INTEGRATING

GLOBAL PULSE

INTO THE MONITORING AND EVALUATION OF DEVELOPMENT PROGRAMMES

GLOBAL
PULSE

The
ROCKEFELLER
FOUNDATION

INTEGRATING BIG DATA INTO THE MONITORING AND EVALUATION OF DEVELOPMENT PROGRAMMES

DISCLAIMER

This report was prepared by an independent consultant at the request of UN Global Pulse. The opinions expressed and arguments employed herein do not necessarily reflect the views or policies of UN Global Pulse.

UN Global Pulse encourages the use, reproduction and dissemination of the material in this information product.

Please cite this publication as: UN Global Pulse, 'Integrating Big Data into the Monitoring and Evaluation of Development Programmes,' 2016

© UN Global Pulse, 2016

ACKNOWLEDGMENTS

This report was produced for UN Global Pulse by Michael Bamberger. UN Global Pulse thanks the Rockefeller Foundation for its generous financial support as well as its technical advice in the preparation of the report. The report also benefitted from the experience and knowledge of many colleagues from UN agencies including UNICEF, UNFPA, UN Women, the UN Office of Internal Oversight Services (OIOS), the World Bank as well as partners such as The Impact Lab and Social Impact.

This document drew inspiration from a report by Sally Jackson (bit.ly/2gQUSej), a consultant who worked with the Global Pulse Lab in Indonesia. Appreciation also goes to Global Pulse colleagues in New York, Uganda (Kampala) and Indonesia (Jakarta) who tirelessly assisted in the process.

It was edited by Tamara Karaica and Felicia Vacarelu and designed by Ana Petak.

ABOUT THE AUTHOR

Michael Bamberger has over forty five years of experience in the evaluation of international development programmes, including ten years working with NGOs in Latin America, twenty years with the World Bank and over fifteen years consulting on development evaluation with UN agencies as well as development banks, bilateral development agencies, NGOs and foundations.

CONTENTS

19 Context

20 Executive summary

26 PART I DEVELOPMENT EVALUATION IN THE AGE OF BIG DATA

28 CHAPTER 1 THE DATA REVOLUTION: IMPLICATIONS FOR INTERNATIONAL DEVELOPMENT

29 1.1 The dramatic expansion of digital technologies

31 1.2 Monitoring and evaluation in the new international development context: opportunities and challenges

34 1.3 Defining big data

38 1.4 Current applications of big data in international development

42 1.5 The experience of UN Global Pulse: moving from proof–of–concept to operational use

43 1.6 Collaboration between data scientists and development evaluators is key to successfully applying new tools and technologies

50 CHAPTER 2 THE PROMISE OF BIG DATA FOR PROGRAMME MONITORING AND EVALUATION – AND THE CHALLENGES

51 2.1 The main types of development monitoring and evaluation

54 2.2 Identifying challenges to current evaluation methodologies and areas where big data analytics may be able to contribute

57 2.3 The characteristics of big data analytics and potential applications for programme monitoring and evaluation

62 2.4 The complementarities between big data analytics and current evaluation practice

63 2.5 How big data can help address common evaluation challenges

68 2.6 How well can big data analytics address the challenges facing programme evaluation

72 PART II GUIDELINES FOR INTEGRATING BIG DATA INTO THE M&E FRAMEWORKS OF DEVELOPMENT PROGRAMMES

74 CHAPTER 3 INTEGRATING BIG DATA INTO THE MONITORING, EVALUATION AND LEARNING (MEL) SYSTEM

75 3.1 The components of an integrated MEL system

76 3.2 Policy evaluation, programme evaluation and project evaluation

78 3.3 The stages of a typical project, programme cycle

80 3.4 Monitoring, evaluation and learning at different stages of the project/programme cycle

82 3.5 A framework for incorporating big data into MEL systems

86 3.6 When integrating big data into M&E systems it is important to follow evaluation best practices

92 CHAPTER 4 BUILDING BIG DATA INTO PROGRAMME MONITORING

93 4.1 Uses of monitoring

95 4.2 Using big data for policy and programme monitoring

96 4.3 Current sources of information for project and programme monitoring and their limitations

97 4.4 How big data and ICT can strengthen monitoring systems

104 4.5 Steps in the integration of big data into programme monitoring

108 CHAPTER 5 BUILDING BIG DATA INTO PROGRAMME EVALUATION

109 5.1 Elements of a dynamic evaluation system

109 5.2 Potential applications of big data for policy and programme evaluation

110 5.3 Potential applications of big data and ICTs in project evaluation

111 5.4 When and how big data can be integrated into programme evaluation

111 5.5 Integrating big data and ICT into the design stage of the programme evaluation

117 5.6 Other design issues

119 5.7 Case studies illustrating the integration of big data and ICTs into impact evaluations

122 CHAPTER 6 MANAGING BIG DATA INCLUSIVE EVALUATIONS

123 6.1 The critical role of the evaluation manager

124 6.2 Special challenges in managing big data–inclusive M&E

125 6.3 Defining the levels at which the evaluation should be conducted and the evaluation approach

125 6.4 The role of the evaluation manager at each stage of the project cycle

TABLES

40 1–1 Examples of how big data can be used for development and humanitarian action

44 1–2 Big data applications for sustainable development from UN Global Pulse

53 2–1 The main types of development monitoring and evaluation

55 2–2 Six sets of challenges to current evaluation methodologies where big data analytics may be able to contribute

60 2–3 Kinds of big data analysis with potential applications for programme monitoring and evaluation

65 2–4 Examples of data visualization that make the findings of complex analysis easily accessible to managers and local communities

80 3–1 Typical monitoring, evaluation and learning activities at different stages of the project cycle

83 3–2 Some of the key questions for each of the 4 types of evaluation

86 3–3 Checklist for assessing the quality and appropriateness of big data and ICT–generated indicators

90 3–4 Checklist for assessing evaluation methodology when big data and ICT are integrated

97 4–1 Case studies illustrating how big data can be used to strengthen programme monitoring

98 4–2 Potential contributions of big data and ICT in programme monitoring

112 5–1 Potential ways that big data and ICT can strengthen programme evaluation

115 5–2 Factors determining the applicability of big data and data analytics in programme evaluation

119 5–3 Case studies illustrating how big data and ICTs were used to strengthen each of the main evaluation designs

126 6–1 The role of the evaluation manager in guiding the planning and focus of big data inclusive M&E systems at each stage of the project cycle

FIGURES

35 1–1 The three dimensions of the big data framework

36 1–2 Overview of data analytics applications for development

39 1–3 The data continuum

48 1–4 Bridge building between data scientists and development evaluators

59 2–1 The four steps in applying data analytics to programme evaluation

66 2–2 Dimensions of complexity in development evaluation

77 3–1 The three levels at which evaluations are conducted

79 3–2 The monitoring, evaluation and learning [MEL] system of a typical project/programme cycle

85 3–3 Steps in the selection of Big Data and ICT generated M&E indicators

94 4–1 Information and outputs at each stage of a project/programme monitoring system

105 4–2 Setting up the big data/ICT–based monitoring system

118 5–1 Trajectory analysis: Different scenarios for how programme effects evolve over time

BOXES

30 1–1 The data revolution and the implications for international development

31 1–2 Using new information technology to conduct a national census in a country with high security concerns. The 2016 Islamic Republic of Afghanistan census

32 1–3 The 'digital dividends' from the data revolution

33 1–4 Challenges for the evaluation of new and complex international development programmes

37 1–5 The data continuum: from small to big and big to small

62 2–1 An example of social media analytics: Rapid assessment of communicable diseases in Indonesia based on the analysis of Tweets

100 4–1 Using satellite imaging to monitor changes in forest cover

101 4–2 Monitoring migration and labour market shocks using mobile data

102 4–3 Mining tweets to monitor food price crises in Indonesia

130 GLOSSARY OF TECHNICAL TERMS

136 REFERENCES

140 ENDNOTES

ABBREVIATIONS AND ACRONYMS

ATM – automated teller machine
FAO – Food and Agriculture Organization of the United Nations
GEF – Global Environment Fund
GIS – Geographic Information System
GPS – Global Positioning System
HDR – Human Development Report
ICT – Information and Communication Technology
M&E – Monitoring and Evaluation
MEL – Monitoring, Evaluation and Learning
NGO – Nongovernmental organization
OECD – Organization for Economic Co–operation and Development
OECD/DAC – Development Assistance Committee
OIOS – United Nations Office of Internal Oversight Services
UN – United Nations
UNFPA – United Nations Population Fund
UNICEF – United Nations Children's Fund
UNDP – United Nations Development Programme
UNHCR – United Nations High Commissioner for Refugees
PA – Predictive analytics
RCT – Randomized Control Trials
SDG – The Sustainable Development Goal
TOC – Theory of Change

FOREWORD

The use of data analytics to inform and implement smart, agile and adaptive projects and programmes has passed beyond the inflection point and is now accelerating within development and humanitarian practice.

Over the past seven years, Global Pulse has been working across the UN and in partnership with academia and the private sector to advance research and development on big data. Part of these efforts have focused on looking at the opportunities and challenges of integrating big data in the M&E of development programmes within our own work and across the UN system.

The adoption of the Sustainable Development Goals makes these efforts even more current. Insights from digital data have a role to play in evaluation, not only because of their potential to complement traditional data sources, but also because many aspects of development progress now occur primarily online.

It is, therefore, imperative that evaluators can familiarize themselves with new data sources, technologies and methodologies and begin integrating them into their work.

Together with the Rockefeller Foundation, we hope this report may serve not only as an introductory guide to big data, but also as an urgent call to action.

Robert Kirkpatrick
Director, UN Global Pulse

There is no longer any doubt that the explosion of available data and the speed with which it can be provisioned will revolutionize the way global challenges are solved. Practitioners and institutions engaged in international development will need to embrace this as the new norm, and begin to responsibly shape the way these trends influence their work.

This report brings together two distinct schools of thought – from the data sciences and social sciences – to explore how big data can be used to support development assistance interventions.

New forms of tech-enabled data such as big data have the potential to complement conventional monitoring and evaluation (M&E) approaches – and lend greater insight into the impact of development programs on poor and vulnerable people.

The Rockefeller Foundation has a history of pushing the envelope on innovation – with a view to driving greater impact. To this end, we are pleased to have supported the United Nations Global Pulse to produce this report. We are thankful to the Global Pulse team for managing this important report through completion, and especially to Michael Bamberger and Sally Jackson for their technical advice and leadership.

We hope that this report will encourage fresh thinking around what is possible in the realms of big data and M&E.

Veronica Olazabal
Director - Measurement, Evaluation, and Organizational Performance, The Rockefeller Foundation

EXECUTIVE SUMMARY

This report provides guidelines for evaluators, evaluation and programme managers, policy makers and funding agencies on how to take advantage of the rapidly emerging field of big data in the design and implementation of systems for monitoring and evaluating development programmes.

The report draws on interviews conducted with a sample of international development experts from UN agencies, bilateral aid agencies, multilateral development banks, and civil society, as well as data analysts specializing in development applications. It also draws on a review of the existing literature as well as an active participation in a number of recent conferences and workshops related to the use of new data sources. Furthermore, interviews were conducted with UN Global Pulse technical staff from Headquarters and from the data innovation labs in Uganda (Kampala) and Indonesia (Jakarta).

The report is organized in two parts. **Part I: Development evaluation in the age of big data** reviews the data revolution and discusses the promise, and challenges this offers for strengthening development monitoring and evaluation. **Part II: Guidelines for integrating big data into the monitoring and evaluation frameworks of development programmes** focuses on what a big data inclusive M&E system would look like. The report also includes guidelines for integrating big data into programme monitoring and evaluation. A final chapter discusses issues in the management of big data inclusive M&E systems.

A CALL TO ACTION

This report is only a first step in trying to align data innovations with the monitoring and evaluation of development programmes; a broader range of approaches are currently being developed and tested. The report is intended as a Call to Action to inspire development agencies and particularly evaluators to collaborate with data scientists and analysts in the exploration and application of new data sources, methods, and technologies.

Most of the applications of big data in international development do not currently focus directly on monitoring, and even less on evaluation. Instead they relate more to research, planning and operational use of big data. Many development agencies are still in the process of defining their policies on big data and it can be anticipated that applications to the monitoring and evaluation of development programmes will start to be incorporated more widely in the near future. This report includes examples and ways that big data, together with related information and communications technologies (ICTs) are already being used in programme monitoring, evaluation and learning.

PART I DEVELOPMENT EVALUATION IN THE AGE OF BIG DATA

Chapter 1: The Data Revolution: implications for international development. The dramatic expansion and potential benefits of new data sources is reviewed. The world is becoming more connected and interdependent. Information is now available on a scale that most people could hardly have imagined even a few years ago. One implication for international development is that new sources of real–time information about people are for the first time available and accessible.

Digital data is part of a broader technology revolution, which can potentially produce 'digital dividends' for development in the areas of inclusion, efficiency, innovation and empowerment, voice and security. The opportunities and challenges for M&E in the new international development context are reviewed. The Sustainable Development Goals (SDGs) approved by the UN General Assembly in 2015 are used to illustrate the need to rethink current approaches to M&E.

The data revolution and the rapid growth of big data are discussed. In practice, most agencies work with a data continuum that combines **big data, open data, ICTs** and often **small–scale qualitative data sets**. The combination of different types/sets of data also stresses the importance of **integrating the human dimension** at different stages of the monitoring and evaluation process: (i) defining the questions to be addressed, (ii) selecting the right mix of data collection and analysis tools, and (iii) often working with communities to ensure they are involved in the data collection process, as well as the interpretation and use of the findings.

Examples are presented to illustrate how big data is currently being used in international development to help inform disaster relief, mine citizen feedback or map population movement to support response to disease outbreaks. None of these examples directly involve programme evaluation, and evaluation offices tend not to be in the forefront in most agencies with experimentation with big data. However, the experience of UN Global Pulse in collaboration with other UN agencies and developing country governments shows that big data is starting to be built into monitoring and evaluation systems.

Later chapters provide examples that illustrate how big data is being applied to programme evaluation. The message is that there are many areas where big data can potentially be used for monitoring and evaluation, but most agencies have not yet exploited many of these opportunities – and the challenge is to find ways to scale–up these promising pilot programmes.

One of the challenges for incorporating big data into development evaluation is the difference in how data scientists and evaluators collect, process and analyse data. Data science and evaluation are also grounded in different approaches to theory. The chapter discusses the **need for bridge building between data scientists and evaluators to allow for the development of a common language and to identify promising areas where big data analytics can be applied in development evaluation contexts.**

Chapter 2: The promise of big data for programme monitoring and evaluation – and the challenges. The characteristics of big data and potential applications for M&E development are discussed. Four sequential **steps (descriptive and exploratory analysis, predictive analytics, detection and evaluation/prescription)** are identified for implementing a big data strategy. These are derived from the broader data analytics field, but the chapter focuses on how these four steps could apply to development and particularly to monitoring and evaluation, followed by examples.

Some of the common methodological and logistical challenges (design, data collection and data analysis) facing development monitoring and evaluation are identified. How big data can contribute to

addressing these problems is also discussed. The chapter concludes by identifying some new methodological challenges using big data, such as: comparability over time, biases introduced through bots, representative and selection bias, spatial auto–correlation and attribution and spurious correlation. There are also a number of important political, ethical and logistical issues concerning big data.

While most of the discussion of potential applications of data analytics focuses on summative (impact) evaluations, it is important to recognize that there are at least four different types of development evaluations: policy and programme evaluation, formative evaluation, developmental evaluation and summative evaluation. Each of these has different purposes, addresses different questions, and often uses different methodologies. Big data can make important contributions to all four.

PART II GUIDELINES FOR INTEGRATING BIG DATA INTO THE M&E FRAMEWORK OF DEVELOPMENT PROGRAMMES.

Chapter 3: Integrating big data into the monitoring, evaluation and learning system. Many agencies now include learning as a part of their M&E strategies, recognizing the need to ensure that lessons from their studies are systematically disseminated and used. Consequently, the report will refer to MEL (Monitoring, Evaluation and Learning) rather than simply M&E. Seven stages of a typical project/programme cycle are identified and key monitoring and evaluation activities are discussed for each stage.

A framework comprising of a six–step approach is then proposed for incorporating big data into the MEL cycle.

The chapter concludes by identifying a number of evaluation best practices, which must also be considered when using big data. These are classified into approaches concerning design, data collection, sample selection and data analysis and dissemination.

Chapter 4: Building big data into programme monitoring. The chapter discusses the main uses of a programme monitoring system, which include: producing data for a results framework, accountability, proposing actions to address problems identified during project implementation, identifying negative outcomes and groups which are not receiving programme services and benefits, providing inputs to programme evaluation and providing inputs to the evaluation of complex programmes.

Focus is also given to the identification of the limitations of current data sources. Examples of big data and ICT tools that can help address the limitations of current monitoring systems are presented.

The chapter concludes with an eight–step process for integrating big data into programme monitoring.

Chapter 5: Building big data into programme evaluation. The elements of a dynamic programme evaluation system are described, and examples are given of how big data could strengthen evaluation at each stage of the programme cycle.

The chapter discusses three main ways that big data can be integrated into a programme evaluation: (I) incorporating big data indicators into a conventional evaluation, (II) using big data to strengthen a conventional evaluation design and (III) using a big data integrated design based on larger amounts of data than can be handled by a conventional design.

Furthermore, the chapter describes six types of widely used evaluation designs, all of which can potentially be used for big data–inclusive evaluations. These designs include: experimental and quasi–experimental, statistical, theory–based, case–based methods, participatory methods and review and synthesis. In addition, two complexity–inclusive designs are included for the evaluation of complex programmes.

A number of additional design issues are identified. These include: the need to understand the time trajectory over which project impacts are expected to be achieved (trajectory analysis); special issues and challenges in the evaluation of complex programmes; the importance of sustainability analysis; and equity–focused evaluation.

The chapter concludes with the presentation of case studies, illustrating how big data was incorporated into each of the eight evaluation designs described earlier.

Chapter 6: Managing big data–inclusive evaluations. This chapter stresses the critical role of the evaluation manager in ensuring that all evaluations address the key questions of concern to stakeholders and that the kinds of information generated can be used by a wide range of stakeholders and for different purposes.

The critical functions of the evaluation manager are outlined and discussed. The chapter emphasizes the key role of the evaluation manager in building the trust and common understanding required to conduct a multi–disciplinary evaluation.

The chapter concludes with a checklist identifying the roles and responsibilities of the evaluation manager at each stage of the programme cycle.

PART 1

DEVELOPMENT EVALUATION IN THE AGE OF BIG DATA

CHAPTER 1
THE DATA REVOLUTION: IMPLICATIONS FOR INTERNATIONAL DEVELOPMENT

'Data are the lifeblood of decision–making and the raw material for accountability. Without high–quality data providing the right information on the right things at the right time, designing, monitoring and evaluating effective policies becomes almost impossible.'
(United Nations IAEG on a Data Revolution, *A World that Counts*, 2014)

1.1 THE DRAMATIC EXPANSION OF DIGITAL TECHNOLOGIES

The world today is more connected, interdependent, and data–rich than at any time in human history. Exponential growth in the volume of data produced globally means that 90 per cent of all the data in existence today – back to the invention of the Phoenician alphabet – has been generated during the past two years alone. The explosion of digital services over the past decade has allowed many new actors to become producers, owners and consumers of data. Between 2005 and 2015 the number of Internet users has more than tripled from 1 billion to 3.2 billion, and more households now own a mobile phone than have access to electricity or clean water[1]. This explosion of information and its many new applications for development is often referred to as the Data Revolution. (Box 1–1).

An important consequence for development is that more data is becoming available on *difficult–to–access populations*. One example is the recent census conducted in the Islamic Republic of Afghanistan by combining an on–going demographic survey, satellite imagery, other remote sensing data, urban data and geographic information system (GIS) statistical modelling. Data analytics was used to integrate the different data sources into a common platform, which was then used to generate information on the country's population. (Box 1–2).

The growth of digital data brings also a trove of *real–time information* on many issues including the cost of food, availability of jobs, access to health care, quality of education, and reports of natural disasters. The 2016 World Development Report is dedicated to the analysis of these 'digital dividends', the potential benefits they offer and the major challenges of the continued digital divide (World Bank, 2016). While the potential benefits relate to inclusion, efficiency and innovation (Box 1–3), the risks

BOX 1–1 THE DATA REVOLUTION AND THE IMPLICATIONS FOR INTERNATIONAL DEVELOPMENT

Since it was coined in 2013, the data revolution has come to mean many things to many people. The United Nations Secretary–General's Independent Expert Advisory Group (IAEG) on *A Data Revolution for Sustainable Development* takes it to mean the following:

- An explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the 'internet of things', and from other sources, such as qualitative data, citizen–generated data and perceptions data;
- A growing demand for data from all parts of society.

The data revolution for sustainable development is:

- The integration of these new data with traditional data to produce high–quality information that is more detailed, timely and relevant for many purposes and users, especially to foster and monitor sustainable development;
- The increase in the usefulness of data through a much greater degree of openness and transparency, avoiding invasion of privacy and abuse of human rights from misuse of data on individuals and groups; and the usefulness of data in minimizing inequality in production, access and use of data;
- Ultimately, more empowered people, better policies, better decisions and greater participation and accountability that will lead to better outcomes for people and planet.

Source: United Nations IAEG on a Data Revolution, *A World that Counts*, 2014 (cited in Jackson, 2015)

include control, inequality and concentration. These are the same challenges this report will address in the chapters dedicated to the data revolution.

Digital technologies such as mobile phones, GPS devices, sensors (to name but a few) have been used extensively by communities, NGOs, governments, and international development agencies to improve the delivery of their projects and programmes.[2] These developments are closely linked to the equally exponential growth of big data and smart data analytics, which can provide information that would have been unimaginable even a few years ago. Such new approaches can help to identify development needs, provide early–warning signals on potential emergencies or crises, plan, implement and evaluate development programmes.

It is likely that in the near future many of the data sources used for programme M&E will be generated passively through the use of new technologies, rather than being collected through the stand–alone M&E studies that are commonly used today. Future M&E systems are likely to be closely linked to broader systems encompassing programme identification, design and management.

However, despite the development of these technologies, there are a number of challenges that can limit the extent to which the promise is fulfilled. It is not yet clear to what extent these technologies can

BOX 1–2 USING NEW INFORMATION TECHNOLOGY TO CONDUCT A NATIONAL CENSUS IN A COUNTRY WITH HIGH SECURITY CONCERNS. THE 2016 ISLAMIC REPUBLIC OF AFGHANISTAN CENSUS

The last census in the Islamic Republic of Afghanistan took place in 1979, and due to security concerns, no recent censuses could be conducted. The United Nations Population Fund (UNFPA), in collaboration with Flowminder, was able to generate maps of the population combining an on–going demographic survey, satellite imagery, other remote sensing data, urban data and GIS statistical modelling.

Source: Flowminder, 2016. Cited in UNFPA, 2016

promote a more inclusive social development framework in which benefits reach the poorest and the most vulnerable groups, or the degree to which these groups will be involved in the decision–making process and empowered to hold governments and development agencies accountable. The concern is that, unless adequate political, regulatory and social controls are put in place, these benefits will still serve only a portion of the population. There is also a concern that new information technologies will be used extractively by governments, large development agencies and corporations, resulting in poor and vulnerable groups having less, rather than more, information and control over decisions and policies affecting their lives[3,4]. Finally, unrecognized biases arising from methods of data collection, analysis or usage by policy makers can lead to all manner of unintended harmful consequences.

This report addresses the question of how big data and ICTs can be used to strengthen the systems used by governments, development agencies and civil society to monitor and evaluate the effectiveness of development programmes in an open and participatory manner.

1.2 MONITORING AND EVALUATION IN THE NEW INTERNATIONAL DEVELOPMENT CONTEXT: OPPORTUNITIES AND CHALLENGES

In 2015 the UN General Assembly unanimously approved the Sustainable Development Goals (SDGs)[5]. The 2030 Agenda defines the vision and goals of national governments, civil society and the international development community for the period 2015–2030 for creating 'the world we want' through a simultaneous focus on People, Planet, Prosperity, Peace and Partnership, in which 'No–one is left behind.' The 2030 Agenda includes 17 broad goals and 169 targets with complex patterns of interaction among them. They also involve multiple donor agencies, implementing partners and stakeholders at national,

regional and local levels. The SDGs reflect the expanding goals and the increasing scope and complexity of international development.

The SDGs illustrate the challenges facing the international community when developing a framework for the evaluation of increasingly broad and complex development initiatives. For example, the goal of SDG–4 is to 'Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.' In addition to the challenges of determining whether 'equitable education services' have been provided, the education outcomes can be significantly affected by the general state of the economy, the quality of infrastructure, access to health services and safe drinking water. Box 1–4 illustrates some of the

BOX 1–3 THE 'DIGITAL DIVIDENDS' FROM THE DATA REVOLUTION

The growth of new technologies and new sources of data offers the potential of a number of important dividends for development:

- *Inclusion:* Connectivity reduces costs of transactions making it possible for small and isolated producers (e.g. small farmers, fishermen and informal businesses) to have access to markets. Access to information, ranging from market prices to education is another example. Technology also makes it easier to identify and communicate with socially and economically excluded groups, to ensure their access to development programmes, and to give them voice in defining development priorities.
- Efficiency: Economic and financial transactions become cheaper, faster and more convenient. In addition to benefits to business and government, it also represents major savings of cost and time for low–income populations to access information and to complete transactions with public authorities (such as reporting service break–down, paying bills and conducting financial transactions).
- *Innovation:* In many areas transaction costs are reduced to almost zero through search programmes, e–commerce platforms and digital payment programmes. Auction sites such as eBay are another example. Innovation benefits all groups, including agencies delivering services to low–income groups or working in emergency programmes, and low–income and vulnerable groups themselves.
- *Increase in the generation and use of new sources of data.* The costs of generating data are falling, the types of data are increasing exponentially, and the ability to integrate different sources of data and to find patterns that could previously not be detected is also increasing.
- *Empowerment, voice and security:* Technologies such as mobile phones permit people to communicate more easily, to develop social networks, to organize politically, to make their voices heard and to hold governments accountable. These techniques have been used to target women, youth and different language and ethnic groups. Tools also empower staff in development agencies to be more informed about the organization and to share their views and ideas. They help in preventing violence against women including domestic violence, youth violence and peer pressure within schools.

Source: Adapted by the author from the World Bank, *World Development Report 2016: Digital Dividends,* 2016.

new evaluation challenges that these programmes pose, including, but not limited to: (a) the development of a complexity–responsive evaluation framework, (b) managing massive amounts of data from multiple sources, (c) working with real–time data, (d) assessing programme sustainability over long periods of time, (e) measuring processes of behavioural change, and (f) applying new data analytic approaches to the evaluation of programme outcomes and impacts.

BOX 1–4 CHALLENGES FOR THE EVALUATION OF NEW AND COMPLEX INTERNATIONAL DEVELOPMENT PROGRAMMES

The following are some of the challenges that must be addressed in the evaluation of the development interventions that are now being implemented. While none of these evaluation challenges are completely new, they must often be addressed on a larger scale and at a greater speed:

- The multiple interventions in multiple contexts and with multiple stakeholders involved in large and ambitious programmes such as poverty reduction, rural development as well as programmes designed to contribute to the achievement of the SDGs, make these development programmes complex, and their evaluation requires the use of new, and still evolving complexity–responsive evaluation methods.
- The amount of information that is now available on development programmes and the contexts within which they operate has grown incrementally in recent years. Today most of this information comes from a much wider range of sources. The volume of information is beyond the computing capacity of conventional computers, and the complexities introduced by the need to assess interactions among many components also requires new forms of data analytics.
- Information is now becoming available in near real–time, which requires new technologies for the collection, dissemination and use of this information, and new organizational processes and policies.
- The increased focus on sustainability requires that programmes be assessed over a much longer period of time (the SDGs for example have a 15 year time–horizon). This requires the development of innovative technologies to economically collect information far beyond the project implementation period.
- Monitoring and evaluation data typically demands high quality standards to be acceptable. This can cause evaluators to reject or ignore new sources of data that could potentially provide valuable insights.
- Many programmes seek to produce complex processes of behavioural change, which often require innovative mixed–method approaches for the capture and analysis of new forms of data.
- Data analytics has developed new approaches to impact evaluation using predictive modelling that employs a fundamentally different approach (based on Bayesian probability analysis) from the experimental (randomized control trial) methodologies generally used by development evaluators. These approaches have not yet been (widely) applied in development evaluation, but they can potentially require a rethinking of evaluation approaches.
- National statistics offices are often under–staffed or most of their resources are committed to conducting conventional surveys.

