/working-papers/di/

Educators' Guide from:
http://www.gdi.manchester.ac.uk/research/publications/working-papers/di/educators-guide/

Table of Contents

Measuring the Barriers to Big Data for Development: *Design-Reality Gap Analysis in Colombia's Public Sector*

Liliana Fernández Gómez & Richard Heeks
Centre for Development Informatics, University of Manchester, UK

2016

Abstract

While big data has the potential to make a significant contribution to international development, that potential is currently constrained by a number of barriers. Systematic analysis of those barriers is rare, so this paper applies the design-reality gap model to identify and evaluate barriers to effective use of big data in one context: the Colombian public sector. The model provides a structured framework that exposes a broad set of barriers, and also helps highlight priority areas for action to accelerate the application of big data. The design-reality gap model can also be seen to provide the basis for related analyses such as readiness for big data, and risk identification for big data initiatives in developing countries.

A. Introduction

A new "digital development" paradigm is emerging as – due to geographic, maturational and experiential changes associated with mass diffusion of digital technologies – information and communication technologies change from being a tool to achieve specific development goals to being the platform that mediates development (Heeks 2014). One part of this emergent paradigm is data-intensive development: the growing salience and presence of digital data in development decision-making and processes.

Though typically identified with the "big data for development" label, data-intensive development actually consists of four "V" trends (adapted from Dumbill 2012):

- Volume: the greatly-increased amount of data that is available.
- Velocity: the greatly-increased speed with which data is available, reducing the lag between event and data about that event.
- Variety: the greatly-increased number of forms and types of data.
- Visibility: the greatly-increased availability and accessibility of data.

These trends have led to great hopes for big data, with the potential seen to significantly improve decision-making, actions and outcomes across a whole range of development sectors (Global Pulse 2012, Kshetri 2014, Hilbert 2016). Yet, alongside this promise, there is also widespread recognition of the barriers to application of big data in developing countries. Because of the formative nature of the field, discussion of the barriers to big data for development tends to fall into one of three camps:

- General expressions of concern about the presence of barriers (e.g. HLP 2013, Jerven 2013).
- Lists of a number of issues (e.g. CGD 2014, Kshetri 2014, Piotrowski 2014, Spratt & Baker 2015).
- Identification of issues via a few broad-ranging categories (e.g. Global Pulse 2012, Hilbert 2016).

Three things have therefore been lacking, in general, in analysing the barriers to big data for development. First is systematic analysis, with only one source (Hilbert 2016) providing any sort of framework of barriers. Second is breadth of analysis, with almost all sources focusing on just three barriers: the limitations of the data, skills, and technological infrastructure. Third is context-specificity of analysis, with sources typically undertaking broad-brush descriptions that try to cover all developing countries.

B. Focus and Framework

Seeking to address these three limitations, the current paper utilises the design-reality gap model as the basis for analysing barriers to use of big data for development. The design-reality gap model (e.g. Heeks 2002, Heeks 2006) understands barriers to change as a gap between the assumptions or requirements inherent within the design for that change, as compared to actual on-the-ground realities. The gap is typically assessed as a 0-10 scale

(e.g. 0=no difference between design and reality; 5=some difference; 10=complete and radical difference) along seven "ITPOSMO" dimensions (see Figure 1):

- **I**nformation: includes both information and data, as the precursor to information.
- **T**echnology: mainly focuses on the information and communication technology (ICT) that handles data.
- **P**rocesses: the activities undertaken in generating, capturing, analysing, presenting and using data.
- **O**bjectives and values: the 'objectives' component covers issues of self-interest and politics, and informal and formal strategies; the 'values' component covers culture: what stakeholders feel are the right and wrong ways to do things.
- **S**kills and knowledge: covers both qualitative and quantitative aspects of human competencies for undertaking data-related processes.
- **M**anagement systems and structures: the broader management systems required to organise within and between data-related organisations and networks, plus the way in which those systems are structured, both formally and informally.
- **O**ther resources: time and money.

