



Challenges and issues related to M&E in the era of big data

*Iraj Namdarian - CREA (Council for Agricultural
Research and Economics)*

What is big data?

'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)

What is big data?

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.



What is big data?

Big data is sometimes referred to as ‘a collection of large volumes of data,’ that involves three interlinked components

- **Data generation:** Collection of large volumes of generated data.
- **Data platform:** Platforms that link the multiple organizations and individuals that generate, analyse and use big data.
- **Data analytics:** Organization and integration of multiple sources of 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).



What is big data?

- The three dimension of the Big Data Framework



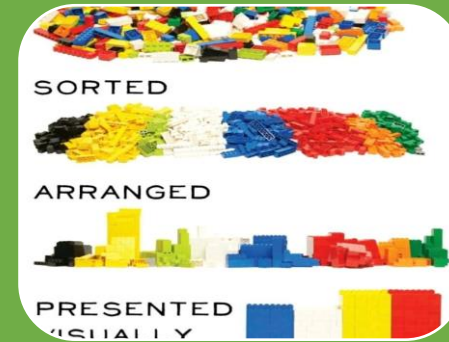
DATA GENERATION

Generation of new
sources of data



DATA PLATFORM

Involving producers,
analysts, users
and regulators
of big data.



DATA ANALYTICS

Organization,
integration, analysis
and dissemination
of big data



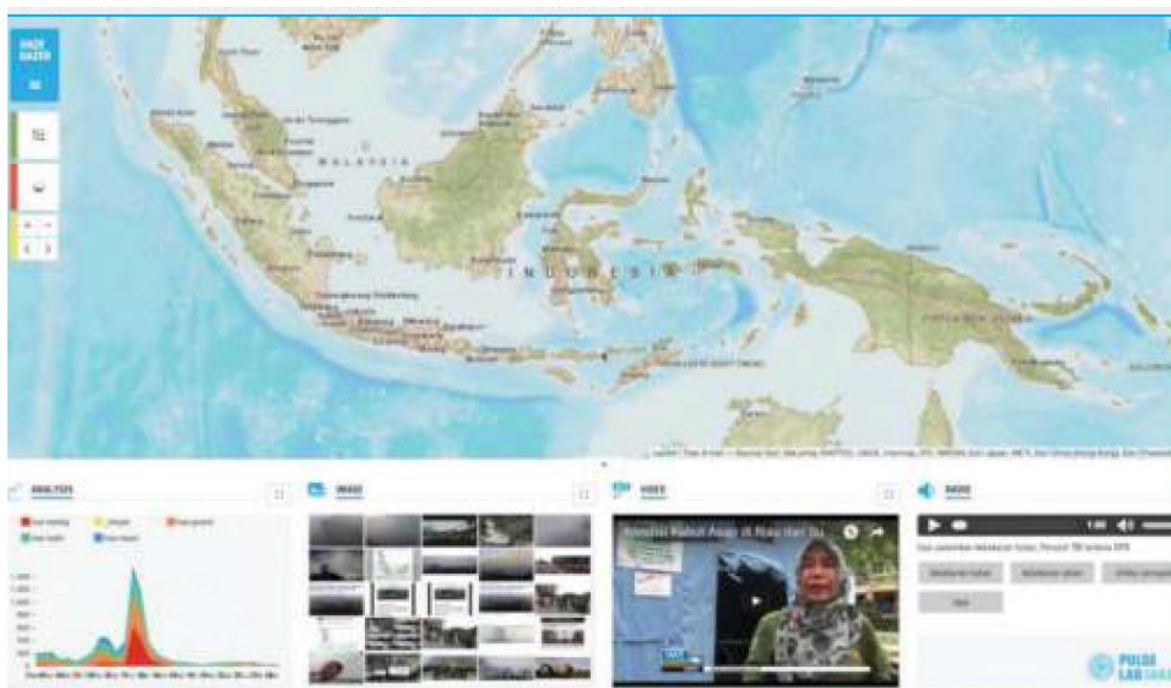
Some examples of the use of Big Data



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

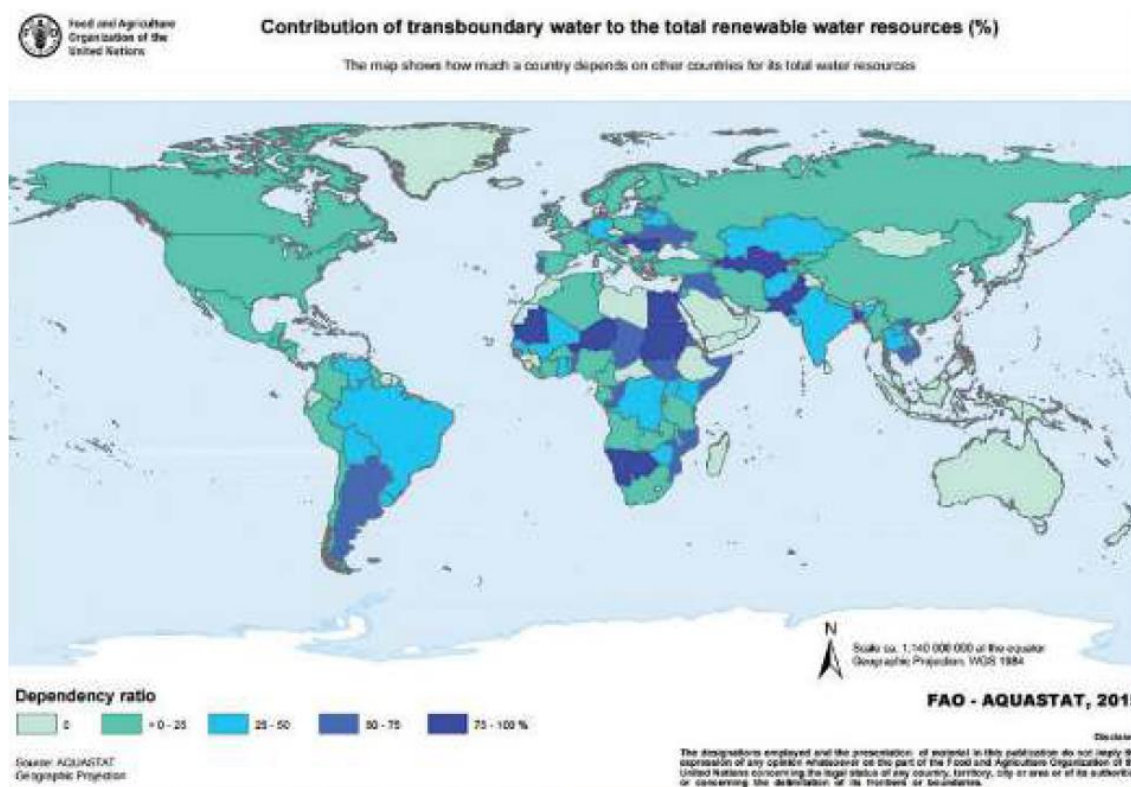
Mapping Population displacement using mobile phone data:
Reports on population displacement in Nepal following the April 2015 earthquake were produced by Flowminder/WorldPop based on mobile operator data.

Some examples of the use of Big Data



Using social media to detect and manage forest 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)