There is increasing recognition that current evaluation methodologies are not well suited to evaluate the outcomes of these complex development programmes. Even before the launch of the SDGs, the evaluation community recognized the limited ability of current evaluation methodologies to gauge complex development programmes (Bamberger, Vaessen and Raimondo, 2016; Patton, 2011; Byrne, 2013; Byrne and Ragin, 2009; Funnell and Rogers, 2011).

All of these factors are creating a greater demand for new complexity–responsive evaluation designs that are also flexible, rapid and cost–effective. At the same time, the rapid and exciting developments in the areas of new information technology are creating the expectation that the reduced cost and ease of collecting and analysing larger and more complex kinds of data is rapidly increasing. There is also a belief among many in the big data community that it will soon be possible to complement (some would argue to replace) current development evaluation methodologies with tools and techniques for data collection, data analytics and prediction that are now widely used by the business community and that are being applied by many new data analytics consulting companies and universities.

The purpose of the present report is to assess the potential contribution of big data, complemented by new information technologies, to developing monitoring and evaluation frameworks that are responsive to the nature of new development programmes and the complex environments in which they operate.

1.3 DEFINING BIG DATA

While big data is sometimes referred to as 'a collection of large volumes of data,' in fact, from the perspective of international development, big data is an 'integrated approach to research and development' (including development evaluation) that involves three interlinked components (Figure 1–1):

- *Data generation:* Generation and collection of large volumes of data.
- *Data analytics:* Organization and integration of multiple sources of data, and the application of data science and data analytics to find previously undiscovered patterns and associations in the data, and to predict outcomes of development interventions. A key element is the presentation of the findings of the analysis in a user–friendly format (data visualization).
- *Data ecosystem:* An ecosystem that links the multiple organizations and individuals that generate, analyse and use big data. There is also a continuum that bridges big data and small data. Most development agencies that use big data are likely to combine this with the kinds of information they are already collecting, and to combine big data analytics with conventional quantitative and qualitative analysis. Finally, successful applications of big data tend to include critical interactions between humans and the big data digital technology.

DIMENSION 1: DATA GENERATION

Smart technologies frequently involve the collection of vast quantities of data, which have a volume, velocity and variety unattainable a couple years ago.

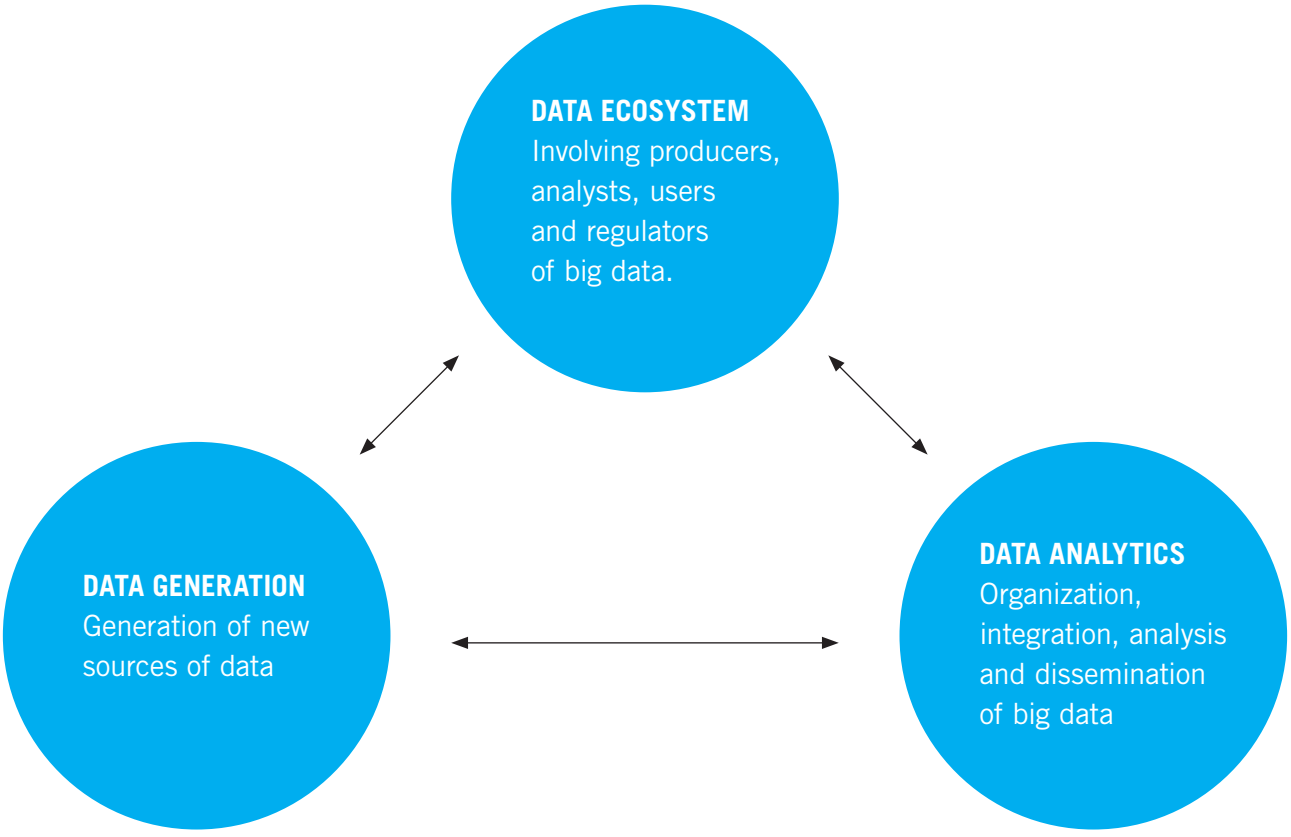
THERE IS NO UNIVERSALLY ACCEPTED DEFINITION OF BIG DATA, BUT MOST DEFINITIONS INCLUDE THE FOLLOWING ELEMENTS:

- Data that is huge in volume and generated very fast;
- Data sets that are so large or complex that they require access to very large serves; and cannot be analysed using conventional data analysis systems; Most big data is considered 'passive' in that it is generated automatically and for a purpose other than the research, monitoring or evaluation applications to which it can be applied;
- Data can be relational in nature: containing common fields that permit the integration of different kinds of data;
- Data that are more granular and permit more detailed disaggregation; Exhaustive in scope, striving to capture entire populations or systems;

DIMENSION 2: DATA ANALYTICS

Data analytics involves the organization and integration of multiple sources of data and the identification of previously undiscovered patterns and associations in the data. Data analytics can be understood in terms of (Figure 1–2):

FIGURE 1–1 THE THREE DIMENSIONS OF THE BIG DATA FRAMEWORK



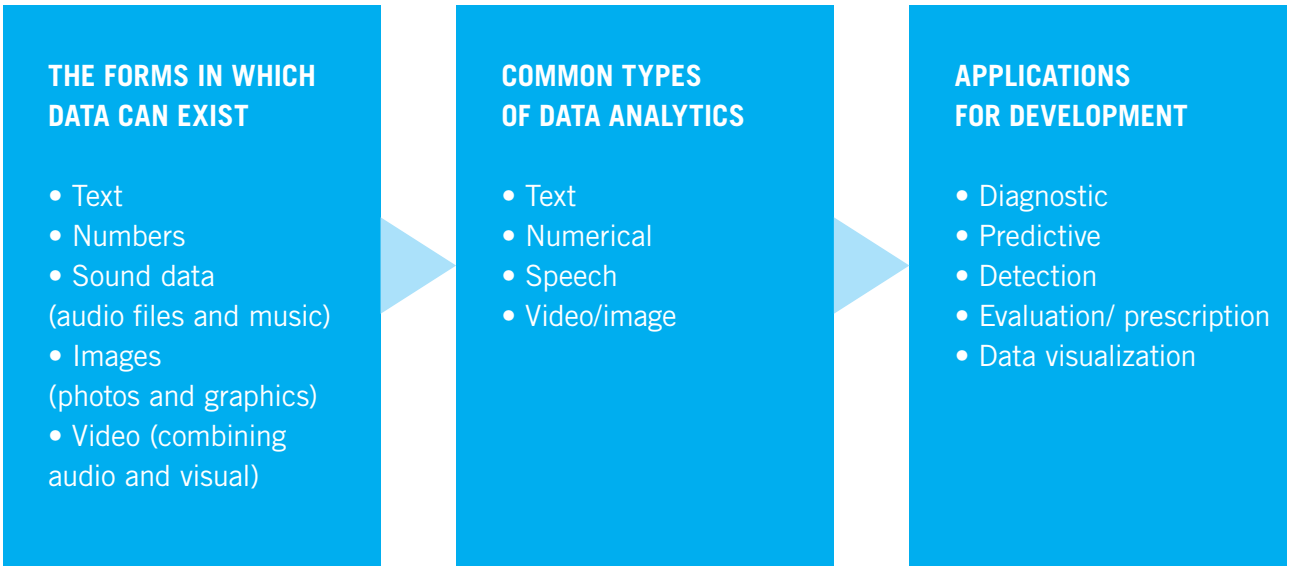
- The forms in which data can exist;
- The main types of analytics;
- The applications for development planning and evaluation.

There are four common types of unstructured data that are generated daily in large quantities today: text, numbers, images, audio/ video. Text can include, among others, PDF documents (including the very large number of reports that have been generated by most development agencies), tweets and other social media documents. Examples of audio data analysis will be given, including phone conversations and radio programmes. Images can include photographs and graphical data. Videos combine audio and visual formats/forms. For example, images captured by smart phones can be used to analyse interactions and processes during different types of household or community activities.

For each of these types of data a set of relatively mature analytical procedures is available. For example, tools for the analysis of text include: text categorization, text clustering, concept extraction, sentiment analysis and document summarization. Speech analysis is another area, which is evolving rapidly, and in addition to the analysis of content, it is also possible to analyse the emotional content of the speech: modern call centre staff may now be notified automatically when a caller's tone becomes irate. Video and image analytics now permit facial recognition, traffic flow analysis, and rapid damage assessment following natural disasters.

Chapter 2 discusses some of the most important applications of data analytics in the development field; descriptive and exploratory analysis, predictive analysis, detection of outliers or groups likely to fail, evaluation prescription and promoting utilization of findings through effective dissemination of findings and data visualization.

FIGURE 1–2 OVERVIEW OF DATA ANALYTICS APPLICATIONS FOR DEVELOPMENT
Source: Adapted From Marr, 2015



DIMENSION 3: THE DATA ECOSYSTEM

Some authors (including Letouzé), argue that big data is not so much 'data sets that are impossible to store and process using common software tools, regardless of the computing power or the physical storage at hand', but rather, 'non–sampled data, characterized by the creation of databases from electronic sources whose primary purpose is something other than statistical inference.'(Letouzé el al, 2016). Therefore, size is only one of the determinants.

However, while some definitions focus on the types of data that are generated, others talk about the big data approach and the community or ecosystem (people, institutions, emitters of data, analysts and users of data) within which the generation, analysis, dissemination and use of big data takes place. A mindset that can make sense of the overwhelming volumes of data is also critical. The concept of community is important because the interpretation of big data requires blending of techniques from different cultures including business, the social sciences and the computer sciences. Big data and data analytics transform disorganized data into actionable information. Consequently, an effective interaction between data users and data analysts is critical. Big data is often combined with, or validated through mixed method techniques and other qualitative approaches that involve human interaction. Some agencies that work with these new data sets prefer to talk about data science approaches and do not use the term big data.

While some forms of data such as satellite imagery have been around for decades, the recognition of the potential applications of big data for the monitoring and evaluation of development programmes may only have begun in 2012 with the publication of the UN Global Pulse White paper on Big Data for Development (UN Global Pulse, *Big Data for Development*, 2012). Therefore, the ideas are still very

BOX 1–5 THE DATA CONTINUUM: FROM SMALL TO BIG, AND BIG TO SMALL

Consider an hypothetical big data project where some data was structured (information in a set format like a spread sheet) and entered using forms. Some data was unstructured free text, and some was financial transaction records. If you analysed each data type in isolation with traditional methods, it would be considered as 'small' data analysis. If you took the data sources and integrated them, then analysed the data as a whole using techniques that could pull insights from both structured and unstructured information simultaneously, then this could be considered the realm of 'big' data analytics. Why? Because it is complex enough that traditional analysis techniques would not be able to analyse the data as a whole.

On the other hand, it is possible to analyse big data using small data analytical methods. A Global Pulse project analysed social media data (in particular Twitter data) to understand online conversations about communicable diseases in Indonesia. An algorithmic approach was used to filter and conduct basic sorting of a relatively small number of Tweets, which were then manually analysed. (Jackson, 2015)

new, particularly in the field of international development and many development agencies are still in the process of defining their big data policies and potential applications. (Box 1–5)

There are also different kinds of analytical procedures for data analysis at each level, but again with considerable overlap and interaction (Figure 1–3). One example is where several different small and large data sets are combined to create an integrated data platform (such as FAO's national water resource data bases). Although each individual data set can be analysed using small or large analytical procedures, the integrated database will often require the use of big data analytics. It is also possible to complement big data analytics with small, often qualitative analysis.

Finally, there are many instances in which digital data collection and analysis will be strengthened when combined with human intervention. One important example of digital–human interaction that will be discussed in the report applies to understanding emergencies or crises situations using sentiment analysis of online conversations. Beyond the fact that social media content is largely perceptual in nature, such data may be biased, or otherwise of poor quality. For example, it may have been obtained from analysis of mobile phone usage patterns within a given mobile operator's network, which always represents a biased sample of the total population; or it may come from the analysis of tweets where key words are used to infer that a crisis is underway. A follow up human investigation is usually required to check on the validity of early warning signals[6].

1.4 CURRENT APPLICATIONS OF BIG DATA IN INTERNATIONAL DEVELOPMENT

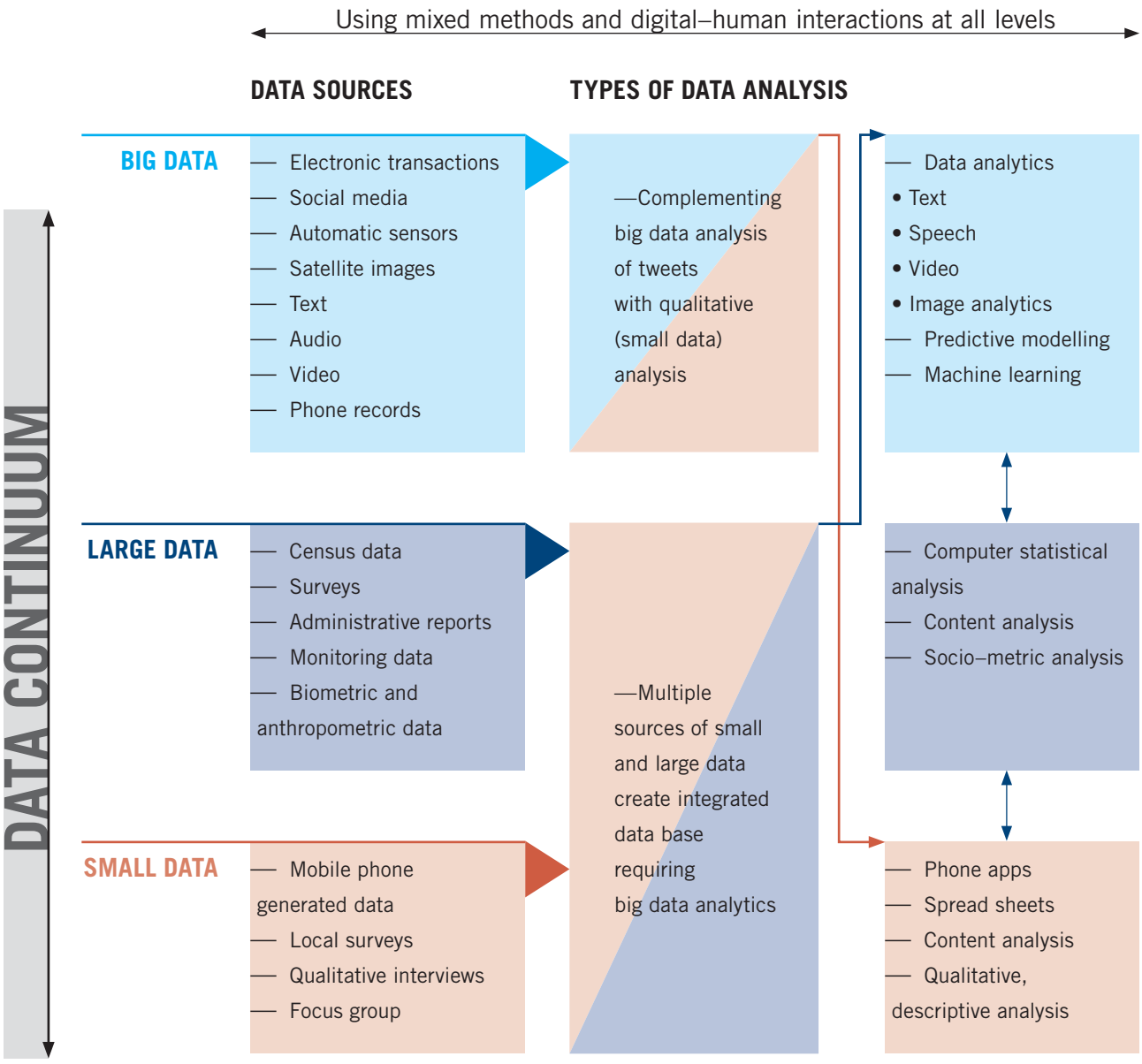
Many forms of big data and data analytics have only come on the scene within the past two to three years, many agencies are still in the early stages of understanding big data and its potential applications in development. However, a number of agencies have already started to apply the big data approach in development research and programme design, monitoring and evaluation. Below are some examples of how big data is currently being used in international development. These examples are only a few of the many different big data applications that are already being tested or that are still at the planning stage.

- Using data analytics to predict the characteristics of sub–groups, like for example school dropout rates, job training or other social welfare programmes (see Chapter 2 Section 2.3);
- Analysis of Twitter and other social media to assess the attitudes of different groups to social problems and issues or their response to different preventive or educational programmes.

The cases in Table 1–1 illustrate:

- *Integrating multiple sources of data into a data platform:* Big data analytics can create standardized data categories into which many different types/sets of information can be fitted so that data are comparable over time and space. The examples cite the compilation of multiple sources of information on national water resources, and the creation of a data platform integrating

FIGURE 1–3 THE DATA CONTINUUM



* Mobile phones and ICT used extensively for data collection

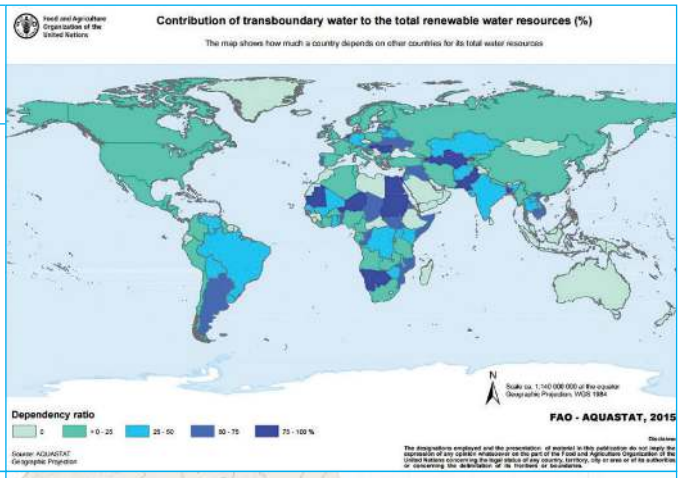
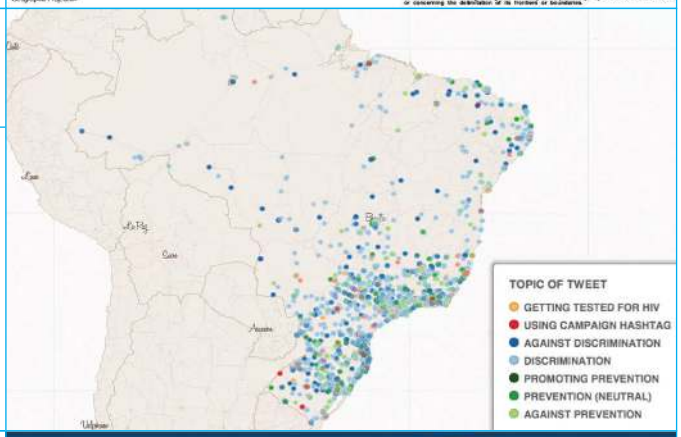
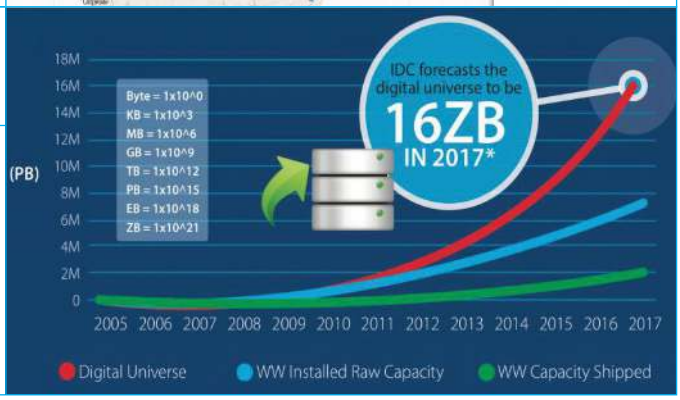
multiple sources of information on disease epidemics such as Ebola.

• *Mapping:* Data visualization is a powerful tool that is used to illustrate the current situation with respect to access to services, crop disease, electoral fraud or other phenomena with a geographical distribution. The examples cite the mapping of for example Ebola outbreaks, the quality of crops and the spread of crop diseases, location of victims in earthquakes, the location and spread of forest and peat fires and the location of rural poverty in China.

• *Monitoring trends:* Big data is often used to understand trends, as multiple sources of data are

TABLE 1–1 EXAMPLES OF HOW BIG DATA CAN BE USED FOR DEVELOPMENT AND HUMANITARIAN ACTION

Applications of big data that do not relate directly to monitoring and evaluation

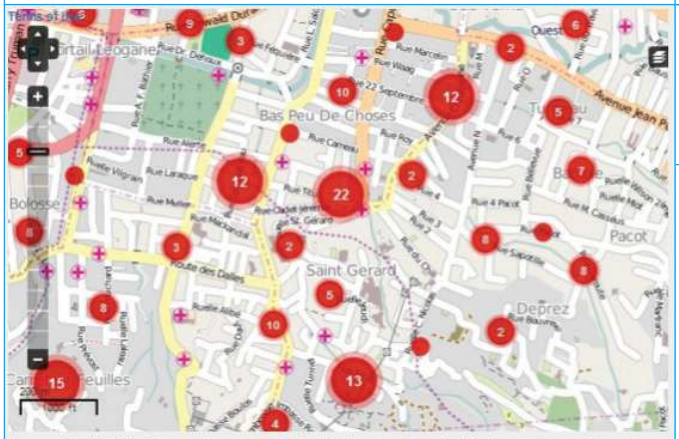
1. SYNTHESIS OF AVAILABLE DATA ON WATER RESOURCES. AQUASTAT is FAO's global water information system, which collects, analyses and disseminates data and information on water resources, water uses, agricultural water management and other information. It combines satellite and other sources of data. [Source: FAO][7] Aquastat platform. Source FAO. >	
2. USING SOCIAL MEDIA TO EXPLORE HIV–RELATED STIGMA. As part of UNAids' Protect the Goal campaign to raise awareness of HIV and Aids during the World Cup in Rio, the project explored whether tweets could be used to measure HIV–related stigma. The goal was to determine whether discrimination makes people less likely to access health services such as condoms, HIV tests and antiretrovirals. [Source: The Guardian][8] HIV tweets in Brazil during World Cup >	
3. MAPPING POVERTY IN CHINA USING CALL DATA RECORDS. Mining phone data to develop proxies for poverty indicators, which could provide a much more economical and continuous source of data on poverty trends. [Source: UNDP China][9] Global digital data created and storage capacity Source Where in the World is Storage. International Data Corporation (IDC) Infographic >	

Among people with homes in Kathmandu Valley, an estimated 500,000 people more than expected had left Kathmandu Valley.

4. MAPPING POPULATION DISPLACEMENT USING MOBILE PHONE DATA.

Reports on population displacement in Nepal following the April 2015 earthquake were produced by Flowminder/ WorldPop based on mobile operator data. Reports were provided to the UN Office for the Coordination of Humanitarian Affairs (UN OCHA). [Source: Flowminder]

< Mapping population displacement using mobile phone data. Source Flowminder.



Close up of the Haiti Map. Each number represents the individual number of reports within the area. Users could zoom in further to see the individual reports. Credit: Ushahidi Haiti Project (UHP).

5. USING SOCIAL MEDIA TO GUIDE EMERGENCY SERVICES IN THE AFTERMATH OF THE HAITI EARTHQUAKE.

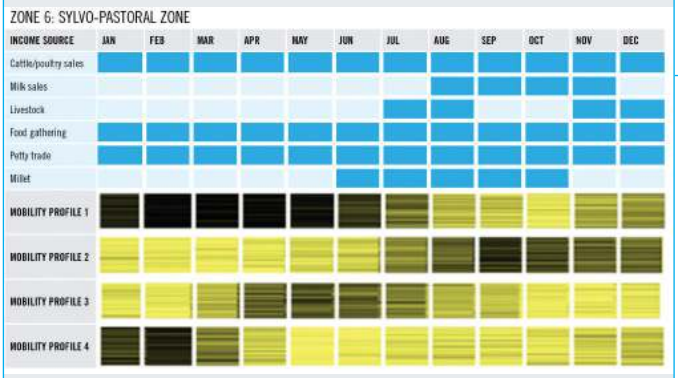
A group of volunteers analysed information from Facebook and Twitter, and later SMS messages relating to victims of the earthquake. The information was located on a rapidly constructed crisis street map. More than 1.4 million edits were made to this map as information was refined. (Source: Patrick Meier (2015), *Digital Humanitarians*.)

< Close up of the Haiti Map. Source Ushahidi Haiti Project (UHP)

6. USING SOCIAL MEDIA TO DETECT AND MANAGE FOREST AND PEAT FIRES IN INDONESIA.

A crisis analysis and visualization tool that provides real–time situational information from various sources of data to enhance disaster managements efforts in regions affected by forest fires and haze. (Source: UN Global Pulse)[10]

< Using social media to detect and manage forest and peat fires in Indonesia



7. USING MOBILE PHONE DATA TO TRACK SEASONAL MIGRATION IN SENEGAL.

The movements of populations in Senegal were quantified using anonymiszed mobile phone data. Movement patterns among populations groups were extracted and visualized, which resulted in a series of mobility profiles from different regions of Senegal. (Source: World Food Programme and UN Global Pulse)[11].

< Annual calendar of income generating activities mapped against the four most representative mobility profiles of population subgroups obtained from mobile phone data in the sylvo–pastoral livelihood zone in 2013

often available at frequent intervals (sometimes on a daily or weekly basis). The examples cite: trends in discrimination against women in the workplace in Indonesia, trends in rural poverty in China, and seasonal migration in Senegal.

• *Real–time early–warning signals:* Data analytics can provide information on trends in almost real–time based on, for example, the analysis of social media. This information can provide early warnings on hot spots of drought, hunger, disease or ethnic conflict. The above examples refer: gender discrimination, female discrimination in the workplace /or formal and informal discrimination against women at work, outbreaks of epidemics such as Ebola, outbreaks of forest fires, crisis mapping in war zones.

Tools like the ones cited above, are now beginning to be used in development programmes and emergency management. However, fewer examples of how big data is being used for programme evaluation currently exist, although there are more cases of applications for programme monitoring. The following section uses the example of UN Global Pulse to illustrate how pilot projects have been developed to test the feasibility of applying the big data approach in a wide range of development fields. The challenge is now to find ways to scale–up and operationalize these models.

1.5 THE EXPERIENCE OF UN GLOBAL PULSE: MOVING FROM PROOF–OF–CONCEPT TO OPERATIONAL USE

In the previous section the report discussed examples of big data applications and data analytics, which are now being widely used in development research and programme planning and design. In the last years UN agencies, multilateral development banks and bilateral development agencies have launched initiatives to explore the potential applications of new sources of data and develop operational applications, including M&E. However, many of these initiatives are managed by researchers with a background in statistics or data science and evaluation department are not directly involved during the initial stage of projects.