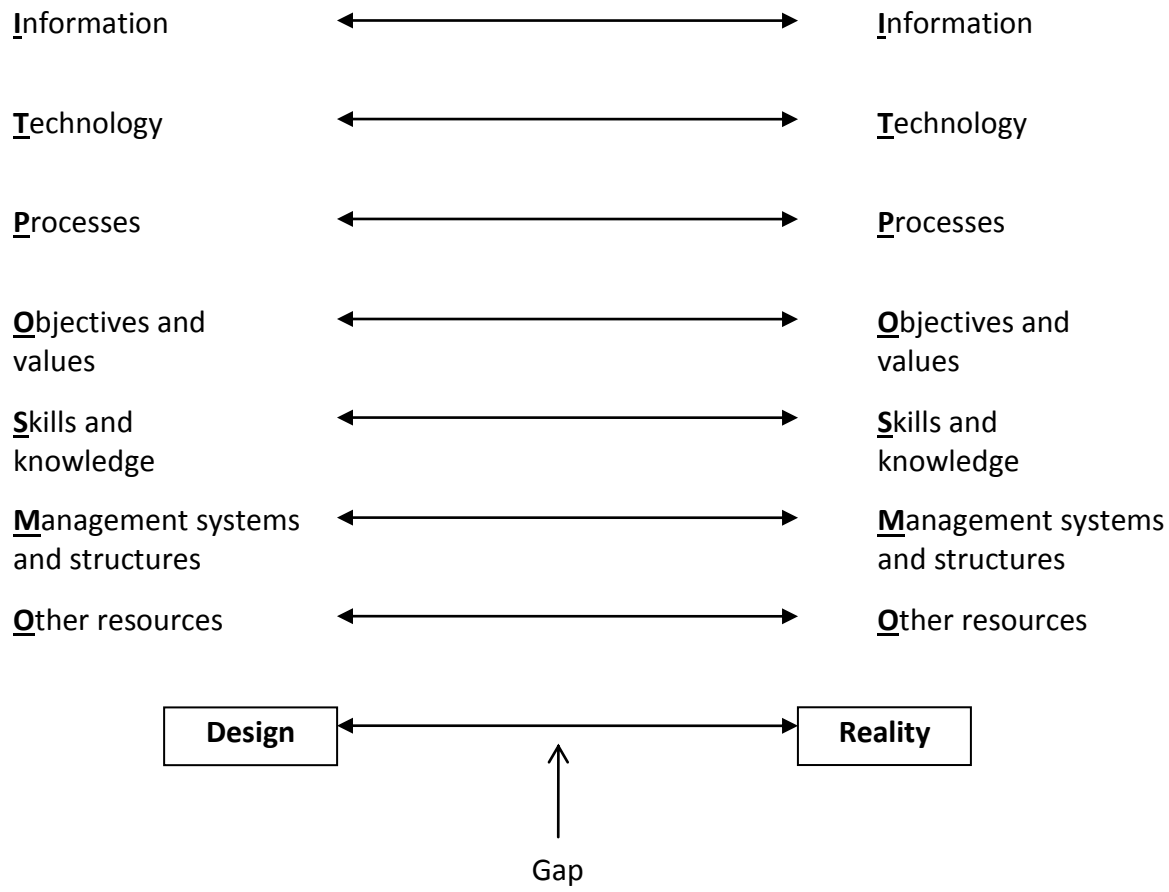

Figure 1: Design-Reality Gap Model

The design-reality gap model can be applied in various different ways: cross-sectionally to ascertain the gap between design and reality at a given time, or longitudinally to assess how implementation of a design materialises in reality; at various points in the lifecycle of

change: pre-, durante- or post-hoc; and for general assessment of a type of change or for specific assessment of a particular change initiative.

