BigDataUN Global Working Group

19th November 2020

Big Data Maturity Matrix

UNGWG Big Data Task Team on Training, Competencies and Capacity Development

What is a Maturity Matrix?

This Maturity Matrix has been designed with the National Statistical Organisation (NSO) or departments contained within the National Statistical Service in mind. A Maturity Matrix is a self-assessment tool to help the organisation understand the extent to which it has developed or implemented, in this instance, big data[1] infrastructure and applications. It therefore aims to help the organisation understand its level of "organisational maturity" with respect to big data development, and also to identify its strengths and weaknesses from which a development plan or road map may be produced.

The Maturity Matrix can be used to find the current level of big data maturity, as well as to define the target level for the organisation. Hence, as a management tool, the Maturity Matrix can help to identify a) where work is needed, and b) what the priorities are for the organisation. To note that the future development of guidance is anticipated, which will provide direction for how to develop the organisation's big data maturity.

There are two tiers:

- Tier 1: High-Level Maturity Matrix – this may be used to set the target level for the organisation across each dimension (this can be different for each dimension)
- Tier 2: Detailed level Maturity Matrix – this section is composed of questions within Excel, which upon answering, will indicate the current level of organisational Maturity.

Example

In the example below, an NSO has used the Tier 1 Maturity Matrix to identify its target maturity level as "Intermediate Level" across all four dimensions. It should also be noted, however, that the NSO may not aspire to reach "Intermediate Level" for *all* of the dimensions and may agree different target levels for each dimension.

Working through the Tier 2 detailed Maturity Matrix, questions are answered and entered into the pre-coded Excel question sheet. Grading is automatically applied (in dark blue in this example), providing a visual display which identifies the current level of maturity within each dimension. We advise that organisational level discussions are used to agree on the answers provided.

The example output below shows that this NSO has already reached "Practitioner Level" for the "IT Infrastructure" dimension, but not for the others. It has most to do in the "Legal & Policy Framework" and "Application" dimensions. The NSO may now use this information to prioritise areas for future work.

	Pre-Foundation	**Foundation**	**Practitioner**	**Expert**	***Target***
Legal & Policy Framework	Description	Description	Description	Description	*Practitioner*
IT Infrastructure	Description	Description	Description	Description	*Practitioner*
Human Resources	Description	Description	Description	Description	*Practitioner*
Application	Description	Description	Description	Description	*Practitioner*

[1] Big data refers to any non-traditional data source, e.g. web-scraped prices data, AIS shipping data, mobile phone data, etc

The NSO may wish to produce a road map that outlines the steps and work required to reach the target maturity level as quickly and as efficiently as possible within the priority dimensions.

Big Data Maturity Matrix

Background

The results from the 2019 **Global Assessment of Institutional Readiness for the Use of Big Data in Official Statistics** provided a view to the different stages of development across NSOs on their big data journeys. These insights have been used to create this maturity matrix, which aims to group the key areas of big data development into four dimensions and four levels:

Dimension 1: Legal & Policy Framework – how well established are the legal framework and associated policies for data access and data sharing at the NSO. This includes the safeguards to maintain privacy and confidentiality of big data and the processes to analyse it.

Dimension 2: IT infrastructure – the existence of an IT infrastructure to enable big data analytics in a secure environment.

Dimension 3: Human resources – this relates to the number of data science posts at the NSO, the skills, the teams and the future for recruitment and growth.

Dimension 4: Application – how data science / big data is being applied / used to solve problems in the NSO

Insights from the NSO Readiness Assessment findings were used to illustrate what an NSO at four different levels of big data/data science maturity might look like in each of the four dimensions:

Level 1: **Pre-Foundation.** This level typically describes an organisation that is at the start of their Big Data journey. They are discussing and considering how to commence big data projects as well as the strategies that they will need to put in place at the organisation to make it happen.

Level 2: **Foundation**. This describes the next level of development, where big data leadership, strategies and frameworks are being developed, and a small number of big data projects are underway to possibly investigate the use of different data sources.

Level 3: **Practitioner.** The next level of development describes an organisation where the appropriate frameworks are established, data scientists are in post and big data projects are underway and are being managed in a strategically coordinated way.

Level 4: Expert. This is the most mature data science level of development. Here, big data/data science is well embedded and data science is extensively applied within the organisation. Staff have the knowledge, skills and experience to lead and undertake big data projects within and across teams. Training, coaching and mentoring is available internally, and the organisation may also offer it externally to others.

Tier 1

High-Level Maturity Matrix – set your desired target

The high-level Matrix sets out the main objectives within each dimension by level. This aims to provide an overview for the organisation and may be used to set the desired target level within each dimension.