The following chapters discuss how monitoring applications are being introduced. However, it is still difficult to find examples where big data is being used for programme evaluation (although a few examples are presented). The challenge is to build on these initial experiences and to develop operational systems for utilizing the power of big data and data analytics to strengthen programme M&E and to apply the approaches to the emerging challenges of how to evaluate complex development initiatives.

UN Global Pulse is an innovation initiative of the United Nations Secretary–General on harnessing big data. Its mission is to accelerate discovery, development and adoption of big data innovation for sustainable development and humanitarian action. Through its Pulse Labs in New York, Jakarta and Kampala, Global Pulse identifies high–potential approaches, designs and develops analytics solutions,

and conducts pilot–based evaluations with users drawn from governments, UN agencies and civil society partner organizations. Table 1–2 gives examples from Asia and Africa of technologies that have been successfully tested for a wide range of development research and operational initiatives. These will be referred to in later chapters, which describe the operationalization of big data based M&E systems.

A number of potential evaluation applications are being now tested by UN Global Pulse.

1.6 COLLABORATION BETWEEN DATA SCIENTISTS AND DEVELOPMENT EVALUATORS IS KEY TO SUCCESSFULLY APPLYING NEW TOOLS AND TECHNOLOGIES

The interviews conducted during the preparation of this report revealed that one of the challenges to integrate big data into development M&E is that most data scientists operate within a very different research and management paradigm than do most evaluators. Both groups use different terminology and have a very different approach to questions; such as how to evaluate interventions, the nature of data and how to assess the quality and utility of different types of data, approaches to data analysis and the role of theory. The limited familiarity of each group with the approaches used by the other sometimes means that questions that could easily be clarified on the basis of a discussion can become confrontational or misunderstood.

'At least on the academic side, the two communities have very different traditions, and generally approach problems very differently. I don't think this necessarily, but it does mean that you cannot just put a social scientist and a data scientist in a room and assume that magic will ensue. There are some very real obstacles that stand in the way of collaborations happening at scale, but also some value that each side can bring to the table. As long as they can learn to get along.' (Source: Josh Blumenstock, the Director of the Data science and Analytics Lab at the University of Washington. Quoted in Catherine Cheney, 2016)[12].

Quoted in the same article, Emmanuel Letouze, Director of the Data–Pop Alliance, presented the broader picture: 'What statisticians, demographers and economists need to realize is that data science is not just a fad, and what computer scientists and engineers need to acknowledge is that they cannot solve global poverty by crunching numbers alone.'

Given these genuine differences of approach and important methodological issues requiring discussion, there is a need for bridge–building to create a space for development of a common understanding (See Figure 1–4). This would seem to be an important step in the process of assessing to what extent and how different big data and data analytics approaches can be integrated into development M&E. The following are some of the issues that need to be addressed:

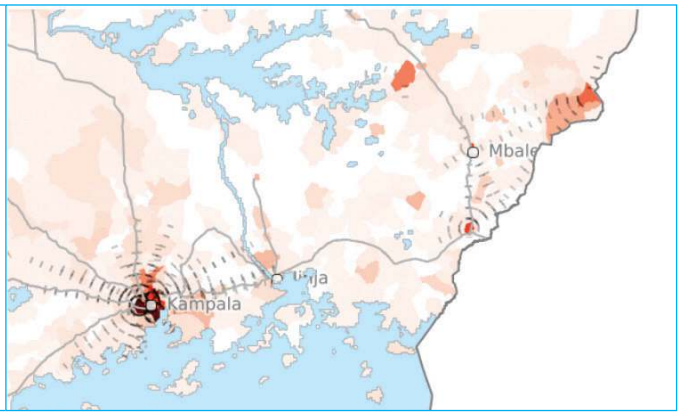
TABLE 1–2 BIG DATA APPLICATIONS FOR SUSTAINABLE DEVELOPMENT FROM UN GLOBAL PULSE

1. USING CROWDSOURCING FOR UNDERSTANDING IN REAL TIME TRENDS IN COMMODITY PRICES (IN COLLABORATION WITH WFP AND FAO)

Price data for vegetables suggesting that they account for a significant proportion of the upward price movement >

2. USING DATA VISUALIZATIONS AND INTERACTIVE MAPPING TO SUPPORT RESPONSE TO DISEASE OUTBREAKS IN UGANDA (IN COLLABORATION WITH GOVERNMENT PARTNERS)

Visualization of sub–country level typhoid incidence and human mobility from high infected areas >

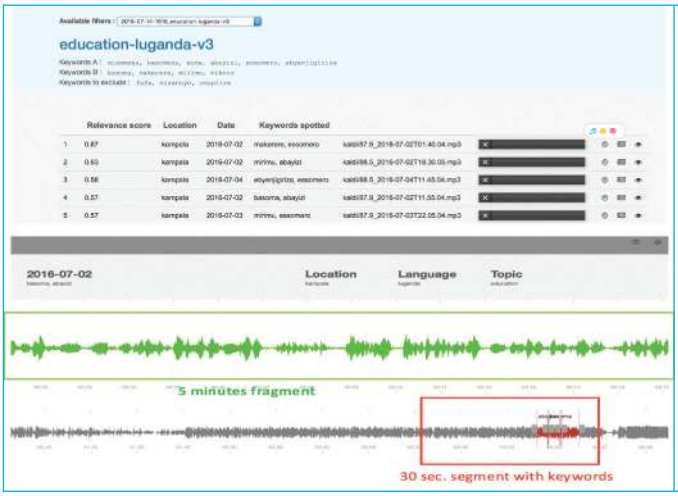


3. USING FINANCIAL TRANSACTION DATA TO MEASURE THE ECONOMIC RESILIENCE OF POPULATIONS TO NATURAL DISASTERS (IN COLLABORATION WITH BBVA BANK)

Visualizing the expected level or transactions in comparison to what really happened as a result of Hurricane Odile >

4. NOWCASTING FOOD PRICES IN INDONESIA USING SOCIAL MEDIA SIGNALS (IN COLLABORATION WITH WFP)

< Fluctuations in price for onion as one of four commodity prices analyzed



5. MINING FM TALK RADIO SPEECH DATA IN UGANDA TO CREATE A BETTER UNDERSTATING OF PEOPLE'S PRIORITIES (IN COLLABORATION WITH GOVERNMENT PARTNERS)

< Dashboard that helps review segments containing relevant keywords

6. EXPLORING THE USE OF SATELLITE IMAGERY IN UGANDA TO TRACK POVERTY TRENDS (IN COLLABORATION WITH GOVERNMENT PARTNERS)

< Visualizing modifications over time of materials roof are made of

7. PROVIDING REAL TIME INSIGHTS ON THE LOCATIONS OF FIRE AND HAZE HOTSPOTS IN INDONESIA USING VARIOUS SOURCES OF DATA (SOCIAL MEDIA, MOBILE PHONES, SATELLITE IMAGERY)

More information on the project available at: http://unglobalpulse.org/projects/haze-gazer-a-crisis-analysis-tool

< Haze Gazer dashboard combines information from social media data and the national complaint system

A. DIFFERENCES IN HOW BIG DATA IS APPLIED IN COMMERCIAL AND IN DEVELOPMENT RESEARCH

Much big data analytics is drawn from commercial applications where it is often sufficient to determine that certain kinds of messages or marketing strategies affect consumers' search or buying behaviour – without needing to explain why. This contrasts with development evaluation, which must understand why a relationship exists and whether, and how, it contributes to the broad development goals of a programme, for example cost effectiveness, who benefits and who does not. Many predictive models accept lower standards of validity and causal analysis because the data is continually being updated and the model revised. In contrast, most development researches require higher standards of inference, as many of the operational decisions based on the findings involve major investments or operational decisions, which are expensive and difficult to change.

B. ASSESSING UNINTENDED OUTCOMES

Another concern of development evaluation is to identify **unintended outcomes**, many of which can have serious consequences (for example, increases in violence against women and domestic violence as a result of programmes designed to promote women's economic empowerment). Development research requires a broader focus and the need to dig deeper.

C. WHAT ARE THE DIFFERENCES BETWEEN THE TYPES OF DATA USED BY DATA SCIENTISTS AND DEVELOPMENT EVALUATORS, AND HOW DIFFERENT ARE THE CRITERIA USED TO ASSESS DATA QUALITY

An important distinction is that data scientists are used to working with data that is generated in real–time and which is updated on a frequent basis. Furthermore, much of the data is generated from sources such as mobile phones or social media posts where there is significant selection bias. Often, the data is of a poor quality where words, whose meaning may not be clear, are used as proxy indicators for concepts such as hunger, ethnic hostility, locating earthquake victims or the early stages of an infectious disease.

In contrast, evaluators spend a great deal of time and effort trying to ensure that their data is unbiased and of a high quality. The important point to understand is that real–time data and evaluation data usually have different purposes. Real–time data frequently provide an important early–warning signal of a potential crisis or emergency (ethnic conflict, disease outbreaks, hunger), which then need to be validated using other methods, often those familiar to evaluators. Once the concept of early–warning signals is understood, then the discussion of data purpose and quality takes on a different perspective.

D. BIG DATA AND M&E DATA ARE OFTEN USED UNDER VERY DIFFERENT DECISION–MAKING TIME HORIZONS

Some applications of data analytics can lead to very rapid decision–making (for example an evaluation of the impact of slight changes in wording of an on–line advertisement on click–rates can result in very rapid changes in the advertisement); whereas in many development programmes monitoring data is often built into a three or six month management decision–making cycle. Consequently, some kinds of real–time data may be difficult to utilize if decision–makers cannot make corrections to their programmes based

on the data for several months. Therefore, understanding the nature of management decision–making processes is important in designing data collection and analysis systems.

E. DIFFERENT APPROACHES TO ATTRIBUTION AND CAUSALITY

Current evaluation approaches consider experimental designs, usually with a pre– test and/or post–test comparison group design to be the strongest quantitative method for assessing the contribution of a given input to observed changes in an intended outcome variable. There are also ranges of other approaches, but for many, the experimental design is considered the methodologically most rigorous. This approach is relatively static in that it is assessing change between two given points in time (usually at the beginning and at the end of a programme). It is also criticized by data analysts for being 'backward looking' as it is assessing changes that took place between two points in the past. Something that is often overlooked in the discussion is that current evaluation approaches do include 'formative evaluations' that could be considered 'forward looking' as their purpose is to provide guidance on how to improve the performance of the on going programme being evaluated.

In contrast, big data analytics tends to advocate predictive analytics based on Bayesian probability models. These draw on all available data to estimate the probability of different future events, so in this sense they are 'forward looking.' However, these models do not include an attribution methodology to assess the extent to which observed changes are due to the programme inputs. Therefore, it is important to discuss and compare predictive and experimental designs and to understand exactly what information each is producing and what lessons can be learned. There should then be a discussion on how the two approaches can complement each other to address a broader range of questions.

F. THE ROLE OF THEORY

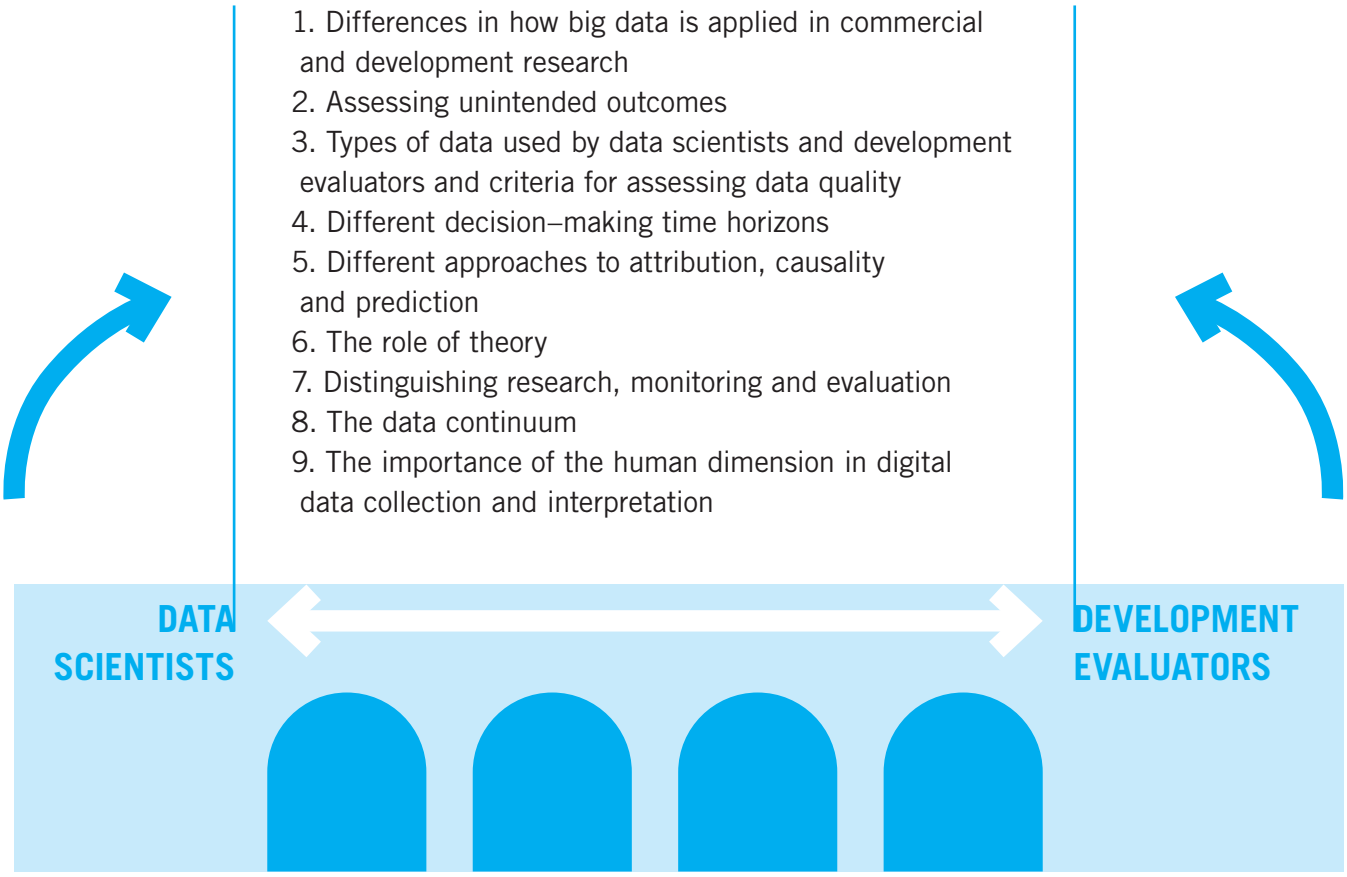
Theory has a less prominent role in data analytics than in evaluation methodology, and in fact some big data advocates claim that big data represents the end of theory. In a provocative and much cited article in *Wired*, Chris Anderson (2008) argued that the 'data deluge makes the scientific method obsolete' and presages the 'end of theory.'

Petabytes of data allow us to say: 'Correlation is enough. We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot. (p. 4)'

However, subsequent discussion argues that this is an oversimplification, that all data analytics is based on some kind of model and that in social, as opposed to natural science contexts, theory is required to define what kinds of variables will be included in the analysis and to give meaning to the findings. However, while it is clear that operating with petabytes of data, multiple variables that interact in complex ways, and with data analytical power big data analytics radically changes the approach to evaluation design and analysis.

Some techniques such as data mining troll through large data sets to identify correlations between interventions and outcomes of interest without using a theoretical framework to guide the analysis. While it is argued that it is always necessary to have at least an implicit theory to guide the selection of variables to include in the analysis, theory seems to play a less prominent role.

FIGURE 1–4 BRIDGE BUILDING BETWEEN DATA SCIENTISTS AND DEVELOPMENT EVALUATORS



Finding what works without knowing why may be useful in some contexts (such as assessing the effectiveness of different advertising campaigns). Evaluators argue that this approach is not sufficient for development evaluation where the purpose of the evaluation is to understand causal processes and to draw lessons that can be used to strengthen the design of future programmes.

However, issues relating to the appropriate use of theory will be important when strategies are discussed for building big data and data analytics into development monitoring and evaluation systems.

G. DISTINGUISHING RESEARCH, MONITORING AND EVALUATION

In big data analytics the distinction between monitoring and evaluation is often not clarified. It is often assumed that if large volumes of data can be collected cheaply and quickly on a continuous basis, this will automatically provide the basis for programme evaluation. Evaluators argue that even though monitoring data is an essential component of a programme evaluation, it is not in itself sufficient to determine the extent to which programme inputs have contributed to changes in outcomes. The latter requires defining a counterfactual describing what would have been the situation in the absence of the programme. This question is still an area of discussion between data analysts and programme evaluators.

H. THE DATA CONTINUUM

There is a continuum of data collection methods ranging from big data to small data. While the characteristics of each type of data can be categorized, there is considerable overlap. Similarly, there is a continuum of data analysis approaches ranging from the data analytics, which are required to handle big data through the methods used to analyse large and small data. Again there is a considerable overlap. For example, different small and large data sets, when combined into an integrated data platform, may require the use of big data analytics. Further, as data collection and data analytical speed increases, what might be considered big data today could be considered smaller data in the near future.

The implication is that for most practical applications, an integrated approach must be used, frequently drawing on data and analytical approaches at different points on the continuum.

I. THE IMPORTANCE OF THE HUMAN DIMENSION IN DIGITAL DATA COLLECTION AND INTERPRETATION

Finally, it is important to recognize that many applications of digital technology require a human dimension. In some cases this is to validate data, or to dig deeper through the preparation of case studies, in–depth interviews or focus groups. Sometimes a human intervention is needed to prepare the way for digital interventions, such as visiting communities to convince husbands or mothers–in–law to allow spouses or daughters–in–law to participate in a digital survey.

While the debates and disagreements tend to capture the headlines, throughout this report examples of successful and promising collaboration of what can be achieved when data scientists join forces with development researchers and evaluators will be cited. The challenges for building bridges are to identify and strengthen the areas of consensus. Jake Porway, the Founder and Executive Director of DataKind points out that ‘They do have at least one thing in common. Neither group can resist a fascinating question that might help improve the world and that can be a great way to bring them together.’ (Quoted in the previously cited article by Catherine Cheney). This shared interest may provide a good foundation on which to build future cooperation between data scientists, development researchers and evaluators.

CHAPTER 2
THE PROMISE OF BIG DATA FOR PROGRAMME MONITORING AND EVALUATION – AND THE CHALLENGES

The present, and two following chapters identify some of the current and potential applications of big data and data analytics for development monitoring and evaluation.

2.1 THE MAIN TYPES OF DEVELOPMENT MONITORING AND EVALUATION

Before discussing the potential contributions of big data analytics it is useful to classify the different kinds of development evaluation. This is important because much of the discussion on the benefits of big data analytics has focused on the relative merits of experimental evaluation designs and predictive analytics (defined below as 'summative evaluation'). In fact, there are at least four distinct types of development evaluation and big data analytics can potentially contribute to each of these. In order to maximize the benefits of big data and data analytics, M&E systems must be closely integrated with programme design and management systems. Much of the data used for monitoring and evaluation will in fact be generated for the big data systems used for programme design and management. This contrasts with most current evaluation systems, which tend to generate their own data.

Table 2–1 identifies four main types of evaluation and describes the purpose and use of each as well as the stages of the project/programme cycle at which they are used, the common data collection and analysis methods and the broad development goals that each assesses. For the purposes of this typology, monitoring and evaluation are combined. The four types are:

A. POLICY AND BROAD–BASED PROGRAMME EVALUATION

This assesses how well policies and broad programmes (such as country programmes and multi–donor collaborative programmes) are designed and implemented and how well they achieve their development objectives. These evaluations focus on the upstream development and planning. Many of the evaluations are conducted retrospectively, often at the end of a country programme cycle (typically lasting four to five years). Many of these evaluations use the OECD/DAC evaluation criteria (e.g. relevance, efficiency, efficacy, impact and sustainability), but many other policy evaluation methodologies can be used.[13]

B. FORMATIVE EVALUATION

The purpose of formative evaluation is to provide regular feedback to management and other stakeholders to help strengthen the implementation of programmes and projects. There is a close linkage between monitoring and evaluation to ensure that maximum use if made of monitoring as a tool for agile management and not just for accountability. Formative evaluation combines quantitative and qualitative methods, often combined into a mixed–methods approach. There is also a focus on evaluation as a learning tool. Formative evaluation is used throughout the programme and project cycle. The approach is based on close collaboration between management and the evaluation team. Organizational approach is distinct from many summative evaluations, which often stress that 'objectivity' can only be achieved by maintaining a distance between managers and evaluators.

Most of the formative evaluations also include a rights–based approach, which employs qualitative and participatory approach/access through which voice is given to poor and vulnerable groups, and which enables promotion of social justice. Many forms of gender evaluation fall into this category.

For projects that are defined as complex, a special evaluation approach will often be required that uses complexity–responsive evaluation tools. Ideally, complexity–focused formative evaluations will begin at the project identification and design stage and will continue after project completion. However, in practice, the duration of the evaluation will be more limited.

C. DEVELOPMENTAL EVALUATION

In recent years many agencies include Michael Patton's developmental evaluation (2011) as a fourth type of evaluation. This has many similarities with formative evaluation in that the purpose is to help managers and other stakeholders to improve programme performance and to learn lessons for the selection and design of future programmes. Developmental evaluation focuses on innovative programmes and those that operate in complex environments where an adaptive approach to design and implementation must be used. Interventions evolve and adapt and often do not have any completion point. The approach is based on a very close collaboration between managers and evaluators, where the latter are closely involved in programme implementation and adaptation.

D. SUMMATIVE EVALUATION

The purpose of summative evaluation is to assess the extent to which observed changes in outcome variables (the intended project goals) can be attributed to the effects of the project intervention. These evaluations can either be quantitative, estimating the size and statistical significance of the changes, or they can adopt a more qualitative approach – where one of the main sources of evidence is the opinions of the affected population and other stakeholders. Traditionally, summative evaluations have been used for accountability and to provide guidance on the potential replicability of programmes. Traditionally, the main application of summative evaluation was to assess the replicability on a larger scale of pilot programmes.

The most widely used tool for summative evaluation has been Randomized Control Trials (RCT). RCTs are one of the most widely used, and most criticized evaluation methodologies. This is also the evaluation approach most widely criticized by the big data analysts (see Chapter 1 Section 6). An exclusive focus

TABLE 2–1 THE MAIN TYPES OF DEVELOPMENT MONITORING AND EVALUATION

TYPE	PURPOSE/USE	STAGES OF PROGRAMME CYCLE	METHODS	DEVELOPMENT GOALS
A. Policy and programme evaluation	— Assessment whether programmes contribute to broad policy objectives	— Upstream and end–of–project	— OECD evaluation criteria — Pipeline designs and natural experiments	— How effectively do policies contribute to development goals
B. Formative evaluation	— Improving programme selection and design — Agile programme management identification and correction of problems — Evaluating complex programmes	— Diagnostic/ planning stage — Throughout project/programme implementation — Starts during project design and continues after project completion	— All standard evaluation tools with a focus on qualitative and process analysis — Complexity-responsive evaluation	— Empowerment, rights based, and gender–responsive outcomes — How well do programmes achieve goals in a complex world
C. Developmental evaluation	— Designing and implementation of innovative programmes — Learning	— Throughout project/ programme cycle	— All standard evaluation tools can be applied	— Empowerment, rights based, gender evaluation
D. Summative evaluation	— Accountability — Learning — Decisions on future programmes — Evaluating complex programmes and projects	— Baseline — Mid–term review — End–of–project	— Experimental and quasi–experimental — Statistical — Theory–based — Case based — Systematic reviews — Complexity responsive evaluation	— Assessing the quantitative contribution to development outcomes — Assessing replicability and scaling–up

on RCTs is also widely challenged within the evaluation community due to, among other things, a narrow focus on one or a small number of usually quantitative outcomes, a lack of attention to the process of project implementation and to the context within which programmes are designed, implemented and evaluated. RCTs are also challenged by rights–based evaluators who stress the need to listen to multiple voices and who argue that there is no one way to identify or assess programme outcomes. An important development is the 'RCT+' approach, which combines experimental evaluation designs with qualitative approaches (Bamberger, Tarsilla and Hesse–Biber, 2016).

The most important recent development for summative evaluation (as well as for other kinds of evaluation) is the recognition that most programmes must be considered 'complex' due to the nature of the programme itself, the multiple contextual factors that affect design and implementation, the complex relations among the multiple agencies involved, and the non–linear patterns of causality. All of these factors seriously challenge the validity of conventional evaluation designs that assume a linear relationship between programme inputs and outcomes. Consequently, 'complexity–responsive' evaluation designs will often be required.

2.2 IDENTIFYING CHALLENGES TO CURRENT EVALUATION METHODOLOGIES AND AREAS WHERE BIG DATA ANALYTICS MAY BE ABLE TO CONTRIBUTE

Table 2–2 summarizes six sets of challenges that many current evaluation methodologies face. This will provide a framework for discussing potential areas where big data analytics may be able to contribute:

A. UNDERSTANDING THE PROGRAMME CONTEXT

The design, implementation and outcomes of development programmes are affected by multiple contextual factors, which may include: economic, social, political, cultural, demographic and ecological factors among others. Information on the factors is often too expensive and difficult to capture for many evaluations. Sometimes these are ignored or only captured in informal or anecdotal manner.

B. DATA COLLECTION

Data collection is expensive. Many evaluations are only able to collect data on a smaller sample than would ideally be required to ensure a satisfactory level of statistical power. Cost may also limit the types of data that can be collected. Many kinds of data may also be time–consuming to collect and analyse so that feedback to decision–makers may be much slower than, ideally required, for agile decision–making.

TABLE 2–2 SIX SETS OF CHALLENGES TO CURRENT EVALUATION METHODOLOGIES WHERE BIG DATA ANALYTICS MAY BE ABLE TO CONTRIBUTE

TOPIC	CHALLENGES
A. Understanding the programme context	— Collecting information on a large number of contextual variables
B. Data collection – including on sensitive topics	— Data is expensive to collect so that sample sizes are often smaller than desired — Data can be time–consuming to collect — Many groups are difficult to reach — Delays in providing feedback to decision–makers — The time–period over which data is collected may be limited
C. Monitoring processes and behavioural change	— Difficult to collect information on sensitive topics — Difficult to observe and measure behavioural change – both because a longer period is often required to observe the changes and because many of the processes are subtle, difficult to capture and often not recognized by the informant. — Different and expensive to collect continuous data required to monitor processes of project implementation and change
D. Capturing different voices and using evaluation to empower vulnerable groups	— Expensive and difficult to capture the voices of multiple stakeholders, particular those who are vulnerable or who do not have channels to express their views
E. Evaluating complex programmes	— Information must be collected on many more variables — Information must be collected and processed more quickly — Interactions among multiple variables must be analysed — Processes must be monitored with continuous observation — Non–linear causal chains must be monitored
F. Disseminating evaluation findings	— Findings must be disseminated in different ways to different audiences — Findings must reach all sectors of the target population — New feedback mechanisms must be developed to involve more groups

Data on qualitative or sensitive topics may also be difficult to collect and there may be legal or other restrictions for the collection of some kinds of data. Cost, the duration of the evaluation and other considerations may also limit the time–period over which data can be collected. Many evaluations also require the collection of information on sensitive topics such as intra–household dynamics, sexual orientation or sexual preferences, sexual harassment, domestic violence, illegal economic activities, situations of risk. Sometimes, such informal types of information are often very difficult and expensive to collect.

C. MONITORING PROCESSES AND BEHAVIOURAL CHANGE

Monitoring processes, such as how project implementation changes over time, how groups created by the project evolve or how the project affects different community activities, require the continuous collection of data – often on a large scale. This can be expensive, time consuming, difficult and often beyond the resources of the monitoring and evaluation systems.

Many programmes seek to promote behavioural change in, for example, relations among household members, the role of women in community organizations, or the level of violence within communities. Documenting these processes and changes often require continual observation, the collection of information on sensitive topics or capturing subtle and difficult–to–capture processes of change. Much of this information also requires in–depth observation over a long period of time, and is consequently expensive and time–consuming to collect.