Here, we apply it cross-sectionally, largely pre-hoc, and at a general level. In this form, it acts as a tool for general assessment of readiness for big data. Given the point raised about context-specificity, we apply it not to developing countries overall, but to the public sector in Colombia specifically. Colombia was selected because the eGovernment Directorate within the Ministry of Information and Communication Technologies (MINTIC) identified expanded use of big data in the public sector as part of its forward strategy from 2014.

Our research question was therefore: What are the barriers to use of big data in the Colombian public sector? In answering this, we also hoped to provide some guidance on actions to help overcome the barriers.

Evidence to address the research question was gathered from three sources: participant-observation in two July 2014 IT summits held in Bogota; semi-structured interviews with three government officials with specific responsibilities for public sector data: two in MINTIC and one in Bogota City government; and a set of secondary literature such as reports and news items produced by MINTIC. Given the time and scope limitations of the research, the analysis that follows should be regarded as illustrative rather than definitive.

C. Analysis of Barriers to Big Data for Development in Colombia

The analysis below draws on the evidence sources to assess the size of gap on each of the ITPOSMO dimensions in turn.

Information. An effective big-data-for-development (BD4D) *design* requires that the four Vs are in place: with significant amounts, speed, forms, availability and accessibility of data that meets typical data quality standards (Beyer & Laney 2012, Buhl & Heidemann 2013). But in *reality*, the volume and visibility of data being produced in Colombia is relatively limited: for example, from more than 1,300 public sector entities in the country, only 620 datasets are available (CIO Summit 2014). Analysis of these datasets (via MINTIC 2014b) exposed other characteristics. In relation to velocity, most datasets were not dynamic (as one example, there are hydrometeorology datasets but they do not provide real-time data from monitoring stations). Variety of data was mixed: there were many different categories of formats of public datasets but there was a high concentration on basic administrative data such as details of public sector contacts or procedures. Likewise, data quality was mixed with issues such as accuracy, representation and usability still needing to be worked on (BCC 2014). The gap between design and reality on this dimension was judged to be: 7.

Technology. BD4D *design* requires a strong ICT infrastructure that supports all stages of the data chain: data generation, capture, analysis, storage and presentation including visualisation and dissemination (UN 2014a). That ICT infrastructure should be widely-distributed and interoperable (Helbing & Balietti 2011, Howard 2012). In *reality*, data

sources in the form of feature mobile phones were quite widespread – the 2014 penetration rate was 107 per 100 inhabitants – but multi-data-chain-function smartphone penetration was much lower (CA 2014). Internet of things and related data source penetration rates were also very low, though applications were emerging around use of video surveillance and environmental sensors in the capital city (BCC 2014, GSMA 2014). In relation to general use of data-chain ICTs within the public sector, private sector and by citizens, Colombia finds itself towards the top of middle-income country rankings, and somewhere close to the bottom of the top third of countries overall (e.g. UN 2014b, WEF 2014). There is, however, still quite a significant digital divide within the country leading to inequalities and exclusions along, for example, geographic, educational and age dimensions (though not gender) (Velasquez 2013). Interoperability was being improved but, for example, some public sector datasets were still being gathered by exchange of physical media so even basic interconnection was lacking in some cases (BCC 2014, CIO Summit 2014). The gap between design and reality on this dimension was judged to be: 6.

Processes. BD4D *design* requires all of the information value chain processes to be enacted on a wide scale. The information value chain (see Figure 2) includes the steps of the data chain but also the extraction of value from big data by turning it into information that feeds development decisions then actions then results.

