Using big data for evaluating development outcomes: a systematic map

Francis Rathinam, Sayak Khatua, Zeba Siddiqui, Manya Malik, Pallavi Duggal, Samantha Watson, Xavier Vollenweider
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL methods working papers

The CEDIL methods working paper series offers innovative research methods to develop impact evaluation and evidence synthesis work in low- and middle-income countries.

About this working paper

This paper, ‘Using big data for evaluating development outcomes: a systematic map’, discusses the methodological, ethical and practical constraints relating to the use of big data for measuring and evaluating development outcomes.

Please direct any comments or queries to the corresponding author, Sayak Khatua, at skhatua@3ieimpact.org


Cover design: PhilDoesDesign

Copyright: © 2020 This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

Using big data for evaluating development outcomes: a systematic map

Authors
Francis Rathinam
Athena Infonomics

Sayak Khatua
3ie

Zeba Siddiqui
3ie

Manya Malik
3ie

Pallavi Duggal
3ie

Samantha Watson
Flowminder

Xavier Vollenweider
Flowminder
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

Contents

List of abbreviations vi
Executive summary 1

1. Introduction 3
   1.1 The issue: data challenges 3
   1.2 Objectives and the research questions 4
   1.3 Scope of work 5
   1.4 Overview 6

2. Methodology: definitions, inclusion and exclusion 7
   2.1 Methodological approach 7
   2.2 Criteria for including the studies 7
   2.3 Sources of big data 8
   2.4 Outcomes 10
   2.5 Evaluation of potential methodological issues 12
   2.6 Evaluating reporting on privacy and ethical considerations 13
   2.7 Exclusion criteria 14

3. Results 15
   3.1 Search, screening and coding 15
      3.1.1 Searching and screening 15
      3.1.2 Coding of included studies 16
      3.1.3 Mapping of the studies 16
   3.2 Characteristics and trends of the evidence base 17
      3.2.1 Distribution of studies across data sources 20
      3.2.2 Distribution of studies across development themes 23
      3.2.3 Units of observation 24
      3.2.4 Studies using mixed methods 25
      3.2.5 Studies with a rural or urban focus 25
      3.2.6 Studies by fragile context 26
   3.3 Appraising potential methodological biases, risks and limitations 26
      3.3.1 Reporting on methodological challenges and data transparency 26
      3.3.2 Reporting on privacy and ethical challenges 28

4. Key findings, discussion and lessons 29
   4.1 Evidence base and key gaps 29
   4.2 Analytical, ethical and practical challenges 31

5. Using big data in IEs: potentials and challenges 34
   5.1 Using satellite data in IEs 34
   5.2 Using CRD analytics to inform disaster management 36

6. Limitations of this study and potential next steps 39
   6.1 Next steps 40

7. Conclusions 41
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

8. References 42
Appendix 1 The OECD definition of fragile states 46
Appendix 2 Details of sub-maps 49
Appendix 3 Search strategy and the databases searched 51
Appendix 4 Data extraction tool 56
Appendix 5 ML and manual screening 61
Appendix 6 SR appraisal tool and summary 62
Appendix 7 Additional tables and figures 65
Appendix 8 Systematic map of big data sources and outcomes 69
Appendix 9 List of included studies 70

List of figures, tables and boxes

Figure 1: The PRISMA flowchart 16
Figure 2: Number of studies published per year 17
Figure 3: Distribution of studies over regions 18
Figure 4: Geographical distribution of studies 19
Figure 5: Number of studies by income classification 20
Figure 6: Number of studies per different type of big data 21
Figure 7: Number of studies against development outcomes 23
Figure 8 Units of observation 25
Figure 9: Distribution of population sub-groups 25
Figure 10: Number of studies in fragile contexts 26
Figure 11: Number of IEs and MS against data quality and transparency 27
Figure 12: Number of studies reporting on ethics issues 28
Figure 13: Steps in CRD processing for displacement and return/resettlement/recovery pattern analysis 38
Figure 14: Order of relevance of studies at the title and abstract screening stage 61
Figure 15: Systematic map of big data sources and outcomes 69

Table 1: Selection criteria for studies 7
Table 2: Sources of big data 9
Table 3: Outcome categories and definitions 11
Table 4: Number of IEs and measurement studies across data sources 22
Table 5: Distribution of studies across development themes 24
Table 6: Fragile contexts provided in the OECD fragility framework 2018 (decreasing order of severity) 47
Table 7: Outcome categories and definitions 49
Table 8: Data extraction tool 56
Table 9: Top 20 countries with maximum number of studies 65
Table 10: Distribution of studies across the region 66
Table 11: Studies using multiple sources of big data 66
Table 12: Big data source wise studies in fragile context 67
Table 13: How studies have dealt with data quality and transparency 68

cedilprogramme.org5
Box 1 Transparency in data analysis, use and sharing
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRD</td>
<td>Call Record Details</td>
</tr>
<tr>
<td>DFID</td>
<td>Department for International Development</td>
</tr>
<tr>
<td>DM</td>
<td>Data Mining</td>
</tr>
<tr>
<td>DMSP-OLS</td>
<td>Defense Meteorological Satellite Program-Operational Linescan System</td>
</tr>
<tr>
<td>EC</td>
<td>Environmental Clearance</td>
</tr>
<tr>
<td>EPPI</td>
<td>Evidence for Policy and Practice Information</td>
</tr>
<tr>
<td>EVI</td>
<td>Enhanced Vegetation Index</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>IDP</td>
<td>Internally Displaced Person</td>
</tr>
<tr>
<td>IE</td>
<td>Impact Evaluation</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>L&amp;MICs</td>
<td>Low and Middle Income Countries</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>MNO</td>
<td>Mobile Network Operator</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised Controlled Trial</td>
</tr>
<tr>
<td>SDG</td>
<td>Sustainable Development Goal</td>
</tr>
<tr>
<td>SEDAC</td>
<td>Socioeconomic Data and Applications Center</td>
</tr>
<tr>
<td>SR</td>
<td>Systematic Review</td>
</tr>
<tr>
<td>SRS</td>
<td>Satellite Remote Sensing</td>
</tr>
<tr>
<td>SSCI</td>
<td>Social Sciences Citation Index</td>
</tr>
</tbody>
</table>
Executive summary

Data challenges in international development are stark, especially in developing country contexts. Traditional data collection can be costly, target populations may be inaccessible, phenomena cannot always be directly observed and interviewing people could be unethical, dangerous or impossible. Budget constraints can limit the available sample size, information on covariates, the level of aggregation and the frequency of data collection.

Spatially and temporally relevant ‘big data’ that does not require data collection in the field has the potential to provide insights into people’s economic, social, behavioural and political lives, and hence could be used in measuring key development outcomes. Big data consists of human-generated data including online searches, social media, citizen reporting or crowdsourced data, process-mediated data such as mobile phone call record details (CRD), commercial transactions data and machine-generated data from satellites, sensors or drones. The primary value of big data is that it is possible to measure outcomes that could not previously be measured using household surveys at the required temporal and spatial scale. The potential of big data to answer causal attribution, however, is still not widely understood, especially in low- and middle-income countries (L&MICs).

The report is based on a map of the studies using big data and its objective is to discuss methodological, ethical and practical constraints relating to the use of big data. The systematic map includes impact evaluations (IEs) that use big data to evaluate development outcomes, systematic reviews (SRs) of big data IEs and other measurement studies that innovatively use big data to measure and validate any development outcomes. This study also explores the sectoral and geographical spread of big data’s use in international development.

This map includes studies written in English and published between 2005 and 2019, regardless of the target country’s income level or population’s status. We provide detailed breakdowns on the map for different country income classifications, fragile contexts and population characteristics. From the initial list of 17,393 studies we arrived at a final list of 437 studies, which included 48 IEs, 381 measurement studies and 8 SRs.

Key findings and lessons

- **There is considerable potential for measuring various development indicators using big data.** The number of measurement studies across development themes indicates the potential for big data to measure development outcomes. These measurement studies serve as a proof-of-concept for evaluators who look for innovative ways to measure development outcomes.

- **There is potential for more IEs on development interventions.** The map shows that the number of IEs that use big data to measure outcomes or control variables is growing fast. However, there are fewer IEs than measurement studies, and the extent of their thematic and geographical coverage is limited. IEs seem to be concentrated around environmental sustainability, economic development and urban development.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

- **Satellite data is used the most.** The use of satellite data for IEs and measurement studies has been facilitated by the availability of pre-processed satellite data, new machine learning (ML) techniques and increased computational capacity to process the satellite images into meaningful measures of development outcomes. However, despite a number of high-profile measurement studies, CRD data has not been used to rigorously evaluate any development outcome. Similarly, other human-sourced data and process-mediated data have been used only sparingly in IEs. This is a notable gap and a potential area for future exploration.

- **There are potential sectors and themes where SRs will be useful.** Although the number of IEs is small, the map highlights a few potential thematic areas where SRs will be informative. An SR of all IEs that have used satellite data across the sectors will help understand the potential of and challenges in using satellite data for IEs. Similarly, there is a concentration of IEs on forest management. An SR with reference to the data sources used in rigorously evaluating forest cover, and the advantages and challenges thereof, may be useful.

- **There is unrealised potential to conduct studies in fragile contexts.** A number of studies using big data have been conducted in fragile contexts such as conflicts, humanitarian crises, disease outbreaks, natural disasters and areas situated in difficult terrain. However, the IEs are concentrated around conflict and difficult terrain. The number of measurement studies indicate the potential for more IEs in fragile contexts.

- **Ethical concerns and transparency issues are substantial.** Ethical issues related to informed consent, data privacy, data security and unintended exclusion are severe for some of the sources of big data. However, this report shows that very few studies report on ethical issues related to using big data. This report calls for Institutional Review Board (IRB) review specifically designed for ethical issues related to big data. Similarly, very few studies have reported on data quality and fewer studies have data publicly available for replication.

- **Some capacity constraints are acute.** Computational capacity is constrained and technical expertise on large-scale big data analysis is siloed. This report calls for donors to facilitate more interaction among data scientists and development evaluators for collaborations and learning.

This map shows that big data can contribute to the evidence base in development sectors where evaluations are often infeasible due to data issues. One of the key ‘absolute gaps’ that the map has identified is that there are fewer IEs than measurement studies. Given the fast-growing availability of big data and improving computation capacity, there is great potential for the use of big data in future IEs. However, several analytical, ethical and logistical challenges may hinder the use of big data in evaluations. This report calls for standards to be set for the reporting of data quality issues, data representativeness and data transparency.

More interaction is needed between big data analysts, remote sensing scientists and evaluators.
1. Introduction

1.1 The issue: data challenges

Policymakers need access to reliable data to evaluate development outcomes and decide on future resource allocation. Governments, multilateral organisations and other development players in L&MICs use censuses, nationally representative household surveys, other household surveys and administrative data to evaluate development programmes and policies. With the increasing complexity of development programmes, there is a need to collect a vast array of output, outcomes and contextual variables to robustly assess impact. However, significant data collection challenges remain. Data challenges for IEs include limitations on sample size and power due to budget constraints, inaccessible or difficult-to-reach sections of target populations, measurement errors due to recall bias, inadequate frequency and level of aggregation, inadequate information on controls and covariates, data collection lag times and difficulties in measuring long-term impact (Wassenich, 2007). Further, in some contexts, like conflicts and humanitarian emergency situations, data collection is often impossible. The data gaps and challenges are particularly significant for the populations and countries where the need for evidence-informed policy decisions are perhaps the greatest (Gaarder and Annan, 2013). Another key shortcoming of survey data is inadequate aggregation at sub-national administrative units such as districts, counties or villages, inhibiting evaluation of programmes with spatial attributes.

Big data offers great potential for answering some of these data needs. More importantly, it answers the causal questions around which policies or interventions work, including in contexts where traditional methods of data collection are challenging. The UN Global Pulse (2013) defines big data as being digitally generated (as opposed to digitised manually), passively produced (a by-product of digital services, transactions and interactions), automatically collected and geographically and temporally trackable. Although there is no formal definition for big data, currently the term is characterised by the three Vs: high volume, velocity and variety. Satellite images, sensors and drones, mobile phone CRDs, commercial transactions data, online searches, social media, citizen reporting or crowdsourced data are the sources of big data.

Integrating big data with traditional household surveys and administrative data can complement data availability, quality, granularity, accuracy and frequency, as well as help measure development outcomes temporally and spatially in a number of new ways (York and Bamberger, 2020; BenYishay et al., 2018; Salganik, 2017; Lokanathan et al., 2017; and UN Global Pulse, 2016). For example, satellite images and mobile CRD have been used in mapping

---

1 The gap in data availability at the country level is partially driven by a lack of resources, limited capacity within governments and logistical difficulties in collecting the data. For example, the total cost of collecting data on all the 169 SDG targets was estimated to be around USD 254 billion, which is about 12.5% of total official development assistance to be committed for the post-2015 period (Jerven, 2014). A recent UN survey shows that there is existing capacity to collect data on only 40 SDG indicators and data sources for another 47 indicators are available in principle. There is little capacity and resources for collecting data on the remaining indicators (UN, 2018).
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

poverty (Jean et al., 2016; Blumenstock et al., 2015), disaster response (Lu et al., 2012; Wilson et al., 2016) and food security (Decuyper et al., 2014). Web searches and social media were used in predicting unemployment and crime instances (Xu et al., 2013; Gerber, 2014).

While big data is increasingly used for tracking indicators and monitoring development progress on Sustainable Development Goals (SDGs) (UN Global Pulse, 2012; Vaitla, 2014; Lokanathan et al., 2017), available data is less often utilised to address causal questions about the effects of specific policies and programmes. Big data can contribute to answering some of the causal questions around which interventions work. Big data prediction models can generate proxy estimates for key development outcomes such as wealth, human development, infrastructure quality, forest cover and more, which can be used in experimental (Jayachandran et al., 2016; Pellegrini, 2019) and quasi-experimental studies (BenYishay et al., 2018; Jaiswal, 2019). Satellite images such as night light, crop intensity, water availability, land use, proximity to services and physical attributes such as elevation or slope can be used in IEs as a direct measure of outcomes or as covariates. Furthermore, big data can be used for measuring and evaluating the long-term impacts of policies and programmes, conducting ex-post evaluations and estimating spatial heterogeneity. For example, satellite data is available at least as far back as 1993 for all places (high-resolution pictures are available for the entire globe at a granularity as low as 1×1 metre), allowing measurement of long-term impacts. This can help fill the gaps in evidence that cannot be addressed by traditional data sources.

The potential of big data to answer causal attribution, however, is still not widely understood or used, especially in L&MICs (York and Bamberger, 2020). In this context, a systematic collection of various sources of big data and ways of measuring and evaluating development outcomes will be a great value addition to the development community’s contribution to evidence-informed policymaking.

In this paper we look at IEs, SRs and measurement studies² that use big data to evaluate development outcomes with a special focus on fragile contexts. The study highlights the new sources of data; how these new data sources can be used for measuring development outcomes innovatively; and how these new measures can be used in IEs. We map different sources of big data onto development outcomes based on SDGs to identify the current evidence base and its gaps.

² For the purpose of this report, big data measurement studies are defined as the studies that have innovatively used big data to measure and validate any development outcome such as poverty measurement, crop productivity, employment, mobility, forest cover, etc. These are not impact evaluations but can inform future evaluations.
1.2 Objectives and the research questions

The overarching aim of this report is to inform policymakers and evaluators of existing evaluations based on big data and to provide a database of big data-based IEs and studies that could inform future IEs. Specifically, the objectives of the research are to:

- Identify rigorous IEs, SRs and the studies that have innovatively used big data to measure any development outcomes, with special reference to fragile contexts;
- Summarise current understanding of potential uses, pros and cons, reliability, biases, risks and ethical issues in using big data for measurement and evaluation of development outcomes; and
- Generate interest and awareness among key stakeholders (evaluators, researchers, donors, practitioners, implementers and policymakers) of the potential as well as challenges of using big data.