D. CAPTURING DIFFERENT VOICES AND USING EVALUATION TO EMPOWER VULNERABLE GROUPS

Many data collection methods only collect information from, and give voice to certain groups (e.g. only men or dominant ethnic groups). It is difficult and time–consuming to capture the voices of vulnerable and less accessible groups. Also, non–conventional methods may be required to document these different voices.

E. EVALUATING COMPLEX PROGRAMMES

The recognition of the complexity of most development programmes limits the effectiveness of current evaluation approaches. The need for complexity–responsive evaluation approaches presents a number of challenges for development evaluation:

- Information must often be collected on a much larger number of variables – including multiple contextual variables;
- Information must be collected and processed more rapidly – often including real–time feedback;
- Interactions among multiple variables must be analysed;
- Processes must be monitored with continual observation of multiple variables; Analysis of non–linear causal chains must be traced.

F. DISSEMINATING EVALUATION FINDINGS

Due to the time required for analysis and publication of evaluation findings, there are often long delays before findings reach stakeholders and frequently they only reach certain priority groups. Often many

interested NGOs, civil society and community organizations never receive the findings. Often everyone receives the findings in the same format, which can be difficult for many groups to understand. Many evaluation teams find it difficult to customize the findings for different audiences. Feedback on evaluation reports is often only received by select groups invited to a limited number of briefings. Consequently, the views of large sectors of the population affected by different interventions are never received.

2.3 THE CHARACTERISTICS OF BIG DATA AND POTENTIAL APPLICATIONS FOR PROGRAMME MONITORING AND EVALUATION

This section describes a basic framework for the application of big data. The framework only presents a few illustrations of current applications and it is recognized that there are many more ways that each of the steps can be applied (see for example, Siegal; Marr; Meier; Letouzé, 2012; Letouze et al, 2016; Bruce and Koler, 2016; ICTWorks.org, UNGlobalpulse.org.) The call to action is to develop collaborative mechanisms to identify, test and operationalize these exciting opportunities.

The focus of the section is on a four–step data analytics approach that can be applied to the monitoring and evaluation of development programmes. While the approach provides powerful tools for the analysis of big data, it is also useful for the analysis of conventional data sets, and for merging small and large data sets into integrated data platforms, which then require the use of data analysis tools. Figure 2–1 describes the four steps. While they will often be used sequentially, it is also possible to use a single application.

It should be noted that various authors classify these steps in different ways[14], but they all cover similar ground. Data visualization and dissemination is an integral part of each step. The potential applications of each step for development M&E are illustrated in Table 2–3.

A. DESCRIPTIVE AND EXPLORATORY ANALYSIS: DOCUMENTING WHAT IS HAPPENING, OFTEN IN REAL–TIME

Descriptive and exploratory analysis describes the characteristics of a programme or intervention and the context within which it operates. These approaches have proved very useful to organizations that have data that has not been analysed. A common situation is where different departments or units of an organization use data sets that are directly relevant to their particular activities, but no one has ever integrated and merged data from different units to find patterns across departments. For example, the analysis may find that there are big differences in how well programmes operate in different regions or when working with groups that have different socio–economic characteristics.

These approaches are also useful when organizations are generating volumes of data that are too large to process using conventional data collection and analysis tools (such as spread–sheets). Examples

include: tweets, massive data from satellite images with repeat observations over time, or electronic financial transactions.

Table 2–3 Section A illustrates some of the potential applications of descriptive and exploratory analysis. These kinds of exploratory analysis will frequently identify questions that require the use of the more sophisticated kinds of analysis in steps 2, 3 and 4.

B. PREDICTIVE ANALYTICS: WHAT IS LIKELY TO HAPPEN

Predictive analytics (PA) uses patterns of associations among variables to predict future trends. The predictive models are usually based on Bayesian statistics and identify the probability distributions for different outcomes. When real–time data is available predictions can be continuously updated.

Siegel (2013)[15] illustrates some of the ways that predictive analytics are currently used by commercial organizations and government agencies in the USA.

- Which are the on–line advertisements on which customers are likely to click?
- Which mortgage holders will prepay within the first 90 days?
- Which employees will resign within the next year?
- Which female customers are most likely to be pregnant in the near future?
- Which voters will be positively persuaded by political campaign contacts?

Typical public sector applications include:

- The most likely locations of future crimes in a town?
- Which soon–to–be released prisoners are likely to be recidivists?
- Which questions is a student most likely to get right on a test?

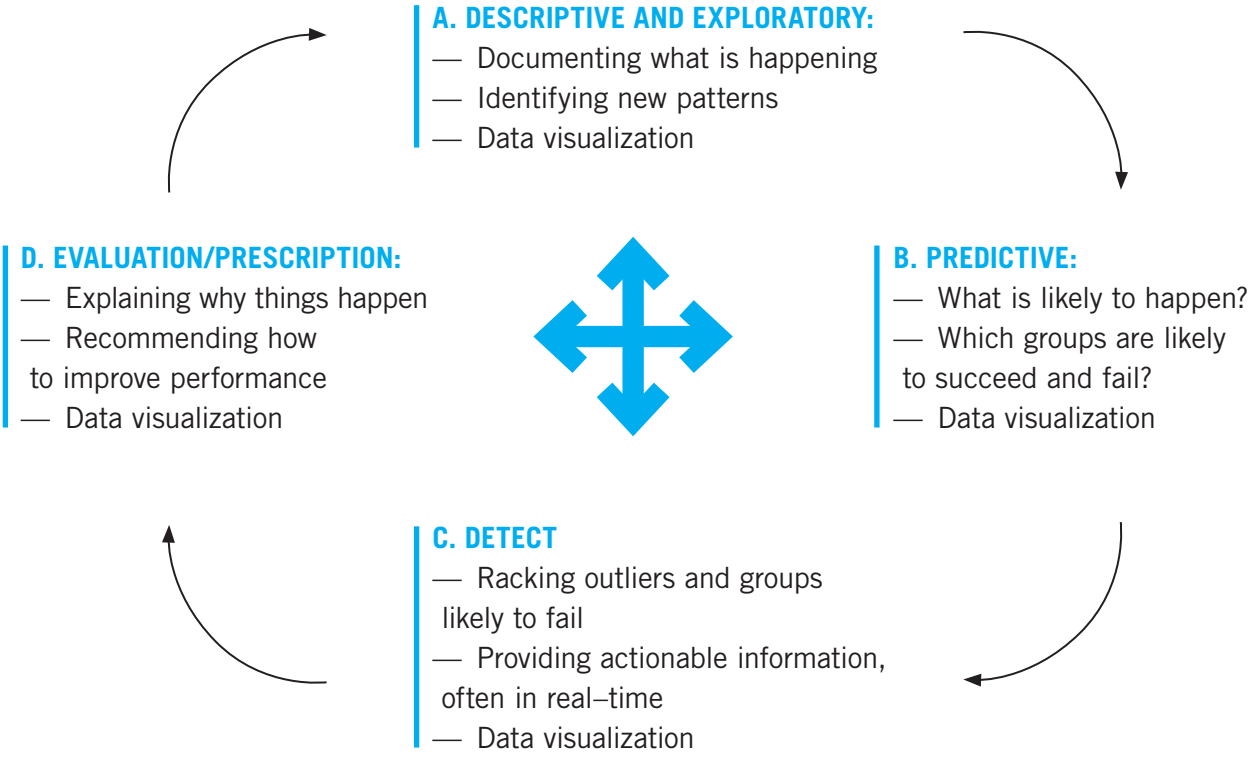
While some data analytics are based on the mining of very large data sets with very large numbers of cases and variables, it is also possible to apply many of the techniques such as predictive modelling with smaller data sets.

While predictive analytics are well developed, much less progress has been made on causal (attribution) analysis. Commercial predictive analytics tends to focus on what happened, or is predicted to happen (e.g. click rates on web sites), with much less attention to why outcomes change in response to variations in inputs (e.g. the wording or visual presentation of an on–line message). From the evaluation perspective, a limitation of predictive analysis is that it is not normally based on a theoretical framework, such as a theory of change, which explains the process through which outcomes are likely to be achieved. This is an area where there is great potential for collaboration between big data analytics and current impact evaluation methodologies.

C. DETECTION: TRACKING WHO IS LIKELY TO SUCCEED AND WHO WILL FAIL

The **step one** of descriptive analysis helps in identifying some of the important issues and challenges, which must be addressed in the programmes being studied. Nevertheless, as much of this analysis is exploratory, at this stage the information is not usually available to identify and target specific problem groups, such as youth most likely to drop out of programmes. In **step two**, using predictive analytics,

FIGURE 2–1 THE FOUR STEPS IN APPLYING DATA ANALYTICS TO PROGRAMME EVALUATION



it is possible to generate more detailed data on groups who are likely to perform well and those likely to drop out. In **step three**, detection methods can be used to identify any anomaly within a data set (for example areas where a crop is growing slower than in others). Detection methods have been used in social welfare programmes to monitor, for instance, the experience of students in a classroom or to understand if they are likely to drop out.

D. EVALUATION AND DATA DIAGNOSTICS: EXPLAINING HOW OUTCOMES WERE ACHIEVED AND PROVIDING RECOMMENDATIONS ON HOW TO IMPROVE PROGRAMME PERFORMANCE

Smart big data analytics using techniques such as data mining, machine learning and natural language analysis are designed to manage large and complex data sets and to identify unseen patterns. Data analytics permit the analysis of unstructured textual material, sound and video materials. Furthermore, combined analysis of the above mentioned categories also bring together different types/sets of data[16]. These techniques are used to help understand how outcomes were achieved.

Some promising analytical tools are being developed to process massive data sets to model complex emergency and other humanitarian situations such as the migration crisis in Europe, the complex dynamics of slavery and human trafficking and forced population movements as a result of massive natural disasters. Systems analysis is also being used to identify the most effective ways to reduce malnutrition and stunting in a particular region.

Big data also has the ability to present complex data in maps and graphs that are easily understandable to managers and local communities, and that permit users to drill down on specific geographical locations or topics of interest (see Table 2–4).

TABLE 2–3 KINDS OF BIG DATA ANALYSIS WITH POTENTIAL APPLICATIONS FOR PROGRAMME MONITORING AND EVALUATION

A. DESCRIPTIVE (EXPLORATORY): DOCUMENT AND CONVEY WHAT IS HAPPENING, OFTEN IN REAL–TIME. SEEK PREVIOUSLY UNIDENTIFIED PATTERNS.	
1. Collecting large volumes of data beyond the capacity of conventional data collection methods	— Can incorporate more contextual factors and capture broader trends
2. Identifying patterns that were previously difficult to identify	— This often involves merging organizational data sets, which were previously not linked
3. Real–time data collection that can be continually updated	— Dynamic monitoring and the generation of actionable data on project problems and new opportunities
4. The benefits of speed in rapidly changing circumstances	— Valuable in emergency situations and dynamic situations such as rapid urban growth or population movements
5. Early warning	— Valuable in emergency situations and to identify potential ethnic, work–related and other kinds of conflict
6. Integration of multiple sources of data	— Permits the use of mixed–methods approaches
7. Processing of unstructured data (free text documents, images, motion pictures, sound recordings, physical objects)	— Valuable for the analysis of qualitative data including large volumes of agency reports and operational documents
8. Allowing for systems mapping, socio–metric analysis and the analysis of complex adaptive systems	— Tools for monitoring changes in communities or other kinds of organization and the interactions among different parts of a system (see later discussion of complexity)

B. PREDICTIVE: WHAT IS LIKELY TO HAPPEN? WHO IS MOST AT RISK, WHO MIGHT DROP OUT?	
1. Analysing data sets too large and complex to be analysed using conventional methods	— Predicts what is likely to happen but without any underlying theory so that it is not possible to explain 'why' it will happen — Does not usually identify and test underlying assumptions of the model
2. Predicting opportunities (groups likely to succeed)	— Predicts probability of success and failure for different groups
3. Predicting groups at risk	
C. DETECTION: FOCUS ON ANOMALIES AND OUTLIERS: TRACKS ISSUES IDENTIFIED IN THE DESCRIPTIVE ANALYSIS	
1. Tracking outliers and groups at risk	— Builds on prediction and develops ways to track different groups – often in real–time.
2. Identifying unintended outcomes	— Identifies possible unintended outcomes and impacts, especially negative impacts (that make things worse not better) that should also be investigated and tracked
D. EVALUATION: DIAGNOSTIC/ANALYTIC/PRESCRIPTIVE]: SHEDDING LIGHT ON WHY THINGS HAPPEN – WHO IS MOST AT RISK, WHO MIGHT DROP OUT	
1. Powerful data analytics that can conduct analysis beyond the capacity of conventional computing systems	— Basic analytics: [breaking down data into smaller units, data visualization, monitoring]
2. Displaying and dissemination of large data sets	— Advanced analytics: [predictive modelling, pattern matching techniques]
3. Analysis of complex adaptive systems	— Data mining

Source: Adapted by the author from Letouzé, Areias and Jackson, 2016; Peng & Matsui, 2015–16; and Gee, 2015

2.4 THE COMPLEMENTARITIES BETWEEN BIG DATA ANALYTICS AND CURRENT EVALUATION PRACTICE

There is a lot of discussion in the big data literature and conferences on the differences between big data analytics and development evaluation, often claiming the superiority of the big data approaches. In contrast, at this point most of the evaluation literature and conferences includes very little discussion of big data.[17] Consequently, it is useful to briefly consider the many complementarities between the two approaches:

- Both are concerned with collecting and using available evidence in the most effective way;
- The four phases of the big data evaluation cycle draw on experiences with the evaluation of development programmes, in fields such as education, criminology, youth programmes. Both big data analytics and evaluation practice seek to identify factors affecting programme performance and high–risk groups;
- Both seek to predict how well development programmes will operate in future situations. While big data seeks to do this through predictive analytics, evaluation usually relies on experimental and quasi–experimental designs;
- Both apply modelling to the analysis of large data, although the approaches tend to be different;
- Both seek to monitor behavioural change;
- Most evaluations combine data from different points on the data continuum discussed in Chapter

BOX 2–1 AN EXAMPLE OF SOCIAL MEDIA ANALYTICS: RAPID ASSESSMENT OF COMMUNICABLE DISEASE CONTROL PROGRAMMES IN INDONESIA BASED ON THE ANALYSIS OF TWEETS

Although Internet penetration is Indonesia is relatively low (18 per cent in 2014) compared to other South–East Asian countries, the analysis of tweets provided a rapid and cost–effective way to conduct a rapid assessment of communicable disease incidence and control. It was recognized that this analysis only provides an initial indicator, and that this is biased given the fact that only a small proportion of the population have access to Internet. However, the analysis provided near real–time information for disease surveillance (both early detection and continuous monitoring). It also helped understand community perceptions regarding Information, Education Communication (IEC).
The study report describes the methodology in detail as well as discusses how to compensate for potential bias in the estimates.

Source: Article available at: http://unglobalpulse.org/middle-east-respiratory-syndrome

1, and many combine big data analytics with mixed method approaches which often include in–depth qualitative analysis. Consequently, there is much more overlap in the kinds of data and data analytic approaches than is often assumed;
• Many authors also stress the importance of the human dimension of big data. In addition to the human role in the interpretation of findings and often negotiating access to data, human intervention will often be necessary to prepare the way for the use, and sometimes the collection of big data;
• Finally, both are concerned to improve the dissemination and utilization of evaluation findings.

While there are differences, and sometimes disagreements in terms of how the approaches are used in each of these contexts, there is also a broad base of shared concerns and methods on which to build a collaborative approach.

2.5 HOW BIG DATA CAN HELP ADDRESS COMMON EVALUATION CHALLENGES

There are three types of widely occurring evaluation challenges that all evaluations must address: design, data collection and data analysis. This section discusses how big data can contribute to addressing these challenges. Given the considerable overlap between big data and ICT generated data, included is also ICT supported data collection and analysis that can complement big data.

A. DESIGN CHALLENGES

STRENGTHENING THE COUNTERFACTUAL

For area–based sampling (e.g. environmentally protected areas) it is sometimes possible to use satellite images to identify characteristics on which the project and comparison group samples can be matched (distance from the protected areas boundaries, density of roads and services, density of forest cover). In some cases density of phone coverage can also be used as a proxy for level of economic development. Matching can be strengthened when satellite data is complemented by ICT (GPS mapping, remote sensors) or survey data. The different sources of indictors can be combined using propensity score matching to strengthen matching.

EVALUATING COMPLEX PROGRAMMES

The evaluation of complex programmes normally requires the collection of information on a large number of programme components, contextual factors, interactions among multiple stakeholders, inte-

grating often inconsistent monitoring data collected by different agencies and tracking complex, non–linear processes of change (Bamberger, Vaessen and Raimondo, 2016).

Figure 2–2 summarizes five dimensions of complexity: the intervention being evaluated, the context within which the programme operates, interactions among stakeholders and implementing agencies, non–linear causal relations and challenges in conducting the evaluation[25]. All of these require collection of larger and more complex data sets than what is required for the evaluation of programmes.[26]

Big data, often complemented by ICTs can contribute in a number of ways:

- Combining a range of big data and ICT techniques to collect a wider range of contextual data;
- Using systems mapping to map the interactions among the different components of the intervention and its context;
- Using social media to track attitudes and behavioural change;
- Using software to develop scales and indices (e.g. concept mapping) for the different dimensions of complexity

IDENTIFYING UNINTENDED OUTCOMES

- Many widely used evaluation designs fail to capture unintended outcomes. Most quantitative designs, including randomized control trials, are designed to test whether intended outcomes have been achieved (e.g. Is there a statistically significant difference in the change, in specific outcomes between the project and control groups over the life of the project?). However, they are not designed to identify outcomes and they are not included in the original project design (and the research hypothesis). While qualitative designs, such as a theory of change can potentially identify unintended outcomes, often evaluation clients are only interested in knowing whether their project has achieved intended outcomes (Bamberger, Tarsilla and Hesse–Biber, 2016).
- Big data and ICTs can potentially provide real–time or rapid feedback on changes in a range of key indicators so that the process of project implementation – the time when many unintended outcomes occur – can be tracked. Furthermore, big data and ICTs can also provide feedback on the influence of a wider range of contextual factors that can also contribute to unintended outcomes. More importantly, actionable feedback can be provided to managers and other groups so that early signals of potential problems can be explored. An on–line theory of change provides a useful framework for identifying, tracking and updating unintended outcomes. This can be complemented by the analysis of Twitter and social media to track potential problems that might produce unintended outcomes.

B. DATA COLLECTION CHALLENGES

THE HIGH COST AND TIME REQUIRED FOR DATA COLLECTION

Big data can provide large volumes of data quickly and can be cost effective. As most big data has already been collected for different purposes, it can usually be accessed at a relatively lower cost for monitoring and evaluation purposes. Much of the data can also be delivered in near real–time and updated continuously.

TABLE 2–4 EXAMPLES OF DATA VISUALIZATION THAT MAKE THE FINDINGS OF COMPLEX ANALYSIS EASILY ACCESSIBLE TO MANAGERS AND LOCAL COMMUNITIES

1. Obtaining feedback from more than one million women in India through tablet–based surveys that the women design and interpret through data visualization, (Source: World Bank)[18]
2. Feedback to contributors through data visualization maps where results on the ground can be checked directly. Grantees can also provide feedback directly, (Source: Proving It)[19]
3. Responding to disease outbreaks, (Source: FAO in collaboration with Global Pulse)[20]
4. Automating crop disease detection with maps that allow farmers and agencies to drill–down on specific locations, (Source: Makerere University in collaboration with UN Global Pulse)[21]
5. Conducting radio mining in Uganda. Maps permitting users to pin–point locations where particular problems were identified through analysis of radio programmes, (Source: UN Global Pulse)[22]
6. Compilation of data on Boston disaggregated by city vitality, culture, economy, education, environment, health, housing, public safety, technology and transportation, (Source: Boston Indicators Project)[23]
7. Mapping earthquakes and other emergencies based on crowdsourcing and analysis of social media, (Source: Patrick Meier)[24]

COLLECTING DATA ON DIFFICULT–TO–REACH GROUPS

Certain groups may be difficult to reach either because of security situations or because of their remote and inaccessible locations. There are number of ways that new sources of data can be used to contact these groups, for example by interviewing people over the phone rather than in person, or by automatic monitoring of whether people received automated phone messages (e.g. with reminder for medical appointments) and how they followed–up. Women or some other groups who do not have voice in a particular community may be able to speak more freely on over the phone or on social media. People in high risk zones can sometimes send out video and audio recordings of the situation in these zones and satellites can also track population groups that would otherwise be difficult to locate or contact (e.g. refugees)

MONITORING PROJECT IMPLEMENTATION AND PROCESSES OF BEHAVIOURAL CHANGE

Big data can often provide real–time and continuous data, which is helpful for observing the processes through which a programme evolves. Studying behavioural change also requires capturing information on processes (rather than just comparing two points in time). There are a variety of big data and ICT resources that can assist. Mobile devices can capture video and audio records of meetings, work groups and different aspects of community life that can be helpful. Social media are also a prosperous, wealthy source of information. Social network analysis can be another valuable tool.

FIGURE 2–2 DIMENSIONS OF COMPLEXITY IN DEVELOPMENT EVALUATION
Source: Bamberger, Vaessen and Raimondo (2016) Chapter 1

EMBEDDEDNESS AND THE NATURE OF THE SYSTEM
— Historical, economic, political, sociocultural, administrative, and ecological, legal and regulatory contrast
— Norms and beliefs
— Interconnectedness, boundaries dynamics (e.g. path dependence, system shock)

INTERVENTION
— Design and purpose (e.g. initial logframe, logic model, theory of change)
— Size and scope (e.g. number and types of intervention activities, levels of intervention)
— Data coverage, quality and accessibility

INSTITUTIONS AND STAKEHOLDERS
— Governance, funding coordination, implementation system
— Number and diversity of stakeholders (e.g. implementing agencies, donors, politicians, beneficiaries, evaluators)
— Stakeholder expectations, demands and 'theories–in–use'
— Conflict, cooperation, evaluation culture

CAUSALITY AND CHANGE
— Causality (e.g. non–linearity, emergence, feedback loops, multiple pathways)
— Attribution and contribution
— Theories, mechanisms, models of behavioural change
— Implementation
— Direct, indirect, intended, unintended, positive, negative effects

Challenges in delimitation, sense–making, consensus–making, design, implementation and use of evaluations

EVALUATION
— Purpose
— Time, resources and data
— Methodology
— Participation and process
— Values and ethics

COLLECTING QUALITATIVE DATA

High quality data is often difficult to collect and the recording and interpretation process often introduces a level of subjective interpretation that is difficult to control. Smart phones can now collect high quality audio and visual data and the software for the analysis and interpretation is improving rapidly. This can help in removing certain kinds of reporting bias or subjective interpretation.

COLLECTING AND INTEGRATING DIFFERENT SOURCES OF DATA

Both big data and ICTs offer a range of ways to integrate data from multiple sources and in multiple formats. While big data can do this for very large data sets, ICTs can do the same for smaller data sets.

ENHANCING QUALITY CONTROL OF DATA COLLECTION AND ANALYSIS AND REDUCING THE COST OF THESE CONTROLS

ICT software is now available to control the quality of data at all stages of the collection and analysis process. There are a series of consistency checks on how data is input. For example, GPS can ensure that the right household is being interviewed and when random route sampling is used mobile devices can ensure that appropriate selection procedures are used.

COLLECTING INFORMATION ON THE SPATIAL DIMENSIONS OF PROGRAMMES

Satellite images, remote sensors and GPS mapping can help analyse the spatial dimensions of programmes. These tools also make it much easier to include a more in–depth analysis of the contextual factors (transport networks, access to services and markets, population movements, soil quality and crop production) that are essential for a full understanding of the wide range of factors affecting programme outcomes.

SAMPLE DESIGN CHALLENGES

A challenge for all non–experimental evaluation designs is the question of sample selection bias. Post–project differences between the project and comparison groups that are assumed to be due to the effects of the intervention are often due to differences in how the two groups were selected. A related challenge, which receives less attention, is that sampling frames often do not cover all of the sample population, and frequently there are important differences between the population that is sampled and the population that is excluded. Very often the excluded group is poorer or has less access to services. There are number of ways that big data and ICTs can help address these problems.

Satellite images and GPS maps can provide images of the total target population that can be overlaid with the population that is actually sampled to determine if there are important differences. Phone companies keep detailed information on their customers and this can be used to ensure that selected samples of phone users are representative of all phone users. It is also possible to determine how closely a sample of phone users matches the total population. The previous chapter discussed how satellite images could be combined with GPS mapping data and information from households, farms and other

kinds of surveys, to improve the match of the comparison and project samples using techniques such as propensity score matching.

C. DATA ANALYSIS CHALLENGES[27]

Big data analytics offer a number of powerful tools for the analysis of data sets that are too large and complex for analysis using conventional data analysis programmes. These can be broadly classified into:

- **Basic analytical tools** use data mining to break data down into smaller units that are easier to explore. Data visualization is used to present the findings in an easily understandable manner. This can provide data for programme monitoring. These basic analytical techniques are often used to identify trends, relationships and patterns that can later be explored with more advanced analytics.
- **Advanced analytics:** include predictive modelling and text analytics (analysing unstructured text and transforming it into structured information that can be quantitatively analysed).
- **Operationalizing data analytics for an organization or set of organizations.** Models must be developed to design the particular applications required by a particular organization.

Further details of how new information technologies and data analytics can help address common evaluation challenges available at: bit.ly/2gLA6k0

2.6 HOW WELL CAN BIG DATA ANALYTICS ADDRESS THE CHALLENGES FACING PROGRAMME EVALUATION?

Section 2.5 identified many areas in which big data analytics can potentially strengthen programme monitoring and evaluation. However, at this point in time it is difficult to find systematic assessments of the extent to which this promise has been fulfilled and the challenges limiting the utility and widespread application of the data analytic approaches. The following section briefly reviews some of the new methodological challenges that the application of big data involves, logistical and organizational issues and some of the political and ethical challenges.

SOME OF THE NEW METHODOLOGICAL CHALLENGES THAT BIG DATA ANALYTICS INTRODUCES

Letouze, Areas and Jackson (2016) identify a number of methodological issues that are largely unique to the use of big data analytics:

a. Comparability over time: when data comes from third parties, it is difficult to know if it is consistent and comparable over time, as companies such as Google frequently update their algorithms.

b. Non–human Internet traffic: Bots are computer programmes that are designed to post automatically and act as humans. It is estimated that over 60 per cent of Internet traffic is generated by bots, so this can skew results.

c. Representativity and selection bias: How large data sets are selected means that the sample is frequently not representative of the total population being studied. Some readers are mislead by the 'fallacy of large numbers' into assuming the sample must be broadly representative as it is so large.

d. Spatial autocorrelation: Ownership of mobile devices is often concentrated in certain geographical areas of a population, so that responses, unless weighted, will be biased towards information from these areas.

e. Attribution and spurious correlation: According to Taleb, 2013 (cited in Letouzé et al ,2016) the larger the number of variables in a big data analysis, the higher the risk is of spurious correlations when using data mining procedures.

LOGISTICAL AND ORGANIZATIONAL ISSUES

a. In many development agencies, the introduction of big data tends to be through new offices run by data scientists and where developing evaluation applications have not been a priority.

b. A practical consideration in many development contexts concerns the much more limited data availability compared to developed countries, where big data analytics are widely applied. Even when data such as phone records, ATM transactions may be available, the data may be less representative due to the small proportions of the population who use such services.

c. A major practical concern at this point is the relatively limited interest in, and demand for big data analytics in many countries and sectors. An important cause is the limited awareness of many evaluation agencies about the potential benefits of big data for evaluation. To date, there has been more interest from planning offices and emergency relief agencies.

d. The limited access of many agencies to big data is another major consideration.

e. At this point in time, it is also probably the case that few systematic assessments have been made of the practical benefits of big data and what are its benefits compared to current monitoring and evaluation tools and techniques. Consequently, there are few examples or models that evaluation offices could draw on.

POLITICAL AND ETHICAL CONSIDERATIONS

In addition to these methodological questions, there are also a number of important political and ethical issues that must be addressed:

a. Access to big data is often limited and data may only be available to governments, UN agencies and a few bilateral and multilateral agencies. This can reduce the control that local communities have over important information affecting their lives. This is particularly important as it is often claimed that big data can promote participatory and inclusive development, whereas it may achieve the opposite.
b. Access to big data can also be expensive, again excluding many groups who wish to use it.
c. A related concern is that commercial survey research agencies may collect information on and about poor and vulnerable groups which they then sell to private companies and without any benefit or compensation to the communities to which the information refers.
d. There are also important privacy issues, as much of the big data and ICT generated data contains sensitive information that could fall into the hands of security agencies or on–line hackers. As the amounts of data collected increases, and when uploaded to remote on–line central locations, it becomes technically more difficult to ensure data privacy and data protection.

