Big data readiness assessment
National guide

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity **Dignified Life**

VISION

ESCWA, an innovative catalyst for a stable, just and flourishing Arab region

MISSION

Committed to the 2030 Agenda, ESCWA's passionate team produces innovative knowledge, fosters regional consensus and delivers transformational policy advice. Together, we work for a sustainable future for all.

Distr.
LIMITED
E/ESCWA/CL4.SIT/2020/TP.11
3 November 2020
ORIGINAL: ENGLISH

Economic and Social Commission for Western Asia (ESCWA)

Big data readiness assessment
National guide

United Nations
Beirut, 2020

20-00412

© 2020 United Nations
All rights reserved worldwide

Photocopies and reproductions of excerpts are allowed with proper credits.

All queries on rights and licenses, including subsidiary rights, should be addressed to the United Nations Economic and Social Commission for Western Asia (ESCWA), e-mail: publications-escwa@un.org.

The findings, interpretations and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of the United Nations or its officials or Member States.

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Links contained in this publication are provided for the convenience of the reader and are correct at the time of issue. The United Nations takes no responsibility for the continued accuracy of that information or for the content of any external website.

References have, wherever possible, been verified.

Mention of commercial names and products does not imply the endorsement of the United Nations.

References to dollars ($) are to United States dollars, unless otherwise stated.

Symbols of United Nations documents are composed of capital letters combined with figures. Mention of such a symbol indicates a reference to a United Nations document.

United Nations publication issued by ESCWA, United Nations House,
Riad El Solh Square, P.O. Box: 11-8575, Beirut, Lebanon.

Website: www.unescwa.org.

Key messages

- ✓ Public sector organizations should consider starting small, with single-function big data applications aligned with their current systems. Once realized and digested, scale-up is possible;
- ✓ Readiness and effectiveness capabilities are enhanced by focusing on institutional processes with dedicated data science expertise and information technology (IT) governance;
- ✓ Inadequate applications of big data before mature readiness can backfire and cause serious long-term operational and institutional damage.

The present guide aims to assist organizations, especially public institutions, in assessing their national readiness to appropriately harness the advantages of big data, and reduce the risks of data mismanagement so as to serve society more effectively and protect human rights.

Contents

I. Background on big data requirements

The abundance of data from numerous sources, the ubiquity of mobile phones, decreased storage costs, increased processing power, and advances in cloud computing have democratized big data analytics.

The following factors are necessary to ensure the feasibility of any big data solution:

1. Data quality.
2. Infrastructure and human resources.
3. Legal regulations.

A. Data quality

The Big Data Quality Framework of the United Nations Economic Commission for Europe (UNECE) established a hierarchical structure composed of three hyper dimensions: source of the data, the metadata and the data itself. The quality aspects are nested within each dimension. Timeliness and frequency are the two important quality aspects of big data.[1]

The American Association for Public Opinion Research (AAPOR) provided guidelines on estimation in the generation and preparation of data processes.[2] Data error was modelled by: Extract/Transform/Load **(**ETL) errors introduced by several sources and operations: specification (metadata), matching, coding, editing, and data integration.

The International Monetary Fund (IMF) suggests that big data indicators must pass through a quality assessment to minimize risks for policymaking across the following seven features: relevance, accuracy, credibility, timeliness, accessibility, interpretability, and coherence.[3]

B. Infrastructure and human resources

Big data applications require efficient infrastructure, talented human resources, and effective management. Agencies aiming to harness benefits in productivity, speed and cost must incorporate a robust infrastructure for storage, processing and networking, in addition to analytics software.[4] According to IBM, data significance is based on the following three aspects: **speed, access and availability**.[5]

---

[1] A Suggested Framework for the Quality of Big Data Deliverables of the UNECE Big Data Quality Task Team December, 2014.