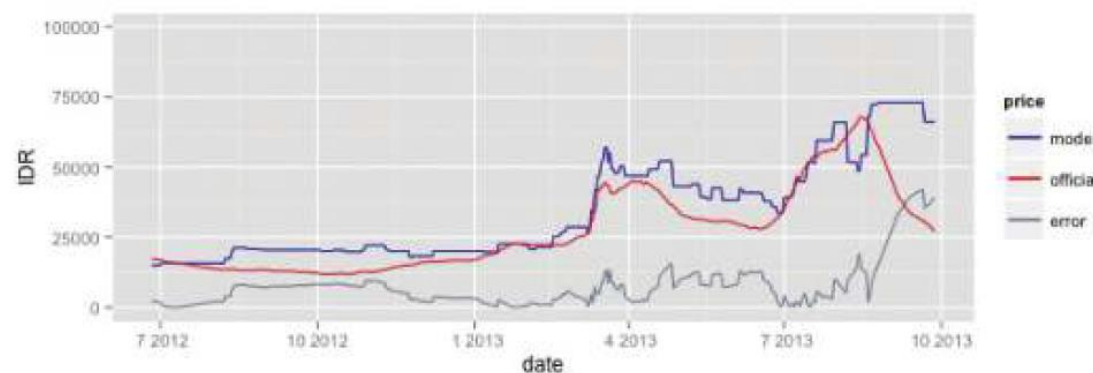
Some examples of the use of Big Data



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]

Some examples of the use of Big Data

1,954 tweets contained price quotes on onions.