PART 2

GUIDELINES FOR INTEGRATING BIG DATA INTO THE M&E FRAMEWORKS OF DEVELOPMENT PROGRAMMES

CHAPTER 3 INTEGRATING BIG DATA INTO THE MONITORING, EVALUATION AND LEARNING (MEL) SYSTEM

3.1 THE COMPONENTS OF AN INTEGRATED MEL SYSTEM

This chapter presents a framework for an integrated MEL system for development programmes. In many operational contexts big data will be combined with other types of data, often drawing upon data generated through mobile devices and other ICT tools as well as from conventional evaluation data collection. Chapters four and five will discuss the monitoring and evaluation components in more detail and Chapter six will describe the role of the evaluation manager in the development of a big data M&E responsive system.

MEL systems are required at the project, programme, agency and national levels. Many agencies now incorporate 'learning' into their M&E systems, recognizing that M&E systems are not intended only for accountability but that an important function is to disseminate lessons that can improve the design and performance of on– going and future programmes. All projects, programmes and broader interventions require an integrated MEL system that:

- Provides regular and rapid feedback on how a programme is progressing compared to intended outcomes;
- Provides feedback throughout the process of programme implementation and identifies deviations from the intended implementation model and other potential problems;
- Provides information required by stakeholders in a form they can understand and use;
- Assesses whether intended outcomes have been achieved and the extent to which these outcomes can be attributed to the effects of the programme;
- Identifies unintended outcomes and proposes ways to address them;
- Understands the dimensions of complexity affecting a programme and how these can be incorporated into the M&E systems;
- Is transparent, accessible to all stakeholders, affordable, technically viable, sustainable and ethical;
- Synthesizes and disseminates in timely manner lessons from the on–going monitoring and evaluation studies.

Monitoring is defined as:

'A continuous internal management activity, whose purpose is to ensure that the programme achieves its defined objectives within a prescribed time frame and budget. Monitoring involves the provision of

regular feedback on the progress of programme implementation, and the problems faced during implementation. Monitoring consists of operational and administrative activities that track resource acquisition and allocation, production or the delivery of services, and cost records.' (Valadez and Bamberger, 1998: 12).

In contrast, evaluation is defined as:

'An internal or external management activity to assess the appropriateness of a programme design and implementation methods in achieving both specified objectives and more general development objectives and to assess programme results, both intended and unintended and factors affecting the level and distribution of benefits produced.' (Valadez and Bamberger, 1998: 13).

While M&E systems usually operate independently of each other, as both have different purposes, an effective programme management system requires that the two are treated as part of an integrated system that provides the different kinds of short, medium and long–term information required by managers, planners, policymakers and other stakeholders. Even when the stated purpose of an evaluation is to assess programme impacts, it is essential to know how the way in which the programme was implemented has affected outcomes. When intended outcomes are not achieved, is this due to design failure or to implementation failure?

3.2 POLICY EVALUATION, PROGRAMME EVALUATION AND PROJECT EVALUATION

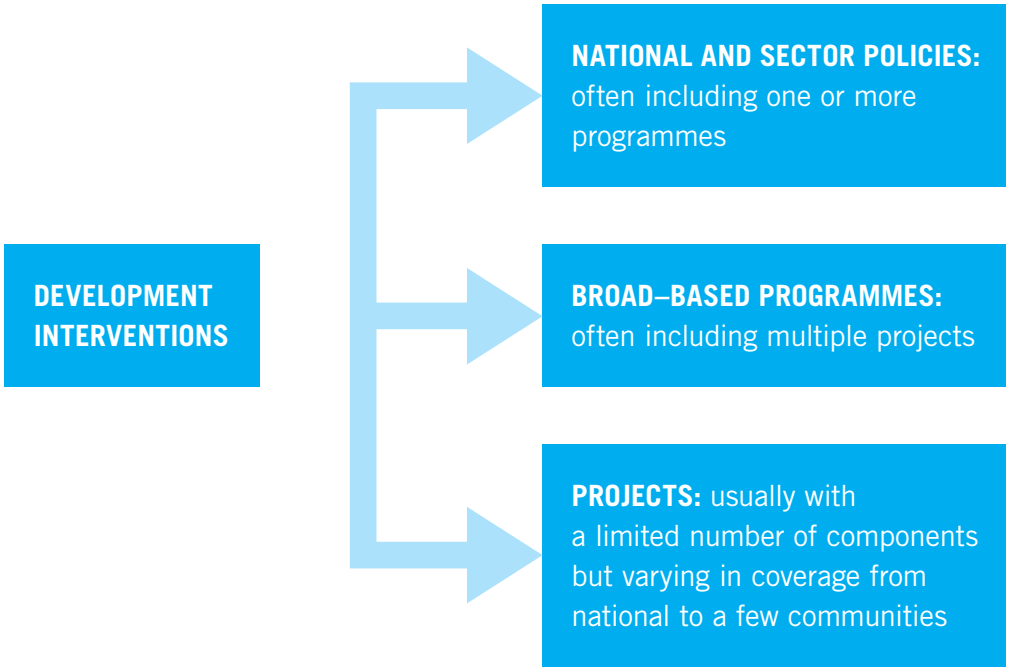
While much of the evaluation literature tends to focus on the monitoring and evaluation of stand–alone projects that have a limited number of components and clearly defined objectives, it is important to recognize that an important function of development evaluation is to also evaluate broad–based programmes that may include many individual projects, together with development policies that are implemented at the national and sectorial levels. Different evaluation approaches are used at each of these levels, and big data analytics can bring potential contributions to each of these levels.

Level refers to projects, programmes and policies, while the term development interventions will be used to cover all three levels. Similarly evaluation refers to project, programme and policy evaluation, again using development intervention evaluation to cover all three types of evaluation (Figure 3–1).

Much of the discussion in this and subsequent chapters will focus on *project evaluation* as most of the tools used for project evaluation form the building blocks for evaluations at the two higher levels. However, there are a number of special challenges and approaches used at the programme and policy levels.

Policy evaluation – Policy evaluations face a number of unique challenges. First, many policies are not based on a clearly articulated theory of change which can be identified and assessed. Second, policy change requires many different kinds of behavioural and attitude change from many different actors and agencies. These are often very difficult to monitor. Third, the outcomes of many policies cannot be observed for a number of years. For example, policies may not be implemented until the next four or five year development plan, and results may not be observed until the next plan has been underway. So it could be five or more years before the assessment can be completed. Fifth, given the nature of most policy interventions, it is extremely difficult to identify a counterfactual, and consequently it is difficult

FIGURE 3–1 THE THREE LEVELS AT WHICH EVALUATIONS ARE CONDUCTED



to determine the extent to which changes can be attributed to the new policy. Finally, when assessing the impacts of donor initiated policy reforms, an additional challenge can be that many governments do not wish to acknowledge that their policy reforms were based on advice from donors.

Programme level evaluation – Programmes typically involve a number of different components and often operate at several different levels such as national, regional, district and local or community level. Often the evaluation strategy will involve separate evaluations of different programme components and levels, and then trying to assemble the findings of the different component evaluations to assess the overall programme performance on dimensions such as: relevance, efficiency, effectiveness, impact and sustainability.

Programme evaluations often face three challenges. First, not all programme components are implemented in all areas, so it can be difficult to assess overall effectiveness and outcomes when the input components are not applied in a similar way in all areas. Second, as programmes are intended to cover all of the target population, it is difficult to identify a comparison group (counterfactual) with similar characteristics. Bamberger (2016) discusses a number of strategies that can be used to identify and approximate counterfactual. Most of these rely upon the fact that greater number of programmes is either implemented in phases, or do not cover the whole target population. Therefore, it is often possible to identify similar groups to programme beneficiaries who have not received benefits, or who will experience a delay before the services are received. Many of these approaches are similar to pipeline designs or naturalistic experiments described in the evaluation literature.

The third challenge relates to the common situation where each individual component of a programme receives satisfactory assessment, but at the same time the overall programme may have had little or no impact on its broader development objectives. For example, a programme to promote women's economic

empowerment might involve training workshops on empowerment related topics, assistance in marketing and microcredit. It might also involve skills training on accountancy, marketing and other related topics. Each of these activities might receive a positive rating, but there may be no measurable impact on development outcomes such as poverty reduction, increasing community security or increasing women's control over household and community productive resources. There may be a number of reasons for this, which must be assessed. For example, there may be problems of coordination between the different components, the scale of the programme may be too small to affect broad development outcomes, the design and logic of the programme may not be relevant to addressing broader themes, or some important additional components may be required.

3.3 THE STAGES OF A TYPICAL PROJECT, PROGRAMME CYCLE

The project/programme cycle typically has seven stages (Figure 3–2):

1. Exploratory and diagnostic studies, project identification and appraisal. Different options are reviewed that could potentially achieve the intended development objectives (e.g. protect forests, reduce poverty, promote equality of economic opportunities between women and men). Potential projects are compared in terms of their estimated efficiency in achieving intended objectives.

2. Planning and design and stakeholder consultations. Once the best design option has been selected the next stage involves planning and design. This usually involves developing a theory of change (TOC) or similar programme theory model explaining how the programme is intended to achieve its objectives. The TOC also helps identify the indicators that will be incorporated into the monitoring system. For many programmes these indicators will be built into the results–based M&E system. Most programmes involve large groups of stakeholders, and part of programme design involves ensuring that the objectives and priorities of each stakeholder are addressed and that there is agreement on how the programme will be implemented, monitored and evaluated.

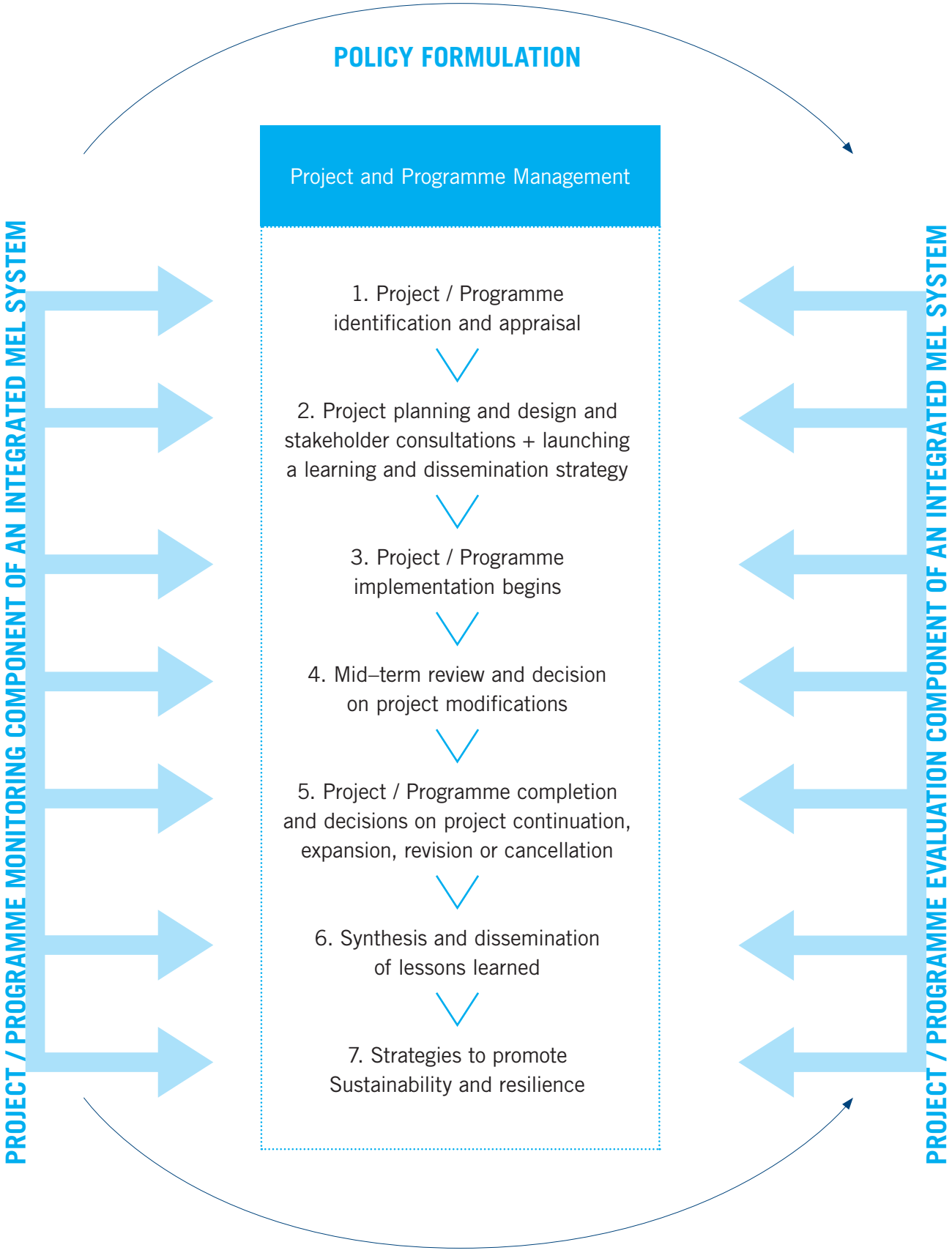
Evaluation theory stresses the importance of a mixed methods design combining both quantitative and qualitative indicators. Almost all indicators of project inputs, processes, outputs, outcomes and impacts must measure both the quantitative dimension (how much? how many? who benefits and who does not?), and the qualitative dimension (the quality of the services, were they culturally responsive, and how were programme outcomes affected by the processes of behavioural change through which project interventions are transformed into outcomes and impacts).

At this stage, learning and dissemination strategy should be put in place to ensure that findings and issues from the M&E studies are fed–back to stakeholders and other interested parties on a regular basis.

3. Project/Programme implementation. This also involves the launch of the M&E systems.

4. Mid–term review and decisions on any programme revisions. Many projects that last for more than three years tend to have a mid–term review during which decisions are made concerning any revisions to the project design, scope, financing or implementation strategy. Frequently, rapid evaluations of progress on the different project components are commissioned.

FIGURE 3–2 THE MONITORING, EVALUATION AND LEARNING [MEL] SYSTEM OF A TYPICAL PROJECT/PROGRAMME CYCLE



5. Project/Programme completion. A project completion (end of project) review is prepared which assesses programme performance and the extent to which key objectives have been achieved. Sometimes a preliminary impact assessment will be included in the Project Change Request (PCR).

6. Synthesis and dissemination of lessons learned. Findings, lessons and recommendations are disseminated to stakeholders and often to a wider audience.

7. Planning for and assessing the sustainability of the programme. Many funding agencies only focus on the implementation phase of a programme (the phase directly financed and supervised by the donor agency) and less attention is given to building in financial, management and organizational mechanisms to ensure sustainability – that the programme continues to operate and to deliver the intended services over its expected lifetime.

3.4 MONITORING, EVALUATION AND LEARNING ACTIVITIES AT DIFFERENT STAGES OF THE PROJECT/PROGRAMME CYCLE

Table 3–1 Identifies the M&E activities typically conducted at each stage of the project cycle. While monitoring and evaluation have different functions and responsibilities the two should be considered as parts of an integrated system. A well–designed monitoring system should provide essential information for the evaluation, in particular assessing to what extent failure to achieve intended outcomes is due to the way in which the project was implemented. These activities are described in Chapter 4 (monitoring) and Chapter 5 (evaluation).

TABLE 3–1 TYPICAL MONITORING, EVALUATION AND LEARNING [MEL] ACTIVITIES AT DIFFERENT STAGES OF THE PROJECT CYCLE

STAGE OF PROJECT / PROGRAMME IMPLEMENTATION	INTEGRATED MEL SYSTEM [LEARNING ACTIVITIES ARE BUILT INTO BOTH M&E]	
	MONITORING	EVALUATION
1. Project/Programme identification and appraisal	— Identify issues and indicators to be included in monitoring — Define target population — Identify contextual factors — Identify potential unintended outcomes	— Diagnostic studies — Mapping the programme context — Checklist to assess the level of complexity — Technical, economic, financial and social appraisal — Evaluability analysis

2. Planning and design	— Defining programme inputs, processes and outputs to be monitored [drawing on the TOC] — Designing the monitoring framework (results based evaluation) — Recruiting and training monitoring staff — Develop the programme monitoring system — Assessing strengths and weaknesses of existing data sets and need to incorporate big data and ICT–approaches	— Developing a theory of change or similar theory–based approach or adapting the existing operational TOC — Identify possible evaluation designs and select the one that is most appropriate to address the evaluation questions — Consider the need and feasibility to incorporate big data based evaluation designs — Recruiting and training evaluation staff — Stakeholder mapping — Identify the key evaluation questions
	— Defining information needs and assessing the quality and relevance of different sources of conventional and big data — Incorporating a mixed methods design — Ensure unintended outcomes are identified and incorporated into the TOC — Designing the integrated monitoring, evaluation and learning system (MEL)	
3. Project/Programme implementation	— Launching the monitoring system — Establishing systems for the regular dissemination and discussion of monitoring reports and for deciding follow–up actions — Rapid feedback monitoring reports	— Baseline study for impact evaluation — Periodic evaluation studies as required — Designing contribution analysis — Process and formative evaluations that draw on monitoring data — Periodic review and updating of the theory of change
4. Mid–term review and decisions on modifications	— Compilation of monitoring data to assess progress towards programme goals	— Mid–term evaluation review
5. Project/programme completion and decisions on future directions	— Synthesis of all monitoring data collected over the life of the project	— Project completion report — Pre–test and post–test evaluation — Ex–post evaluations
6. Synthesis of lessons learned and putting in place a learning and dissemination strategy	— Identify deviations from intended implementation strategy and consequences for the achievement of intended programme outcomes	— Synthesis of lessons on factors affecting the achievement of programme outcomes
7. Planning and implementing a sustainability and resilience strategy	— Continue monitoring to assess programme sustainability over time	— Sustainability and resilience analysis

3.5 A FRAMEWORK FOR INCORPORATING BIG DATA INTO MEL SYSTEMS

When considering the use of big data it is important to understand where the data comes from, how it is stored, how it could be used and what are the limitations (including legal) concerning access to data (see Figure 3–3).

STEP ONE: IDENTIFICATION OF REQUIRED INFORMATION FOR PROGRAMME MONITORING AND EVALUATION

Table 3–2 provides examples of some of the key questions that must be addressed for each of the four kinds of evaluation described in Chapter 2 (Policy evaluation, Formative evaluation, Developmental evaluation and Summative evaluation).

Each question will normally require a different evaluation design and it is essential to clearly define the priority questions before selecting the appropriate design. An important role for the evaluation manager (see Chapter 6) is to ensure the evaluation questions are clearly defined and understood by the team designing the evaluation. Potential indicators are then identified, including both conventional and big data.

STEP TWO: ASSESSMENT OF POTENTIAL SOURCES OF BIG DATA

Data is classified into (1) data coming from conventional sources such as surveys, government publications, focus groups; (2) data generated through ICTs; and (3) data generated by new sources of data (social media, mobile phones, financial records).

STEP THREE. CLASSIFICATION OF TYPES OF BIG DATA

Data is categorized according to how broadly it is available and complete.[28]

- Ubiquitous data is generated everywhere at the same time (e.g. meteorological data on rainfall, temperature);
- Non–ubiquitous data is not generated everywhere and can be classified into two types;
- System specific– generated within a specific system (such as a data only referring to a particular organization or geographic context);
- Outside of the system (e.g. external contextual factors such as the economic climate in the region or presence and quality of infrastructure throughout a state).
- Complete (homogenous) or incomplete (heterogeneous) and only providing partial information on some aspects of a programme.

Given the importance of mixed methods approaches it may also be useful, in some cases, to classify data according to whether it is qualitative or quantitative.

STEP FOUR: SIX TYPES/SETS OF QUESTIONS THAT DATA ANALYTICS ADDRESS ACCORDING TO THE TYPE OF EVIDENCE IT PROVIDES

- Descriptive analysis describes, for example, the characteristics of a programme or the context in which it operates or a needs assessment (as stated by the project population). Descriptive and exploratory analysis have benefitted greatly from the smart data analytics software packages now

TABLE 3–2 SOME OF THE KEY QUESTIONS FOR EACH OF THE 4 TYPES OF EVALUATION

A. Policy evaluation	a. Rating policies in terms of: — Relevance — Efficiency — Effectiveness — Impact — Equity — Sustainability b. To what extent can observed outcomes be attributed to the effects of the policy? c. How influential was donor agency advice in the formulation and implementation of national development strategies? d. For policies through which full results will not be visible for a certain period of time (until after the evaluation is completed), what indicators can be used to estimate the success after a shorter period of time?
B. Formative evaluation	a. How likely is project design to achieve the different development objectives? b. How effectively is the project being implemented? c. Are any sectors of the target population being excluded or receiving less access to project benefits? d. Are there any unintended outcomes (negative but also positive) that management must address?
C. Developmental evaluation	a. Are there mechanisms to ensure that all sectors of the target population are consulted? b. Are project services and benefits reaching all sectors of the target population? c. Does the evaluation design identify and address all of the complexity dimensions of the project/programme? d. Does programme implementation have the flexibility to adapt to the changing context within which the programme operates?
D. Summative evaluation	a. To what extent can a specific impact be attributed to the intervention? b. Did the intervention make a difference? c. How has the intervention made a difference? d. Will the intervention work elsewhere? e. What are the key factors (contextual, design, organization and coordination) that are important for successful replication? f. How simple or complex are the different dimensions of the programme on a complexity rating scale? (Bamberger, Vaessen and Raimondo Table 1–2, 2016) g. Is it necessary to use a complexity–responsive evaluation design? h. What are the main contextual factors affecting different programme outcomes? i. How are programme outcomes affected by problems of coordination among different stakeholders?

available for portable computers/notebooks.[29]

- Exploratory analysis seeks to find new patterns, trends and relationships within the data. Often one of the most useful applications is to merge different data sets within an organization into a standard platform where data can be compared using standard categories.
- Inferential analysis uses a relatively small sample to say something about the characteristics of a larger population. Often a hypothesis will be tested on a different, more representative sample to see whether the observed relationship still holds (Peng and Matsui, 2015:17).
- Predictive analysis uses the data on some objects to predict values for other objects. Often outcomes will be predicted for different sub–groups within the target population.
- Causal analysis involves the assessment of the extent to which observed changes in an outcome variable can be attributed to the effects of programme input or other variable. There is an important distinction between simple programmes where a direct causal relationship can be observed between an input and an outcome, and a complex programme where direct, simple causal relationships between an input and a particular outcome usually cannot be observed or inferred.
- Mechanistic analysis focuses on how changes in one variable (e.g. a diet high in fresh fruit) lead to a change (e.g. reduction in viral infection) in another variable.

STEP FIVE: THE SEVEN STAGES OF THE PROJECT/PROGRAMME AND M&E CYCLES PRESENTED IN FIGURE 3–1 ARE CONDENSED INTO FOUR PHASES WHERE EVIDENCE IS USED

- Design (stages one to two)
- Implementation and monitoring (stages three to five)
- Evaluation and learning (stage six)
- Sustainability (stage seven)

STEP SIX: ASSESSING POTENTIAL INDICATORS IN TERMS OF THE INDICATOR ASSESSMENT CHECKLIST (TABLE 3–3)

Based on the assessment of these criteria, a preliminary set of indicators will be defined. These will be subject to further assessment during the evaluability analysis (see Chapter 5) and the periodic assessments of the M&E systems.

For more detail on the methodology for identifying and assessing big data and ICT generated indicators see Jackson, 2015 Section 3.4

FIGURE 3–3 STEPS IN THE SELECTION OF BIG DATA AND ICT GENERATED M&E INDICATORS

STEP 1

Identification of required information for programme M&E

STEP 2

Data classified into conventional, ICT generated and big data sources

STEP 3

Potential sources of big data categorized according to availability and completeness

STEP 4

Data categorized into six types according to the type of question and type of evidence:
- Descriptive
- Exploratory
- Inferential
- Predictive
- Causal (attribution) analysis
- Mechanistic

STEP 5

Four phases of the project cycle
- Design
- Implementation and monitoring
- Evaluation and learning
- Sustainability

STEP 6

Assessing potential indicators on the indicator quality checklist (Table 3–3)

TABLE 3–3 CHECKLIST FOR ASSESSING THE QUALITY AND APPROPRIATENESS OF BIG DATA AND ICT–GENERATED INDICATORS. Source: Adapted from Morra–Imas and Rist (2009) with additional categories added.

		Rating 1–5*
Relevant	Directly relates to the topic being measured	
Comprehensive	Covers all important aspects of the topic	
Clear and unambiguous	Easy to interpret and use without any confusion or misinterpretation.	
Cost effective	The data is inexpensive to collect	
Accessible	The data is easy to access (physically and in terms of the format) and does not require difficult administrative procedures to access.	
Monitorable	The data can easily be collected over time	
User friendly	Uses concepts that are widely accepted and easy to use	
Stable over time	The definitions or collection methods do not change over time	
Comparable across agencies and locations	All agencies define the indicator in the same way. Does not vary in different regions	
Ethical	Collecting or using the data does not pose ethical concerns	

* Rating: 1 = This dimension is ignored or inadequately addressed. 5 = fully addressed

3.6 WHEN INTEGRATING BIG DATA INTO MONITORING AND EVALUATION SYSTEMS IT IS IMPORTANT TO FOLLOW EVALUATION BEST PRACTICES

When planning the integration of big data into monitoring or evaluation systems it is important to ensure that evaluation best–practice guidelines are followed. This is particularly important for big data and ICT generated data, as enthusiasm for the speed and ease with which huge volumes of data can be generated, can sometimes lead researchers to overlook some of the basic evaluation principles. The

report discusses best practices and challenges with respect to four categories: data collection, sample selection, evaluation design and data analysis and dissemination. Most of the following arguments are common to all kinds of evaluation, although some of the points are particularly important for the analysis of big data. It should also be recognized that some of the questions might need to be adapted to the special characteristics of big data.

Table 3–4 presents a checklist for assessing evaluation methodologies when big data and ICTs are integrated. The checklist can also be used to assess standard evaluation designs that do not incorporate big data or ICTs.

A. EVALUATION DESIGN CHALLENGES

Evaluation design options are discussed in Chapter 5 Section 5. Some of the common challenges that potentially affect most evaluations include:

a. Failure to clearly define the key questions that the evaluation must address. Many evaluators implicitly assume that all evaluations are addressing the same question as evidenced by the frequently discussed question, 'which is the best evaluation design.' In fact, there are at least four different questions of interest to clients and stakeholders.[30]

- To what extent can a specific (net) impact be attributed to the intervention?
- Did the intervention make a difference?
- How has the evaluation made a difference?
- Will the intervention work elsewhere?

b. Need to review all of the different evaluation design options before selecting the evaluation Chapter 5 Section 5 lists eight widely used evaluation designs, all of which should be considered before selecting the design.

c. Preference for a particular evaluation design that is considered the most rigorous and is applied irrespective of the evaluation question. Advocates of randomized control trials are often criticized for always trying to use an RCT, even in situations where it may not be the most appropriate. Similar criticisms can be levelled against other evaluators who always try to use, for example, focus groups.

d. Failure to adapt the evaluation design to the programme and the context within which it operates. There are at least 11 factors that have an important influence on how an evaluation should be designed and implemented.[31]

e. Failure to recognize when a programme and the context within which it operates is complex, and when a complexity responsive evaluation design is required.

f. Overlooking the importance of a mixed method evaluation design. There is a growing recognition in the evaluation community of the importance of mixed methods designs that can capture both the quantitative and qualitative dimensions of programmes. Even if mixed method designs are not always used, this should always be considered as a design option.