This systematic map addresses the following questions:

- How have different types of big data and methods been used for measuring and evaluating development outcomes?
- How dispersed or concentrated is the use of big data across development goals and geographies?
- What are the potential biases, measurement reliability issues, pros and cons, risks and ethical issues in using big data for measuring and evaluating development outcomes?
- What are some of the unexplored but promising applications of big data for IEs?

1.3 Scope of work

For the purpose of this research, we define big data sources as digitally generated, passively produced and automatically collected data, as defined in UN Global Pulse (2013). The sources of big data include satellite images, sensors and drones, mobile phone CRDs, commercial transactions data, online searches, social media, citizen reporting or crowdsourced data. See Table 2 for more details on various sources of big data adapted from UN Global Pulse (2012 and 2013), United Nations Economic Commission for Europe (2014) and Blazquez and Domenech (2018).

3ie evidence gap maps compile IEs and SRs. However, in this study, we include IEs and SRs as well as measurement studies: the studies that have innovatively used big data to measure and validate any development outcome. These are multidisciplinary studies that use state-of-the-art methods from computer science and statistics to collect, clean and analyse big data for measuring development outcomes. For example, Jean et al. (2016) use transfer learning techniques as well as daytime and night light data from satellite images to estimate consumption expenditure at the cluster (village) level to map poverty in five African countries: Nigeria, Tanzania, Uganda, Malawi and Rwanda. While a number of such studies have used big data for measuring various development outcomes, few IEs have used these innovative big data-based outcome measures. These measurement studies, we hope, would serve as
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

proofs of concept for innovative use of ML and big data that can be used in future evaluations.

3ie defines an IE as a ‘study of the attribution of changes in the outcome to the intervention’⁶. For the purpose of this systematic map, we define big data-based IEs as any experimental or quasi-experimental studies that use any form of big data to measure the outcomes of interest and/or the confounding variables.

We use the Organisation for Economic Co-operation and Development (OECD) definition of fragile contexts, which includes conflicts, institutional fragility, social fragility, environmental risks, health risks and climatic risks. This list is more inclusive than the list used by the UK Department for International Development (DFID) and the World Bank, which includes conflict and institutional and social fragility (DFID, 2016). Please see Appendix 1 for more details on the classifications and country list. We use the OECD definition for classifying fragile contexts based on:

- Difficult terrain
- Natural disasters
- Conflict or humanitarian crisis
- Chemical or radio-nuclear issues
- Disease outbreaks or epidemics.

Using big data in evaluation poses a number of analytical challenges on issues including data quality, transparency, generalisability, and privacy and ethical challenges such as consent for using data and anonymisation of the data. This report also explores how the included studies dealt with these challenges.

1.4 Overview

Section 2 of the report provides an overview of methodology, inclusion and exclusion criteria, and defines big data sources and development goals used in the report. Section 3 provides a detailed description of the trends and distribution of the included studies and how they dealt with the analytical and ethical challenges. Section 4 summarises the key findings and pointers for future studies. Section 5 provides an overview of how effectively big data can be used in IEs. Section 6 discusses the limitations of the study and the possible next steps for the evaluators, practitioners and donors. Section 7 presents concluding remarks.

---

2. Methodology: definitions, inclusion and exclusion

2.1 Methodological approach

We follow 3ie's methodology and process for evidence gap maps (Snilstveit et al., 2017). To create this map, we used systematic methods to identify any completed and ongoing IEs, SRs and big data measurement studies relevant to our research objectives. We conducted systematic searches and data extraction as described in Appendices 3 and 4. The studies identified are mapped on to the framework of big data sources and SDG outcomes to provide a visual display of the volume of and the trends in the evidence base. We also coded how the included studies have dealt with ethical and transparency related challenges. The systematic map is available through an online interactive platform on the 3ie website and allows users to explore the available evidence through different filtering options. There are links to study summaries in the 3ie repositories (wherever applicable) and confidence ratings for the SRs.

2.2 Criteria for including the studies

Table 1 summarises the criteria we used for searching, screening and including the studies for the map.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>This map includes all population from all countries but we provide breakdowns for rural areas, urban areas, conflicted-affected persons and ethnic minorities. We also provide breakdowns for L&amp;MICs and fragile contexts separately.</td>
</tr>
<tr>
<td>Sources of big data</td>
<td>Big data may originate from any of the following sources:</td>
</tr>
<tr>
<td></td>
<td>● Human-sourced information:</td>
</tr>
<tr>
<td></td>
<td>o Social networks</td>
</tr>
<tr>
<td></td>
<td>o Internet searches</td>
</tr>
<tr>
<td></td>
<td>o Mobile data content</td>
</tr>
<tr>
<td></td>
<td>o Citizen reporting or crowdsourced data</td>
</tr>
<tr>
<td></td>
<td>● Process-mediated data (traditional business systems and websites):</td>
</tr>
<tr>
<td></td>
<td>o Data produced by public agencies</td>
</tr>
<tr>
<td></td>
<td>o Data produced by businesses</td>
</tr>
<tr>
<td></td>
<td>o Mobile phone CRD</td>
</tr>
<tr>
<td></td>
<td>● Machine-generated data (automated systems):</td>
</tr>
<tr>
<td></td>
<td>o Data from fixed sensors</td>
</tr>
<tr>
<td></td>
<td>o Data from mobile sensors (tracking)</td>
</tr>
</tbody>
</table>

4 The online map can be accessed here: [https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map](https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map)
### Outcomes

1. Economic development and livelihoods
2. Agriculture and food security
3. Health and well-being
4. Quality of education
5. Governance and human rights
6. Water and sanitation
7. Energy, industry and infrastructure provision
8. Urban development
9. Environmental sustainability
10. Partnerships for goals

### Study design

**IEs:**
- Randomised controlled trial (RCT)
- Regression discontinuity design
- Controlled before-and-after study using appropriate methods to control for selection bias and confounding, such as propensity score matching or other matching methods
- Instrumental variable estimation or other methods using an instrumental variable such as the Heckman two-step approach
- Difference-in-differences
- A fixed-effects or random-effects model with an interaction term between time and intervention for baseline and follow-up observations
- Natural experiments
- Other quasi-experimental studies inducing synthetic control studies
- Survey, laboratory or lab-in-the-field type experiments
- Cross-sectional or panel studies with an intervention and comparison group using methods to control for selection bias and confounding as described above

**Measurement studies**

We included the studies that innovatively used big data to measure and validate any development outcomes. These studies use big data to measure components that would have been difficult to measure using survey data.

**SRs**

We include only the reviews that specifically looked at studies that used big data to measure development outcomes and explicitly described the search, data collection and synthesis methods according to a standard SR protocol, such as the 3ie SR protocol.
2.3 Sources of big data

Innovations in the type of devices available for measurement (satellites and sensors); daily personal use (mobiles, wearables, Internet of Things, etc); social interaction (blogs, Facebook, Twitter, WhatsApp, etc); and recording business transactions digitally (CRD, e-transactions, mobile money, credit card payment, etc) have led to an explosion of automatically collected data. However, there is no official definition of big data. McKinsey defined it broadly as data ‘whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse’ (Manyika et al., 2011).

UN Global Pulse (2013) defines big data for the purposes of development as being digitally generated (as opposed to digitised manually), passively produced (a by-product of digital services, transactions and interactions), automatically collected and geographically or temporally trackable. While the size, velocity and veracity are all defining characteristics of big data, the definition relevant for IE is that these are non-sampled data, passively left behind by humans using digital devices and services or automatically collected by the services providers for the purpose other than statistical inference (Letouzé, 2015; UN Global Pulse, 2016). Hence, unlike the conventional survey data where the respondents say what they do or feel, big data captures what people actually do. The implication of this is that big data is non-reactive: in other words, there is less likelihood of social desirability bias (Salganik, 2017). The other key characteristic of big data that matters for evaluation is that it is near real-time and can be available across multiple frequencies (e.g. hourly, daily) over a long period. Table 2 provides the types of big data, sub-classifications, definitions and sources.

Using these definitions, we include the following broad classification of big data⁵:

- Human-sourced information from social networks that is provided voluntarily by users;
- Process-mediated data from traditional business systems and websites that includes digitally recorded business activities;
- Machine-generated data from automated systems includes information from sensors and machines that measure and record events and situations in the physical world.

Table 2: Sources of big data

<table>
<thead>
<tr>
<th>Data type</th>
<th>Source of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-sourced information</td>
<td>Text, metadata and location data from social networking sites, opinion platforms, blogs, pictures, videos, etc such as Twitter, Facebook, LinkedIn, YouTube, Wiki pages</td>
</tr>
</tbody>
</table>

⁵ There are also other classifications of big data based on its structure: structured, semi-structured and unstructured data. Structured data is in the standard columns and rows form such as what sensors provide; semi-structured data contains free texts but can be tagged, such as Twitter; and the unstructured data includes images and videos (Desouza and Jacob, 2017). Big data can also be classified as open and proprietary data (Maaroof, 2015).
**Internet searches** | Internet text and internet search queries, i.e. Google Trends; web logs (open data)
---|---
**Mobile data content** | Text messages
**Citizen reporting or crowdsourced data** | OpenStreetMap, Humanitarian Data Exchange platform, etc. Collected from a large group of people who voluntarily provide information. This is a useful source for recording events and receiving people’s opinion and feedback on issues. Collected through hotlines, user-generated maps, marking of instances, events, etc

**Process-mediated data**

| Data produced by public agencies | Medical records; postal data; tax data, etc |
| Data produced by businesses | E-commerce transaction records and credit card transaction records, ATM transactions, mobile money transfers, savings and loan repayments (collected by the service provider as a part of regular business operation and monitoring; proprietary and commercially sensitive data) Tolls and public transport card use data |
| Mobile phone CRD | Mobile CRDs that provide metadata on when the call took place, the cost, the time and the recipient of the call; location of the caller and of the recipient; and the users’ mobility, social interaction and airtime transaction details; top-up data |

**Machine-generated data**

| Data from fixed sensors | Home automation, weather/pollution sensors, traffic sensors/webcams, security/surveillance videos/images and activity records such as electricity meters (mostly administrative data collected by the authorities as a part of regular monitoring) |
| Data from mobile sensors (tracking) | Community or privately owned drones (common property or privately held data) Mobile phone global positioning system (GPS) (open data available from Google) Vehicle GPS and self-tracking apps (proprietary data) |
| Data from satellites | Open data available from a number of sources, for example: ● Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) ● Visible Infrared Imaging Radiometers ● Landsat ● European Space Agency Land Cover, etc |

Source: Adapted from UN Global Pulse (2012 and 2013), United Nations Economic Commission for Europe (2014) and Blazquez and Domenech (2018)
2.4 Outcomes

We use the SDGs as the basis for identifying the outcome categories, similar to Phillips et al. (2017). We have regrouped them into smaller thematic groups based on complementarities between the sectors (White et al., forthcoming). Table 3 provides the definition of each of the outcomes as defined in UN (2019). Appendix 2 provide more details on the sub-maps where further breakdowns have been provided for key sub-classifications where relevant.

Table 3: Outcome categories and definitions

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic development and livelihoods (SDGs 1 and 8)</td>
<td>End poverty in all its forms everywhere Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all</td>
</tr>
<tr>
<td>Agriculture and food security (SDG 2)</td>
<td>End hunger, achieve food security and improved nutrition and promote sustainable agriculture</td>
</tr>
<tr>
<td>Health and well-being (SDG 3)</td>
<td>Ensure healthy lives and promote well-being for all ages</td>
</tr>
<tr>
<td>Quality of education (SDG 4)</td>
<td>Ensure inclusive and equitable quality education and promote life-long learning opportunities for all</td>
</tr>
<tr>
<td>Governance and human rights (SDGs 5, 10 and 16)</td>
<td>Achieve gender equality and empower all women and girls Reduce inequality within and among countries Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels</td>
</tr>
<tr>
<td>Water and sanitation (SDG 6)</td>
<td>Ensure availability and sustainable management of water and sanitation for all</td>
</tr>
<tr>
<td>Energy, industry and infrastructure provision (SDGs 7 and 9)</td>
<td>Ensure access to affordable reliable, sustainable, and modern energy for all Build resilient infrastructure promote inclusive and sustainable industrialisation and foster innovation</td>
</tr>
<tr>
<td>Urban development (SDG 11)</td>
<td>Make cities and human settlements inclusive, safe, resilient and sustainable</td>
</tr>
</tbody>
</table>

*Our map refers to SDGs since they are globally acknowledged and fairly inclusive list of development outcomes. All the studies relevant to this map are expected to fall under at least one of the 169 targets listed within SDGs.*
Evaluation of potential methodological issues

While big data can help resolve many data related challenges, there are considerable methodological, analytical, logistical and ethical challenges in the way of using it in measuring development outcomes (Lokanathan et al., 2017; Salganik, 2017; Olteanu et al., 2019; Letouzé, 2016). This section briefly discusses some of the prominent methodological challenges. As big data is varied in type, quality and composition, we also discuss if the challenges are specific to any particular type of big data.

We have grouped all the big data challenges that may affect measuring development outcomes credibly and lead to questionable internal and external validity of the studies.

**Non-representativeness of data and selection bias**: Big data may unintentionally exclude certain sections of the population or marginalised communities, thereby making the sample unrepresentative of the population being analysed. Large samples do not solve this systematic bias. This, however, is not a challenge when using satellite data that has universal coverage but, with human-generated and CRD data, non-representativeness is a serious challenge. Human-generated data such as Twitter, Facebook or web searches, as well as mobile phone use that generate CRD data, are not representative as the usage is limited by income, education, infrastructure, etc. However, clarity on what is the sample frame (i.e. who is included and who is excluded) will help interpret big data results appropriately. Non-representative sample is still useful for within-sample comparisons, but may lead to erroneous out-of-sample generalisations (Salganik, 2017; Olteanu et al., 2019).

**Construct validity**: Construct validity is whether the proposed measure actually measures what it claims to be measuring. This becomes important when the construct is unobservable and has to be operationalised via some observed attributes (Olteanu et al., 2019). For example, does night light data truly reflect local GDP and other development outcomes such as health and education? What is it in the CRD data that reflects people's income or employment status? In many cases, the big data-based measures may not be straightforward, and it is good practice to clearly state construct validity and provide necessary support to back the claim in the papers. Development measures based on social media are particularly challenging due to
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

different communication styles, special usage of terms and differences in language proficiency.

Data quality issues

- **Comparability of data over time:** Since most of these data are collected routinely as a part of business, the nature and quality of data may change with the technology and business requirements. This may happen because the underlying technology has changed or because the people who use it have changed. For example, satellite data is not readily comparable across the years as there is a vast quality difference (Jain, 2020); good flu trends based on online searches peaked comparing to officially reported data when the underlying Google algorithm started prompting people to query more and broke the relationship between Google searches and flu prevalence (Archie et al., 2018).

- **Lack of completeness:** Most big data is a by-product of peoples’ everyday action and/or result of system logs of the government and businesses. It may not contain all the necessary information, such as demographic characteristics. However, combining multiple sources of data, especially big data and administrative data, can help resolve this problem (Salganik, 2017).

**Generalisability:** Generalisability or external validity refers to the applicability of the findings of a study to population or context other than it was produced. In the context of big data, generalisability would mean the applicability of the model to a setting different from the setting of the data that the model was trained on. For example, a model trained on satellite data from a specific geographical region may not be generalised (Jean et al., 2016; Head et al., 2017). It is good practice for studies to report on the representativeness of training data.

**Data transparency:** Transparency in this context refers to publishing all relevant materials, including the data and code, used in a study in the public domain for independent verification. Sharing of raw data in the public domain is often crucial for establishing confidence and reliability in the results. There are two challenges here: first, some of the major sources of big data (such as CRD) are proprietary and sharing may not be permitted beyond the closed group of researchers; and second, the data has to be de-identified before it can be shared and it is crucial to check whether there are variables or a combination of variables that can be used to re-identify research subjects.