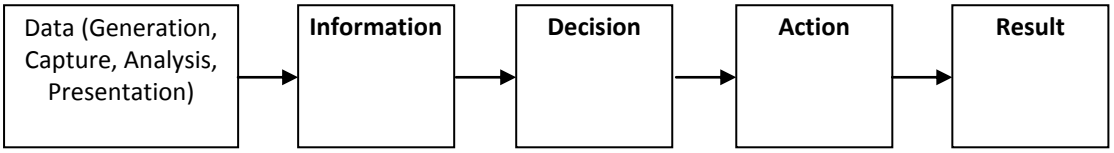

Figure 2: Basic Information Value Chain

In *reality*, the extent of foundational data chain processes was constrained by the relatively-limited availability of big data for the public sector, as noted above. It was also constrained by relatively-limited sharing of large datasets within and outside the public sector in Colombia (BCC 2014). But there also appeared to be quite a disjuncture within the chain, with presence of datasets being much stronger than their utilisation (Torres 2014). The gap between design and reality on this dimension was judged to be: 7.

Objectives and Values. Effective BD4D *design* requires a whole set of informal and formal institutions to be in place, of which two will be highlighted here. First, big data needs a foundation of laws and regulations covering issues such as data security and privacy (MGI 2011, UN 2014a). Second, big data needs to be embedded within a data-intensive culture that trusts and values the role of data (Laney et al 2013). In *reality*, the foundations for some data policy issues are in place in Colombia – such as a data security commission, data security regulations and laws on rights of access to data (CIO Summit 2014, Rojas et al 2014) – but legislation and implementation of the full range of necessary policies was still a work in progress, with no recognition of big data exploitation in Colombia's main ICT strategy, "Vive Digital 2014-2018" (MINTIC 2014c). Likewise, these policies did not explicitly address big data-specific issues such as personal information discovery from anonymised datasets, data rights relating to location-specific and surveillance data, use of private big datasets, and commercialisation of public sector big data (eGD1 2014). There appears to be some

lack of a culture of data-sharing within the public sector, partly due to fears of then being held to account on the basis of what the data shows (BCC 2014). More generally, the lack of a 'big data culture' in Colombia is cited as a key issue (Torres 2014), with many managers stating they believe intuition and experience rather than data to be the most important basis for good decision making (Gomes 2014). Alongside this in the general population, a majority of individuals surveyed reported they did not trust the public (and private) sector to only use big data for authorised, anonymised purposes (UdlA 2014). There are individual big data projects in Colombia (e.g. Clark 2014, Krchnak 2015, Velasco 2015) but these have relied on drivers from international organisations or exceptional public sector leaders, rather than representing an institutionalised system. The gap between design and reality on this dimension was judged to be: 7.

Skills and Knowledge. BD4D *design* requires firstly awareness of knowledge of big data, and then a set of capabilities that run right across the information value chain (see Figure 2). In particular, one can identify three key capability sets that are needed: data science capabilities (more-technical skills to support big data analytics); data application capabilities (more-"business"-oriented skills to make use of big data for better decisions and actions); and data hybrid capabilities (e.g. analytic consultants who can bridge the technical-business divide) (MGI 2011, Brown et al 2014). In *reality*, although some big data workshops were starting to be held, there was a lack of both awareness and knowledge of big data in Colombia (Torres 2014). Evidence on foundational capabilities of relevance to big data – e.g. ICT skills and maths skills – showed a mixed picture with thousands trained as part of the government's "Talento Digital" programme but fairly-poor enrolments and performance in relation to maths (MINTIC 2014a, WEF 2014). Looking specifically at capabilities in the public sector, foundational technical capabilities were present but specific big data capabilities were largely absent bar one or two pilot projects (MoF 2013, BCC 2014). Analysis of the capability profiles of public sector ICT leaders showed a focus on infrastructure rather than value generation (MINTIC 2012b). No data science or data analytics postgraduate programmes were identifiable within Colombia (MoE 2014). The gap between design and reality on this dimension was judged to be: 7.