B. DATA COLLECTION CHALLENGES

The following are some of the concerns and challenges for data collection.

a. High data collection costs reduce sample size and make it hard to include difficult–to–reach groups.

b. Delays in data collection and distribution to operational staff and other stakeholders make it difficult to act on findings in a timely manner.

c. Failure to collect information on the process of project implementation (data often only collected on outcomes).

d. Mono–method bias. Many evaluations rely upon a single approach to data collection (for example, paper questionnaires, on–line surveys, focus groups or case studies). Every method has strengths and weaknesses, so the reliance on a single method increases the risk of misleading or creating over–simplified indicators of important multi–dimensional constructs (such as poverty, wellbeing, empowerment).

e. Preference for numerical indicators. Related to the previous point is the preference for quantitative, numerical indicators as compared to qualitative indicators. Numerical data is easier to collect and analyse but exclusive reliance runs the danger of presenting a one–dimensional picture of phenomena that must be assessed in terms of both quantity and quality. For example, many educational assessments measure the number of schools, teachers, but fail to assess the quality of education. In some cases, the schools are not even operating much of the time, or the new teaching resources never arrived or are not at all being used. This is again a challenge for big data as much of the data is numerical. A related challenge is that many programmes seek to produce behavioural change (as well as numerical outcomes). Behaviour is difficult, but not impossible, to measure numerically.

f. Weak construct validity. Data is used to construct indicators that are intended to measure constructs (inputs, processes, outputs, outcomes and impacts). Constructs are abstract concepts (poverty, well–being, vulnerability, domestic violence, school performance, health of ecosystems) and the validity of the analysis and evaluation findings will be dependent to a significant degree on how well the individual indicators reflect the underlying construct. While conventional evaluations must often rely on proxies that do not adequately capture the underlying construct, the risks are potentially greater for big data as many of the indicators that are used were generated for a completely different purpose.

g. Failure to collect data on difficult to reach groups and those unwilling to be interviewed. A weakness of many evaluation designs is that due to cost and time constraints some of the more remote or difficult–to–reach groups may be left out of the study. In other cases, emergency or security situations make it difficult to reach important sectors of the sample population. Big data and ICTs may present an advantage in reaching these groups. Nevertheless, big data may face other challenges, as remote data collection does not provide the same opportunities to track difficult to reach groups, which are more accessible to the researchers on the ground.

h. Difficulties in collection of information on contextual factors affecting programme implementation and outcomes.

i. Failure to capture information on processes of behavioural change.

j. Need for mixed method designs that collect both quantitative and qualitative data.

k. Failure to collect gender responsive data. Many studies do not disaggregate important indicators by gender. Even when there is a gender breakdown, very few evaluations collect information on important gender dimensions such as time–use, access to and control of productive resources.[32]

l. Low–income and vulnerable groups are usually not involved in design of surveys and interpretation of findings.

C. SAMPLE SELECTION CHALLENGES

One of the biggest challenges for evaluations that do not use experimental designs (which is the vast majority of evaluations) concerns sample *selection bias*. There are two main causes of selection bias: **(a) how participants are selected and (b) how the sample is selected for the evaluation.** The two most common causes of participant selection bias are self–selection and administrative selection of beneficiaries. In the first case, subjects who self– select tend to have attributes that make them more likely to succeed; while in the second case planners or implementing agencies tend to select individuals, communities or institutions that are most likely to be successful. With respect to bias in the sample selection process there are number of factors:

a. The sample frame (list/directory) that is used may not include all units in the population (e.g. Illegal squatters may not be included in the list of addresses used to select the sample);

b. The sample selection procedure may introduce bias (e.g. If the intended respondent is not at home the interviewer may interview a different household member or may find a replacement from a different household). In both cases this may introduce a systematic sample selection bias against people who are less likely to be at home (e.g. long–distance truck drivers or fishermen);

c. Another bias may result from how respondents are defined. Many surveys interview the person defined as 'household head.' This will often mean that wives, or other household members are under–represented;

d. Problems in selecting a well–matched comparison group.

D. DATA ANALYSIS AND DISSEMINATION OF FINDINGS

Examples of common challenges affecting data analysis include:

a. Too little time is often allowed for data analysis. This is often due to the analysis taking longer than expected;

b. Unexpected findings (such as outliers or inconsistent results) that cannot be explained are often ignored. Most surveys do not budget time or resources to return to the field to verify/authenticate inconsistencies;

c. While many evaluations state that triangulation will be used to strengthen the consistency and validity of findings, in practice this is rarely done;

d. While many evaluations state that a mixed methods design (with systematic integration of quantitative and qualitative data) will be used, in practice it is common for quantitative and qualitative data to be collected and analysed independently with little integration.[33]

Examples of challenges facing dissemination of evaluation findings include:

a. Limited audience – evaluation reports are often disseminated mainly to donors and government agencies, and often the reports do not reach important groups such as civil society, affected communities or parliamentarians.

b. Format and Dissemination Strategies– reports often have a standard format and are not designed to be accessible to a wider audience. Very few evaluations employ creative dissemination strategies such as video, data visualization, or coordination with mass media.

c. Timing – reports are produced too late when decisions have already been made on future policies or programmes.

d. Language – reports are often not available in local languages.

TABLE 3–4 CHECKLIST FOR ASSESSING EVALUATION METHODOLOGY WHEN BIG DATA AND ICT ARE INTEGRATED [CAN ALSO BE USED TO ASSESS STANDARD EVALUATION DESIGNS]

A	Evaluation design challenges	
A–1	Have key evaluation questions been clarified?	
A–2	Have all possible design options been reviewed?	
A–3	Has uncritical selection of a preferred design option been avoided?	
A–4	Has the evaluation design been adapted to the programme characteristics and the context within which it operates?	
A–5	Has the need for a complexity–responsive evaluation design been assessed?	
A–6	Has the benefit of a mixed method design been considered?	
A–7	Have opportunities for integrating big data been identified and assessed?	
B	Data collection challenges	
B–1	Are there ways to reduce the cost and time of data collection?	
B–2	Can M&E data reach programme staff more quickly?	
B–3	Is data on programme implementation being collected for the evaluation?	
B–4	Has mono–method bias in choice of indicators been avoided?	
B–5	Has over–reliance on numerical indicators been assessed?	
B–6	Have ways been found to strengthen construct validity?	
B–7	Is data being collected on difficult–to–reach groups?	
B–8	Is data collected on key contextual factors affecting programme performance?	
B–9	Is data collected on processes of behavioural change?	
B–10	Is mixed method data being collected on key indicators?	
B–11	Is gender–responsive data being collected?	
B–12	Are vulnerable groups involved in the design and interpretation of surveys?	
B–13	Have opportunities for integrating big data been assessed?	
C	Sample selection challenges	

C–1	Has the adequacy of the sample frame been assessed?	
C–2	Have potential sample selection biases been assessed?	
C–3	Have biases in respondent selection been addressed?	
C–4	Are the best procedures used for selecting the control/comparison group?	
C–5	Have opportunities for integrating big data been assessed?	
D	Data analysis and dissemination challenges	
D–1	Is sufficient time allowed for data analysis?	
D–2	Are unanticipated findings adequately addressed?	
D–3	Is triangulation systematically used?	
D–4	Is an integrated mixed method design used?	
D–5	Is there a dissemination strategy that targets all stakeholder groups?	
D–6	Are flexible formats used to appeal to all groups?	
D–7	Is there a clear understanding when the reports must be delivered?	
D–8	Are reports available in local languages?	
D–9	Have opportunities for integrating big data been assessed?	

CHAPTER 4
BUILDING BIG DATA INTO PROGRAMME MONITORING

4.1 USES OF MONITORING

In Chapter three monitoring is defined as:

'A continuous internal management activity whose purpose is to ensure that the programme achieves its defined objectives within a prescribed time frame and budget. Monitoring involves the provision of regular feedback on the progress of programme implementation, and the problems faced during implementation. Monitoring consists of operational and administrative activities that track resource acquisition and allocation, production or the delivery of services, and cost records.' (Valadez and Bamberger, 1998: 12).

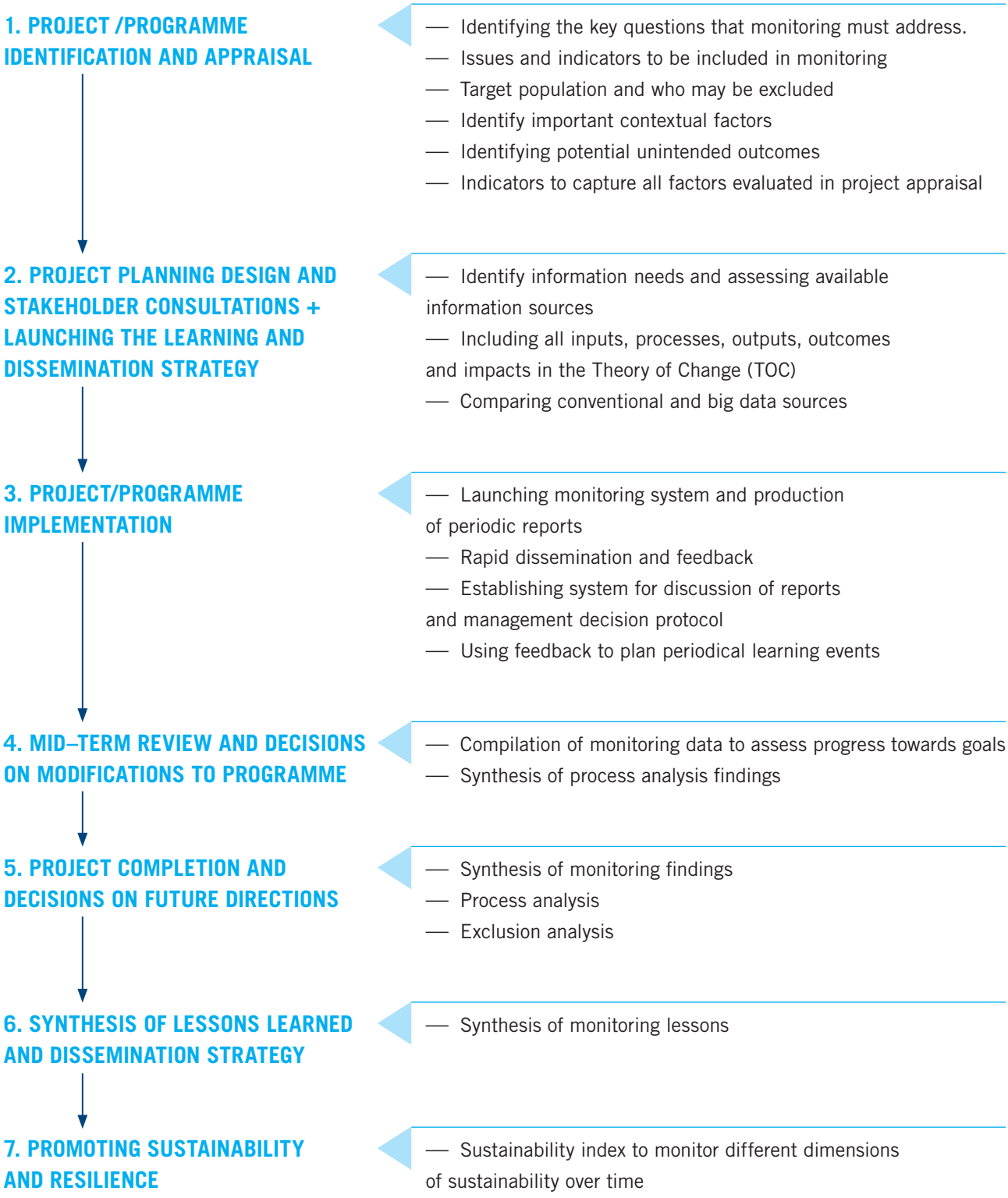
A monitoring system normally has most of the following functions:

- Producing data for a results framework;
- Accountability: did the programme achieve its outcomes in a timely manner and within budget;
- Actionable information on problems detected during project implementation;
- Identifying negative outcomes or groups who are not receiving programme benefits and services;
- Providing data inputs to the programme evaluation;
- Providing inputs for the evaluation of complex programmes. Complex evaluations require a broader range of monitoring indicators than those required in a conventional monitoring system. In particular, more information is required on what happens during project implementation (including behavioural changes at the individual or organizational level), interactions among stakeholders and how these affect programme implementation, and the influence of contextual factors.

Figure 4–1 summarizes the main kinds of monitoring information that are typically required at each stage of the programme cycle. A key requirement is to ensure that the design of the monitoring system begins with a definition of the key questions for which monitoring must provide answers. The system must be *demand driven* (responding to the information needs of stakeholders) and not *supply driven* (focusing on questions of interest to researchers).

At each of the seven stages the monitoring system is required to produce different kinds of information. During stages one and two the focus is on identifying the different types/sets of monitoring information that will be required, as well as assessing possible sources of information in terms of their relevance, coverage, reliability and quality. Often a theory of change will be used to identify the indicators that will be required to measure inputs, implementation process, outputs, outcomes and impacts. It is

FIGURE 4–1 INFORMATION AND OUTPUTS AT EACH STAGE OF A PROJECT/PROGRAMME MONITORING SYSTEM



also essential to identify potential unintended outcomes and to define how they can be monitored and measured. During stage three, the monitoring system becomes operational and begins to generate data on a regular basis to respond to the stakeholders' information needs. The system continues to generate information throughout the project cycle, ending with stage seven, which monitors the extent to which the programme is sustainable and continues to deliver the intended services and benefits throughout its intended lifetime.

4.2 USING BIG DATA FOR POLICY AND PROGRAMME MONITORING

The following chapter focuses on project monitoring, as this is the level at which monitoring is most widely used and because programme and policy monitoring use the same tools as project monitoring. However, there are a number of different monitoring challenges at these levels.

POLICY MONITORING

Most policy assessments tend to focus more on evaluation, seeking to assess the changes that policies have produced over a relatively long period of time. However, there are a number of ways that monitoring is used.

First, some policy assessments develop a TOC that identifies the processes through which change is intended to be produced and the multiple actors who must be influenced. Monitoring can be used to track how effectively the strategy identifies, involves and influences the different agencies and actors. There may also be a number of mileposts (e.g. meetings to involve all stakeholders, issue and review of draft documents, submission of legislation or regulations) that can be monitored.

Second, many policies involve communication campaigns to explain and gain support for the policy. These may involve face–to–face lobbying or mass media campaigns. These campaigns can be monitored to assess how many people are reached, changes in the level of knowledge and change of attitudes. Increasingly, these can be monitored through social media analysis.

Finally, many policies involve physical changes that can be monitored using satellites and remote sensing. Examples include, changes in land–use patterns, transport, population location and movement, access to public services.

Big data analytics can be a powerful tool at the policy level as many policies involve changing operating procedures and relationships among different agencies, and consequently will often require integrating many different data sets, that are normally used separately. Policy outcomes are also affected by many different contextual factors that big data analytics can help analyse.

PROGRAMME MONITORING

As discussed earlier, programmes frequently combine a number of different projects, each of which can be monitored separately using project monitoring tools. Programmes are usually implemented in many different districts, provinces or states and often the documentation is quite weak on exactly which components have been implemented in which areas. Consequently, a major challenge is to develop an effective monitoring system to track exactly what has been implemented where and how effectively the different components were coordinated among the multiple agencies involved in the programme. Big data can often provide valuable, cost and time–saving tools for monitoring and organizing this information.

4.3 CURRENT SOURCES OF INFORMATION FOR PROJECT AND PROGRAMME MONITORING AND THEIR LIMITATIONS

Some of the limitations of widely used sources of monitoring data include:

- Information is incomplete, out–of–date or of poor quality;
- Information is compiled from different sources that are often not consistent or compatible;
- Information is expensive or time–consuming to collect;
- Some kinds of information are difficult to collect;
- Delays in data analysis and dissemination to project staff and other stakeholders;
- Difficult to collect information either on groups or contextual factors beyond the direct scope of the project;
- Difficult to capture information on how stakeholders feel about the project;
- Difficult to measure processes;
- Difficult to capture information on unintended outcomes.

The following section discusses how big data and ICTs can contribute to addressing these limitations, and to strengthening the operational utility of monitoring.

More information on Conventional sources of information for programme monitoring and their limitations is available at: bit.ly/2gdoaE1

TABLE 4–1 CASE STUDIES ILLUSTRATING HOW BIG DATA CAN BE USED TO STRENGTHEN PROGRAMME MONITORING

a. Using satellite imaging to monitor changes in forest cover in protected areas [Source: GEF and UNDP] See Box 4–1.
b. Monitoring migration and labour market shocks using mobile phone call data [Source: World Bank] See Box 4–2
c. Mining tweets to monitor food price crises in Indonesia [Source: UN Global Pulse]. See Box 4–3
d. Estimating migration flows using online search data [Source: UNFPA in collaboration with UN Global Pulse] Project summary available at: http://unglobalpulse.org/projects/migration-search-data
e. Using crowdsourcing for real–time tracking of trends in commodity prices [Source: WFP and FAO in collaboration with UN Global Pulse] Project summary available at: http://unglobalpulse.org/projects/high-frequency-food-price-data-indonesia
f. Training communities to use GPS mapping to create maps of the services in their communities Source: [UN Global Pulse Data Innovation Competition. Indonesia]. Project summary available at: http://www.unglobalpulse.org/blog/data-action-when-communities-engage-mapping-urban-villages-together
g. FAO's AQUASTAT integrated data base: a reference source for many potential monitoring activities [Source: FAO] More information available at: http://www.fao.org/nr/water/aquastat/main/index.stm
h. Using real–time monitoring to track how countries are coping with crises using available national databases [Source: UNICEF in collaboration with UN Global Pulse] Project summary available at: http://www.unglobalpulse.org/projects/unicef-contribution-global-pulse-establishment-real-time-monitoring-pilots

4.4 HOW BIG DATA AND ICT CAN STRENGTHEN MONITORING SYSTEMS

Table 4–1 gives examples of ways big data and ICTs have, or potentially could be used to strengthen programme monitoring. In a number of the examples, big data were used for feasibility or planning studies, where data was only collected at one point in time or for analysing trends based on data up to the time of the planning/feasibility phase. However, these are all scenarios in which it would have been possible to continue real–time (or rapid) data collection once the project was launched, and to build these methods into a monitoring system. While the cases mainly describe how the big data was generated, all of the examples also involve the use of smart data analytics for the integration and analysis of the data. This involves the combination of different data sets into an integrated data platform where advanced statistical analysis using techniques such as time–series analysis and predictive analytics can be used.

TABLE 4–2 POTENTIAL CONTRIBUTIONS OF BIG DATA AND ICT IN PROGRAMME MONITORING

STAGE OF THE PROJECT CYCLE	POTENTIAL SOURCES OF BIG DATA AND ICT–GENERATED DATA	
	BIG DATA	ICT
1. Project identification and appraisal	a. Analysis of tweets to identify potential social and political conflicts b. Analysis of tweets to identify the spread of infectious diseases c. Satellite images to identify poor and vulnerable groups d. Satellite images to identify protected areas — Satellite images to detect, for example: areas of drought, flooding, deforestation, quality of road maintenance — Using satellites to detect illegal land invasions, or increasing population concentrations in communities that have received drinking water. e. Using social media to detect increases in violence against woman and domestic violence f. Crowdsourcing[34]	— Using mobile phones to reduce the costs of collecting data on a random sample of the population
2. Project planning and design and stakeholder consultation	a. Using the framework presented in Table 5–1 to identify available data and its quality b. Assessing the possible availability and utility of big data to fill some of the gaps in current monitoring data. c. Analysis of spatial data[35] d. Integrating multiple sources of data, including satellite images, remote sensors, national surveys to provide an integrate database on topics such as water resources that can be used planning, or potentially monitoring[36].	— Using mobile devices to facilitate stakeholder consultations
3. Project implementation	a. For large programmes: use of satellite data to monitor trends in forest coverage, land use, water resources b. Integrating data from different sources to provide more robust, multi–dimensional monitoring data c. Analysis of social media (details)[37] d. Using mobile phone activity[38] e. Using Twitter to monitor public sentiment[39] f. Online search data[40] g. Using digital signals[41] h. Nowcasting using social media signals[42] i. Real–time monitoring[43]	— Building real–time feedback from smart–phones into monitoring systems — Mobile phone–based surveys[44]

4. Mid–term review	a.Inputs from integrated databases covering a range of indicators b.Data on a broad range of contextual factors not accessible from conventional data sources c.Analysis of social media d.Using all of the above to incorporate data on vulnerable and difficult to reach groups	— Inputting data from digital monitoring systems — GPS mapping data to track changes in access to services, population density, frequency and location of crime and traffic accidents
5. Project completion	a. Similar to Mid–term review (above)	— Similar to Mid–term review (above)
6. Synthesis of lessons learned and dissemination	a. Similar to mid–term review (above) b. Create easy to grasp data visualizations for different groups of stakeholders	— Similar to mid–term review (above)
7. Programme continuation and sustainability	a. Periodic feedback from satellite images, automatic sensors and integrated data bases on a range of sustainability indicators	— Feedback from smart–phone generated data on sustainability indicators

Table 4–2 lists some of the potential contributions of big data and ICTs at each stage of the programme monitoring cycle. However, there is considerable overlap as the same sets of data could be used for planning, programme monitoring or mid–term and final reviews. The following are examples on how big data can be applied to monitoring.[45]

A. SATELLITES AND REMOTE SENSING (OFTEN IN REAL–TIME)

(Box 4–1 > Using satellite imaging to monitor changes in forest cover in protected areas)

Satellite images are becoming increasingly refined so that different levels of resolution can be combined for different purposes. For example, lower resolution data can use thermal images to detect types of crops while higher resolution captures more detail on other types/sets of data. These can be used to:

- Monitor the movement of populations, changes in forest cover and land use;
- Monitoring the impact of man–made and natural disasters;
- Provide more economical and faster estimates of economic growth and poverty[46]
- Monitor depletion of biodiversity.

BOX 4–1 USING SATELLITE IMAGING TO MONITOR CHANGES IN FOREST COVER

The Global Environmental Fund (GEF) and UNDP are collaborating on a worldwide programme to support biodiversity conservation by looking at protected areas and protected area systems. One of the three central evaluation questions concerned the impacts of the GEF and UNDP interventions:

What have been the impacts and contributions of GEF and/or UNDP to support biodiversity conservation?

The evaluation used a quasi–experimental design as it was assessing programmes that had been operating for a number of years, therefore it was not possible to use random assignment. Protected areas (PA) where GEF/UNDP were working were compared with matched PAs where the organizations were not working. Satellite images were used to measure forest coverage and other indicators. In countries such as Mexico, where the quality of non–satellite data was good, the quality of matching was improved through propensity score matching using indicators that could be obtained from surveys, agricultural records and other sources. Time series data could be obtained from the satellite images so that changes in, for example, forest coverage could be compared over relatively long periods of time. Smart data analytics were used to integrate the different data sources into an data platform, and to permit time–series analysis on the integrated data.

B. MOBILE CALL DATA RECORDS

(Box 4–2 > Monitoring migration and labour market shocks using mobile phone call data)

Phone companies keep very detailed records on all calls, including the duration of the call and the location of the caller and the amount of airtime purchase. While there are often limitations on access to these records, properly anonymized and aggregated mobile data can provide a valuable source of monitoring data.

- Monitoring population displacement;
- Capturing seasonal and temporary migration (often overlooked in regular surveys);
- Detecting impacts from small scale violence;
- Using trends in air–time purchase as an indicator of poverty;
- Understanding instances of violence against women and domestic violence.

C. ANALYSIS OF SOCIAL MEDIA DATA

(Box 4–3 > Mining tweets to monitor food price crises in Indonesia)

Platforms such as Facebook and Twitter are publicly available and are being widely used to analyse peoples' attitudes and sentiments. When used in emergency relief (such as locating victims trapped by

earthquakes or floods) a challenge is to identify valid information from irrelevant data[47]. The following are examples of applications.

- Identify potential conflicts and emergencies using sentiment analysis;
- Monitor the spread of diseases;
- Identify trends in poverty and food prices;
- Predict increases in unemployment or crisis–related stress;
- Monitor violence against women and domestic violence;
- Analysis of social media for nowcasting.[48]

D. INTERNET/TEXT

('Estimating Migration Flows Using Online Search Data ', Global Pulse Project Series available at: http://www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Search_Migration_2014_0.pdf)

Internet searches can be used to analyze the frequency with which words or phrases appear over time and in different geographic locations. UN Global Pulse and UNFPA conducted a study to explore how online search data could be analysed to understand migration flows. Using Australia as a case study, Google search query data from around the world was disaggregated by country and compared to historical official monthly migration statistics provided by UNFPA. Correlations were observed between relevant search queries (for example, searching for 'jobs in Melbourne') and official migration statistics (number of people who migrated to Melbourne).

- Internet queries monitor the frequency of key words to understand trends and identify potential issues;

BOX 4–2 MONITORING MIGRATION AND LABOUR MARKET SHOCKS USING MOBILE DATA

The purpose of this study was to monitor internal population migration responses to changes in the labour market in different regions. The study had access to phone company call data records (CDRs) over several years. These were complemented by census data, labour market statistics, satellite data (to assess the effect of weather conditions on migration) and information on domestic and international commodity prices. The analysis of CDRs is extremely time– consuming and tedious, as millions of unstructured individual files have to be processed, anonymzed and aggregated to ensure that the privacy of individuals is observed. These then have to be compared on a daily basis with labour market statistics and weather records.

The analysis was conducted in three steps: (1) isolate local shocks to labour demand such as weather or commodity prices and identify the resulting labour market outcomes, (2) draw on the identified shocks to labour demand and then use CDRs to relate this to migration patterns and (3) estimate the effects of migration dynamics on labour market equilibrium.

Source: World Bank (undated) big data in action for development. Latin American and Caribbean Regional Office in collaboration with Second Muse

- Identifying proxy indicators to monitor social–economic data in real time that is normally collected through expensive and time–consuming surveys;[49]
- Sentiment analysis;
- Lexical analysis to understand elements of culture or financial literacy to help design micro–finance programmes;
- More complex analysis of unstructured text data.

BOX 4–3 MINING TWEETS TO MONITOR FOOD PRICE CRISES IN INDONESIA

This project explored Twitter conversations in Indonesia to understand how the volume of chatter relates to macro–level events. In particular, the project monitored food–price related tweets between January 2011 and December 2012 to see if variations in their volumes could be connected with food and fuel price inflation.

The first step in the project was to create a taxonomy of relevant keywords and phrases in Bahasa Indonesia to extract tweets relevant to the prices of food and fuel. As a second step, the researchers defined categories in which to classify the tweets, depending on the sentiment they expressed (i. e. 'positive', 'negative', 'confused'). A representative, hand–labelled sample of tweets was then used to train a monitor to classify the tweets in the correct category and detect the sentiment of new tweets being published in real time.

Finally, all relevant data was analysed to calculate the proportion of tweets related to each theme, and to determine the statistical pattern of conversation for each category. The general volume of relevant tweets, independent from the conversation, was analysed, and three spikes in volume of tweets were observed in 2012, corresponding to three real–world events:

- **July 2012:** a global soybean price rise, which affected the prices of tempeh and tofu, two dishes made of soybeans consumed by many Indonesians as affordable protein–rich options.
- **March 2012:** a proposal by the Indonesian Government that its fuel subsidy would be cut by 33 per cent, which caused violent protests and raised concerns that food prices may eventually be affected.
- **November 2012:** approval of a law establishing a new food agency with policymaking authority to help Indonesia reach self–sufficiency in staple foods, including rice and soybeans.

The initial research results showed that around the same time when these real–world events occurred, conversations related to food prices also spiked dramatically among Indonesian Twitter users, illustrating the potential value of employing regular social media analysis for early warning and impact monitoring.

Source: UN Global Pulse (2013). Mining Indonesian Tweets to understand food prices crises. Available from http://unglobalpulse.org/projects/social–media–social–protection–indonesia

E. CROWDSOURCING

('Feasibility Study: Crowdsourcing High- Frequency Food Price Data in Rural Indonesia', Global Pulse Project Series, available at: http://www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Crowdsourcing_Food_Prices_2015_0.pdf)

Crowdsourcing obtains feedback from large numbers of people or groups. (e.g. A feasibility study done in Nusa Tenggara Barat, one of Indonesia's poorest provinces, which involved recruiting a trusted network of local citizen reporters to submit food price reports via a customized mobile phone application). This can combine direct responses from individuals with data collected during community or group meetings and sent as a summary of group agreements. Examples include:

- Compiling opinions of communities and individuals on development priorities;
- Monitoring trends in food prices and other indicators;
- Obtaining feedback from particular groups such as young people, people eligible to vote in a particular country or people providing feedback on emergencies such as floods, earthquakes or conflict.

F. GPS MAPPING

('Data in Action: When communities engage in mapping urban villages,' article available at: http://www.unglobalpulse.org/blog/data-action-when-communities-engage-mapping-urban-villages-together)

GPS–enabled mobile phones can identify and record the location from which a call is placed or an audio or visual recording is made. This can be used to create maps locating particular features such as public services or to permit the location from which calls were made. Applications include:

- Mapping the location of services such as water supply, bus stops or stores selling food or other items of interest;
- Mapping problem areas (e.g. poor quality services, high conflict areas, or traffic accidents, election abuses or the location of victims of earthquakes);
- Monitoring traffic density or routes travelled by, for example motorcycles or women collecting water or fuel.