We assessed whether the studies included in the systematic map asked the following questions:

- Is the data representative of the population of interest?
- Is the construct validity explained (i.e. is there a discussion on how the big data-based indicator measures what the study claims to measure)?
- Are there data quality issues in the dataset used and how are they addressed?
- Are the results generalisable? For example, are the research findings generalisable to other situations, such as other platforms (data source) or communities, or over time?
- Are data and codes publicly available for replication?
2.6 Evaluating reporting on privacy and ethical considerations

There are concerns over data access, privacy, consent and ethics in using big data. Although these are foundational issues for both small and big data studies, the challenges posed by big data have greater repercussions.

When using big data sources such as mobile data, most mobile operators have ‘inform and consent’ policies that mandate disclosure of all relevant information to potential participants who can then evaluate this information and give explicit permission. However, these policies often contain legal language that is generally not discernible, and it is not clear if explicit consent is obtained to repurpose the data. This kind of informed consent may be completely absent in research leveraging social media data due to the impracticality of obtaining consent from millions of users.

Mobile phone user data and social media data are some of the most used sources of big data that can inform researchers about individuals’ behaviour. Even if the data is de-identified, concerns still remain over the consent and ethics of sharing such data with researchers. It is thus imperative to have an ethics approval process in place that lays down the conditions under which such research can take place. There is a need for clear ethical standards for big data research and studies should be monitored by the IRBs.

Another ethical criterion when using big data can be concerned with the assessment of risks, the most common being privacy breaches leading to identity theft or other cybersecurity risks. The possibility of the re-identification of any individual user from poorly anonymised datasets adds to the concerns over anonymity of subjects. When combined with other sources, such datasets can be used to gain detailed insights about people without their knowledge. Such precise inferences may create the capacity for discrimination or mass manipulation.

Sometimes data obtained for one purpose in social data research is used for secondary analyses, but the associated risks may not be well understood. For example, Facebook data in the past has been used for ad targeting, as well as for tailoring propaganda (Horowitz et al., 2018).

Big data may also inadvertently exclude certain sections of the population. For example, this bias can be observed in the case of ‘Street bump’, a mobile app that notifies the Boston City Hall whenever the user hits a bump on the road (Carrera et al., 2013). The data includes information only from the app users who often use both their cars and the app; this might inadvertently exclude poorer parts of the city that app users may not frequent. Policy based on such big data sources may have unintended consequences for the people who are excluded.

We assessed the studies on the following:

- Ethical approval obtained
- Consent for secondary and other use of data discussed explicitly
- Discussions on data privacy
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

- Discussions on data security and governance arrangement
- Discussion of any potential unintended exclusion
- Discussion of potential unintended consequences for any group of people or individuals.

2.7 Exclusion criteria

This report includes papers written in English and published between 2005 and 2019. This map does not include studies that develop algorithms or methodologies for using big data without explicit application to measuring or evaluating a development outcome. This map does not include studies that describe how big data and ML have been used in development programming to help programme implementation, coordination and management for designing and scaling new development solutions; neither does it include studies that show how big data methods are used in RCTs to identify the differential impact of sub-groups and in improving survey data collection, such as defining sampling frames.
3. **Results**

3.1 **Search, screening and coding**

3.1.1 **Searching and screening**

We systematically searched academic databases with the help of an information specialist (see Appendix 3 for the list of academic databases and search strings) and manually searched specialist organisational websites and grey literature sources. The initial searches resulted in 17,393 studies, of which 17,008 studies were identified through bibliographic databases search and 385 studies were identified through hand-searching specialist databases and IE and grey literature repositories (Figure 1). All the results were uploaded onto Evidence for Policy and Practice Information (EPPI)-Reviewer 4 for screening and coding. Screening of the studies was done in three stages.

We first employed EPPI-Reviewer’s built-in text mining, an ML technique, to sort the studies based on the inclusion and exclusion criteria at the title and screening stage (see Appendix 4 for more details on how text mining was used for screening). This reduced the number of studies to be screened at the title and abstract level to 9,720. At the second stage, three researchers screened the studies for eligibility based on inclusion and exclusion criteria defined in the study protocol. At the end of stage two, we had identified 1,348 studies to be screened at the full-text level.

At stage three, four researchers coded the studies at the full-text level. We followed a single screening with safety first approach, where only one reviewer screened each record and coded doubtful studies as ‘to discuss’ to avoid excluding the studies that are potential includes. About 25% of the studies at this stage were double-coded for consistency and any disagreements were resolved with the lead author. About 20% of the studies were spot-checked by the lead author. At the full-text screening stage, we excluded 1,292 studies. We excluded 251 studies based on irrelevant data sources, 408 studies on outcomes and 633 studies on study designs. Studies excluded on design were primarily policy papers or computer science technical research papers that aimed at developing new algorithms rather than measuring development outcomes.

---

7 See Rathinam et al. (2019a) for more details on single screening with safety first approach.
### 3.1.2 Coding of included studies

The three-stage process of screening resulted in a final list of 437 studies including 48 IEs, 381 measurement studies and 8 SRs. Through a consultation process, we identified the metadata to be extracted using a standardised data extraction tool and defined them in the study protocol. During the final studies coding, we collected metadata such as outcomes studies, outcome sub-categories, data sources used, geographical location of the intervention, country, evaluation design, target population, data transparency, ethics and other bibliographic information from the included studies (see Appendix 4 for the data extraction tool). We also critically appraised the SRs (see Appendix 5 for the SR appraisal tool and summary of the included SRs). Due to the size and the nature of included studies we did not conduct critical appraisal of IEs or measurement studies.

### 3.1.3 Mapping of the studies

The studies were mapped using 3ie’s evidence gap map platform, which is organised into rows and columns. Various big data sources are placed in the rows and the development themes are placed in the columns. Any intersecting cell represents the development outcome measured or evaluated using the particular type of big data. Different colour bubbles represent the type of study: grey bubbles denote IEs, blue bubbles denote measurement studies, green bubbles denote high-quality SRs and red bubbles denote low-quality SRs. Hovering over the bubbles will show the links to studies. There are also filters for different regions, countries, study design, fragile context and the target population of the studies.

Development themes (such as environmental sustainability, economic development and livelihoods) contain a large number of studies. We have provided maps within the main map to show how the studies are distributed across the sub-themes under these broad themes. In
the sub-maps, the big data sources are mapped against level 2 or level 1 sub-classification provided in the SDGs as relevant. For example, the sub-map for economic development and livelihoods has been coded against the level 2 indicators of eradicating poverty (SDG 1) and employment and economic growth (SDG 8), and the sub-map for environmental sustainability is coded against the level 1 classification of SDGs 12, 13, 14 and 15.

The following development themes have sub-maps:

- Economic development and livelihoods
- Health and well-being
- Governance and human rights
- Urban development
- Environmental sustainability

### 3.2 Characteristics and trends of the evidence base

The map contains 437 studies, of which 48 are IEs, 381 are measurement studies and 8 are SRs. Of the 48 IEs, 8 are RCTs and the remaining are quasi-experimental studies.

Figure 2 displays the number of studies published each year from 2005 to 2020. The light blue bar shows the measurement studies, the dark blue bar shows IEs and the line indicates the cumulative number of measurement studies, IEs and SRs. The number of measurement studies has grown gradually from 2005 to 2012, increasing substantially every year since then with maximum numbers in 2017 and 2018. The past five years alone have accounted for more than 60% of the studies, indicating the increasing availability of big data, improved computational capacity and greater interest among researchers and journals.

The figure also shows that applying big data to IEs is a new phenomenon. The first IE using big data was published in 2009. While almost all the IEs were published after 2013, more than three-quarters of the IEs were published in the last five years. We expect that the measurement studies will be proofs of concept, leading IEs to adopt to these approaches to innovatively measure development outcomes in evaluations. The map also points to the gap between the growth of measurement studies and use of big data in IEs.
Figure 2: Number of studies published per year

Source: Authors’ own calculation

Figure 3 shows the geographical distribution of the included studies. About 50% of the studies (n = 210) are from Asia and close to 30% (n = 132) are from Sub-Saharan Africa. The distribution of IEs and measurement studies are roughly similar to the overall distribution. One notable exception is Latin America and the Caribbean where the region accounts for 15% of total studies (n = 65), but substantially more IEs (38%, n = 18).

Figure 3: Distribution of studies over regions

Source: Authors’ own calculation

China (43) and India (34) are the most-studied countries, and Kenya has highest number of studies (15) in Sub-Saharan Africa (Figure 4). While East Africa is very well-represented on the
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

map, other African countries have fewer entries and several countries in West and North Africa have no studies at all. About 35 of the studies are multi-country studies and 11 studies did not specify the country name primarily to conceal the identity of the data provider. The distribution of IEs and measurement studies are again roughly similar to the overall distribution but Latin America countries, particularly Mexico (5), account for a higher proportion of IEs. See Table 9 for a list of top 20 countries with the maximum number of studies and Table 10 for the geographical distribution of studies across the regions in Appendix 7.
Figure 4: Geographical distribution of studies

Source: Authors’ own calculation
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

We find that most of the studies are concentrated in middle income countries. There are 232 studies (53%) in the middle income group, followed by 103 studies (24%) in the high income group and 71 studies (16%) in the low income group. Overall, about 69% (n = 303) of the total studies are from L&MICs, but the IEs are distributed more in favour the L&MICs as 83% of them (n = 40) are from L&MICs. One of the notable features of the studies on the map is that about 82% (n = 359) of the total studies are published in peer-reviewed journals and the remainder are working papers (18%, n = 78).

Figure 5:  Number of studies by income classification

![Bar chart showing the number of studies by income classification](source: Authors' own calculation)

3.2.1 Distribution of studies across data sources

As discussed in section 2.2, big data can be generated by human interaction on social media, process-mediated data recorded by governments and the business and machine-generated data that is recorded by the automated systems. Figure 5 shows that machine-generated data is used the most. Of the total number of studies, close to 84% (n = 380) of the studies used some form of machine-generated data, while 12% (n = 53) of the studies used human-generated sources and 17% of the studies (n = 77) used process-mediated data.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

Figure 6: Number of studies per different type of big data

Table 4 below provides a detailed breakdown of the number of studies per data source.

Data from satellites and fixed sensors: Satellite data is the most used source of big data as it accounts for 71% of the measurement studies (n = 210) and close to 81% of the IEs (n = 39). Data from fixed sensors (such as weather and pollution sensors, traffic sensors and electricity meters that provide high-frequency, localised measurements) could also be readily used in IEs. This is the second most used data source, with 15% of the IEs using these sources. This shows that the data from satellites and in situ sensors that help measure spatial outcomes are used most in IEs. Other big data sources have been seldom used for IEs despite measurement studies showing proof-of-concept.

Mobile phone CRD: A good number of measurement studies have used CRDs (17% n = 65) for measuring population movement, migration, disease spread and even to understand the literacy level of the subscribers. Surprisingly, we found no IEs that used this source of big data despite the availability of a good number of proof-of-concept papers in measuring key development outcomes.

Table 4: Number of IEs and measurement studies across data sources

<table>
<thead>
<tr>
<th>Data source</th>
<th>IEs</th>
<th>Measurement studies</th>
<th>SRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-sourced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social networks</td>
<td>0</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Internet searches</td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>
Human-sourced data: Social networks including Facebook, Twitter and Wiki pages were used to measure development outcomes in 18 measurement studies. Internet searches like Google trends and other search engine queries were used in 11 studies. Mobile data content and crowdsourced data were used in ten studies, primarily to measure disease outbreak, price data or opinion on issues like development services. This source has not been used in IEs, with two notable exceptions using crowdsourced data (Van der Windt, 2016; Edjekumhene, 2019).

Complementarity between data sources: There are about 57 studies on the map that have combined at least two sources of data and about seven of them have combined three or more sources (Table 11 in Appendix 7). Data from fixed sensors and satellites data seem to complement each other well: 13 measurement studies and 2 IEs have combined these two sources. Mobile phone CRD and satellite data is the other combination that has been used repeatedly. About ten of the measurement studies have combined CRD data and satellite data in their analysis (see section 5 for a discussion and example on how satellite data and CRD can be combined together for better results). However, IEs seem not to have exploited this complementarity. See Table 11 in Appendix 7 for a list of studies using multiple sources of big data.
3.2.2 Distribution of studies across development themes

Section 2.3 identifies ten broad development themes based on SDGs. Figure 7 and Table 5 show the number of studies across the development themes. About 50% of studies (n = 217) focus on environmental sustainability, which includes sustainable consumption and production, climate change, underwater life, and life on land. Economic development and livelihoods accounts for about 26% of the total studies (n = 114). Urban development and health account for 16% each (n = 68). Governance and human rights (7%, n = 30) and energy, industry and infrastructure (7%, n = 30) account for the remaining studies.

Figure 7: Number of studies against development outcomes

While the distribution of IEs and measurement studies across the development themes remains the same as the overall distribution, there are a substantial number of IEs on economic development and livelihoods (17%) and energy, industry and infrastructure (17%). While most of the SRs looked at cross-sectoral themes, health is the most-studied sector (n = 5), followed by urban development (n = 3) and environment sustainability and economic development (n = 2). See Appendix 6 for critical appraisal of the SRs.

Table 5: Distribution of studies across development themes

<table>
<thead>
<tr>
<th>Development themes</th>
<th>IEs</th>
<th>Measurement studies</th>
<th>SRs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic development and livelihood</td>
<td>8</td>
<td>105</td>
<td>1</td>
<td>114</td>
</tr>
</tbody>
</table>

Source: Authors’ own calculation
3.2.3 Units of observation

The unit of observation (or unit of analysis) is the class of elemental unit that constitutes the population and the units of measurement. Typically, in IEs, the units of observation are individuals, households, facilities (in facility surveys) or various levels of administrative units such as villages, counties or districts. We have classified the unit of observation as population (including both individuals and households) or administrative units (villages, land parcels or any other units with a spatial element). The unit of observation seems to be an important element in analysing the use of big data in measuring development outcomes. Figure 8 (Panel 1) shows that about 70% of the measurement studies (n = 267) and 65% of the IEs (n = 31) have administrative units as their unit of observation. There is a clear distinction between different sources of big data, as shown in Figure 8, Panel 2. The unit of analysis for satellite data-based studies is predominantly administrative units (n = 259, 83%), while CRD-based studies are usually based on population units (n = 53, 82%). This difference shows that satellite data is more applicable when the outcome of interest has some spatial dimension such as local economic development, agricultural land productivity, forest cover or urban development.

Figure 8 Units of observation

<table>
<thead>
<tr>
<th>Category</th>
<th>Administrative unit</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustainable agriculture and food security</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Health and well-being</td>
<td>5</td>
<td>58</td>
</tr>
<tr>
<td>Education</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Water and sanitation</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Governance and human rights</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>Energy, industry and infrastructure</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Urban development</td>
<td>3</td>
<td>69</td>
</tr>
<tr>
<td>Environment sustainability</td>
<td>25</td>
<td>190</td>
</tr>
<tr>
<td>Global partnership</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Source: Authors’ own calculation
### 3.2.4 Studies using mixed methods

Mixed-methods IEs that combine qualitative and quantitative analyses help assess the quality implementation and reliability of data and understand the mechanism of programme impact (Bamberger, 2012). Big data IEs can be combined with qualitative methods. However, only three IEs and five measurement studies reported using mixed methods.

### 3.2.5 Studies with a rural or urban focus

Figure 9 shows the proportion of studies focused on rural areas or urban areas, or both. Most studies looked at both rural and urban areas (74%, n = 325). About 9% of the studies (n = 41) focused on rural areas; 12% (n = 51) focused on urban areas. Among the remaining studies, 14 studies looked at conflict affected population, 4 were studies of ethnic minorities and 2 studied refugees.