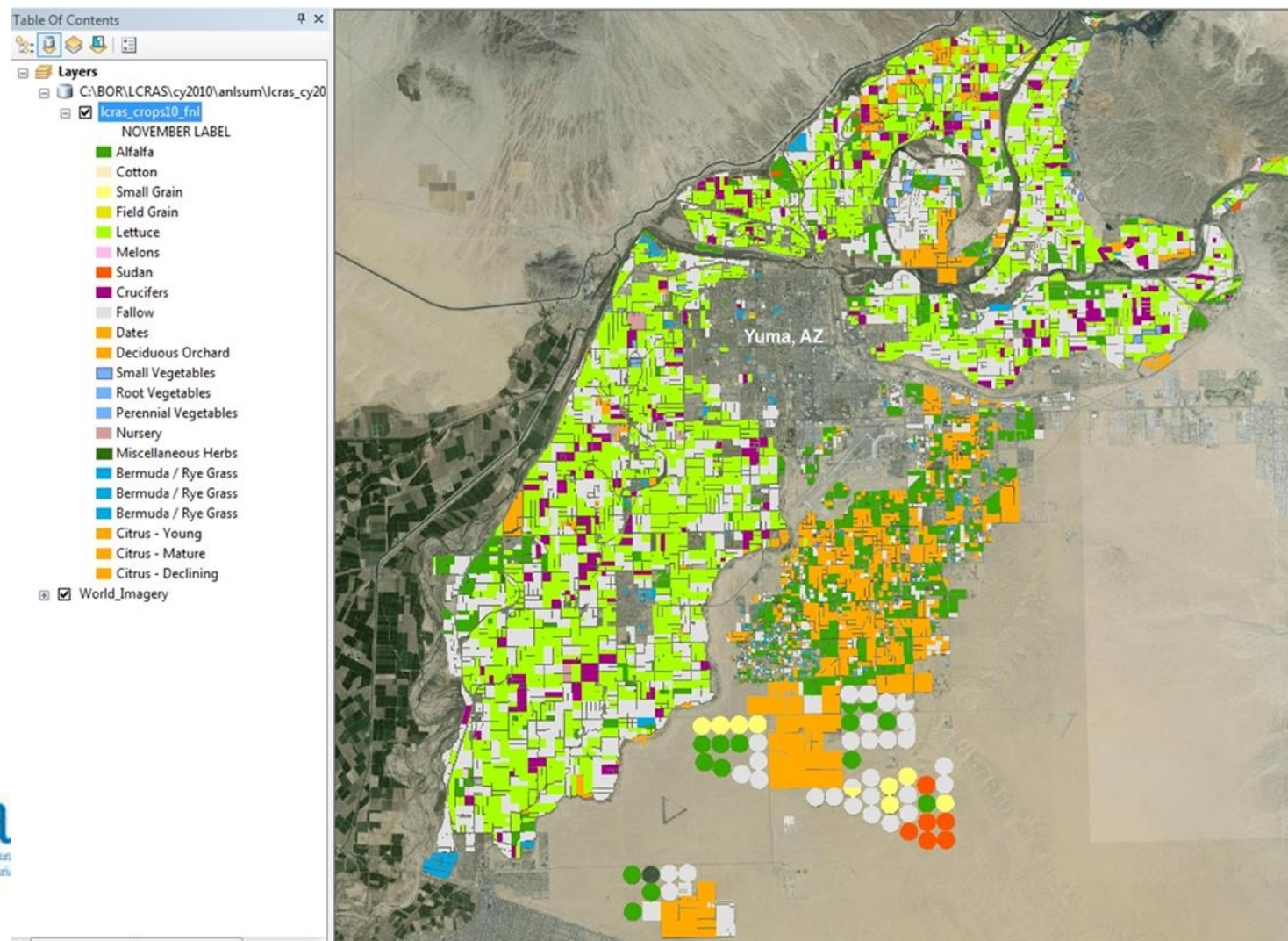


Nowcasting food prices (onion) in Indonesia using social media signals (WFP)

Big data: GIS/Remote Sensing applications



Big data: The classification of winter crops



Big data: Monitoring of irrigation needs

gLOCAL
EVALUATION WEEK
2021



GEI
Global
Evaluation
Initiative

 **crea**
Consiglio per la ricerca in agricoltura
e l'analisi dell'economia agraria

mipaaf
ministero delle politiche
agricole alimentari e forestali



-Challenges to the current evaluation methodologies

-Areas where Big Data analytics could contribute:

- Understanding the programme context
 - i. Collecting information on a large number of contextual factors, which may include: economic, social, political, cultural, demographic and ecological, etc.
- Data collection – including on sensitive topics
 - i. Data is expensive to collect so that sample sizes are often smaller than desired
 - ii. Data can be time-consuming to collect
 - iii. Many groups are difficult to reach
 - iv. Delays in providing feedback to decision-makers
 - v. The time-period over which data is collected may be limited



-Challenges to the current evaluation methodologies

-Areas where Big Data analytics could contribute:

- Monitoring processes and behavioural change
 - i. Difficult to collect information on sensitive topics
 - ii. Difficult to observe and measure behavioural change because a longer period is often required to observe the changes
 - iii. Different and expensive to collect continuous data required to monitor processes of project implementation and change