G. DEVELOPING INTEGRATED DATA PLATFORMS

(Aquastat, FAO's integrated data platform: a reference source that could potentially be used for programme monitoring)

Big data makes it possible to combine data from conventional sources such as censuses, national household surveys and farm surveys with data generated in real–time from sources such as satellite and drone images, social media, mobile phone records and digital financial transactions to broaden the range of data that can be incorporated into a database. There are a number of challenges to be addressed as agencies used to working with conventional, static data learn how to integrate this with real– time data that is constantly changing and the validity of which is more difficult to assess. Examples include:

- Integrating data for different departments and agencies to permit comparisons of indicators across agencies and time;
- Integrating conventional data platforms with big data from sources such as Twitter;
- Real–time monitoring selecting indicators that are available from public sources and which are

comparable across agencies, and often across countries and which permit the tracking of trends over time.

POTENTIAL CONTRIBUTIONS OF ICTS

Table 4–2 identifies some of the ICT data collection and analysis tools that can be used, usually in coordination with big data to strengthen programme monitoring. These include:

- Using mobile devices to reduce the costs and time of data collection and analysis;
- Using GPS–enabled mobile phones to monitor population movements;
- Using smart phones to facilitate stakeholder consultations;
- Using smart phones for on–line development of theories of change;
- Real–time feedback on monitoring indicators through smart–phones;
- Conducting surveys through phone–based SMS, and automatic surveys;
- GPS mapping to monitor availability of services;
- Using smart phones to collect sustainability indicators (including through photos which can be analysed automatically).

4.5 STEPS IN THE INTEGRATION OF BIG DATA INTO PROGRAMME MONITORING

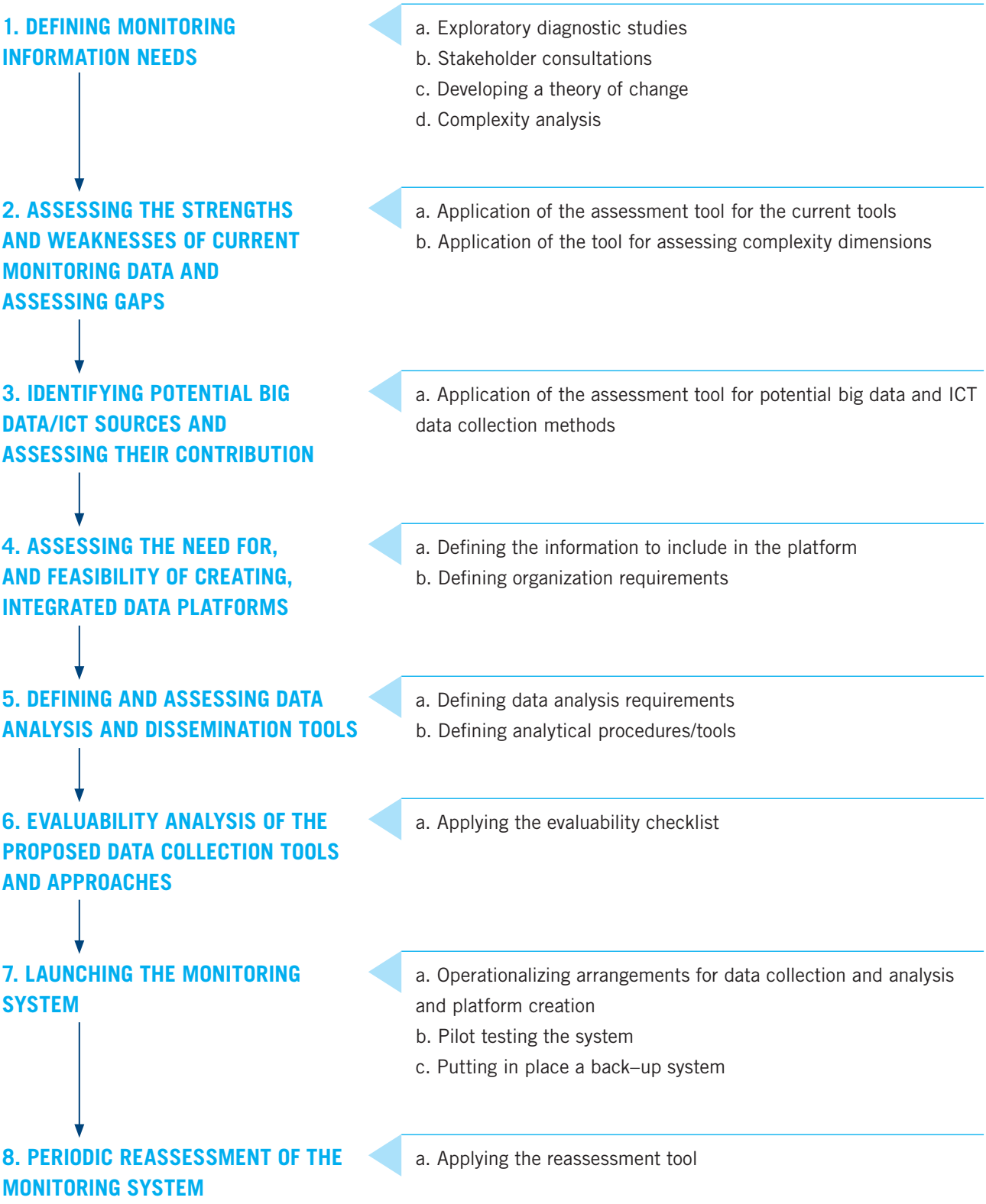
Figure 4–2 identifies eight steps for the integration of big data into a programme monitoring system.

Step one: Defining monitoring information needs

The first step is to define the information that is required to cover the monitoring information needs of all key stakeholder groups, as well as to provide information on all stages of the programme design. The information to be collected must come from the key questions of concern to stakeholders (demand driven) and is not based on data, which is available from different big data sources (supply driven). The information needs are defined based on:

a. Exploratory diagnostic studies to understand the programme and the context in which it operates;
b. Stakeholder consultations;
c. Review of the design of previous monitoring systems for similar projects and assessments of how well they worked;
d. Analysis of the programme design and how this is translated into the theory of change
e. Information requirements for the results–based management system;
f. Application of complexity analysis to determine whether complexity M&E approach is required. If some dimensions of a programme are defined as complex, it will usually be necessary to increase the volume and sophistication of the monitor data to be collected;

FIGURE 4–2 SETTING UP THE BIG DATA/ICT–BASED MONITORING SYSTEM



g. If a programme is intended to contribute to the achievement of one or more of the new Sustainable Development Goals [SDGs] then this may increase the monitoring data requirements.

Step two: Assessing the strengths and weaknesses of current monitoring data

Often an initial monitoring design will be developed based on the approaches used in similar projects. This will then be assessed using the checklist of indicator quality and suitability [Table 3–2]. Examples of the assessment criteria include: the quality and completeness of the information, and whether there are gaps in the information needs, identified in step one.

Step three: Identifying potential big data and ICT data sources and assessing their potential utility

Table 4–2 identifies some of the potential big data and ICT sources that can be considered for each stage of the monitoring cycle. Examples of applications of the different techniques are given in footnotes to the table and in case studies given in boxes throughout this chapter.

Step four: Assessing the need for, and the creation of integrated data platforms and hybrid information systems

Big data can bring together different data sets and integrate them into a single platform, which permits comparisons across time and across different kinds of analysis[50]. A hybrid approach is required in which cleaned, tightly structured and well understood data sets are combined with less structured and less clean and less understood multiple sources of data that are becoming available through big data.

The Aquastat platform is an example of how FAO is developing and using large, integrated databases to synthesize all available national–level data on topics such as water resources and water utilization.

Step five: Defining and assessing data analysis and dissemination tools

Modern data analytics may draw on a wide variety of techniques for analysis of a wide variety of data types. An assessment should be made of the potential opportunities for using these different techniques as well as some of the challenges and limitations.

Step six: Evaluability analysis

Once the proposed system has been designed, it is important to conduct an evaluability assessment to determine if it is technically, organizationally, politically and economically feasible. It is of course also necessary to determine whether all of the required information is being collected.

Step seven: Launching the system

Launching the new monitoring system involves:

- Defining coordination arrangements with all of the different agencies that will provide information and use outputs;
- Defining the organizational and management structure;
- Defining and putting in place the additional management and technical systems for the big data components. These can be quite different from the management of conventional monitoring systems;
- Where possible, the monitoring system should be pilot–tested on a small scale to identify and correct bugs and other problems.

A mistake that evaluators should avoid is to assume the new more sophisticated and complex systems will work from the start, and often the previous conventional systems are disbanded. Even when the new systems are fundamentally sound, they can take a significant amount of time to become fully operational and to be able to provide all of the information that management and stakeholders require. Consequently, it may be useful to consider keeping in place a back–up system until the new system is completely functional. Often, this may be to continue with the previous system during the transition period.

Step eight: Periodic assessment of the monitoring system

Periodic assessment should be made to evaluate how well the new system is working. This should involve all stakeholders and not just technical experts.

CHAPTER 5
BUILDING BIG DATA INTO PROGRAMME EVALUATION

5.1 ELEMENTS OF A DYNAMIC EVALUATION SYSTEM

A dynamic evaluation system can:

- Evaluate the outcomes of policy, programme and project interventions;
- Address the key evaluation questions for each of the four types of evaluation (policy, formative, developmental and summative) described in Table 3–2;
- Provide robust estimates of the extent to which the observed changes in outcomes can be attributed to the programme interventions;
- Open up the 'programme black box' and assess the extent to which failure to achieve an outcome is due to design failure or to implementation failure;
- Assess the outcomes of complex programmes operating in complex contexts;
- Design evaluations operating under real–world budget, time and data constraints;
- Provide rapid feedback on outcomes;
- Provide predictive and well as retrospective analysis;

5.2 POTENTIAL APPLICATIONS OF BIG DATA FOR POLICY AND PROGRAMME EVALUATION

Most of the discussion in this chapter will focus on project evaluation, both because this is the type of evaluation most widely discussed in the literature, and because most of the evaluation tools are also used for policy and programme evaluation. However, there are a number of special issues that affect policy and programme evaluations. For more detail on the topics covered in this chapter see: bit.ly/2grvcbm

POLICY EVALUATIONS

A wide range of economic, political, socio–cultural, demographic and other national and international factors affect the outcomes of policy interventions. Much of this information is difficult to capture and analyse with conventional evaluation tools, and consequently this is an area where big data can potentially contribute.

Other areas in which big data and data analytics can contribute include: analysis of the processes of policy implementation, assessing behavioural change, and information and attitudes to the policy of different sectors of the target population. Data analytics can also help combine different data sets into an integrated database. Finally, data visualization can also help communicate progress and findings in a user–friendly way to different sectors of the target population.

PROGRAMME EVALUATION

Programme evaluation usually involves identifying and organizing information of large numbers of project components that often operate in different, and not well documented, combinations in different communities and locations. There are also a large number of contextual variables that affect outcomes. These are both areas where big data can contribute.

A comprehensive evaluation should combine analysis of both processes of implementation and behavioural change, as well as measuring outcomes. A major challenge, where big data can contribute, is the identification of a counterfactual. This often requires the identification of communities or groups that are similar to groups receiving programme benefits. As most programme evaluations will use a quasi–experimental design, it will often be necessary to use techniques such as propensity score matching to improve the match of project and comparison groups. Big data and data analytics are well suited for the collection and organization of these kinds of multi–variable data platforms.

5.3 POTENTIAL APPLICATIONS OF BIG DATA AND ICTS IN PROJECT EVALUATION

Table 5–1 illustrates some of the ways that big data and ICTs can strengthen evaluation designs at each stage of the project cycle. For example, during the stage of project identification and appraisal, satellite images can provide information on migration patterns, economic status of different villages or regions, rainfall and access to infrastructure. Similarly, social media can provide information on issues of concern to different groups and attitudes towards particular issues.

Frequently, big data and ICT–generated data will be combined (see early discussion of the data continuum). This permits the use of mixed methods designs that combine multiple–sources of data and strengthen the scope and quality/validity of data.

5.4 WHEN AND HOW BIG DATA CAN BE INTEGRATED INTO PROGRAMME EVALUATION

There are three different ways that big data–responsive designs can be incorporated into evaluation designs:

a. Big data can be incorporated into a conventional evaluation design. For example, mobile phone data or remote sensing can be used to complement survey and key informant data used to estimate poverty trends in a particular region.
b. Big data can strengthen a conventional evaluation design. For example, remote sensing can be used as proxy for an evaluation of the effect of interventions in maintaining forest coverage in protected areas. Data analytics can also be used to combine several conventional data sets into an integrated data platform.
c. Using an evaluation design based on collection and analysis of large data sets that cannot be analysed using conventional computer systems. For example, understanding sentiment on issues such as the use of biofuels through the analysis of large amounts of social media data.

When discussing potential applications of big data it is important to clarify the conditions under which big data are most applicable. Table 5–2 identifies seven dimensions in terms of which the applicability of big data can be assessed.

5.5 INTEGRATING BIG DATA AND ICT INTO THE DESIGN STAGE OF THE PROGRAMME EVALUATION

a. Understanding the evaluation questions to be addressed

Many evaluators have a preference for a particular evaluation design (e.g. RCT, regression discontinuity, focus groups, case studies), which they try to apply to all evaluations. However, there is no one–size–fits–all evaluation design that is appropriate in all situations. In fact, the choice of evaluation design is largely determined by two sets of factors: (i) the characteristics of what is being evaluated and the context (economic, political, socio–cultural, environmental) within which the programme is being implemented, and (ii) the questions that the evaluation is asked to address.

TABLE 5–1 POTENTIAL WAYS THAT BIG DATA AND ICTS CAN STRENGTHEN PROGRAMME EVALUATION

EVALUATION ACTIVITIES AT EACH STAGE OF THE PROGRAMME	BIG DATA	ICT
1. Project identification and appraisal		
Initial diagnostic studies and defining affected populations	a. Anaysis of Remote sensing b. Analysis of social media data and internet queries to identify potential issues and problems	a. Doing mobile phone surveys
2. Project planning and design		
Developing a theoretical framework	a. Theory of Change from passive observations	a. Theory of Change from active data collection
Selecting the evaluation design	a. Identifying potential big data contributions to each design option b. Building in mixed methods	a. Using ICTs (e.g. excel files, PDF documents) to incorporate both quantitative and qualitative methods
Designing evaluation of complex programmes	a. Using predictive analytics and systems analysis to model complex systems and causal pathways	a. Running regression models
3. Project implementation		
Developing early warning systems	a. Using social media data, call phone records and remote sensing b. Creating real time data visualizations	a. Conducting mobile phone surveys
Data collection	a. Remote sensing b. Social media data c. Analysis of other forms of social media d. Remote sensors e. Integrated data platforms f. Developing ontologies for collection of multiple sources of data on a common theme	a. Mobile surveys b. Micro narratives c. Biometric data d. GPS mapping e. Incident reports via phone and internet
Process analysis	a. Real–time feedback on project implementation (dynamic data platforms) b. Satellite tracking of population movements, growth of human settlements	a. Smart phone video and audio recording during meetings, work groups b. Web–based M&E platforms allow for better documentation of processes

Collecting qualitative data	a. Automatic analysis of text–based data, sentiment analysis	a. Using people to extract information from audio or video recordings
Collecting contextual data	a. Satellite images can track physical changes over large areas b. Crowdsourcing provides feedback on natural disasters, political protects and spread of disease	a. Web based access to archive repositories
Quality control of data collection	a. Bias assessment based on comparisons with ground truth data b. Data consistency checks	a. GPS enabled phones/ tablets can check location of interviewers b. Internal consistency checks on phone surveys c. Randomly activated audio recorder can listen–in to interview
Monitoring behavioural change	a. Social media data b. Analysis of phone records and financial transaction records c. Large scale surveys of household purchases (using smart phones to record food labels)	a. Human analysis of video and audio–recordings at project locations, in the community or households improve capacity to monitor behaviour directly
Sample selection	a. Using remote sensing for area sampling b. Using satellite imagery to select samples based on physical conditions of houses (e.g. thatched roofs used as indicator of low–income household) c. Calibration of big data sources with ground truth data	a. Assessment or selection of random samples b. Rigorously selected automatic dialled samples (combined with human follow–up)
4. Mid–term review		
Data analysis and data visualization	a. Real time big data analytics results b. Data visualization dashboards	a. Mobile phone surveys b. Internet surveys (e.g. Survey monkey)
5. Project completion		
Data analysis and interpretation	a. Management of multi–dimensional data sets b. Smart big data analytics c. Analysis of complex QCA case studies d. Synthesis of all data sources	a. Similar to Mid–Term review

6. Synthesis of lessons learned and dissemination strategy		
Dissemination of findings	a. Interactive data visualizations	a. Dissemination via phones, tablets and internet
7. Planning and implementing a sustainability strategy		
Evaluating sustainability	a. Longitudinal big data sets	a. Periodic cell– phone and internet surveys

It is also important to decide which of the four evaluation approaches identified in chapter (policy evaluation, formative evaluation, developmental evaluation and summative evaluation) is being used – as the key questions to be addressed are different for each approach (See Table 3–2).

b. The range of possible evaluation designs

There are at least eight possible evaluation designs that should be considered, as well as many sub–variations. The first six are standard designs that can be applied to most programme evaluations.

Design one – Experimental and quasi–experimental designs: These designs compare changes over time (usually between project launch and project completion) for intended outcome variables for the project group and for a comparison group, which is either selected randomly or through matching (statistical or judgmental).

Design two – Statistical designs: These frequently use econometric techniques at the national level to assess the effects of a policy or national programme (for example, on the production of low–cost housing or the effects of tax reform on access to education) by comparing indicators with other similar countries while controlling for macro–level indicators. Big data can potentially increase the range of indicators that can be used in the analysis.

Design three – Theory–based designs: These develop models describing how the programme is intended to achieve its outcomes and impacts and the random chains through which they will be achieved. The models can also include contextual factors that might influence outcomes. The effectiveness of the programme is assessed, by comparing intended outcomes with observed outcomes. Big data can potentially provide a broader range of data on contextual factors, and can also provide real–time feedback on implementation indicators and behavioural change. This makes it possible to continually test and update the programme theory based on continuous feedback.

Design four – Case based methods: These methods take the case (individual, household, school, community, country) as the unit of analysis. Cases can be selected to illustrate findings of the quantitative analysis or they can be used as stand–alone evaluation design. In this latter scenario, it is possible to match cases that did and did not receive project services or inferences can be made only on the basis of a sample of beneficiary case studies. Recently there has been an increasing use of qualitative comparative analysis (QCA) where a matrix is prepared for each case listing household characteristics. An analysis is then conducted to determine which set of factors (configuration) is associated with the achievement of outcomes and which configuration did not contribute to the achievement of the outcome. Normally the

TABLE 5– 2 FACTORS DETERMINING THE APPLICABILITY OF BIG DATA AND DATA ANALYTICS IN PROGRAMME EVALUATION

HIGH APPLICABILITY	LOW APPLICABILITY OF BIG DATA
1. Programmes where conventional evaluation designs are considered to be methodologically weak	1. Programmes where conventional evaluation designs are considered perfectly adequate and there is no obvious need for a new approach
2. Programmes that use easily measurable (and readily available) physical measurement such as climate change, urban growth, traffic patterns	2. Programmes that rely on social and behavioural indicators such as violence against women and domestic violence, community organization where digital data is not easily available
3. Availability of big data indicators with high construct validity [indicators were collected for a purpose relevant to the evaluation]	3. Big data indicators with low construct validity [proxy indicators generated for a different purpose and proxy validity is not clearly validated]
4. Programmes with a relatively long duration and where (real– time) time series data can be generated	4. Programmes where time series data with high granularity do not add value
5. Programmes that will continue to operate after the initial proof of concept so that prediction is possible	5. Experimental projects the sole purpose of which is to test a theory
6. Programmes where there are large numbers of potential variables that might affect outcomes and where there is no articulated theory of how outcomes are expected to be achieved	6. Programmes where simple models based on existing data correctly articulate the theory of change
7. No data privacy and security considerations	7. Where risks regarding privacy and security may overweight potential benefits

analysis is conducted with samples of 50 cases or less, but it might be possible with big data analytics to increase the sample size or the number of variables included in the matrix for each case.

Design five – Participatory methods: These methods seek to collect the perspectives of communities and groups affected by projects through different types of group consultations (such as PRA and most significant change), or through in–depth observation where the researcher tries to observe communities and become part of their activities rather than conducting structured interviews. These methods assess programme effects based on the opinions and perspectives of the affected populations, often combined with in–depth observation. Participatory methods are usually conducted on a relatively small scale. However, organizations such as the World Bank supported Social Observatories in India that worked with communities to develop and test survey instruments, which could then be administered to as many as one million households.

Design six – Review and synthesis: All evaluations, which have been conducted on a particular topic (such as the effects of good drinking water on children's health) that achieved acceptable standards

of methodological rigor are reviewed and the findings are synthesized. The average effect size of the average findings on programme outcomes provides an estimate of the range of potential effects of a well–designed programme. Synthesis reviews can be used both to provide a more macro–level assessment of the effects of a particular intervention in many different contexts, or the findings can be used to guide the design of a new evaluation by helping identify the range of factors that affect outcomes. The findings can also be used to estimate the required sample size. The larger the expected effect, the smaller the required sample to measure the effect is. Big data may be able to increase the number of variables included in the analysis, while data analytics can assist with the analysis when large numbers of studies are being synthesized.

There are two other sets of designs that have been developed specifically to evaluate complex programmes.

Design seven – Holistic designs: These designs use methods such as systems analysis and sociometric analysis to study a complex system in its entirety. While these approaches have permitted insightful descriptive analysis, and while they offer great promise for complexity evaluation, there are not yet many examples where these kinds of evaluation have been conducted.

Design eight – Unpacking complex programmes into a set of components, each of which is easier to evaluate: After conducting a holistic analysis to understand how a programme is affected by the broader systems within which it is embedded, an assessment is made as to whether it is feasible to break the programme into a set of elements (such as the main services it provides, different levels of the theory of change or different regions or levels at which it operates). After each element has been evaluated separately, the findings of the different component evaluations are reassembled to assess its overall impact in the real–world context in which it is embedded.

For a more detailed discussion of issues in the integration of big data into evaluation planning, design and implementation, see Jackson, 2015 Sections C and D. For more detail on the different evaluation design see Vaessen, Raimondo and Bamberger, 2016.

5.6 OTHER DESIGN ISSUES

a. Trajectory analysis

Programme effects can occur over different periods of time and evolve according to different trajectories (See Figure 5–1). While some projects produce steadily increasing outcomes over the project lifetime (Scenario two), in other cases effects may reach a maximum and then gradually decline (Scenario three). This often happens when projects require a high level of maintenance (e.g. irrigation canals and pumps). When funding is no longer available for maintenance, or this ceases to be a priority (e.g. after the completion of donor involvement), it is common for maintenance to deteriorate and the level and quality of services declines. In other cases, most effects may be produced at a particular point in time (Scenario three), for example when a road is completed. Understanding the expected trajectory of outcomes is critical for determining when the evaluation should be conducted.

b. Complexity–responsive evaluations

Chapter two described some of the main dimensions of complexity that must be addressed in a programme evaluation. When projects are considered 'complex' it is difficult to evaluate outcomes using conventional evaluation designs. In these cases a complexity–responsive evaluation design will normally be required. (see http://unglobalpulse.org/sites/default/files/Annex%204_Big_Data_and_ME_Report_0.pdf)

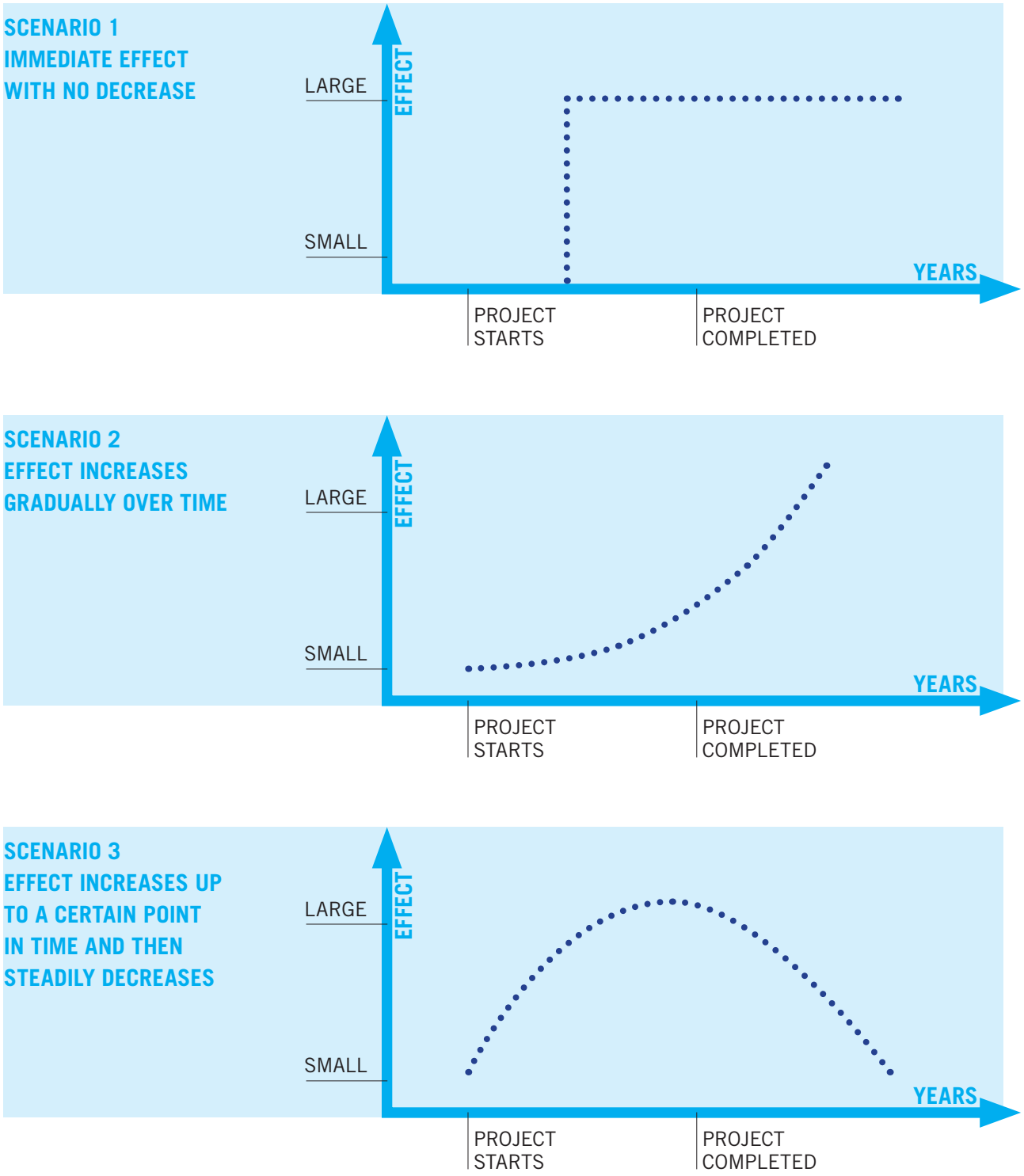
c. Sustainability analysis

For operational reasons, many impact evaluations are conducted around the time that project implementation is complete and the programme moves into the operational phase. The reason for this is that many development agencies only fund the implementation phase and for accountability purposes they require an evaluation of the project phases they funded. The implementation phase often ends soon after the schools have been constructed, the road or irrigation system has become operational etc. Consequently, an evaluation conducted at this point in time is too early to assess whether the financial, institutional, organizational and political mechanisms are in place to ensure that it will continue to deliver services. Therefore, a sustainability responsive design should be put in place.

d. Equity–focused evaluation

One of the central development objectives of most international development agencies is to promote equity, to ensure that programme benefits reach the poorest and most vulnerable groups and to ensure that programmes contribute towards the achievement of broader equity goals. However, many evaluations only measure aggregate outcomes (e.g. on average a higher proportion of children attend school or that the proportion of the population below the poverty line has been reduced). There is an extensive body of research showing it is quite common to achieve aggregate improvements while the gap between the poorest, for example, 20 per cent and the rest of the population may not have been reduced or

FIGURE 5–1 TRAJECTORY ANALYSIS: DIFFERENT SCENARIOS FOR HOW PROGRAMME EFFECTS EVOLVE OVER TIME
Source: adapted from Bamberger, Rugh and Mabry, 2012:204 and Woolkock ,2009



may even have increased (Bamberger and Segone, 2011). Consequently, many evaluations may need to incorporate an equity focus.