Management Systems and Structures. Usually associated with organisation-level systems, BD4D *design* in this case relates at least as much to public sector-wide arrangements. It involves a requirement for big data strategy and governance across the public sector, as well as within specific public agencies (Laney et al 2013). These need to focus on issues such as data standards, sharing/interoperability, security and privacy. In *reality*, open data is identified as a priority for government as part of its 2012 eGovernment Strategy (MINTIC 2012a), and a Framework for IT Architecture in the public sector launched in 2014 and addressing issues of data standards, good practices and interoperability was about to be rolled out (Bohorquez 2014, CIO Summit 2014). However, both of these were general rather than having big data-specific considerations built into them, and it was recognised that the focus for a number of policies is on compliance rather than value generation; a focus that could be somewhat orthogonal to the needs of big data (Guzman 2014). There were also recognised challenges for public sector CIOs in fulfilling their existing roles, even before adding big data issues to their responsibilities – something which has not yet happened for most of them (eGD2 2014). The gap between design and reality on this dimension was judged to be: 7.

Other Resources. Creating a big-data-for-development ecosystem brings with it a *design* requirement for heavy investments of money and time (Gale 2013). At the time of analysis, the *reality* in Colombia was that investments had only been set aside for a relatively few specific big data projects (eGD2 2014). However, investment plans were in place for more substantial activity, and budgets had been set aside for this. The gap between design and reality on this dimension was judged to be: 5.

D. Recommendations and Conclusions

The potential for big data to contribute to development has undoubtedly been over-hyped: a universal issue for all digital and development innovations (e.g. Algoso 2014, Gartner 2015). Nonetheless, when the hype is stripped away, big data is still left with a demonstrably-valuable – even if not central – role to play in the future of development. But that role will only be fulfilled if the barriers to big data can be overcome.

Reviewing the design-reality gap evidence – summarised in Figure 3 – there remains a broad swath of serious barriers to big data in the Colombian public sector. Judgement on size of the gaps/barriers is subjective; and would be improved through an inter-subjective approach involving some of the key stakeholders. But it is a reasonable conclusion to say that all seven ITPOSMO dimensions represent significant challenges. A simple summary would be that Colombia's public sector has a number of the foundations or precursors for big data in place, but very few of the specific components that make up a big data ecosystem.

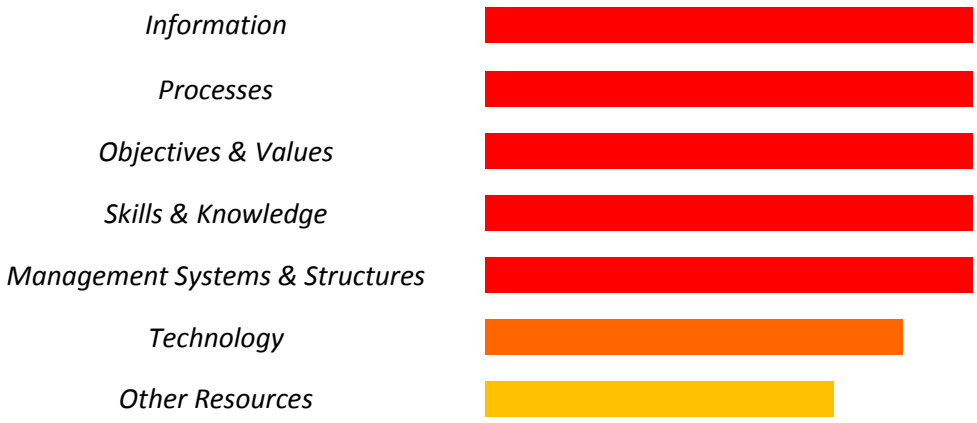

Figure 3: Overview of Big Data Design-Reality Gaps

This was a cross-sectional analysis at one point in time, and design-reality gaps change continuously. Reality in particular is changing to close the gap to design, thus slowly reducing the barriers to effective use of big data in Colombia's public sector. For example, the government supported a number of big data events and workshops during 2015. And, in late 2015, it began work on a Centre of Excellence in Big Data and Data Analytics, and held a World Congress on big data.