[2] AAPOR, AAPOR Report on Big Data, 2015. Available at www.aapor.org/Education-Resources/Reports/Big-Data.aspx#7.1%20Relative%20Advantages%20of%20Survey%20Data%20and%20Big%20Data%20to%20Support%20Research.

[3] International Monetary Fund, IMF's Data Quality Assessment Framework, 2010. Available at https://unstats.un.org/unsd/accsub/2010docs-CDQIO/Ses1-DQAF-IMF.pdf.

[4] Fedtech, How agencies can prepare their infrastructure for big data initiatives, 2016. Available at https://fedtechmagazine.com/article/2016/12/4-infrastructure-requirements-any-big-data-initiative.

[5] IBM, 3 key IT infrastructure requirements for big data and analytics, 2014. Available at www.ibm.com/downloads/cas/AYWDRYLW.

1. **Speed:** in addition to analysing voluminous and varied data, the infrastructure must provide the necessary speed to support timely decision-support and data usefulness as strategic asset for the organization.

2. **Access:** is important since providing unsecured data would put the organization at risk. Moreover, access should be smooth with all types of data, regardless where it is stored.

3. **Availability:** system fails, or crashes hinder availability, resulting in loss of valuable insights. Insights are only useful if they are delivered at the right time to the right people.

In addition to infrastructure requirements, human expertise is critical. Special and well-defined scientific skills are needed to complement classical statistitians and data miners.

Most big data applications require the following four distinct support roles:

(a) Domain expert pertaining to the application area (poverty, agriculture, climate, etc);
(b) Data-science researcher;
(c) Computer scientist;
(d) IT system administrator.

C. Legal regulations

Data accessibility, availability, ownership, proper utilization and management are challenging subjects for legal framework and related regulations. The opportunities for big data applications are based on data being a valuable asset. For many years, the World Economic Forum has been actively encouraging markets to upgrade regulatory frameworks to properly harness the development benefits of frontier technologies driven by data.[6] Policymakers are advised to develop these frameworks using an inclusive consultative approach. The approach should address the following most commonly faced policy challenges:

1. **Data availability:** is the starting point and requires an appropriate legal framework and policies that allow public agencies to disseminate their data. This is related to the issue of open government and data.

2. **Data ownership:** can cause conflict among data producers, analysts and users. This is undoubtedly related to the difficulties in assigning clear ownership across the wide range of application areas.

3. **Data stewardship:** is an extremely sensitive topic with the adoption of Internet of Things and high levels of connectivity in society, Government, and industry. The situation creates obstacles and possible abuse if data is not used for the declared intended purposes.

[6] World Economic Forum, Delivering digital infrastructure: how regulatory policy can keep up, pp. 25-26, 2014. Available at www3.weforum.org/docs/WEF_TC_DeliveringDigitalInfrastructure_InternetEconomy_Report_2014.pdf.

4. **Data collection authority:** is not always obvious and cannot be invariant., This becomes more ambiguous for non-personal data, crowdsourcing and event-based sensors' data, location and smart grid data.[7] The guidelines and principles should be clear to avoid legal controversies.

5. **Privacy and reidentification:** has been trade-off challenge for statisticians for decades: striking an ethical balance between protective aggregation and the embedded value of detailed data.

II. Adopted models for assessing big data readiness

Big data is considered a standing technology like any other. This is an important realization since big data is the methodical integration of several technologies that comprise measurement, collection, transmission, storage, computing and displaying. The technology's maturity is based on well-known standards. For big data, there are two types of technology readiness assessment that complement each other: organizational readiness, and technological readiness.

Several standardization and technology evaluation frameworks have mapped their methodology on big data to assess and classify readiness and maturity. The present guide reviews four popular approaches, and proposes a "best practice" scheme.