**Figure 9: Distribution of population sub-groups**

Source: Authors' own calculation

### 3.2.6 Studies by fragile context

We used the OECD definition of fragile context that includes conflict, institutional, social fragility, environmental, health and climatic risks (OECD, States of Fragility (2018)). Figure 10 shows that 91 studies included on the map (21%) are from countries considered to be fragile. About 39 studies were conducted in a conflict or humanitarian crisis context; 22 studies each were conducted in contexts of difficult terrain and natural disasters; and 15 studies were conducted in the context of epidemics or disease outbreaks. There was one measurement study in the context of a chemical/radio-nuclear disaster. IEs follow the same pattern, except for one notable gap: there are no IEs in the context of epidemics or disease outbreaks despite a reasonably good number of measurement studies.
Figure 10: Number of studies in fragile contexts

![Number of studies in fragile contexts](image)

Source: Authors’ own calculation

Table 12 in Appendix 7 shows that satellite data is the most used in fragile contexts, followed by CRD data and then the sensor data. The table also shows that almost all the big data sources have been used in one or two fragile contexts, indicating the importance of big data in fragile contexts.

3.3 Appraising potential methodological biases, risks and limitations

3.3.1 Reporting on methodological challenges and data transparency

Figure 11 shows that very few studies meet any of the following methodological quality markers.

- Is the construct validity explained (i.e. is there a discussion on how the big data-based indicator measures what the study claims to measure)?
- Are data and codes publicly available for replication?
- Are data collection methods discussed?
- Are there data quality issues in the dataset used and how are they addressed?
- Is the data representative of the population of interest?
- Are challenges in the analysis and reporting process discussed?
- Are the results generalisable? For example, are the research findings generalisable to other situations such as other platforms (data sources) or communities, or over time?
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

Only 95 studies (22%) have reported on at least one of the above transparency criteria. For example, only 20% (n = 91) of the total studies reported on data collection methods, 6% (n = 25) on data quality issues, 8% (n = 36) on data representativeness, 14% (n = 64) on construct validity and 7% (n = 30) on generalisability. Only 4% (n = 19) of the studies have data and codes publicly available or available upon request.

Figure 11: Number of IEs and MS against data quality and transparency

There is, however, considerable difference between IEs and measurement studies in terms of reporting on data quality issues and transparency. Table 12 in Appendix 7 shows that IEs report a lot better on all these parameters. Of the total 48 IEs on the map, 46 of them report at least one aspect of transparency. Almost all the IEs report on data collection methods, 90% (n = 43) report on construct validity, 60% (n = 29) discuss representativeness of data, 54% (n = 26) discuss generalisability and 45% (n = 22) discuss various data quality issues. However, only 23% (n = 11) make data and codes available and 13% (n = 6) discuss key data analysis and reporting challenges.

Source: Authors’ own calculation
3.3.2 Reporting on privacy and ethical challenges

Figure 12 shows that most studies (81%) do not report on ethical challenges and privacy issues. Of the few that do discuss such challenges, the most frequently discussed issue is consent for data use.

**Figure 12: Number of studies reporting on ethics issues**
4. Key findings, discussion and lessons

The use of big data in measuring development outcomes has been on the rise over the past five years. This rising trend is powered by the availability of (and our capacity to process) big data. In this section, we discuss the key findings, some of the notable gaps and the potential for future SRs.

4.1 Evidence base and key gaps

There is a considerable potential for measuring various development indicators using big data. We identify a significant and growing evidence base of measurement studies that use some form of big data to measure a development outcome. Some outcomes are more amenable to the use of big data than others; environmental sustainability, economic development and livelihoods, health and well-being and urban development are where the majority of studies are concentrated. Education, sanitation, governance and human rights seem to be less responsive to big data use.

Multiple entries for most development theme indicate the potential of big data in contributing to measuring development indicators. Identifying measurement studies will be a valuable addition to development evaluators who look for innovative ways to measure a development outcome that was difficult to measure at all required spatial and temporal scales using conventional data collection methods.

There is potential for more IEs using big data on development interventions. The map contains 48 IEs. Use of big data measures in IEs as main outcomes or for controlling key covariates is fast-growing, but the IEs are fewer in number compared to measurement studies as well as in terms of the extent of their thematic and geographical coverage. IEs seem to be concentrated more around environmental sustainability, economic development and urban development. This complements existing efforts to build the evidence base in international development, as these sectors have much less rigorous evaluations (Sabet and Brown, 2018). The IEs also concentrate on using satellite data.

Satellite data is used most. The map shows that 71% of the measurement studies and 81% of the IEs used satellite data. This is also one of the sources that has been used since the early 2000s. The prominence of satellite data is primarily due to the fact that satellite images offer unique possibilities for measuring and evaluating development outcomes. Given the vast number of satellites covering almost every location on earth, it is possible to collect data at a high granularity (spatial resolution) and for multiple temporal frequencies for the past 30 years. Satellite data is freely available from several sources (such as NASA's Landsat and MODIS and the European Space Agency's Sentinel); more importantly, several pre-processed databases are available (such as AidData's Geoquery, Yale University's G-Econ Project, the United Nations Environment Programme's Environmental Data Explorer, NASA's

---

8 http://geo.aiddata.org/query/
9 https://geodata.grid.unep.ch/
Methods working paper: Using big data for evaluating development outcomes: a systematic map

Socioeconomic Data and Applications Center\textsuperscript{10} [SEDAC], Global Forest Change 2000–2018 [Hansen et al., 2013], and several others. This pre-processed data or the image data could then be processed and converted into meaningful outcomes to measure economic activity at local level, urban development, forest cover, land productivity, distribution of the population, etc. These indicators can also be used for controlling for covariates.

**Spatial dimension matters.** One of the key findings of the map is that most of the big data studies are applied in the context where the phenomenon studied has a spatial dimension, meaning the outcome and other covariates are measured on a spatial scale. Close to 70% of the studies on the map report using administrative units as their unit of measurement. This is particularly true for satellite and sensor data-based studies, as 82% have administrative units as their unit of measurement (such as local economic development, agricultural land productivity, forest cover or urban development). This is referred to as geospatial IE (BenYishay et al., 2017). However, there is considerable difference across data sources as CRD data is used to measure changes at the population level.

**CRD data has great potential for measuring and evaluating development outcomes but is not yet used in IEs.** CRD data is one of the most widely used sources in measurement studies. This is used for measuring population movement, migration, disease spread, etc. Despite a number of high-profile measurement studies, our systematic search did not find even one IE that used CRD data for rigorously evaluating a development outcome. This is a notable gap and a potential area for future exploration. It should be noted that CRD data is also fraught with multiple methodological challenges (such as non-representativeness, lack of completeness, etc) and ethical challenges (such as consent, unintended exclusion, etc). Further, CRD data has been difficult to obtain as it is proprietary and hence it is difficult to maintain data transparency.

**Other big data sources such as human-sourced and process-mediated data have good proof-of-concepts.** Human-sourced data (such as social networks, internet searches, mobile data content citizen reporting or crowdsourced data) and process-mediated data (such as data produced by public agencies and by businesses) have a good number of measurement studies as proof-of-concept for using these sources to measure various development outcomes, but not many IEs use these sources. This also shows the possibility of potentially using these sources in future IEs. Similar caveats on methodological and ethical challenges discussed above in relation to CRD data will apply.

**East Africa is well-represented, but not the rest of Africa.** The geographical distribution of measurement studies and IEs show that the studies are evenly spread across the continents. However, Ethiopia, Kenya, Rwanda, Tanzania and Uganda are well-represented in terms of number of measurement studies and IEs, but the only non-East African country that seem to have well-represented on the map is Nigeria. There are very few studies in the rest of Africa. This gap is particularly serious given Africa’s data challenges (Serajuddin et al., 2015).

\textsuperscript{10} https://sedac.ciesin.columbia.edu/
Big data holds great potential for conducting IEs in fragile contexts, including during conflicts, humanitarian crises, epidemics and natural disasters. Conducting rigorous evaluation in fragile contexts (such as natural disasters, disease outbreaks and other crisis contexts) can be costly, risky to the beneficiaries and the evaluators, and in some cases outright infeasible. We identified 73 measurement studies, 17 IEs and one SR in such fragile contexts. Measurement studies are spread evenly across conflict or humanitarian crisis, disease outbreaks or epidemics, natural disasters and difficult-to-reach terrain. However, the IEs are concentrated around conflict and difficult terrain. The number of measurement studies indicate the potential for more IEs in fragile contexts.

There are potential sectors and themes where SRs will be useful. Though the number of IEs are fewer, the map highlights a few potentials thematic areas where SRs will help answer key questions on policy and research methods. For example, there is a concentration of IEs using satellite data, referred to as geospatial IEs, but we know little about how satellite data can help evaluate development programmes better, where it can add value, what type of interventions could be better evaluated and the technical challenges involved in using satellite data. An SR of all geospatial IEs that have used satellite data across the sectors will help understand the potential and challenges in using satellite data for IEs.

Similarly, there is a concentration of IEs on environmental sustainability and within that climate action and forest management. Though there a few SRs and evidence gap maps on forest management (Pelletier et al., 2016; Puri et al., 2016), a new review with reference to innovative, new data sources used in rigorously evaluating forest cover and the advantages and challenges thereof may be useful.

4.2 Analytical, ethical and practical challenges

Satellite data also presents misclassification problems. Researchers, however, point to several technical challenges in using satellite images that may provide misleading conclusions. For example, Jain (2020) argues for the need for ground validation of satellite data as sometime the images could be misclassified (e.g. flood irrigation may be classified as flooded area); often this misclassification is systematic (i.e. forest cover is almost always misclassified as agriculture, which will bias the study results). This can be rectified with a field visit. Further, there could be differences in the data coming from different satellites and from the same satellite constellation but using different sensors (e.g. Normalised Difference Vegetation Index varies for different satellite sources and for same satellites across different versions, such as Landsat 5 and Landsat 8). Another example of what qualitative field visits could contribute to improving the interpretability of satellite data is the indicator ‘quality of roof construction’ as a proxy for economic development. Straw roofs from satellite images are generally classified as low-quality; zinc and other hard roofs are classified as a sign of development. However, several factors may bias this classification: in hot climates, straw roofs may be preferred for their ability to keep cool compared to other hard roofs; poor families may have received donations of high-quality roofs; or lack of tenured security may discourage people invest in immovable roof-tops despite increase in income.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

**Data quality and transparency is paramount.** The map also points to the need to set standards for better reporting, as about 87% of the measurement studies did not report on data quality issues, representativeness, construct validity and generalisability. This would lead to questioning the internal and external validity of the findings. There is also a need to set standards for data transparency, taking into consideration the challenges in sharing proprietary data, data storing and the capacity of the Dataverse (see Box 1 for more details on data transparency).

**Box 1: Transparency in data analysis, use and sharing**

In recent times, the use of big data for in-depth comprehension of developments in the social sector has gained traction. Transparency refers to publishing all relevant materials, including data and code, used in a research study in the public domain for independent verification. Transparency in research encompasses a number of elements that are no different while using big data as opposed to traditional data sources. Certain challenges that might arise are discussed below.

1. **De-identification**: De-identification is related to preserving the identity of the study subject before it is made available for any sort of analysis. A few concepts used in de-identification of big data are K-anonymity, L-diversity and T-closeness. A dataset is said to have K-anonymity if each person in the dataset cannot be identified by information of the other K-1 individuals in the dataset. In contrast, L-diversity is a group-based anonymisation technique that reduces the granularity in the dataset. T-closeness is a refinement of L-diversity and is used to decrease granularity over and above L-diversity. However, there are several examples of combining different source to re-identify the respondents in the data set (Archie et al., 2018).

2. **Scale and storage**: In today's world, storage of a high volume of data is not a challenge owing to developments in cloud computing. A fairly new system that provides solution to the scalability and storage issue is storage virtualisation. In simple terms, this is a network of storage devices that are combined to create a single storage space. A few ways of safeguarding the data storage is encrypting all processes and the usage of hybrid clouds. Data repositories such as Harvard dataverse and figshare have limited capacity to handle big data and are often restricted by the size of the data uploaded. Cloud storages such AWS and other similar such storage will be a better option.

**Ethical concerns are substantial.** Ethical challenges such as consent, data privacy, data security and unintended exclusion are well documented in the literature (Lokanathan et al., 2017; York and Bamberger, 2020). A brief analysis of the studies on the map shows that very few studies report on any of these ethical challenges. However, the challenges are different for different sources of big data. For example, satellite data that involves little human interaction may not need an IRB review but most other big data source that use human-generated data without explicit consent for secondary use should be reviewed by IRBs. We also recommend more mixed-method big data evaluations to mitigate the potential disconnect between development stakeholders and big data researchers. Any mixed-method research needs to be reviewed by IRBs.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

The map shows that most IEs have done well on reporting data quality issues but not on ethical issues. Since big data involves ethical issues (such as consent for secondary data use and unintended exclusion) that are new to conventional ethical standards, there is a need to update the current ethical standards practice to include big data use as well.

Big data may be growing in use and popularity, but the need for independent auxiliary data for ‘ground-truthing’ remain. Many sources of big data are partial in terms of coverage and prone to biases that are difficult to measure, control and correct for in the absence of secondary data. Despite growing awareness and acknowledgement of its limitations, the household sample survey remains the dominant source of development policymaking. Big data often require survey data as ‘ground-truth’ data to validate the findings. Demographic Health Surveys and Living Standards Measurement Studies are the two main surveys used in ground-truthing. There is considerable scope for merging the income and expenditure surveys, and food surveys conducted in several developing countries with big data to assess food shortages, poverty hotspots, etc.

Some capacity constraints are acute. Development organisations need to build staff capacity in order to use big data as a strategic asset (Perera Gomez and Lokanathan, 2017). They need to build multidisciplinary teams consisting of data experts and subject matter professionals, and also compete with the private sector to recruit the staff. Other major costs involve scaling up the technical infrastructure to enable data storage and processing on a large scale and data accessibility costs. The latter can be more difficult to predict considering that big data sources that are currently public may involve licensing in the future. Besides, ensuring the sustainability of data can be a cause of concern. As suggested by Hammer et al. (2017), as most of the big data is produced as a by-product by the private sector, continuity of data provision cannot be guaranteed in this age of evolving technology and market conditions. These concerns will call the wisdom of committing resources upfront to build capacity into question.

Need for better coordination between data scientists and evaluators. Big data analysts and evaluators use different framework and analytical tools. In particular, the big data measurement studies look for hidden patterns in the data with little support from theory and aim at prediction rather than causality (York and Bamberger, 2020). Further, the expertise needed to analyse big data remains largely localised and siloed. Outside of a small and highly specialised group of data scientists, there is uncertainty about how best to carry out large-scale big data analysis. The degree of technical specialisation combined with strict access restrictions to many types of big data has hindered big data applications in development evaluation. Hence, there is a need to promote interaction between development evaluators and data scientists for better cross-learning and adoption of big data in measuring and evaluating development outcomes.

The cost of collecting, analysing, storing and reporting big data is largely unknown. There is very little publicly available information on the cost of collecting, analysing and reporting big data. Blumenstock, Cadamuro and On (2015) reported that the phone survey for ground-truthing the CRD data costed USD 12,000 and took four weeks to administer. This is,
however, only the variable cost of data collection in this study. There are multiple hidden costs such as staff costs and the cost of the necessary computing infrastructure (including storage); in addition, the opportunity cost of time involved in developing partnerships with data providers in some cases is not known. BenYishay et al. (2017) report that the cost of geospatial IE is around USD 150,000. One of the 3ie funded studies using satellite data is reported to have spent USD 3,300 on data collection, which is about 1% of the total study budget, but have spent USD 103,864 on data analysis and reporting (about 32% of the total cost\textsuperscript{11}). Similarly, another 3ie funded study that used in situ fixed sensors reported spending USD 6,152 on data collection or acquisition (11%) and USD 54,444 on staff costs for analysis and reporting (55%). However, studies that have combined satellite data with household survey have reported higher costs of data collection (USD 171,582; 43%) and analysis and reporting (USD 109,495; 27%). This is, however, roughly comparable to the data collection cost of an average 3ie funded multi-year, multi-round survey IE, which costs about USD 176,000 (Puri and Rathinam, 2019).