Nonetheless, we can analyse the evidence in order to pinpoint specific gap-closing measures that will help propel big-data-for-development in Colombia. Examples include:

- Information: make greater use of existing big datasets, such as those available via social media and mobile call detail records and "dark data" already within the public sector.
- Technology: make incremental investments in systems to digitally capture data from existing sources.
- Processes: prioritise enactment of value-generation rather than data-generation processes.
- Objectives and values: enact or amend legislation to deal specifically with big data issues such as data privacy, intellectual property rights, and commercialisation.
- Skills and knowledge: incorporate big data into curricula for public administration training and degree programmes.
- Management systems and structures: appoint big data champions and taskforces in each main public agency, and incorporate champions into the proposed Centre of Excellence to enable cross-agency sharing of good practice.
- Other resources: develop public-private partnership funding for big data initiatives.

Beyond the specifics of the particular case, this research provides a proof-of-concept for use of the design-reality gap model in assessing barriers to big data for development. Rephrasing the focus for the exercise, the model could equally be used to measure readiness for big data; BD4D critical success and failure factors; and risks for specific big data initiatives.

As noted, resource limitations meant this could only be an illustrative rather than definitive account. Greater depth of use of the model would involve, for example, a greater number and range of stakeholders including group rating of design-reality gaps and group discussion of gap-closure measures. The ITPOSMO dimensions are generic and have been handled in a broad-brush manner: an alternative is to revise them – either whole or part dimensions – to make them more attuned to the specifics of big data.

We hope other researchers and consultants will make use of the design-reality gap model for future assessments of big-data-for-development readiness, barriers and risks.

References

Algoso, D. (2014) Hype cycle for development ideas: 2014 edition, *Find What Works*, 2 Jan

BCC (2014) *Personal Interview with Bogota City Council-related Data Manager*, Bogota, 12 Aug

Beyer, M.A. & Laney, D. (2012) *The Importance of "Big Data"*, Gartner, Stamford, CT

Bohorquez, E. (2014) Entidades del Estado estarán más sincronizadas [Public entities will be more synchronised], *El Espectador*, 17 Jun

Brown, B., Court, D. & McGuire, T. (2014) Views from the front lines of the data-analytics revolution, *McKinsey Quarterly*, March

Buhl, H.U. & Heidemann, J. (2013) Big data, *Business & Information Systems Engineering*, 55(2), 65-69

CA (2014) *Telefonía Móvil en la Comunidad Andina [Mobile Telephony in the Andean Community of Nations]*, Comunidad Andina, Lima, Peru

CGD (2014) *Delivering on the Data Revolution in Sub-Saharan Africa*, Center for Global Development, Washington, DC

CIO Summit (2014) *Third National CIO Summit 2014*, CIO Summit, MINTIC, Bogota, 29-30 Jul

Clark, H. (2014) How big data is helping farmers save millions, *gizmag*, 27 Oct

Dumbill, E. (2012) What is big data?, *O'Reilly*, 11 Jan

eGD1 (2014) *Personal Interview with eGovernment Directorate Senior Policy Manager*, Bogota, 8 Jul

eGD2 (2014) *Personal Interview with eGovernment Directorate Senior Practice Manager*, Bogota, 12 Aug

Gale, S.F. (2013) *Data Doesn't Lie: Special Report*, Project Management Institute, Newton Square, PA

Gartner (2015) Gartner's 2015 hype cycle for emerging technologies identifies the computing innovations that organizations should monitor, *Gartner Newsroom*, 18 Aug

Global Pulse (2012) *Big Data for Development: Challenges and Opportunities*, UN Global Pulse, New York, NY

Gomes, L. (2014) Snapshot of big data trends in Latin America, *The Bridge*, 44(4), 46-49

GSMA (2014) *From Concept to Delivery: The M2M Market Today*, GSMA, London

Guzman, F. (2014) National eGovernment Index 2013 results, presentation at *CIO Summit 2014*, Bogota, 29-30 Jul