A. Assessing big data maturity model

The plan was proposed by the private consulting group Strategy&.[8] The assessment assumes the presence of an initial big data transformation, and that related business intelligence tools and data warehousing exist already. The methodology for assessing maturity requires the review of the following main aspects:

1. **Environment readiness:** This is based on government legal and ICT infrastructure that address privacy and data transfer, a certain level of public awareness of big data's benefits, and the education system incorporating data science related courses or degrees.

2. **Organization's internal capabilities:** This pertains to the capacity of subject organizations having the necessary technical capabilities, human resources, and orderly data availability.

3. **Sophisticated use of big data:** The aim towards having valuable assets in data is a journey starting from silo business intelligence tools to a completely integrated business model based on big data analytics.

The four maturity stages of big data depict the various ways in which data can be used, from selective adoption to large-scale institutional implementation, as follows:

(a) **Stage one: Performance management.** Internal data is used against key performance indicators to monitor stated goals;

[7] Katherine Strandburg, Monitoring, datafication, and consent: legal approaches to privacy in the big data context, pp. 5-43. In Privacy, Big Data, and the Public Good: Frameworks for Engagement, edited by J. Lane, V. Stodden, S. Bender, and H. Nissenbaum. Cambridge: Cambridge University Press, 2014.

[8] Booz & Company&, Big data maturity: An action plan for policymakers and executives, 2014. Available at www3.weforum.org/docs/GITR/2014/GITR_Chapter1.3_2014.pdf.

(b) **Stage two: Functional area excellence.** Experiment with internal and external data to improve parts of the business;

(c) **Stage three: Value proposition enhancement.** Start to monetize big data as a new source of competitive advantage;

(d) **Stage four: Business model transformation.** Big data infiltrate the whole organization's processes and decision-making.

B. Big data gap model

This readiness assessment method is simple and universal. The Design vs Reality Gap Model measures the gap between the requirements for the design of a specific change versus actual realities in the seven dimensions measured on a scale of 10, where 0 means no difference and 10 means complete change.[9] For big data transformation or sought-after change, the dimensions are as follows:

1. **Information:** this includes both information and supportive data.
2. **Technology:** pertains to ICT technologies needed to handle data at all stages.
3. **Processes:** represent the activities undertaken in generating, capturing, analysing, presenting, and using data.
4. **Objectives and values:** cover the culture associated with the adopted strategies.
5. **Skills and knowledge:** represent the human competencies and talents needed for undertaking data-related processes.
6. **Management systems:** describe the structures within data-related organizations.
7. **Other resources:** adequate time and money.

C. Technology readiness level model

The National Institute of Standards and Technology (NIST) in the United States has a universal Technology Readiness Level (TRL) chart for any technology.[10] Applying the TRL metric to big data is useful in assessing the overall maturity of a technology in a market or organization. The various scales are generic and represent the readiness state of a certain technology in markets.

Specifically, the NIST-TRL scales represent the following levels:

➢ **Emerging**
- Technology is largely still in research and development;
- Access is limited to the developers of the technology;
- Research is largely being conducted within academic or commercial laboratories;
- Scalability of the technology is not assessed.

[9] Richard Heeks, Information systems and developing countries: failure, success and local improvisations, The Information Society, vol. 18, No. 2, pp. 101-112.

[10] National Institute of Standards and Technology, NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements, 2015. Available at https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf.

- **Incubating**
  - Technology is functional outside laboratory environments;
  - Builds may be unstable;
  - Release cycles are rapid;
  - Documentation is sparse or rapidly evolving;
  - Scalability of the technology is demonstrated but not widely applied.

- **Reference implementation**
  - One or more reference implementations are available;
  - Reference implementations are usable at scale;
  - The technology may have limited adoption outside of its core development community;
  - Documentation is available and mainly accurate.

- **Emerging adoption**
  - Wider adoption beyond the core community of developers;
  - Proven in a range of applications and environments;
  - Significant training and documentation are available.

- **Evolving**
  - Enhancement-specific implementations may be available;
  - Tool suites are available to ease interaction with the technology;
  - The technology competes with others for market share.