-Challenges to the current evaluation methodologies

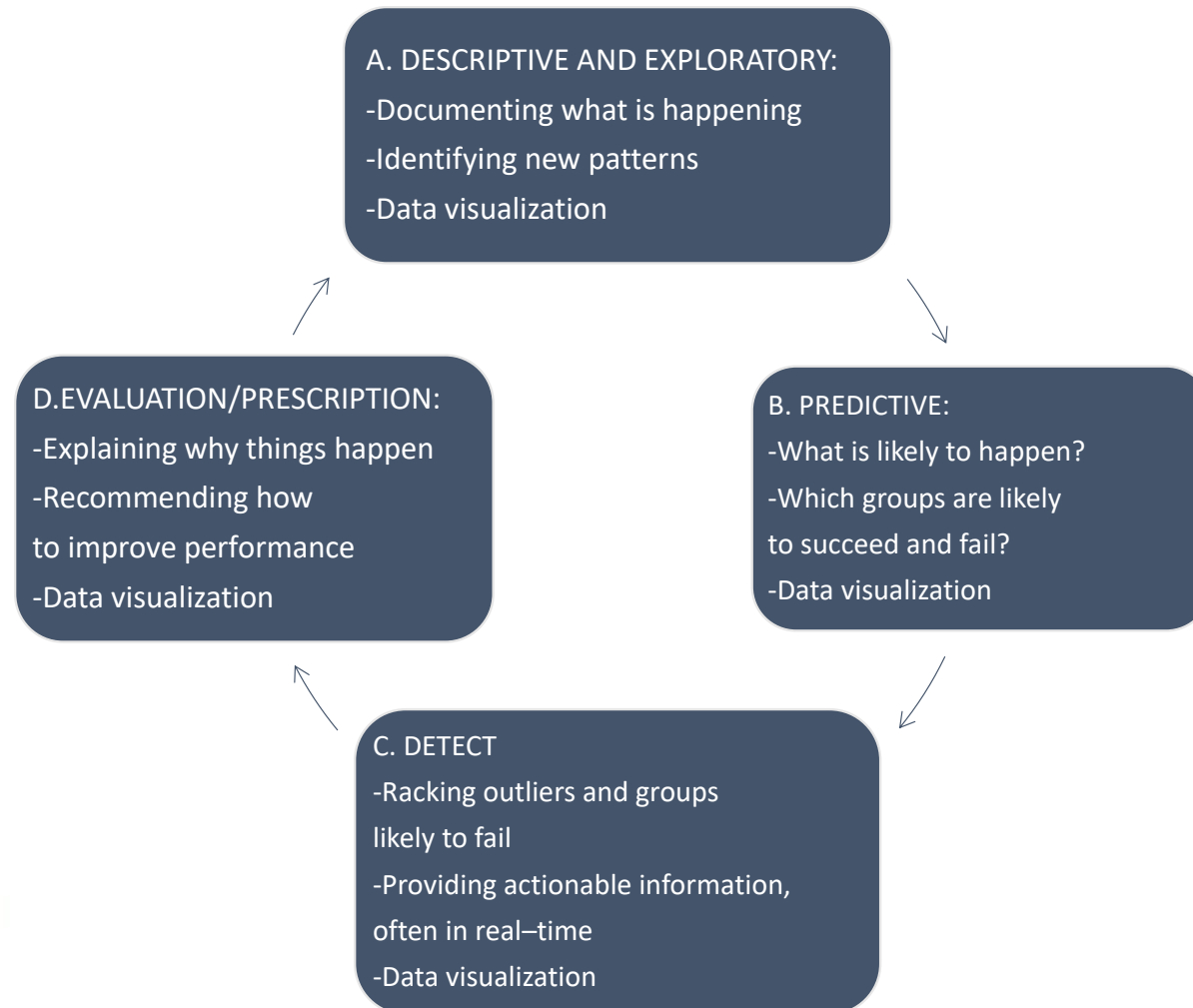
-Areas where Big Data analytics could contribute:

gLOCAL
EVALUATION WEEK
2021

- Capturing different voices and using evaluation to empower vulnerable groups
 - i. 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
- Evaluating complex programmes
 - i. Information must be collected on many more variables
 - ii. Information must be collected and processed more quickly
 - iii. Interactions among multiple variables must be analysed
 - iv. Processes must be monitored with continuous observation

The characteristics of Big Data and potential applications for programme monitoring and evaluation:

- The four steps data analytics approach that can be applied to the monitoring and evaluation of development programmes



New methodological challenges that Big data analytics introduces:

- i. 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.
- ii. 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.
- iii. 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 misled by the 'fallacy of large numbers' into assuming the sample must be broadly representative as it is so large.



New methodological challenges that Big data analytics introduces:

iv. 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.

v. Attribution and spurious correlation: 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:

- i. 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.
- ii. **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.
- iii. 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.



Logistical and organizational issues:

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

v. **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.



Political and ethical considerations:

i. **Access to big data is often limited** and data may only be available to governments, international 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.

i. Access to big data can also be **expensive**, again excluding many groups who wish to use it.



Political and ethical considerations:

- iii. A related concern is that **commercial survey research agencies** may collect information which they then sell to private companies and without any benefit or compensation to the communities to which the information refers.
- iv. 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 third parties or online hackers. As the amounts of data collected increases, and when uploaded to remote online central locations, it becomes technically more difficult to ensure data privacy and data protection.



Checklist for assessing the quality and appropriateness of Big Data and ICT-generated indicators

gLOCAL
EVALUATION WEEK
2021

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) with no 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

What is big data?



Big Data

**Eric Schmidt,
CEO of Google, 2011**

"There was 5 exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days, and the pace is increasing."

But not only the size matters !!!



References

- Integrating Big Data into the monitoring and evaluation of development programmes. UN Global Pulse
- United Nations Population Fund (2016) Big data and the sustainable development goals. Background paper prepared jointly by UNDP, UNFPA (coordinator), UNOPS, UNICEF, 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.