Heeks, R. (2002) Information systems and developing countries: failure, success and local improvisations, *The Information Society*, 18(2), 101-112

Heeks, R. (2006) *Implementing and Managing eGovernment: An International Text*, Sage Publications, London

Heeks, R. (2014) *From ICT4D to Digital Development*, background paper for UN Commission on Science and Technology for Development, Geneva

Helbing, D. & Balietti, S. (2011) From social data mining to forecasting socio-economic crises, *The European Physical Journal-Special Topics*, 195(1), 3-68

Hilbert, M. (2016) Big data for development, *Development Policy Review*, 34(1), 135-174

HLP (2013) *A New Global Partnership*, High-Level Panel of Eminent Persons, United Nations, New York, NY

Howard, R. (2012) *Big Data Forces Big Change in the Age of Smart Government*, Gartner, Stamford, CT

Jerven, M. (2013) What kind of 'data revolution' do we need for post-2015?, *Post2015*, 5 Nov

Krchnak, K. (2015) Big data improving ecosystems, from Chesapeake Bay to Colombia, *Huffington Post*, 9 Sep

Kshetri, N. (2014) The emerging role of big data in key development issues, *Big Data & Society*, 1(2), 1-20

Laney, D., Buytendijk, F. & Linden, A. (2013) *Predicts 2014: Innovating With Information Will Demand New Data, Organizations and Ideas*, Gartner, Stamford, CT

MGI (2011) *Big Data*, McKinsey Global Institute, New York, NY

MINTIC (2012a) *Estrategia Gobierno en Linea 2012-2015 [eGovernment Strategy 2012-2015]*, MINTIC, Bogota

MINTIC (2012b) *Lineamientos Para La Implementacion De Datos Abiertos En Colombia [Guidelines for Implementation of Open Data in Colombia]*, MINTIC, Bogota

MINTIC (2014a) *Talento Digital [Digital Talent]*, MINTIC, Bogota

MINTIC (2014b) *Catalogo de Datos Abiertos [Catalogue of Open Data]*, MINTIC, Bogota

MINTIC (2014c) *Plan Vive Digital 2014-2018*, MINTIC, Bogota

MoE (2014) *Sistema Nacional de Información de la Educación Superior [National Information System on Higher Education]*, Ministry of Education, Bogota

MoF (2013) ISAAC indicators, *Notas Fiscales*, 1(22)

Piotrowski, J. (2014) Big obstacles ahead for big data for development, *SciDev.Net*, 15 Apr

Rojas, L., Bermudez, G. & Lovelle, J. (2014) Open data and big data: a perspective from Colombia, in: *Knowledge Management in Organizations*, L. Uden, D. Fuenzaliza, I.-H. Ting & D. Liberona (eds), Springer, Cham, Switzerland, 35-41

Spratt, S. & Baker, J. (2015) *Big Data and International Development: Impacts, Scenarios and Policy Options*, Evidence Report no. 163, IDS, University of Sussex, Falmer, UK

Torres, A.G. (2014) *Big Data y Analitica en Colombia [Big Data and Analytics in Colombia]*, TechTarget, Newton, MA

UdlA (2014) *Resultados de la Encuesta Privacidad y Datos Personales (Colombia) [Results of the Survey on Privacy and Personal Data (Colombia)]*, Universidad de los Andes, Bogota

UN (2014a) *Big Data and Modernization of Statistical Systems*, United Nations, New York, NY

UN (2014b) *E-Government Survey 2014*, United Nations, New York, NY

Velasco, R.G. (2015) Big data are reducing homicides in cities across the Americas, *Scientific American*, 1 Oct

Velasquez, A. (2013) Digital divide in Colombia, *International Journal of Communication*, 7

WEF (2014) *The Global Information Technology Report 2014: Rewards and Risks of Big Data*, World Economic Forum, Geneva