\textsuperscript{11} It should be noted that the cost discussed here includes only the variable cost of data collection and the staff time, but may not include the cost of fixed infrastructure and equipment.
5. **Using big data in IEs: potentials and challenges**

Rigorous IEs require a valid counterfactual. Randomising programme placement ensures pre-programme comparability of the treatment and control groups in most cases and quasi-experimental studies employ statistical procedures to identify a valid comparison group. In either case, evaluators collect require a vast array of data on the outcomes, covariates and other contextual factors. There is almost always a trade-off between collecting a complete array of necessary data and cost-effectiveness, and in a few cases, it may not be feasible to collect some of the covariates and confounders.

Big data, with the help of improved ML techniques and analytical capacity, can now be manipulated to evaluate development outcomes. The potential advantages for big data are, to date, most discernible in contexts where the immediacy, scale and/or reach of data is highly prized and alternative sources of data are absent or inadequate to the task. The ability to ‘zoom in’ on particular zones of interest, and to produce estimates for small areas, is an oft-cited advantage of many types of big data (e.g. satellite and building footprint data, mobile phone CRD and signalling data and app-based location data) and one with particular relevance to evaluative contexts and SDG-related urbanisation, climate change and infrastructure. This holds particular promise for settings where census data renders small area estimation methods unsuitable. Big data has also been shown to be particularly advantageous for the analysis of disaster-induced displacement and disease outbreaks. In each of these cases, the advantage of big data is that it can support rapid appraisal and introduction or adjustment of policies/interventions on the basis of near real-time information.

In this section, we highlight a few examples from the map to show the steps involved in collecting, processing and using satellite and CRD data for measuring development outcomes. We draw on recent projects from 3ie and Flowminder to illustrate the processes.

5.1 **Using satellite data in IEs**

In a 3ie funded evaluation conducted by the Institute for Financial Management and Research, Pande and Sudarshan (2019) evaluated the recent environmental clearance (EC) reforms in India. Before the 2006 reform, mines of area over 25 hectares were required to hold a public hearing before approval. The new EC reform required mines of area between 5 and 25 hectares to hold a public hearing as well. This study exploits this historical discontinuity in clearance requirements to evaluate the impact of public hearings on mines’ environmental compliance. Apart from rigorously evaluating the EC process in India, this study also provides a proof-of-concept for the use of remote sensing data and other publicly available data to monitor mines’ environmental compliance. Using satellite data to assess the impact of EC process requires data on the timing of the intervention, the geographical scope of the intervention (i.e. the individual mines in this case), the outcome of interest (such as air pollution, land cover and water quality for the corresponding intervention) and control areas for the years before and after the intervention.
The following were the key steps involved in the big data IE.

**Step 1:** the researchers used web scraping techniques to collect information on the mines from their EC application for the years from 2006 to 2016, available online in a database published by Ministry of Environment, Forests and Climate Change. These are all mostly scanned PDF documents. The researchers scrapped for the information on project name and location (the tehsil and village where the mine is located); the dates of key EC stages of submission, review and approval; and mine characteristics such as minerals mined, mine production capacity and size of the mine. They also scrapped the clearing letters available in the same database for cross checking the data. They collected information about all 934 relevant mines and used 134 of them in their regression discontinuity analysis. Finally, the researchers hired ML Infomap, a local company, to geocode all the mines identified.

**Step 2:** Satellite data on various key environmental outcomes such as air pollution, land cover and water quality were collected from different sources.

- The researchers used the data provided by Dalhousie University on the fine particulate matter concentration as a proxy for air pollution. This database contains average annual particulate matter concentration for every one kilometre cell for the study period;
- They have used the Enhanced Vegetation Index (EVI) data from NASA's MODIS satellite to measure deforestation around the mining areas. EVI is available at a resolution of 250 metres for the entire globe and the researchers calculated annual maximum, median and mean EVI at mine sites. EVI data was used to measure the extend of and the date of beginning of deforestation (i.e. structural break in the time series) for each mine; and
- Data on water quality from the site monitor nearest to each mines was collected from the Central Pollution Control Board's ENVIS database. They used Biological Oxygen Demand, a measure of organics pollution, as a proxy for water pollution.

**Step 3:** The researchers then linked the geocoded mine sites to the corresponding cells of environmental outcomes. Of the total 934 mines in their database, they could link 889 of 1km cells of EVI data and 882 of 250 metre cells with corresponding geocoded mine sites. They could also link 538 site monitors to the mines.

**Step 4:** The new EC reform required the mines of between 5 and 25 hectares in area to hold a public hearing that had not been considered big enough to hold public hearings during the previous regime (i.e. only the mines of above 25 hectares in area were required to hold the hearings). This study exploited the discontinuity around the 25 hectare mark and compared

---

12 There are several ways to collect the required satellite data. For a few key variables such as night lights, air pollution, land cover and water quality, elevation, slope, distance from certain services or infrastructure, geocoded data for various granularity and frequency is readily available in several databases such as Aiddata DataQuary, SEDAC, etc. This study has utilised data from differences sources that provide readily useable data. Alternatively, the researchers use ML techniques to analyse satellite images and predict development outcomes (Jean et al., 2016). Another useful source of areal images come from custom built drones that can provide very high resolution data for the exact spatial and temporal frequency (Pellegrini, 2019).
the mines marginally above 25 hectares with the ones marginally below 25 hectares. The final sample included 134 mines, of which 68 were treatment mines (less than 25 hectares) and 66 were control (greater than 25 hectares). Using data before and after the EC applications, they estimated a difference-in-difference model.

This study, utilising web scraping to collect data on project characteristics and various sources of satellite data for measuring the outcomes of interest, is an excellent example of innovative data collection methods in a sector where the evidence base is very small (Rathinam et al., 2019b).

5.2 Using CRD analytics to inform disaster management

In this section, we briefly outline the process for undertaking CRD analytics to measure, characterise and predict population displacement and returnee/resettlement patterns in post-disaster settings. While applications of CRD data analytics to date have lacked an evaluative component, their potential in this regard is evident. We draw on a recent project at Flowminder, which revisited three sudden-onset disaster events to investigate drivers of displacements (individual and contextual) and the feasibility of predicting displacement locations from CRD data and data on disaster intensity and damage, on population density and on the humanitarian response. The three events were the 2010 earthquake in Haiti, the 2015 Gorkha earthquake in Nepal and the 2016 Hurricane Matthew in Haiti.

Flowminder has long-established partnerships with the major mobile phone network operators in Haiti and Nepal. Historically, data access has been a major barrier to the scale-up of CRD analytics for humanitarian and development applications. Mobile network operators (MNOs) are justifiably hesitant to authorise third-party access, given the need to safeguard subscribers’ personal data.\(^\text{13}\)

Prior to analysis, MNO data underwent a long series of cleaning and pre-processing steps as part of quality assurance and to support the generation of standardised metrics. A first stage of analysis was undertaken to structure the data in a usable format and to detect data anomalies. Once data was cleaned, quality assured and converted into an analysable format, a number of preliminary processing steps were undertaken, including:

- Clustering of cell tower locations
- Assessment of each subscriber’s phone usage behaviours (number of events, frequency and regularity)

\(^{13}\) Flowminder’s original ‘data partnership’ model developed lasting collaborations with individual MNOs. The priority was countries where substantial potential gains were available from novel, digital data given the existing data landscape. Lengthy negotiations with MNOs followed, often spanning many years and consuming extensive organisational resources, with no guarantee of a successful outcome. When this model ‘worked’, it led to strong and sustained partnerships with MNOs. In pursuit of impact at scale, Flowminder Foundation supplemented its original ‘data partnership’ model with a toolkit-based approach designed to break down silos between data and methods, in effect negating the need for Flowminder to access MNO data by transferring methods expertise to MNOs themselves via the ‘Flowkit’ suite of software.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

- Determination of a pre-disaster ‘home’ location.

**MNO Data: Pre-processing steps**

Here are some commonly occurring issues in MNO data. Once identified, corrections and/or accommodations can be performed prior to and/or during the preliminary processing and analysis phases.

1. **Standard data quality issues applicable to MNO data:**
   - Item missing data (incomplete data records i.e. missing fields)
   - Invalid entries for fields
   - Duplicate records
   - Interrupted data series’ (e.g. no data for a particular time period)
   - Inconsistent values (either in format, or definition) for keys that are used to join multiple datasets together
   - Inconsistent entries for the ‘same’ value (e.g. different spellings of the same place name)

2. **Issues specific to MNO datasets, CRD:**
   - Inconsistencies or errors in method used for ‘hashing’ (a form of pseudonymisation) subscribers’ IDs
   - Inconsistent ‘hashing’ of sender and recipient IDs for communication events (e.g. standard SMS or phone calls)

3. **Issues specific to MNO datasets: cell location and coverage maps:**
   - Cell locations occur outside national borders
   - Updated cell locations are not consistent with previous cell locations
   - Small deviations in updated cell locations, possibly due to inexact GPS measurements
   - Data, projection and coordinate system information are often missing in coverage datasets
   - Inconsistent output formats
   - For best server or cell-in-isolation maps, polygons should be labelled in a manner consistent with cell table and/or CRD dataset

4. **Common data anomalies, indicative of a network issue or a sudden change in subscribers’ behaviour due to an event or ‘shock’:**
   - Individual cell towers have significantly more/less traffic than normal
   - Overall network traffic is significantly higher/lower than expected
   - Traffic from a particular region is significantly higher/lower than expected

The processing steps undertaken to discern at individual level disaster-induced displacements from pseudonymised, time series CRD data are presented below in Figure 13.
The analysis disclosed striking commonalities in IDP return/resettlement rates, with the fraction of IDPs who remain displaced exhibiting a common rate of decay across all three post-disaster settings studied.

In a further step, the team developed new mobility and social network metrics to permit analysis of the relationships between contextual and individual variables and displacement duration, distance and trajectories, controlling for the severity of impacts and humanitarian response. The results suggest that the dispersal of an individual's social contacts and travel history pre-disaster are highly predictive of their post-disaster displacement trajectories. Individuals with localised travel patterns and social contacts were more likely to be displaced in the vicinity of their usual residence compared with those with more dispersed travel patterns and social contacts. A majority of IDPs remained within a 10 kilometre radius of their usual place of residence. Across the three disasters, 60%–70% of long-distance displacements (in excess of 100 kilometres) involved travel to a familiar location and/or proximate to one or more contacts discernible in the pre-disaster CRD data. This pattern holds controlling for the severity of impacts at local area level and is consistent across all three disasters.

Results were validated via comparisons with reports retrospectively quantifying population displacements produced by the International Organisation for Migration, as well as with reference to data on the intensity of each disaster's impact on affected areas. The results indicate that CRD data analysis can be used to predict the estimated number and spatial distribution of IDPs at different time points based on initial estimates of the number of persons displaced in the immediate wake of a disaster, as well as to predict recovery/resettlement timelines. This has important implications for post-disaster
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

humanitarian response and resettlement efforts. The same methods can support disaster resilience assessments and planning and provide a means to compare recovery and resettlement rates across different disaster events.
6. Limitations of this study and potential next steps

This map covers large thematic areas and outcomes corresponding to SDGs. Given the wide scope of the outcomes, the evidence is sparse and bunched around a few themes. The thematic gaps here may not be read as actual gaps, but these areas may be not readily relevant to using big data. This map rather shows what evidence or proof-of-concepts are available to measure and evaluate development outcomes using big data.

This map followed a systematic process of searching, screening and coding of studies based on a predefined set of criteria in the study protocol developed with inputs from key stakeholders. However, despite best efforts in searching and screening the studies, given the wide scope of big data sources and their application across all developmental themes and the pace at which the literature is growing, it is possible that some relevant studies (especially measurement studies) could have been missed out. It was beyond the scope of the study to provide a critical quality appraisal of the IEs or the measurement studies, given the large number of studies included on the map; nor did the report look at the details of ML methods used in the included studies.

Given the wide scope of development applications, it was not possible to code the studies for all sub-classifications. Though the sub-maps (especially for economic development and livelihoods, health and well-being and urban development) provided coding at level 2 indicators, it was not possible to provide granular analysis of development themes corresponding to SDG indicators at level 3. Future systematic maps should aim to produce more granular classifications on the use of big data at the indicator level.

Some studies have used ML techniques for treatment effect heterogeneity in RCTs (Chernozhukov et al., 2019). However, it was beyond the scope of this report to include the role of big data analytical methods in conventional IE designs such as RCT and other quasi-experimental designs. This is a nascent but growing body literature and could be considered for inclusion in future maps.

Several studies suggest that the key advantages of big data sources (especially satellite data) are their long-term availability which will help evaluate the long-term impact of development interventions. The possibility of collecting a vast array of information on several contextual factors using big data can help evaluate complex interventions (Bamberger, 2016). However, this map did not code the studies for long-term impact or for complex interventions. Future maps may code and analyse the role of big data in measuring long-term impact and in evaluating complex interventions.
6.1 Next steps

The systematic map shows that both IEs and measurement studies have dramatically increased in the past five years and are continuing to grow in number. Given the potential for faster growth in the availability and computational capacity, it is very likely that the number of studies will grow faster than we have witnessed over the past five years. Hence, we recommend that this map be updated within the next two years.

The fact that more than 80% of the included studies are peer-reviewed shows the growing number of journals interested in big data application in international development. It will be useful to include a more exhaustive grey literature search to identify the full extent of the literature.

This map shows the potential for big data to measure and evaluate various development themes. However, most of these studies are supported by universities and specialist organisations and conducted by researchers associated with these organisations. Widely disseminating the findings of the map among development researchers, evaluators, practitioners and donors will help promote the adoption of big data measures in future IEs.
7. Conclusions

Big data has great potential to help address questions of relevance to international development, including for evaluating the effects of interventions. This systematic map compiles IEs, SRs and measurement studies that incorporate big data to highlight how this innovative, new data source is being used to evaluate development outcomes and (more importantly) where there is more potential to use big data in the future evaluations. We found 437 studies, of which 48 are IEs, 381 are measurement studies and 8 are SRs. Roughly half the studies are from Asia and another 30% are from Africa; about 70% are from L&MICs. Of the 48 IEs, 8 are RCTs and the remaining are quasi-experimental studies.

Our results highlight considerable potential for using big data for measuring various development outcomes across SDG themes, but big data is more relevant to environmental sustainability, economic development and livelihoods, health and well-being and urban development. This map also highlights that big data can contribute to the evidence base in development sectors where evaluations are not generally feasible due to a lack of data, particularly due to fragile contexts.

One of the key ‘absolute gaps’ the map has identified is that the number of IEs is lower in comparison to measurement studies. Given the fast-growing availability of big data and improving computation capacity, there is great potential for using big data in future IEs. This may not, however, be straightforward as there are several analytical, ethical and logistical challenges that may hinder the use of big data in evaluations. The development community that helps set standards and best practices and development stakeholders (including donors who facilitate rigorous evaluations and learning) have a strong role to play in facilitating this process. The report highlights the need for setting standards for better reporting on data quality issues, representativeness, construct validity and generalisability, as well as the need for data transparency and sharing. The report also calls for facilitating better interaction between big data analysts, remote sensing scientists and evaluators.

One of the key findings of the report is that satellite and sensor data are the most used data sources for both measurements studies and IEs. There are several sources of pre-processed satellite data that could be used directly in evaluations without the evaluators having to process them using complex ML models themselves. Satellite data seems to be particularly useful in the context where the development interventions and the outcomes studied have spatial dimension economic activity at the local level, urban development, forest cover, land productivity and distribution of the population, or where the outcome and other covariates are measured on a spatial scale (i.e. villages, counties, districts, plots or protected areas). CRD data, on the other hand, despite being used widely in measurement studies, is not yet used in IEs. The data deficiency in international development is partly due to fragile contexts such as diseases spread, violence, natural calamities and difficult terrain. This map highlights the potential of big data in fragile contexts: one-quarter of the studies were conducted in such a context.