- **Standardized**
  - Draft standards are in place;
  - Mature processes exist for implementation;
  - Best practices are defined.

D. Scale of organizational readiness model

The scale of organizational readiness for big data applications is intertwined with the technology readiness level. Therefore, it is assessed along with technology maturity.

The state of organizational readiness for big data applications can be described as follows:

1. **No big data:** The organization offers no awareness or efforts around big data

2. **Ad hoc:** Inside the organization, there is an acceptable level of awareness of big data, and some groups are building solutions in a scattered manner

3. **Opportunistic:** There is an approach to building big data solutions that are opportunistically applied, not widely accepted or adopted within the organization

4. **Systematic:** The organizational approach to big data is systematically reviewed and accepted by multiple affected parties, is repeatable throughout the organization, and nearly always followed.

5. **Managed:** Performance metrics are defined and routinely collected and reviewed for big data projects.

6. **Optimized:** The institutional improvement process for big data capabilities is in practice. The guidelines and assets are maintained to ensure relevancy and correctness in a sustained way.

III. Proposed national big data readiness assessment plan

After reviewing big data readiness assessment models, a plan is complied for a national big data readiness assessment exercise benefiting from the review of international models and detailing the activities required.

A. Questionnaire and rating criteria

The proposed national readiness for big data applications can be assessed in the Arab region using a questionnaire targeting various national institutions representing academia, the private sector, and governmental agencies.

The questions are grouped into six categories. To quantify the assessment activity of national readiness for big data, three levels were suggested for each of the questions: basic, medium, or advanced.

The numeric assignment of the levels can be 1, 3, 5, for example. The score 0 is used if the assessed feature is not available. Once assessed, each pillar will have a score based on an arithmetic average of used criteria. The following table shows the evaluation criteria adopted for each of the questions mentioned in the evaluation. The evaluation levels represent directions of improvement in each area.

Categories	Evaluation Level
I. Legal framework	
(a) In your organization, what is the governmental procedure/legal text available for data processing and transfer?	(1) There is a clear data policy, and minimal circulation. (2) Data policy is well known but rarely enforced. (3) Data policy and strategy guides institutional decisions and programs.
(b) In your organization, what is the governmental procedure / legal text available to protect privacy and confidentiality of data?	(1) A data privacy and security policy is available but it is not applied or known. (2) Data privacy and security policy is applied in most transactions. (3) Data privacy and security is being updated in all operations and at all levels, especially the leadership.
(c) In your organization, are there efforts or initiatives to develop administrative frameworks or technical standards related to big data? If yes, please list.	(1) Data management tools and related standards are available. (2) Data and storage management attract resources and interest. (3) Data management and related enterprise software are critical in day-to-day operations and transactions.

Categories	Evaluation Level
II. Infrastructure and applications	
(a) In your organization, what is the technological infrastructure that supports the following: data collection, data transfer, data storage, data processing, data visualization?	(1) Isolated applications exist for analysing big data in a distributed and inconsistent manner. (2) Reporting on big data analysis uses different platforms and standards. (3) Data management is integrated across the organization with periodic backups, security and privacy standards across a unified platform.
(b) In your organization, what are the systems or applications and tools used in the decision-making process?	(1) The data is used from various sources with various processing tools. (2) The various data sources are managed in a unified platform. (3) Advanced tools such as forecasting after collecting standardized data are applied based on a systematic process and harmonized tools.
III. Organizational and institutional structure	
(a) Does your organization have internal technical capabilities, human resources related to digital transformation?	(1) There are data managers as needed to handle the available data. (2) Data scientists are recruited and trained to undertake holistic tasks using available data. (3) A multidisciplinary data science team is available with statisticians and technologists, with initiatives in learning and applications across the enterprise.
(b) In your organization, what are the current administrative systems and organizational structures that handle data?	(1) Forming ad hoc committees to deal with the data challenges presented. (2) Seeking assistance from other similar teams as needed. (3) There is a unit dedicated to data science management in the institution.
(c) In your organization, where exactly is big data used?	(1) Raising awareness of facts related to institutional work. (2) Services and products for marketing the organization. (3) Throughout the business cycle: Using big data in planning, and operations.
IV. Literacy and training	
(a) Does your institution have awareness-raising programmes and activities on the benefits of big data?	(1) Educating employees in the field of information technology. (2) Awareness for all with training of IT personnel. (3) Awareness of leaders and administrators and providing employees with appropriate training according to their backgrounds.
(b) If your institution is in the field of education, are there academic programs	(1) Big Data course is optional for specific majors. (2) Big Data course is mandatory for all university departments.