5.7 CASE STUDIES ILLUSTRATING THE INTEGRATION OF BIG DATA AND ICT INTO IMPACT EVALUATIONS

Table 5–3 presents examples of how big data is being integrated into each of the evaluation designs described in the previous section. Although these examples are still relatively few and hard to find, this shows that progress is being made and that big data approaches have a potentially wide applicability.

TABLE 5–3 CASE STUDIES ILLUSTRATING HOW BIG DATA AND ICTS WERE USED TO STRENGTHEN EACH OF THE MAIN EVALUATION DESIGNS

EVALUATION DESIGN	EXAMPLE
Design 1: Experimental and quasi–experimental designs	Using high frequency metering data for high–quality information about energy consumption and demand in rural solar micro–grids in India. [Source: Poverty Action Lab]
1A. Randomized control trial	Tablet–based financial education in Colombia. Using savings and transaction data combined with survey and telemetric tablet data. [Source: Poverty Action lab]
1B. Strong quasi–experimental design	Pre–test and post–test comparison group design using propensity score matching to strengthen the comparison group. [Source: GEF protected areas evaluation]
1C. Natural experiments	Assessing the effects of a government tax increase on smoking using changes in search query volume to assess the effects of a major increase in cigarette smoking in the USA. Canada, which did not have a similar increase, was used as the comparison group. (Source: Ayers 2011. Cited in Letouze et al 2016, 237–8)
Design 2: Statistical modelling	Evaluating causal interactions between labour market shocks and internal mobility Understanding labour market shocks using mobile phone data. [Source: World Bank, Latin American region]

Design 3: Theory–based evaluation	The Robert Wood Johnson Foundation conducted an evaluation of a 10–year programme to assess the impacts of its programme to improve health and safety outcomes in distressed USA cities. This combined a quasi–experimental design, including comparison cities, with a theory of change. Given the size, complexity and duration of the programme. Very large data sets had to be managed. [Source: cited in Leeuw, 2016]
Design 4: Case based evaluation	QCA country level data assessing factors determining impacts of women's economic empowerment programmes at the national level. [Source: UN Women, *An empowered future: corporate evaluation of UN women's contribution to women's economic empowerment: Independent Evaluation Office*, 2014]
Design 5: Participatory evaluation	The World Bank India social Observatory uses a participatory approach to involve women in the identification of the key questions that should be included in large scale community surveys to identify priority development issues. Community women are then involved in conducting the surveys and in the interpretation of findings. The surveys have currently been administered to over 800,000 households so data analytics are required for the analysis and synthesis of the findings. [Source: World Bank, India Social Observatory][51]
Design 6: Review and synthesis approaches	A review and synthesis study was conducted to assess the effects of micro–credit on women's empowerment. The study used data analytic search mechanisms with customized key–word sequence to cover academic databases and on–line portals. [Source: Vaessen, Rivas and Leeuw, 2016]
COMPLEXITY– RESPONSIVE EVALUATION DESIGNS	
Design 7: Holistic approaches (systems analysis)	Although not yet published there are several evaluations underway that are using big data systems analysis to model and evaluate large–scale programmes for refugees and for people displaced by emergencies such as forest fires, floods.
Design 8: Unpacking complex programmes	See Bamberger, Vaessen and Raimondo, Chapter 7 for a hypothetical example of how to 'unpack' a rural transportation programme

CHAPTER 6 MANAGING BIG DATA – INCLUSIVE EVALUATIONS

6.1 THE CRITICAL ROLE OF THE EVALUATION MANAGER

While it is recognized that the design of big data–inclusive M&E systems requires technical expertise in the fields of both big data analytics and development evaluation, less attention has been given to the critical role of the evaluation manager in ensuring the effective integration of big data and conventional evaluation approaches. Both data scientists and M&E specialists often have a narrow technical focus, which may not fully relate to the needs of the different stakeholders and partners. Therefore, the evaluation manager plays a critical role with respect to:

- Ensuring the design and implementation of the M&E systems respond to the information and operational needs of all stakeholders and partners.
- Ensuring the choice of data and data collection methods are selected to respond to the information needs of the programme and its stakeholders. The choices of data and data collection methods must be *demand driven* (responding to the information needs of the programme) and not *supply driven* (responding to the technical interests of the data scientists and evaluation researchers).
- Ensuring that all proposed data sources are rigorously assessed using a data quality and data appropriateness checklist (See Chapter 3 Table 3–2).
- Ensuring that that the proposed M&E designs and their requirements are consistent with the available financial, technical and data processing capacities of the evaluation of the M&E offices.
- Equally important, but often overlooked is the need to ensure that the many different agencies the cooperation of which will be required in the collection and analysis of data are able and willing to provide the required data in a timely manner and in the required format. Many M&E systems fail to perform at full capacity because the information they assumed is available is in fact not accessible. In some cases, this is due to a lack of willingness of agencies to cooperate, while in other cases the organization of the data may be beyond the resource capacity of some agencies.
- Many researchers have a preferred data collection and analysis methodology and they may be unwilling to consider other methodologies that may be more appropriate for a particular evaluation.

This becomes a challenge for large–scale evaluations that might require the organization of a multidisciplinary team where different members are using a range of different evaluation methodologies.
• Ensuring that M&E reports are delivered on time and in a format that is accessible to different stakeholders.
• The concept of complex programmes and the design of complex evaluations may be unfamiliar to many evaluators, and it will be the responsibility of the manager to ensure that the appropriate procedures are followed to determine whether a complex–esponsive approach may be required.

6.2 SPECIAL CHALLENGES IN MANAGING BIG DATA–INCLUSIVE MONITORING AND EVALUATION

There are a number of challenges in the management of M&E systems that draw on big data and ICT, and it is the responsibility of the manager to ensure that these are addressed:

• There are important gaps in knowledge and understanding between development evaluators and big data specialists (see Chapter 1 Section 1.6).
• There are also differences in the role of theory.
• Access to big data can be a challenge as much of the data is proprietary and may only be available to certain groups and only on a restricted basis. However, data philanthropy, can create the basis for obtaining data from private sector companies.[52]
• Data privacy and data protection. Big data frequently involves the analysis of large amounts of personal data, much of which may be very personal and in some cases put people at risk. Organizations working with new sources of data should have in place sound data privacy and data protection mechanisms that mitigate the risk of harms to individuals and groups of individuals.
• Capacity development and strengthening computing infrastructure are additional challenges requiring upgrading big data knowledge and skills of M&E specialists as well as management and operational staff. Often organizations will also have to make major investments in upgrading their computing capacity, or building relationships with agencies that already have this capacity.
• The incorporation of big data into programme evaluation requires the development of a big data responsive evaluation culture.

6.3 DEFINING THE LEVELS AT WHICH THE EVALUATION SHOULD BE CONDUCTED AND THE EVALUATION APPROACH

A first important role for evaluation is to help determine the level at which the evaluation will be conducted, and the evaluation approach. As discussed earlier, evaluations can focus at the policy level, the programme level and the project level – or in some cases at more than one level. Over the period of a four to five year national development programme, a number of evaluations may be conducted at different levels. In some cases it is obvious which level will be required, but in other cases the evaluation could be conducted at two or more levels, so the managers plays an important role in clarifying the evaluation purpose, the key questions and hence the appropriate level. For example, a national education reform programme will probably have a policy component, a number of national programmes and a wide range of project interventions. Each of these levels could be evaluated, so the manager must work with stakeholders, as well as the evaluation specialists, to define the level or levels at which the evaluation will be conducted.

The manager will also help determine which of the four evaluation approaches discussed in Chapter 2 (policy, formative, developmental and summative) will be used. The choice of approach will also determine the key evaluation questions of interest to different stakeholders. (see Table 3–2).

While decisions on these questions are required for all evaluations, the incorporation of big data can increase the range of options. The evaluation manager plays an important role in maintaining a focus on the key evaluation questions and avoiding that the evaluation design is driven by the new technologies and sources of data.

6.4 THE ROLE OF THE EVALUATION MANAGER AT EACH STAGE OF THE PROJECT CYCLE

This section focuses on the role of the manager in project–level evaluations. Most of these issues are also relevant for programme and policy–level evaluation although some of the management tasks can become more complicated as there are often more stakeholders involved at the programme and often at the policy level.

Table 6–1 identifies the key functions of the evaluation manager at each stage of the design, implementation and use of the monitoring and evaluation systems. In general, the manager serves as a link between the intended users of the M&E information and the technical staff who will design and implement the monitoring and evaluation. Many evaluators have their preferred methodologies (e.g. randomized

TABLE 6–1 THE ROLE OF THE EVALUATION MANAGER IN GUIDING THE PLANNING AND FOCUS OF BIG DATA–INCLUSIVE M&E SYSTEMS AT EACH STAGE OF THE PROJECT CYCLE

STAGES OF THE PROJECT CYCLE	MANAGING MONITORING	MANAGING THE EVALUATION
1. Project identification and appraisal — Identifying key questions of concern to clients and stakeholders — Clarifying budget and resource availability and assessing how this affects possible M&E designs — Assessing the level of complexity in the programme	— Translating stakeholder questions into monitoring questions — Guiding the focus of exploratory studies	— Translating stakeholder questions into evaluation questions. — Guiding the focus of initial diagnostic studies — Ensuring appraisal studies focus on all stakeholder priority questions — Deciding whether a complexity–responsive design is required
2. Project planning, design and stakeholder consultations — Basing both monitoring and evaluation on a theoretical framework — Identifying all possible sources of data for M&E — Assessing the organizational, political, technical and resource implications and feasibility of proposed data collection systems — Designing the project information systems for project design and management	**Monitoring system design** — Identifying the need for, and the feasibility of developing integrated data platforms (combining big data and conventional data sources) — Design a results–based management and monitoring system and ensuring it addresses all of the key evaluation questions — Developing the monitoring information system and linking this to the management information system	**Evaluation design** — Identifying all possible evaluation designs (big data inclusive and conventional) — Comparing conventional different approaches to attribution (experimental, predictive and systems analysis) — [if required] designing a complexity–responsive evaluation • Begin with systems mapping exercise • Defining boundaries • Developing a 5 stage unpacking–reassembling strategy — Evaluability analysis of the proposed evaluation design — Using satellite imaging and GPS data to strengthen matching of comparison group in quasi–experimental designs
3. Project implementation	— Designing a dynamic monitoring system using ICTs and DB to provide real–time feedback so that corrective action can be taken on identified problems.	**Data collection** — Using ICT and big data to incorporate process analysis and the evaluation of behavioural change

3. Project implementation	— Using ICTs to promote periodic consultations with the stakeholders, including the most vulnerable groups — Building–in periodic checks on the validity of indicators and the performance of the M&E systems	— On–line updating of the Theory of Change **Sample selection** — Using Satellite images, remote sensing, GPS mapping, phone company records and other sources to improve quality and reduce costs of sample selection — Using smart phones to build–in quality control checks on sample selection and data collection
4. Mid–term review and decisions on modifications to the programme	— Synthesis of monitoring data for the MTR report — Data visualization systems	— Using ICTs to conduct rapid surveys of beneficiaries and stakeholders — Using big data to collect contextual data
5. Project completion and decisions on future projects	— Synthesis of monitoring data for the PCR	— Using systems mapping and systems analysis to assess potential replicability in different locations
6. Synthesis of lessons learned and dissemination strategy		— Using new information technology to ensure effective dissemination, including modern data visualization techniques
7. Continuation and sustainability		— Identification of resources and organizational structures to permit the continued collection and analysis of sustainability data

control trials, focus groups, phone–based surveys with multiple choice questions) and it is the responsibility of the manager to ensure that all possible evaluation designs are considered before making the final selection. The manager must also ensure that the M&E teams understand the information needs of the intended evaluation users. Why do they need the information and how will it be used? It is also important to understand the level of rigor and detail that is required. Evaluators often generate more information than is required and sometimes with a higher level of methodological rigor than is needed. It is also the responsibility of the evaluation manager to ensure that the M&E findings are disseminated in a user–friendly manner and that they are used.

DECIDING THE LEVEL AT WHICH THE EVALUATION SHOULD BE CONDUCTED

The first step is to decide, in consultation with stakeholders and the evaluation team, the level at which the evaluation should be conducted:

- Policy level
- Programme level
- Project level

This table describes the different stages in the design and implementation of a project level evaluation.

GLOSSARY OF TECHNICAL TERMS

BIG DATA

An umbrella term referring to the large amounts of digital data continually generated by the global population. Can be privately owned or have varying levels of access control. Big Data is characterized by the '3 Vs:' greater volume, more variety, and a higher rate of velocity. A fourth V, for value, can account for the potential of big data to be utilized for development.

BIG DATA ANALYTICS

A type of quantitative research that examines large amounts of data to uncover hidden patterns, unknown correlations and other useful information. Four main approaches can be applied in development evaluation: diagnostics, prediction, detection and evaluation/prescription.

COUNTERFACTUAL ANALYSIS

A comparison between what actually happened and what would have happened in the absence of the intervention.

DATA CONTINUUM

There is a continuum of types of data ranging from big data to small data (qualitative interviews, case studies etc). Most evaluations that use big data tend to combine it with other types of data.

DATA EXHAUST

Passively collected data deriving from daily usage of digital devices as financial services (including purchases, money transfers, savings and loan repayments), communications services (such as anonymized records of mobile phone usage patterns) or information services (such as anonymized records of search queries).

DATA PHILANTHROPY

Partnership by which private sector companies share data for public benefit, taking the initiative to anonymize their data sets and provide them to social innovators to mine for real– time insights, patterns and trends.

DATA VISUALIZATION

Data analytical techniques that permit the representation of complex data in user– friendly and easily understood maps, tables and charts. The data is often interactive so that the user can touch the laptop screen to dig deeper

EVALUATION

A systematic and objective assessment of an ongoing or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, efficiency, effectiveness, impact and sustainability. Two main approaches to evaluation are frequently discussed:

a. Formative evaluation: providing regular feedback to management and other stakeholders to help strengthen the design and implementation of programs and projects

b. Summative evaluation: usually conducted at the end of a project or program to assess the

extent to which intended outcomes have been achieved. Summative evaluation frequently uses experimental and quasi–experimental designs to assess the extent to which observed changes in outcomes can be attributed to the effects of the program interventions.

Two other approaches discussed in this publication are:

c. Complexity–responsive evaluation: used when programs are considered to be complex

d. Developmental evaluation: similar to summative evaluation but focusing on programs that are innovative, complex or emergent (continually evolving).

FALSIFICATION

Data can be false, fabricated with the intention of providing misleading information. Another meaning, in research design, is that a research hypothesis is formulated in such a way that it can be tested and proved false (falsified).

ICT (information and communication technology): Smart phones, tablets, remote sensors and other (largely and–held) devices that can be used for monitoring and evaluation and the field–level.

INDICATORS

Signposts of change along the path to development. They describe the way to track intended results and are critical for monitoring and evaluation.

INTERNET OF THINGS (IOT)

A system of interrelated computing devices, mechanical and digital machines, objects, that are provided with unique identifiers and the ability to transfer data over a network without requiring human–to–human or human–to–computer interaction (Wikipedia).

MEL (MONITORING, EVALUATION AND LEARNING)

Many development agencies, recognizing the importance of learning and dissemination, focus on these three dimensions rather than the traditional M&E.

MONITORING

An ongoing process by which stakeholders obtain regular feedback on the progress being made towards achieving their goals and objectives.

RANDOMIZED EXPERIMENTS

Also called experimental design, are the most rigorous evaluation design, often referred to as the 'gold standard.' Randomization ensures that the intervention and comparison groups are equivalent with respect to all factors other than whether they received the intervention. In other words, the comparison group serves as the 'counterfactual' of what would have happened in the absence of the program—a key requirement in determining whether a program caused a particular outcome.

QUALITATIVE COMPARATIVE ANALYSIS (QCA)

Using case studies to identify configurations of factors that are associated with the presence of absence of a particular outcome. In contrast to conventional ('frequentist') analysis that uses techniques such as regression to identify the significance of individual variables, QCA believes that most contexts

are complex and that program outcomes can only be understood as resulting from interactions among a set (configuration) of factors.

QUASI–EXPERIMENTAL DESIGN (QED)

QED uses an intervention and comparison group, but assignment to the groups is non random. Evaluators must assess the differences at baseline and account for any demographic or behavioural differences in the analysis. Comparison groups in the quasi–experimental design can be identified through matching a process of identifying individuals that are similar to the participants in the intervention group on all relevant characteristics, such as age, sex, religion and other factors associated with program exposure.

NON–EXPERIMENTAL DESIGN

An intervention group only and lacks a comparison/control group, making it the weakest study design. Without a comparison group, it is difficult for evaluators to determine what would have happened in the absence of the intervention. Evaluators choose to use non–experimental designs when there are resource constraints, when they are unable to form an appropriate comparison group, or when a program covers the entire population and thus there is no comparison group, such as with a mass media campaign. In non–experimental study designs, evaluators must have a clear conceptual understanding of how the intervention was intended to influence the outcomes of interest.

OPEN DATA

A term that refers to data that is free from copyright and can be shared in the public domain.

PREDICTIVE ANALYTICS

Use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future. The analysis is usually based on Bayesian statistical models.

RANDOMIZED CONTROL TRIAL (RCT) OR RANDOMIZED IMPACT EVALUATIONS

A type of impact evaluation which uses randomized access to social programmes as a means of limiting bias and generating an internally valid impact estimate.

SAMPLING BIAS

A consistent error that arises due to the sample selection. Sampling bias means that the data collected may not be accurate or represent the group. Sampling bias can occur any time sample is not a random sample. If it is not random, some individuals are more likely than others to be chosen.

STAKEHOLDERS

People who will benefit from the development activity or whose interests may be affected by that activity.

SENTIMENT ANALYSIS (OR OPINION MINING)

Refers to the study of emotions and opinions expressed in digital messages and translating those sentiments to hard data.

SUSTAINABLE DEVELOPMENT GOALS (SDGS)

An interlinked set of 17 development goals that the international development community has set as targets to be achieved for all countries by 2030. Also known as the 2030 Agenda.

TEXT MINING

The analysis of data contained in natural language text.

REFERENCES

Anderson, C (2008) The end of theory: The data revolution makes the scientific method obsolete. Wired Magazine 6/23/08

Bamberger, Vaessen and Raimondo (2015) Dealing with complexity in development evaluation: A practical approach. Thousand Oaks, CA: Sage

Bamberger, M, Raftree, L and Olazabal,V (2016) The role of new information and communication technologies in equity-focused evaluation: opportunities and challenges. Evaluation Vol. 22(2) 228-244

Bamberger, M (2013) Engendering M&E. PREM Notes. Special series The nuts and bolts of M&E systems. Washington DC. World Bank.

Data Revolution Group (2014) A world that counts: Mobilizing the data revolution for sustainable development.

Gee, M (2015) Introduction to data science: models, tools and prototypes. Workshop during the Data for Good Conference. Chicago, September 2015 Unpublished

Giannone, D, Reichlin, L and Small, D (2008) Nowcasting: the real–time informational content of macroeconomic data, Journal of Monetary Economics (Elsevier) 55:4 665-676

Hurwitz, Nugent, Halper and Kaufmann (2013) Big data for Dummies. New Jersey: Wiley.

Kearns, J (2015) Satellite images show economies growing and shrinking in real–time, Bloomberg News July 8 2015.

Kitchen, R (2014) Big data, new epistemologies and paradigm shifts. Big data and Society. April–June 2014: 1–12

Letouzé E, Areais and Jackson S (2016) The evaluation of complex development interventions in the age of Big data. In Bamberger, Vaessen and Raimondo. Op cit

Marr, B (2015) Big data: Using smart Big data analytics and metrics to make better decisions and improve performance. Chichester, West Sussex, UK: Wiley

Meier, P (2015) Digital Humanitarians: How big data is changing the face of humanitarian response. CRC Press

Morra-Imas, L & Rist, R (2009) The road to results: Designing and conducting effective development evaluations. World Bank.

Peng,R & Matsui, E (2015–16) The art of data science: A guide for anyone who works with data. Skybridge Consulting, LLC.

Patton, M (2011) Developmental evaluation. Applying complexity concepts to enhance innovation and use. New York. Guildford

Siegel, E (2013) Predictive analytics: The power to predict who will click, buy, lie or die. New Jersey: Wiley

Stern, E et al (2014) Broadening the range of designs and methods for impact evaluations. Department for International Development. London.

Toyama, K (2015) The Geek Heresy: Rescuing social change from the cult of technology. New York: BBS Publications.

United Nations Population Fund (2016) Big data and the sustainable development goals. Background paper prepared jointly by UNDP, UNFPA (coordinator), UNOPS, UNICEV, UN–Women and WFP (co-coordinator) for the Joint meeting of the executive boards of UNDP/UNFPA/UNOPS, UNICEF, UN–Women and WFP. June 3, 2016 (unpublished)

United Nations (2015) Transforming our World: The 2030 Agenda for Sustainable Development. https://sustainabledevelopment.un.org/post2015/transformingourworld

UNDP: Big Data for development in China (November 2014) http://www.cn.undp.org/content/china/en/home/library/south-south-cooperation/big-data-for-development-in-china.html

Van den Braak, S; Choenni, S & Bamberger, M (2016) Emergent technologies and creative use of multiple sources of information in Bamberger, Vaessen and Raimondo (eds) op. cit.

Woolcock, M (2009) Towards a plurality of methods in project evaluation: a contextualized approach to understanding impact trajectories and efficacy. Journal of development Effectiveness 1(1): 1–14.

World Bank (undated) Big data in action for development. Latin America and the Caribbean and Second Muse.

World Bank (2016) Digital Dividends. World Development Report.

ENDNOTES

[1] World Development Report 2016 *Digital Dividends* p.2
[2] See for example, the ICT-Works blogs which document the extensive applications of these technologies in the development field (ict-works@inveneo.org)
[3] See Bamberger, Raftree and Olazabal (2016) *The role of new information and communication technologies in equity-focused evaluation: opportunities and challenges.* Evaluation (in press) for a discussion of these challenges including why these technologies can lead to governments and donors adopting an *extractive* strategy whereby information on and about poor and vulnerable groups can be collected without their knowledge or involvement.
[4] *Digital Dividends* identifies 3 sets of factors that have limited the achievement of the potential social and broad economic benefits that digital technologies offer: who controls digital technology, inequality of access and of resources to benefit from these technologies and concentration and lack of competition. This report discusses a broader range of constraints.
[5] United Nations (2015). *Transforming our World: The 2030 Agenda for Sustainable Development.* https://sustainabledevelopment.un.org/post2015/transformingourworld
[6] Patrick Meier, *Digital Humanitarians* provides an extensive discussion of the challenges in interpreting digital early-warning signals during earthquakes, floods and other disasters.
[7] http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en
[8] https://www.theguardian.com/global-development-professionals-network/2014/nov/13/can-twitter-be-a-force-against-hiv-discrimination
[9] UNDP: *Big Data for development in China*. November 2014. http://www.cn.undp.org/content/china/en/home/library/south-south-cooperation/big-data-for-development-in-china.html
[10] http://www.unglobalpulse.org/projects/haze-gazer-a-crisis-analysis-tool
[11] http://www.unglobalpulse.org/projects/analysing-seasonal-mobility-patterns-using-mobile-phone-data
[12] Catherine Cheney (2016), *Data driven development needs both social and computer scientists*, Devex July 29, 2016.
[13] Leeuw (2016) describes the following alternative approaches to policy evaluation: realist evaluation, theory of change approaches, contribution analysis, policy scientific approach, strategic assessment approach, and the elicitation approach.
[14] For example: Letouzé et al (2016) only use three categories: Descriptive analysis, predictive analysis and diagnostic analysis; and Peng and Matsui (2015-16) use exploratory, inference and prediction.
[15] Siegel (2013) *Predictive analytics: The power to predict who will click, buy, lie or die.* See pp. 5-9 for a review of the many ways in which predictive analytics are used.
[16] See Marr (2015) *Big data: using smart big data analytics and metrics to make decisions and improve performance.*
[17] Bamberger, Raftree and Olazabal (2016) review some of the leading evaluation journals over the past 3 years and found very few articles with any reference to big data.
[18] http://www.worldbank.org/en/programs/social-observatory
[19] http://www.proving.it/
[20] http://unglobalpulse.org/mapping-infectious-diseases
[21] http://unglobalpulse.org/mapping-infectious-diseases
[22] http://unglobalpulse.org/radio-mining-uganda
[23] http://www.bostonindicators.org/
[24] Patrick Meier (op.cit)

[25] For more details on how to assess the level of complexity of each of these dimensions for a particular program see Table 1.2 Checklist for assessing levels of complexity in Bamberger, Vaessen and Raimondo 2016.
[26] Many evaluators classify development programs into 3 categories: simple, complicated and complex. The first two category using conventional evaluation methods, whereas programs that are complex require the use of complexity responsive evaluation designs see Funnell and Rogers, 2011; Patton (2011) and Bamberger, Vaessen and Raimondo 2016),
[27] This section draws on Hurwitz et all 2013
[28] Steps 3 and 4 are adapted from Jackson (unpublished). Find the report here: http://unglobalpulse.org/sites/default/files/Annex%201%20Big_data_monitoring_and_evaluation.pdf
[29] One of the most widely used desktop packages in Tableau.
[30] See Stern et al (2012) *Broadening the range of designs and methods for impact evaluations*, for the discussion of evaluation questions and the different designs that must be used to address each question.
[31] See Bamberger et al (2012) op.cit Table 11.1 for a discussion of the characteristics of the program and the context within which it operates that together determine the most appropriate evaluation design.
[32] See Bamberger 2013, *Engendering M&E* for a discussion of key gender indicators that should be included in most surveys. http://www.worldbank.org/content/dam/Worldbank/document/Poverty%20documents/ME%20Engendering_finalOct13.pdf
[33] For a discussion of integrated mixed methods designs see Bamberger, Rugh and Mabry (2012) Real World Evaluation Chapter 14
[34] *Crowdsourcing high-frequency food price data in rural Indonesia* (Global Pulse)
[35] *Accessing spatial data to study biodiversity and devise protection strategies in Zimbabwe* (Global Pulse)
[36] See for example, the FAO AQUASTAT national databases.
[37] *Supporting forest and peat-fire management using social media* (Global Pulse)
[38] *Using mobile phone activity for disaster management during floods* (Global Pulse)
[39] *Using twitter to analyse public sentiment on fuel subsidy policy reform in el Salvador* (Global Pulse)
[40] *Estimating migration flows using online search data* (Global Pulse)
[41] *Digital signals and access to finance in Kenya* (Global Pulse) UNDP
[42] *Nowcasting food prices in Indonesia using social media signals* (Global Pulse)
[43] *Real-time monitoring of vulnerable populations coping with crises* (Global Pulse)
[44] Global snapshot of wellbeing via mobile-phone based survey
[45] This section draws on World Bank (undated) *Big Data in Action for Development*.
[46] Examples of proxy indicators for economic development include: vehicles outside shopping malls, trucks travelling to and from markets, the level and spread of illumination at night in different commercial and residential areas, the level of oil in storage tanks (from the angle of the shadow on the cover which changes as the oil gets lower, quality of materials used in house construction (see Kearns 2015)
[47] See Patrick Meier *Digital Humanitarians* for a detailed discussion of these challenges
[48] Now casting, which became popular recently in economics is now being applied more widely. It is based on the recognition that there are long delays in obtaining economic statistics so that algorithms are required to estimate the present (Giannone et al 2008).

[49] On-line search data, Google search is used for terms related to the topic (e.g. if the topic is migration in Australia the search covers terms such as 'Jobs in Melbourne') and results are correlated with, for example, official migration statistics.
[50] Van den Braak, Choenni and Bamberger (2016) describe the steps involved in bringing together different sources of data into data warehousing or data spaces. The process can be technically complicated and time consuming.
[51] http://www.worldbank.org/en/programs/social-observatory/brief/setting-up-the-social-observatory-in-bihar
[52] For a discussion of data philanthropy see Letouzé et al 2016.

© UN Global Pulse, 2016