For evaluators and researchers, the report calls for better reporting on data quality, ethics and transparency. There is also an absolute gap in using mixed methods jointly with big data.
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

and cost-effectiveness. For the donors, this report calls for more efforts on setting up best practices and ethical standards and in facilitating more interaction among remote sensing scientists, big data analysts and development evaluators.
8. References


Blazquez D and Domenech J (2018). Big Data sources and methods for social and economic analyses. Technological forecasting & social change, 130 99-113


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


https://doi.org/10.23846/TW8IE92


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


Vaitla, B. (2014). The Landscape of Big Data for Development: Key Actors and Major Research Themes. UN Foundation/Data2X.


York, P. and Bamberger, M. (2020). Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology, the Rockefeller Foundation. Available at: cedilprogramme.org50
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

The OECD definition of fragile states

OECD defines fragility as the combination of exposure to risk and insufficient coping capacity of the state, system and/or communities to manage, absorb or mitigate those risks. Fragility can lead to negative outcomes including violence, the breakdown of institutions, displacement, humanitarian crises or other emergencies. The OECD fragility framework considers not only current exposure to negative events such as natural disasters and armed conflict but also capacity to cope with likely future negative events.

The new framework considers five dimensions of fragility.

- **Economic fragility** is vulnerability to risks stemming from weaknesses in economic foundations and human capital including macroeconomic shocks, unequal growth and high youth unemployment;

- **Environmental fragility** is vulnerability to environmental, climatic and health risks that affect citizens’ lives and livelihoods. Risk factors can be external or internal, including exposure to natural disasters; air, water and sanitation quality; prevalence of infectious disease; number of uprooted people; and vulnerability of household livelihoods;

- **Political fragility** is vulnerability to risks inherent in political processes, events or decisions; political inclusiveness (including of elites); and transparency, corruption and society’s ability to accommodate change and avoid repression. Risk factors include regime persistence, state-sponsored violence or political terror and levels of corruption;

- **Security fragility** is the vulnerability of overall security to violence and crime, including both political and social violence. Risks are measured by the homicide rate, level of violent organised crime, number of deaths from non-state actors or terrorism, number of battle deaths from conventional warfare, and levels of domestic violence; and

- **Societal fragility** is vulnerability to risks affecting societal cohesion that stem from both vertical and horizontal inequalities, including inequality among culturally defined or constructed groups and social cleavages. Risk indicators include income inequalities (vertical) and social inequalities related to gender, growth in urbanisation and numbers of displaced people.

In the past few years, OECD has moved away from the ‘fragile states list’ and towards measuring each of those five dimensions on a spectrum of intensity for 58 fragile contexts. This comes as a part of their effort to move towards a universal concept of fragility, recognising that it affects not only developing countries but all countries to some extent.

**Table 6: Fragile contexts provided in the OECD fragility framework 2018 (decreasing order of severity)**

<table>
<thead>
<tr>
<th>Economic</th>
<th>Environmental</th>
<th>Political</th>
<th>Security</th>
<th>Societal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central African Republic</td>
<td>Somalia</td>
<td>Eritrea</td>
<td>Syria</td>
<td>South Sudan</td>
</tr>
<tr>
<td>Country</td>
<td>Country</td>
<td>Country</td>
<td>Country</td>
<td>Country</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>South Sudan</td>
<td>Central African</td>
<td>Sudan</td>
<td>Libya</td>
<td>Yemen</td>
</tr>
<tr>
<td>Liberia</td>
<td>South Sudan</td>
<td>DPRK</td>
<td>Yemen</td>
<td>Syria</td>
</tr>
<tr>
<td>Somalia</td>
<td>Chad</td>
<td>Yemen</td>
<td>Somalia</td>
<td>Egypt</td>
</tr>
<tr>
<td>Solomon Islands</td>
<td>DRC</td>
<td>Syria</td>
<td>Afghanistan</td>
<td>Somalia</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>Burundi</td>
<td>Somalia</td>
<td>South Sudan</td>
<td>Sudan</td>
</tr>
<tr>
<td>Niger</td>
<td>Mozambique</td>
<td>South Sudan</td>
<td>Iraq</td>
<td>Burundi</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>Guinea-Bissau</td>
<td>Burundi</td>
<td>Sudan</td>
<td>Eritrea</td>
</tr>
<tr>
<td>Comoros</td>
<td>Niger</td>
<td>Chad</td>
<td>Central African</td>
<td>DPRK</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>Liberia</td>
<td>Congo</td>
<td>Nigeria</td>
<td>DRC</td>
</tr>
<tr>
<td>Mozambique</td>
<td>Burkina Faso</td>
<td>Gambia</td>
<td>Mali</td>
<td>Congo</td>
</tr>
<tr>
<td>Haiti</td>
<td>Guinea</td>
<td>Ethiopia</td>
<td>Pakistan</td>
<td>Chad</td>
</tr>
<tr>
<td>Mali</td>
<td>Sierra Leone</td>
<td>Guinea</td>
<td>Chad</td>
<td>Equatorial Guinea</td>
</tr>
<tr>
<td>Congo</td>
<td>Mali</td>
<td>DRC</td>
<td>DRC</td>
<td>Pakistan</td>
</tr>
<tr>
<td>Libya</td>
<td>Malawi</td>
<td>Mauritania</td>
<td>West Bank &amp; Gaza</td>
<td>Iran</td>
</tr>
<tr>
<td>Syria</td>
<td>Zambia</td>
<td>Bangladesh</td>
<td>Cameroon</td>
<td>Central African Republic</td>
</tr>
<tr>
<td>Yemen</td>
<td>Cameroon</td>
<td>Angola</td>
<td>Egypt</td>
<td>Afghanistan</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>Eswatini</td>
<td>Djibouti</td>
<td>Niger</td>
<td>Zimbabwe</td>
</tr>
<tr>
<td>Djibouti</td>
<td>Côte d’Ivoire</td>
<td>Venezuela</td>
<td>Haiti</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>Gambia</td>
<td>Eritrea</td>
<td>Iran</td>
<td>Ethiopia</td>
<td>Kenya</td>
</tr>
<tr>
<td>Mauritania</td>
<td>Ethiopia</td>
<td>West Bank &amp; Gaza</td>
<td>Kenya</td>
<td>Eswatini</td>
</tr>
<tr>
<td>Iraq</td>
<td>Afghanistan</td>
<td>Zimbabwe</td>
<td>Myanmar</td>
<td>Mauritania</td>
</tr>
<tr>
<td>Honduras</td>
<td>DPRK</td>
<td>Madagascar</td>
<td>Nepal</td>
<td>West Bank &amp; Gaza</td>
</tr>
<tr>
<td>Timor-Leste</td>
<td>Yemen</td>
<td>Afghanistan</td>
<td>Burkina Faso</td>
<td>Angola</td>
</tr>
<tr>
<td>Eswatini</td>
<td>Haiti</td>
<td>Central African</td>
<td>Congo</td>
<td>Iraq</td>
</tr>
<tr>
<td>Pakistan</td>
<td>Syria</td>
<td>Libya</td>
<td>Eritrea</td>
<td>Guatemala</td>
</tr>
<tr>
<td>DRC</td>
<td>Tanzania</td>
<td>Egypt</td>
<td>Burundi</td>
<td>Gambia</td>
</tr>
<tr>
<td>Eritrea</td>
<td>Zimbabwe</td>
<td>Guinea-Bissau</td>
<td>Mauritania</td>
<td>Cameroon</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>Madagascar</td>
<td>Iraq</td>
<td>Iran</td>
<td>Haiti</td>
</tr>
<tr>
<td>Burundi</td>
<td>Congo</td>
<td>Sierra Leone</td>
<td>Comoros</td>
<td>Uganda</td>
</tr>
<tr>
<td>Country</td>
<td>Country</td>
<td>Country</td>
<td>Country</td>
<td>Country</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>Rwanda</td>
<td>Mali</td>
<td>Bangladesh</td>
<td>Rwanda</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Myanmar</td>
<td>Haiti</td>
<td>Venezuela</td>
<td>Lao PDR</td>
</tr>
<tr>
<td>Chad</td>
<td>Uganda</td>
<td>Côte d'Ivoire</td>
<td>Tanzania</td>
<td>Myanmar</td>
</tr>
<tr>
<td>Guinea</td>
<td>Pakistan</td>
<td>Mozambique</td>
<td>Uganda</td>
<td>Venezuela</td>
</tr>
<tr>
<td>Malawi</td>
<td>Comoros</td>
<td>Myanmar</td>
<td>Mozambique</td>
<td>Tajikistan</td>
</tr>
<tr>
<td>Madagascar</td>
<td>Bangladesh</td>
<td>Rwanda</td>
<td>Guatemala</td>
<td>Bangladesh</td>
</tr>
<tr>
<td>Rwanda</td>
<td>Tajikistan</td>
<td>Liberia</td>
<td>Côte d'Ivoire</td>
<td>Guinea</td>
</tr>
<tr>
<td>DPRK</td>
<td>Lao PDR</td>
<td>Nigeria</td>
<td>Sierra Leone</td>
<td>Libya</td>
</tr>
<tr>
<td>Zambia</td>
<td>Djibouti</td>
<td>Pakistan</td>
<td>Liberia</td>
<td>Djibouti</td>
</tr>
<tr>
<td>Uganda</td>
<td>Mauritania</td>
<td>Burkina Faso</td>
<td>Honduras</td>
<td>Guinea-Bissau</td>
</tr>
<tr>
<td>Tanzania</td>
<td>Kenya</td>
<td>Uganda</td>
<td>Guinea</td>
<td>Tanzania</td>
</tr>
<tr>
<td>Papua New Guinea</td>
<td>Papua New Guinea</td>
<td>Kenya</td>
<td>Guinea-Bissau</td>
<td>Honduras</td>
</tr>
<tr>
<td>Nepal</td>
<td>Angola</td>
<td>Honduras</td>
<td>Angola</td>
<td>Mozambique</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>Iraq</td>
<td>Niger</td>
<td>Gambia</td>
<td>Côte d'Ivoire</td>
</tr>
<tr>
<td>Lao PDR</td>
<td>Gambia</td>
<td>Tajikistan</td>
<td>Tajikistan</td>
<td>Madagascar</td>
</tr>
<tr>
<td>Myanmar</td>
<td>Sudan</td>
<td>Equatorial Guinea</td>
<td>Equatorial Guinea</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Cameroon</td>
<td>Nepal</td>
<td>Cameroon</td>
<td>Lao PDR</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Angola</td>
<td>Timor-Leste</td>
<td>Eswatini</td>
<td>Solomon Islands</td>
<td>Zambia</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>Nigeria</td>
<td>Lao PDR</td>
<td>Madagascar</td>
<td>Mali</td>
</tr>
<tr>
<td>Guatemala</td>
<td>Guatemala</td>
<td>Guatemala</td>
<td>Papua New Guinea</td>
<td>Sierra Leone</td>
</tr>
<tr>
<td>Côte d'Ivoire</td>
<td>Solomon Islands</td>
<td>Zambia</td>
<td>Zimbabwe</td>
<td>Nepal</td>
</tr>
<tr>
<td>Kenya</td>
<td>Venezuela</td>
<td>Timor-Leste</td>
<td>Zambia</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>Venezuela</td>
<td>Honduras</td>
<td>Nepal</td>
<td>Timor-Leste</td>
<td>Malawi</td>
</tr>
<tr>
<td>Nigeria</td>
<td>West Bank &amp; Gaza</td>
<td>Tanzania</td>
<td>Malawi</td>
<td>Niger</td>
</tr>
<tr>
<td>Sudan</td>
<td>Iran</td>
<td>Papua New Guinea</td>
<td>Eswatini</td>
<td>Timor-Leste</td>
</tr>
<tr>
<td>Iran</td>
<td>Egypt</td>
<td>Comoros</td>
<td>Djibouti</td>
<td>Liberia</td>
</tr>
<tr>
<td>West Bank &amp; Gaza</td>
<td>Libya</td>
<td>Malawi</td>
<td>Rwanda</td>
<td>Solomon Islands</td>
</tr>
</tbody>
</table>
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

<table>
<thead>
<tr>
<th>Egypt</th>
<th>Equatorial Guinea</th>
<th>Solomon Islands</th>
<th>DPRK</th>
<th>Comoros</th>
</tr>
</thead>
</table>

cedilprogramme.org55
### Table 7: Outcome categories and definitions

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Sub-categories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Economic development and livelihoods (SDG 1 and 8)</strong></td>
<td>End poverty in all its forms everywhere Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all</td>
<td>Poverty mapping/measurement Measuring wealth/GDP Access to basic services Social protection, including financial inclusion Vulnerability reduction Economic growth and productivity Full and productive employment, including youth unemployment Eradicate forced and underage labour, protect labour rights and promote safe workspaces Promote sustainable tourism Strengthen capacity of domestic financial institutions</td>
</tr>
<tr>
<td><strong>Health and well-being (SDG 3)</strong></td>
<td>Ensuring healthy lives and promoting well-being for all ages</td>
<td>Reduce mortality End epidemics of communicable diseases Strengthen prevention and treatment of substance abuse Decrease fatalities due to road accidents Universal access to sexual and reproductive healthcare services Achieve universal health coverage Reduce fatalities due to pollution</td>
</tr>
<tr>
<td><strong>Governance and human rights (SDG 5, 10 and 16)</strong></td>
<td>Achieving gender equality and empower all women and girls Reduce inequality within and among countries Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels</td>
<td></td>
</tr>
<tr>
<td><strong>Urban development (SDG 11)</strong></td>
<td>Make cities and human settlements inclusive, safe, resilient and sustainable</td>
<td>Access to affordable housing Access to affordable transport systems Enhance inclusive and sustainable urbanisation Strengthen efforts to protect cultural and natural heritage Better disaster management Reduce environmental impact of cities Universal access to green public spaces</td>
</tr>
</tbody>
</table>
## Environmental sustainability (SDG 12, 13, 14 and 15)

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure sustainable consumption and production patterns</td>
</tr>
<tr>
<td>Take urgent action to combat climate change and its impacts</td>
</tr>
<tr>
<td>Conserve and sustainable use the oceans, seas and marine resources for sustainable development</td>
</tr>
<tr>
<td>Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation and halt biodiversity loss</td>
</tr>
</tbody>
</table>
Appendix 3  **Search strategy and the databases searched**

We developed a systematic search strategy in consultation with an information specialist after finalising the protocol.