Categories	Evaluation Level
with explicit decisions in the field of data science? If yes, list.	(3) There is an independent degree in Big Data available.
(c) Does your organization offer or facilitate training programs or capacity building courses in big data?	(1) An elective course available online and open source. (2) A compulsory course in the institution's major. (3) A step-by-step whole series of Big Data.
V. Administrative policies and regulations	
(a) Does your organization have a big data policy or strategy? If yes, please state the goals.	(1) There is a consistent big data policy that is not enforced or known. (2) Big data policy/strategy has been updated and published according to a specified schedule. (3) Big data policy/strategy is frequently updated and applied throughout the organization's operating cycles, including the evaluation of the effectiveness of the strategy itself.
(b) Does your organization have a big data implementation action plan? If yes, please list.	(1) Big data plans are only available on paper and for marketing purposes. (2) Big data action plans are selectively executed to serve leadership goals. (3) Big data business plans are implemented across all internal and external applications aimed at improving efficiency and productivity.
(c) In your organization, what resources are allocated to big data (time and funding)?	(1) There is no dedicated human resources or budget for big data, and this is done selectively. (2) The available skills are modest and only small changes are being made with big data. (3) Dedicated talent is committed and is constantly renewed and supported with new skills, and adequate budget is allocated for all activities and tools related to human.
VI. Future initiatives and projects	
(a) Does your organization have a future plan/project/initiative to use big data? If yes, please list.	(1) In case of future plans, they are not public. (2) There are general intentions to use big data in action but without specifying the path. (3) Future plans for harnessing big data are improving. Internal processes and stakeholders were efficiently discussed with the team and published with partners while identifying resources and a schedule for implementation.
(b) What are the challenges that hinder the use of big data in your organization?	There is no evaluation for this question (it is intended to identify challenges only).

B. Assessment reporting

It is recommended to verify the assessment questionnaire results with national stakeholders for irregularities or biases. In addition, it is beneficial to suggest improvement steps for various sectors in the national data landscape with specific action plans.

The reporting must not focus on numeric values, but rather on the general assessment clustered results. The assessment matrix is as rich as the diversified target organizations, and reflects the efforts of focal points to complete the questionnaire. The filled questionnaires should include numerous (at least three) representatives from academia, the Government, and the private sector.

The compiled big data national assessment matrix should identify gaps between design level objectives and counterparts of organizations that can differ between sectors, and among organizations within the same sector.

Continuous improvement cycles are recommended at a frequency not exceeding one year.

It is important to keep in mind certain values while reporting on big data readiness, as follows:

- Organizational alignment by focusing only on big data applications with a good fit within the organization scope;
- Relevant structural collaboration between organizations;
- Strengthening capabilities by focusing on data science expertise, data governance and IT governance;
- Inadequate applications of big data before mature readiness can back fire and cause serious operational and institutional damage lasting for long time;
- Public sector organizations should consider starting small, with single-function big data applications suited to their current organizations. Once realized and digested, scale up is possible.[11]

[11] Bram Klievink and others, Big data in the public sector: uncertainties and readiness, Information System Frontiers, vol. 19, pp. 267-283, 2017.

20-00412