We searched the following general databases:

- **CAB Abst:** [www.cabi.org/publishing-products/online-information-resources/cab-abstracts/](http://www.cabi.org/publishing-products/online-information-resources/cab-abstracts/)
- **Econlit (Ovid):** [www.ovid.com/site/catalog/databases/52.jsp](http://www.ovid.com/site/catalog/databases/52.jsp)
- **Ebsco Discovery:** [https://www.ebscohost.com/discovery](https://www.ebscohost.com/discovery)
- **Scopus:** [https://www.scopus.com/](https://www.scopus.com/)
- **Social Sciences Citation Index (SSCI) (via Web of Science):** [https://library.maastrichtuniversity.nl/collections/databases/ssci/](https://library.maastrichtuniversity.nl/collections/databases/ssci/)

Bilateral and multilateral agencies and general repositories of IEs in international development:

- **3ie Repository of IEs:** [www.3ieimpact.org/en/evidence/impactevaluations/](http://www.3ieimpact.org/en/evidence/impactevaluations/)
- **3ie RIDIE (Registry for International Development IEs):** [http://ridie.3ieimpact.org/](http://ridie.3ieimpact.org/)
- **USAID Evaluation Clearing House:** [https://dec.usaid.gov/dec/content/evaluations.aspx](https://dec.usaid.gov/dec/content/evaluations.aspx)
- **Innovations for Poverty Action:** [www.poverty-action.org/project-evaluations](http://www.poverty-action.org/project-evaluations)
- **J-Poverty Action Lab:** [www.povertyactionlab.org](http://www.povertyactionlab.org)
- **The World Bank:** [www.worldbank.org](http://www.worldbank.org)
- **AEA RCT Registry:** [https://www.socialscienceregistry.org/](https://www.socialscienceregistry.org/)
- **DFID Research for Development:** [http://r4d.dfid.gov.uk/](http://r4d.dfid.gov.uk/)
- **Campbell Collaboration:** [www.campbellcollaboration.org](http://www.campbellcollaboration.org)
- **BREAD:** [http://ibread.org/bread/papers](http://ibread.org/bread/papers)
- **Center for Effective Global Action:** [http://cega.berkeley.edu/evidence/](http://cega.berkeley.edu/evidence/)

We searched specialist organisational databases that might include big data measurement studies and IEs:

- **AidData: A Research Lab at William & Mary:** [www.aiddata.org/](http://www.aiddata.org/)
- **Flowminder:** [https://web.flowminder.org/](https://web.flowminder.org/)
- **Stanford University Center on Food Security and the Environment:** [https://fse.fsi.stanford.edu/](https://fse.fsi.stanford.edu/)
- **UN Global Pulse:** [www.unglobalpulse.org/](http://www.unglobalpulse.org/)
- **Data-Intensive Development Lab:** [http://didl.berkeley.edu/](http://didl.berkeley.edu/)
- **Global partnership for sustainable development:** [www.data4sdgs.org](http://www.data4sdgs.org)
- **Arxiv database:** ([https://arxiv.org](https://arxiv.org))
- **Eldis:** [www.eldis.org](http://www.eldis.org)
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

We also searched grey literature via Google Scholar and checked references of any SR that we find in our searches and met our inclusion criteria.

Given the relative recent state of the evidence base in this field and the fact that most programmes/initiatives started to flourish in the late 2000s we conducted the searches from 2005 onwards.

**Web of Science (SSCI and Science Citation Index), searched 22 September 2019**

# 1 34,746
TS=( "big data" or metadata or "meta data" or meta-data or "macro data" or macrodata or "mass data" or massdata or "large data" or "bulk data" )

# 2 45,418
TS=((satellite* NEAR/3 imag*) or sensor* or surveillance or meter* or drone* or (mobile* or cell) NEAR/3 (phone* or telephone*)) or mobiles or gps or "global positioning" or gis or "global information system*" or self-tracking or "self tracking") NEAR/3 data)

# 3 41,524
TS=(((e-commerce or commerc* or business* or "credit card*" or ATM* or "automated teller" or "cash machine*" or "money transfer**")) and (transaction* or record or records)) or ((savings or loan or loans) NEAR/3 repay*) NEAR/3 data)

# 4 109
TS=((toll or tolls) NEAR/2 (road* or highway* or motorway*)) or "public transport") NEAR/3 data)

# 5 1,036
TS=(((online or internet or web or virtual or google) NEAR/3 (search* or log or logs)) NEAR/3 data)

# 6 14,088
TS=(("social network*" or "social media" or "opinion platform*" or blog* or Twitter or Facebook or LinkedIn or YouTube or Wiki* or Open or "text messag*" ) NEAR/3 data)

# 7 596
TS=(((citizen* NEAR/2 (report* or derived or submit* or communicat* or inform or informed or informing or feedback)) or hotline* or crowdsourc*) NEAR/3 data)

# 8 133,635
#7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

# 9 35,994
TS=((poverty NEAR/5 (eradicat* or measur* or map*)) or ((wealth or GDP) NEAR/3 measur*)
or (access* NEAR/5 (service* or financ*)) or "social protection" or (reduc* NEAR/3 vulnerab*))

# 10 36,506
TS=(((hunger or malnutrition or malnourished) NEAR/3 (eradicat* or decreas* or end*)) or "food security" or (improv* NEAR/3 nutrition*) or "sustain* agriculture" or (access* NEAR/3...
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

Food*) or ((productiv* or income*) NEAR/3 ("small scale" NEAR/2 produc*)) or (sustain* NEAR/3 "food produc*")) or ((maintain* or maintenance) NEAR/2 "genetic diversity"))

# 11 57,651

TS=( (((health or well-being or wellbeing or "well being") NEAR/2 (promot* or ensur*)) or (reduc* NEAR/3 (maternal or child* or "under 5*" or under-5* or "under five") NEAR/2 (mortality or death*)) or ("communicable disease*" or "infectious disease*") NEAR/3 (epidemic* or pandemic*)) or (reduc* NEAR/2 ("premature mortality" or "premature death")) or (prevent* or treatment) NEAR/3 "substance abuse") or (reduc* NEAR/3 (road* or traffic or vehicle*) NEAR/2 (accident* or death* or fatal*)) or (access* NEAR/2 (reproductive or sexual) NEAR/2 service*) or (universal NEAR/2 health NEAR/2 coverage) or (pollut* NEAR/3 (death* or mortality or fatal*)) NEAR/2 reduc* )

# 12 26,746

TS=( (education or "life-long learning" or (skill* NEAR/3 (build* or acquir* or learn*)) or literacy or numeracy) NEAR/3 (access* or complet* or graduat* or gender* or opportunit*))

# 13 40,719

TS=( (gender* or women or girl* or female*) NEAR/3 (equalit* or inequalit* or discriminat* or empower* or violen* or harm* or (unpaid NEAR/2 (domestic or house*)) or (sexual or reproductive) NEAR/3 right*))

# 14 100,434

TS=( (water or WASH or sanitation) NEAR/3 (access* or quality or sustainable or manag* or ecosystem*) )

# 15 139,810

TS=( (energy NEAR/3 (access* or renewable or afford* or clean or sustainable or efficien*)) )

# 16 210,410

TS=( (((econom* NEAR/2 (growth or productivity)) or employment or labour or work or job or jobs or development or "global resource*" or ((youth or "young people" ) NEAR/3 unemploy*) or tourism or (financ* NEAR/3 institution*) or bank or banks or banking) NEAR/3 (sustain* or increas* or promot* or efficien* or reduc* or forced or underage or "under age" or rights or safe or strengthen* or (capacity NEAR/2 build*)) )

# 17 19,679

TS=( (industr* or innovation or infrastructure or (financ* NEAR/2 service*) or ((scien* or techn*) NEAR/2 research)) NEAR/3 (resilen* or sustain* or promot* or foster* or access* or upgrad* or modern*)) )

# 18 97,676

TS=( (inequalit* NEAR/2 (reduc*)) or ((income NEAR/2 growth) or "equal opportunit*" ) NEAR/2 (sustain* or increas* or promot*)) or (inclusivity NEAR/2 (promot* or sustain* or empower*)) or (equality NEAR/2 (increas* or promot* or engender*)) or (global NEAR/2 (financ* or market* or institution*) NEAR/3 (regulat* or monitor* or legislat* or law or laws)) or ("developing countr*" or LMIC* or "low and middle-income" or "third world" or "global south") NEAR/3 (represent* NEAR/3 (enhanc* or ensur* or increas* or promot*)) or (migration
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

NEAR/3 (policy or policies) NEAR/3 plan*) or (different* or special or positive) NEAR/2 (treat* or discriminat*))

# 19 36,884
TS=((city or cities or urban* or communit* or village* or housing or dwelling* or transport* or ((cultural or natural) NEAR/3 heritage) or ((disaster* or emergenc*) NEAR/3 manag*) or "green space") NEAR/3 (sustain* or safe or inclusive or resilien* or afford* or protect* or ((reduc* or limit* or mitigat*) NEAR/3 "environmental impact" NEAR/3 (city or cities or urban*) or access*))))

# 20 149,994
TS=(((consum* or production or "natural resource*" or waste or wastes or procurement or (information NEAR/2 (access* or availab*)) NEAR/3 (sustain* or responsible or efficien* or environment* or reduc*)))

# 21 201,412
TS=("climat* change*" or "global warming" or (greenhouse NEAR/2 (effect* or gas*)) NEAR/3 (action or combat* or mitigat* or impact* or resilien* or adapt* or (national NEAR/2 (policy or policies)) or ((promot* or rais* or improv*) NEAR/3 aware*)))

# 22 54,354
TS=((ocean* or sea or seas or marine or coastal or fishing or fisher* or "small island") NEAR/3 (sustain* or conserv* or manag* or protect* or regulat* or ecosystem* or (reduc* NEAR/2 pollut*) or acidity))

# 23 22,000
TS=(((terrestrial or land or land-based or forest* or mountain* or (natural NEAR/2 (habitat or ecosystem*)) or poaching or poacher* or traffick*) NEAR/3 (sustain* or conserv* or protect* or (degrad* NEAR/3 (neutral or prevent* or protect* or minimi* or reduc*)) or ("genetic resources" NEAR/2 (utilis or utiliz* or use or using or promot*)) or prohibit* or ban or banning or banned) or ("alien species" NEAR/2 (invasive or introduced or imported)) NEAR/3 (prevent* or impact*)) or ((ecosystem* or biodiversity) NEAR/3 (national or local*) NEAR/3 (plan* or policy or policies)))

# 24 2,378
TS=(((societ* NEAR/3 (peaceful or inclusive)) or justice or "rule of law") NEAR/3 (sustain* or access* or effective* or accountab* or strong or robust or promot*))

# 25 105,185
TS=((violen* or corrupt* or ((illicit or illegal) NEAR/2 financ* NEAR/2 (activit* or dealing*)) NEAR/2 (reduc* or mitigat* or eliminat*)) or (institution* NEAR/2 (accountab* or transparen* or effective*))

# 26 17,951
TS=(("decision making" or decision-making or participati*) NEAR/2 (representative or democra*)) or ("global governance" NEAR/3 (participati* or involv* or includ*)) NEAR/3 ("developing countr*" or LMIC* or "low and middle-income" or "global south" or "third

cedilprogramme.org61
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

world") or ("legal identit* NEAR/3 (provi* or promot* or regist*)) or (information NEAR/3 access*))

# 27 141

TS= ("sustainable development goal*" NEAR/3 (partnership* or implement* or revitali* or strengthen* or sustainability or (capacity NEAR/2 build*) or ((development or financial or domestic) NEAR/2 (resource* or assistance)) or ((investment or technolog* or ("trading system*" NEAR/2 universal)) NEAR/3 (promot* or innovation*)) or (knowledge NEAR/2 shar*)) or "technology bank*" or ((exports or market*) NEAR/3 (increas* or promot*) NEAR/3 ("developing countr*" or LMIC* or "low and middle-income" or "third world" or "global south") or "global macroeconomic" or ((policy or policies) NEAR/2 coheren*)) or "global partnership*" or "public-private partnership*" or (progress NEAR/3 (monitor* or assess* or evaluat* or review*)))))

#27 OR #26 OR #25 OR #24 OR #23 OR #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9

# 28 1,196,499

#28 AND #8

# 29 16,624

# 30 4,365

TS=(sdg or sdgs or "sustainable development goal")

#30 AND #1

# 31 35

#31 OR #29

# 32 16,625 (Indexes=SCI-EXPANDED, SSCI Timespan=2005-2019)

#28 AND #8

# 33 3,519 (Indexes=SSCI Timespan=2005-2019)
## Data extraction tool

### Table 8: Data extraction tool

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study ID</td>
<td>Unique ID ascribed to each record</td>
</tr>
<tr>
<td>Title name</td>
<td>Use only the English version of the publication's main title. If paper is not written in English and has the title translated, use the translated version of the title. If the publication does not provide an English version, include the title in its original language. Please enter title in sentence case. Ensure there are no line breaks</td>
</tr>
<tr>
<td>Language</td>
<td>Select full-text language that applies: English</td>
</tr>
<tr>
<td>Open access</td>
<td>If the study’s (full-text) content is available, code as ‘Yes’. If study has paywalls, code as ‘No’. Please save the PDF in the Dropbox folder called ‘Full-Text PDFs’ using the following format Firstauthorsurname_year_record id. If study has multiple versions, in other words, if the study has been published as both a journal article and a working paper, both versions may be included in the IER</td>
</tr>
</tbody>
</table>
| Type of big data | Select one or more from the list based on the type of big data being used:  
   1. Human-sourced information (social networks)  
      1.1. Social networks  
      1.2. Internet searches  
      1.3. Mobile data content  
      1.4. Citizen reporting or crowdsourced data  
   2. Process-mediated data (traditional business systems and websites)  
      2.1. Data produced by public agencies  
      2.2. Data produced by businesses  
      2.3. Mobile phone CRD  
   3. Machine-generated data (automated systems)  
      3.1. Data from fixed sensors  
      3.2. Data from mobile sensors (tracking)  
      3.3. Data from satellites |
| Outcome       | Select ONE outcome that applies according to the intervention being evaluated:  
   Economic development and livelihood  
   Sustainable agriculture and food security  
   Health and well-being  
   Education  
   Water and sanitation |
## Governance and human rights
- Poverty mapping/measurement
- Measuring wealth/GDP
- Access to basic services
- Social protection, including financial inclusion
- Vulnerability reduction
- Economic growth and productivity
- Full and productive employment, including youth unemployment
- Eradicate forced and underage labour, protect labour rights and promote safe workspace
- Promote sustainable tourism
- Strengthen capacity of domestic financial institutions
- Reduce mortality
- End epidemics of communicable diseases
- Strengthen prevention and treatment of substance abuse
- Decrease fatalities due to road accidents
- Universal access to sexual and reproductive healthcare services
- Achieve universal health coverage
- Reduce fatalities due to pollution
- Achieve gender equality and empower all women and girls
- Reduce inequality within and among countries
- Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels

## Sub-outcomes
- Access to affordable housing
- Access to affordable transport systems
- Enhance inclusive and sustainable urbanisation
- Strengthen efforts to protect cultural and natural heritage
- Better disaster management
- Reduce environmental impact of cities
- Universal access to green public spaces
- Ensure sustainable consumption and production patterns
- Take urgent action to combat climate change and its impacts
- Conserve and sustainable use the oceans, seas and marine resources for sustainable development
- Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation and halt biodiversity loss
<table>
<thead>
<tr>
<th>Gender equity focus</th>
<th>Does this study consider gender and/or equity?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Cost</td>
<td>Is cost data provided?</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Population</td>
<td>What population sub-groups does the study target?</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
</tr>
<tr>
<td></td>
<td>Refugees</td>
</tr>
<tr>
<td></td>
<td>Conflict affected persons</td>
</tr>
<tr>
<td></td>
<td>Ethnic minorities</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
</tr>
<tr>
<td>Geographical coverage</td>
<td>Select the continent/region in which the study was conducted:</td>
</tr>
<tr>
<td></td>
<td>Asia</td>
</tr>
<tr>
<td></td>
<td>Africa</td>
</tr>
<tr>
<td></td>
<td>North America</td>
</tr>
<tr>
<td></td>
<td>South America</td>
</tr>
<tr>
<td></td>
<td>Europe</td>
</tr>
<tr>
<td></td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Antarctica</td>
</tr>
<tr>
<td>Fragile context</td>
<td>Select the type based on fragile context definition</td>
</tr>
<tr>
<td></td>
<td>Conflict related difficulty</td>
</tr>
<tr>
<td></td>
<td>Logistical difficulty</td>
</tr>
<tr>
<td></td>
<td>All others</td>
</tr>
<tr>
<td>Evaluation design</td>
<td>Select one of three options defined as:</td>
</tr>
<tr>
<td></td>
<td>1. Experimental</td>
</tr>
<tr>
<td></td>
<td>a) RCT defined as prospective randomised assignment, where randomisation is implemented by researchers (or by decision makers in the context of an evaluation study)</td>
</tr>
<tr>
<td></td>
<td>2. Quasi-experimental</td>
</tr>
<tr>
<td></td>
<td>a) Quasi-random assignment: i) regression discontinuity design (sharp designs) or ii) natural experiment in which exposure to treatment is random</td>
</tr>
<tr>
<td></td>
<td>b) Non-random assignment: i) studies that control for unobservables (DID, FE, IV, Fuzzy RDD, ITS) or ii) studies that control for observables only (eg statistical matching, synth control, regression adjustment)</td>
</tr>
<tr>
<td></td>
<td>3. Measurement studies</td>
</tr>
<tr>
<td></td>
<td>Studies that have innovatively used big data to measure and validate an SDG indicator such as poverty mapping, food security, forest cover, etc</td>
</tr>
</tbody>
</table>
| Evaluation method | If experimental, then select (only if one or two selected under evaluation design):  
|                   | RCTs  
|                   | If quasi-experimental, then select:  
|                   | Sharp RDD  
|                   | DID  
|                   | FE estimation  
|                   | IV estimation  
|                   | Fuzzy RDD  
|                   | Statistical matching (includes PSM)  

| Methods for analysing big data | Select one or more methods used in the analysis of ‘big data’:  
|                               | ML: supervised learning, unsupervised learning, clustering, anomaly detection, random forests, artificial neural network, convolutional neural network, support vector regression  
|                               | Bayesian geostatistical models: latent Gaussian models, integrated nested Laplace approximations  
|                               | Natural language processing: vectorisation, embedding, categorisation, sentiment analysis  
|                               | Any other: please specify  

| Big data validation | Select one or more methods used to calibrate/validate results of big data analysis using survey data:  
|                    | Administrative/survey data used for calibration or validation?  
|                    | Yes/No  
|                    | Type of data  
|                    | ● Administrative data  
|                    | ● Survey data  
|                    | ○ Secondary survey data  
|                    | ○ Primary survey data  
|                    | Survey name  
|                    | ● Demographic Health Survey  
|                    | ● Living Standards Measurement Study  
|                    | ● Other (specify)  
|                    | Mode of data collection  
|                    | ● Face-to-face  
|                    | ● Telephone mode  
|                    | ● Mobile phone telephone mode  
|                    | ● Dual-frame  
|                    | ● Other (specify)  
|                    | Sample size  

<table>
<thead>
<tr>
<th>Type of data</th>
</tr>
</thead>
</table>
| ● Administrative data  
| ● Survey data  
| ○ Secondary survey data  
| ○ Primary survey data  

<table>
<thead>
<tr>
<th>Survey name</th>
</tr>
</thead>
</table>
| ● Demographic Health Survey  
| ● Living Standards Measurement Study  
| ● Other (specify)  

<table>
<thead>
<tr>
<th>Mode of data collection</th>
</tr>
</thead>
</table>
| ● Face-to-face  
| ● Telephone mode  
| ● Mobile phone telephone mode  
| ● Dual-frame  
| ● Other (specify)  

<table>
<thead>
<tr>
<th>Sample size</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Sample design</th>
<th>Link to survey details (if available)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mixed methods</strong></td>
<td>Select YES if study includes quantitative and qualitative analyses, otherwise select NO</td>
</tr>
</tbody>
</table>
| **Transparency in data collection, analysis and reporting** | Choose if the study discusses the following:  
- Are the data collection methods described?  
- Are data quality issues such as completeness (missing or incomplete entries; empty cells) and noise discussed?  
- Is the data representative of the population of interest?  
- Is the construct validity explained (ie is there a discussion on how the big data-based indicator measures what the study claims to measure)?  
- Are the results generalisable? For example, are the research findings generalisable to other situations such as other platforms (data source) or communities, or over time?  
- Are data and codes publicly available for replication?  
- Any other data, analysis and reporting challenge discussed? |
| **Ethical approval and discussion** | Choose if the study discusses the following:  
- Ethical approval obtained?  
- Consent for secondary and other use of data  
- Data privacy  
- Data security and governance arrangement  
- Unintended exclusion  
- Unintended consequence on any group of people and/or individuals  
- Others (specify) |
| **Unit of observation** | Enter all the levels of observation of the variables used for the analysis:  
- Community  
- Cohort (includes schools or clinics)  
- Household  
- Individual  
If more than one, include in separate rows:  
- Country  
- Districts  
- Sub-districts  
- Village/city |
<p>| <strong>Funding agency</strong> | What category of funding agency funded the research? |</p>
<table>
<thead>
<tr>
<th>Note: only code if reported in the study; no need to do additional research to find</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select one of the following:</td>
</tr>
<tr>
<td>- Government agency</td>
</tr>
<tr>
<td>- International aid agency</td>
</tr>
<tr>
<td>- International financial institution</td>
</tr>
<tr>
<td>- Non-profit organisation</td>
</tr>
<tr>
<td>- For-profit firm</td>
</tr>
<tr>
<td>- Academic institution</td>
</tr>
<tr>
<td>- Charitable foundation or private foundation</td>
</tr>
<tr>
<td>- Not specified</td>
</tr>
</tbody>
</table>
Appendix 5  

**ML and manual screening**

All the studies obtained from manual and automatic search were imported to EPPI-Reviewer 4, a software facilitating management of references, identification and removal of duplicates, and screening of studies at both title and abstract and full-text stages. The title and abstract screening was done using the ML functionality available in EPPI in order to make the process more efficient. The model built using this functionality learns from the set of manually screened studies to guess the inclusion criteria and reorganizes the list of studies based on the likelihood of being included. Using a training dataset consisting of 3,300 manually screened studies, the algorithm ranked studies by prioritizing them in order of likelihood of inclusion and excluding more than 1,500 studies in a row. Based on this likelihood, the studies are grouped into ten groups (Figure 14), with the set of excluded studies classified in the 0%-10% range.

**Figure 14: Order of relevance of studies at the title and abstract screening stage**

![Graph showing order of relevance of studies](source: Authors' own calculation)

This systematic map includes studies that are extremely diverse; hence ML may not be accurate. To undertake a more nuanced approach and to account for this diversity in studies we conducted an additional manual screening within the groups. Since this manual screening gave us only 5% includes from the set of studies within 10%-40% range, those were not considered in the subsequent screening process. Once we had a set of studies that were included in the title and abstract screening stage, this was followed by full-text screening. Here, we screened studies that had more than a 50% likelihood of inclusion based on the ML model. A random check on 20% of studies belonging to the 40%-49% group provided a very limited number of relevant studies that could be included, due to which the lower likelihood studies were not screened.
Appendix 6  

**SR appraisal tool and summary**

**Summary of the SRs**

**Yan et al. (2017). Utility and potential of rapid epidemic intelligence from internet-based sources.**

The study aimed to summarise internet-based methods that use freely accessible and unstructured data for epidemic surveillance, exploring their timeliness and accuracy outcomes. The study is based on 84 articles published between 2006 and 2016 relating to internet-based public health surveillance methods. These studies employ search queries, social media posts and approached derived from existing internet-based systems for early epidemic alerts and real-time monitoring. The primary methodology used for this review is the preferred reporting items for SR and meta-analyses. Using this method, the authors can assess the benefits and challenges of a healthcare intervention through an evidence-based minimum set of items. The study does not clearly demonstrate the inclusion criteria for the study designs, making it difficult for the readers to interpret the findings.

**Fung et al. (2016). Ebola virus disease and social media: a systematic review.**

The study is an SR of the existing research pertinent to the Ebola virus and social media, especially to identify the research questions and methods used to collect and analyse social media. The study searched six databases for research articles relevant to Ebola and social media. Twelve articles were included in the main analysis: seven from Twitter and one including Weibo, one from Facebook, three from YouTube and one from Instagram and Flickr. All the studies were cross-sectional. The study uses a standardised form to extract the data. A key challenge of the review is that the methods used by the review authors to analyse the findings of the included studies are not clearly defined.

**Williamson et al. (2019). Satellite remote sensing in shark and ray ecology, conservation and management.**

The study aims to present a summary of the existing state of knowledge on the application of satellite remote sensing (SRS). SRS is seen as a key opportunity to analyse important environmental drivers in elasmobranch ecology and to aid management decisions for the conservation of declining population. The review included 71 papers and made use of several academic tools in order to conduct the bibliographic search. These included ISI Web of Science, Scopus and Google Scholar. One of the limitations of the review is that it examines a wide variety of studies but does not combine the results of these studies explicitly.

This SR of literature aims to determine the scope of big data analytics in the field of healthcare including its applications and challenges in its adoption in healthcare. The review employs a systematic search of articles on five major scientific databases: Science Direct, PubMed, Emerald, IEEE Xplore and Taylor & Francis. Two reviewers independently extracted information on the definition of big data analytics and the sources and applications of big data. A total of 58 articles were selected. The authors did not explicitly state the methods used to analyse the quality of included studies and also did not report results for each of the studies in the review.


This study aimed to present an SR regarding data mining (DM) and ML approaches adopted in the promotion of smart cities. The study seeks to provide, from a literature review in journals belonging to the Web of Science and Scopus databases, the different DM and ML techniques used, as well as to present the sectors most engaged in the promotion of smart cities. A total of 39 studies were included on the map, which were further analysed to assess the most commonly used DM and ML techniques to promote smart cities. While the report describes individual results of the studies clearly, it does not combine the results of all the studies.


The studies in this review demonstrate how social media may be a valuable tool in improving the ability of public health professionals to detect disease outbreaks faster than by using traditional methods and to enhance outbreak response. A social media application was defined for this review as ‘an internet-based application where people can communicate and share resources and information, and where users can activate and set their own profiles, have the ability to develop and update them constantly and have the opportunity to make such profiles totally or partially public and linked with other profiles in the network’. A total of 60 articles were selected for this SR, which addressed the two research questions: can social media be integrated into disease surveillance and can it be used to improve health outcomes?


The aims of this SR were to determine the capability of GPS to collect high-quality data on the location of activities in research on the relationship between physical activity and the environment. Studies were eligible for inclusion if they were undertaken on humans, used GPS to measure the locations where physical activity occurred and included analysis of the relationship between the characteristics of the environment and any form of physical activity behaviour (including leisure-time physical activity, sport or active travel). The capability of GPS
CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map

was expressed in terms of data quality, which in turn was defined as the proportion of GPS data lost in each study. The authors do not mention the proper risk of bias assessment of the included study.

Bennett and Smith (2017). Advances in using multitemporal night-time lights satellite imagery to detect and estimate and monitor socioeconomic dynamics.

This paper reviews progress in using the multitemporal DMSP-OLS, for which digital imagery was available from 1992 to 2013. It also reviews Visible Infraref Imaging Radiometer imagery to analyse urbanisation, economic and population dynamics across range of geographic scales. The review provides an overview of data corrections and processing for a comparison of multitemporal nightlight imagery. The review includes a total of 17 studies. The results of all the studies were not properly combined and reported in the review. A risk of bias assessment of the included study has also not been reported in the review.
### Table 9: Top 20 countries with maximum number of studies

<table>
<thead>
<tr>
<th>Country</th>
<th>Total</th>
<th>IE</th>
<th>Measurement study</th>
<th>SRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>43</td>
<td>6</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>34</td>
<td>4</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>31</td>
<td>3</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Multi-country</td>
<td>31</td>
<td>3</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>Unidentified</td>
<td>14</td>
<td>1</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Kenya</td>
<td>13</td>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>9</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Rwanda</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>7</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Nigeria</td>
<td>7</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>6</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>6</td>
<td></td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

### Table 10: Distribution of studies across the region

<table>
<thead>
<tr>
<th></th>
<th>East Asia and Pacific</th>
<th>Europe and Central Asia</th>
<th>Latin America and the Caribbean</th>
<th>Middle East and North Africa</th>
<th>North America</th>
<th>South Asia</th>
<th>Sub-Saharan Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>IE</td>
<td>12</td>
<td>5</td>
<td>18</td>
<td>3</td>
<td>5</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Measuremen</td>
<td>92</td>
<td>69</td>
<td>42</td>
<td>29</td>
<td>49</td>
<td>82</td>
<td>112</td>
</tr>
</tbody>
</table>
### Table 11: Studies using multiple sources of big data

<table>
<thead>
<tr>
<th>Combination of big data</th>
<th>IEs</th>
<th>Measurement study</th>
<th>SRs</th>
<th>Grand total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile phone CRD</td>
<td></td>
<td>Data from fixed sensors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone CRD</td>
<td></td>
<td>Data from satellites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citizen reporting and crowdsourced data</td>
<td></td>
<td>Data from fixed sensors</td>
<td></td>
<td>Data from satellites</td>
</tr>
<tr>
<td>Citizen reporting and crowdsourced data</td>
<td></td>
<td>Data from mobile sensors (tracking)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citizen reporting and crowdsourced data</td>
<td></td>
<td>Data from satellites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from fixed sensors</td>
<td></td>
<td>Data from satellites</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Data from mobile sensors (tracking)</td>
<td></td>
<td>Data from satellites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data produced by public agencies</td>
<td></td>
<td>Data from fixed sensors</td>
<td></td>
<td>Data from satellites</td>
</tr>
</tbody>
</table>
Data produced by public agencies || Data from satellites
1 || 3 || 4

Data produced by public agencies || Data produced by businesses
1 || 1

Mobile data content || Mobile phone CRD
2 || 2

Mobile data content || Citizen reporting and crowdsourced data
2 || 2

Mobile data content || Data from mobile sensors (tracking)
1 || 1

Social networks || Citizen reporting and crowdsourced data
2 || 2

Social networks || Data from satellites
1 || 1

Social networks || Data produced by public agencies || Data from fixed sensors || Data from satellites
1 || 1

Social networks || Internet searches
4 || 1 || 5
Table 12: Big data source wise studies in fragile context

<table>
<thead>
<tr>
<th>Source Type</th>
<th>Chemical or radio-nuclear</th>
<th>Conflict or humanitarian crisis</th>
<th>Difficult terrain</th>
<th>Disease outbreaks or epidemics</th>
<th>Natural disaster</th>
<th>Others</th>
<th>Total (row)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile phone CRD</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>11</td>
<td>10</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Citizen reporting and crowdsourced data</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Data from fixed sensors</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Data from mobile sensors (tracking)</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Data from satellites</td>
<td>0</td>
<td>25</td>
<td>18</td>
<td>4</td>
<td>10</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>Data produced by businesses</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 13: How studies have dealt with data quality and transparency

<table>
<thead>
<tr>
<th></th>
<th>Data collection methods</th>
<th>Data quality issues</th>
<th>Data representative of the population of interest</th>
<th>Construct validity explained</th>
<th>Results are generalisable</th>
<th>Data and code publicly available</th>
<th>Data, analysis and reporting challenges discussed</th>
</tr>
</thead>
<tbody>
<tr>
<td>IE</td>
<td>47</td>
<td>22</td>
<td>29</td>
<td>43</td>
<td>26</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Measurement study</td>
<td>43</td>
<td>3</td>
<td>7</td>
<td>21</td>
<td>4</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>SRs</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Grand total</td>
<td>91</td>
<td>25</td>
<td>36</td>
<td>64</td>
<td>30</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>
## Systematic map of big data sources and outcomes

### Figure 15: Systematic map of big data sources and outcomes

The online map can be accessed here:
[https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map](https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map)
Appendix 9  
List of included studies


cedilprogramme.org79


SRs


### Measurement studies


<table>
<thead>
<tr>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
</table>


88. Barros, N., Fontes, T., Silva, M.P. and Manso, M.C. (2013). How wide should be the adjacent area to an urban motorway to prevent potential health impacts from traffic


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


186. Head, A., Manguin, M., Tran, N., & Blumenstock, J. E. (2017, November). Can human development be measured with satellite imagery?. In ICTD (pp. 8-1). Available at: https://bit.ly/3kLaMHL


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


CEDIL methods working paper: Using big data for evaluating development outcomes: a systematic map