Dimensions	Levels			
	Pre-Foundation	**Foundation**	**Practitioner**	**Expert**
Legal & Policy Framework	The organisation (or another Government department) is starting to consider the requirements for and the development of a Legal Framework for access to and the sharing of non-traditional/big data at the organisation. Discussions may have commenced to consider disclosure control, policies and sanctions for misuse, data acquisition powers, and the IP and Copy Right requirements.	Legislation is not yet in place, but work has commenced to develop reference documents that outline how non-traditional/big data should be disclosed. A permissible gateway is in place for third parties to share data with the organisation, and agreements are recorded. The organisation uses clear statements in its communications regarding IP and Copy Rights.	The organisation has a policy or guidance in place, for the use of non-traditional/big data although these are not required by Law. Statutory powers are in place that require some third parties to share data with the organisation, and it is registered and adheres to local Laws of IP and Copy Rights.	The organisation is required by Law to have policy or guidance in place. It has a reference document in place that outlines all non-traditional/big data sources and associated disclosure rules. Statutory powers are in place that requires all big-data providers to share data with the organisation and it is registered and adheres to International Laws of IP and Copy Rights.
IT Infrastructure	The organization is starting to consider the required IT infrastructure. It does not yet have dedicated big data infrastructure (or access to it if it is external to the organisation) for working with non-traditional/big data sources. Computing power and storage is limited. Processes and tools for non-traditional/big	Projects using non-traditional/big data at the organisation may have localised infrastructure which is not integrated with normal operations (e.g. localised Hadoop or Spark applications). The reuse of some processes/tools may occur, but not in a readily accessible format. Security improvements	There is a dedicated big-data infrastructure at the organisation that has no limits to the size or complexity of the data, although some manual intervention may be required for deployment of the data. Alternatively, access to the external IT infrastructure is agreed and in place. A platform for analysis is in place with disaster recovery. Processes and tools are created	The dedicated/acquired big-data infrastructure is scalable and can cope with increasingly complex projects. A mature development process is established that uses code and data pipelines for regular statistical processes. Data security and confidentiality is fully integrated, and analysts can access the data easily as part of normal operations.

	data work are developed fresh for each project. There are no data security measures in place for the import/export of big data. The data is not widely accessible to all within the organisation.	may have been made for the import/export of data, although these may be inconsistent. Gaining access to the big data within the organisation may not be clear. Alternatively, the organisation may be in the process of agreeing access to the required infrastructure, which is available externally to the organisation.	once for multiple use across the organisation. Data security measures are in place for the import and export of big data and members of staff/analysts can access the data through a centralised request process.	
Human Resources	There is currently little awareness of big data analytics at the organisation, but senior managers have started to discuss and consider the requirements, in terms of human resources plus the associated financial requirements. It is starting to map out the big data posts that will be required and possible external expertise in this field for project delivery and/or advice. It is developing plans for recruiting and/or developing the skills within the organisation.	Big data awareness is growing within the organisation and a plan for big data positions is being developed. The mechanisms for how to recruit and build big data/data science skills are being put in place. The organisation may rely heavily on external big data expertise at this level. Some analysts may be experimenting with and practicing the use of big data.	Most of the organisation is aware of the benefits of big data analytics. The Recruitment strategy is in place and people are in big data posts, within teams. Capability building, and knowledge sharing is underway. Where required, the organisation may draw on external resources and/or support to aid big data capability building, although some internal training delivery may be established.	The culture for big data analytics is well established at the organisation and has become a normal consideration when undertaking research projects. A Big Data / Data Science profession (or job family) is defined within the organisation. There is an internal pipeline being developed to take on future big data roles, as the organisation increasingly relies on internal expertise for its big data analytics project work. All departments across the organisation work together to enhance and adopt big data approaches. The organisation may offer its expertise to others through, for example, the delivery of training, consultancy, mentorship and / or advocacy.
Application	Big data tasks may be undertaken ad hoc by individual employees, out of their own self-interest. The organisation has not developed procedures for obtaining, processing and sharing big data but it is investigating this. The solutions developed as	Big data tasks are rare but may sometimes be undertaken by teams appointed to carry them out. The procedures for acquiring, processing and sharing big data are created only for the needs of the team's task. The teams operate in an uncoordinated	Big data tasks are carried out by teams constantly operating within the organization. The team may have been recruited specifically for big data tasks at the organisation or is compiled of staff from different parts of the organization. Within the organization there are procedures for handling big data. The solutions	The big data unit has a permanent place in the structure of the organization. Tasks are performed systematically and may involve other units within the organization, or externally. The solutions developed as part of the task are implemented and the methods applied to other statistical

	part of the tasks are not generally implemented into the statistical production. The internal exchange of knowledge and information about subcontracting activities does not exist. The implementation of tasks from planning and analysis of big data is done on non-advanced office packages.	manner, and the solutions developed by them could be implemented into production to a limited extent. Information and knowledge exchange are carried out only within the team and between team leaders. The implementation of the task does not significantly affect the development of knowledge about big data in the organisation. The tools used at work are adjusted only to the achievement of the main goal of the task.	developed by the team are implemented by the organisation. Knowledge exchange takes place between the big data team and the unit responsible for the implementation of big data tasks. Theoretical and practical knowledge resulting from the tasks performed is distributed within the organisation. The implementation of tasks is based on scalable tools that can also be used for other similar tasks.	operations. Knowledge of big data tasks is widely spread throughout the organization. The organization's knowledge of big data activities is monitored on an ongoing basis. The organisation draws conclusions on an ongoing basis from the tasks performed and arising problems. On their basis appropriate solutions are developed. The organization possesses a wide range of tools that can be used for new tasks.

Tier 2

Detailed Questions to establish current level within Maturity Matrix

A. Legal & Policy Framework Dimension

The questions below will allow you to reflect on your Big Data Legal Practices and Policies. Please complete it as honestly as you can. We recommend that you answer the questions with the Legal Department to ensure accuracy and strategic buy-in.

Questions	Answers by Level			
	Pre-Foundation	**Foundation**	**Practitioner**	**Expert**
1. How would you describe the status of the Policies or guidance you have governing the handling of Big Data/Non-traditional sources?	The organisation's current policies do not adequately describe how non-traditional/big data should be used and disclosed.	We are not required by legislation to have a policy or guidance in place, but we are in the process of developing a reference document that describes how non-traditional and/or big data should be used and disclosed.	We are required by legislation to have policy or guidance and we are in the process of developing a reference document that describes how non-traditional and/or big data should be managed, used and disclosed.	We are required by legislation to have a policy or guidance in place and a Reference document is in place that outlines the list of non-traditional and/or big data and how they should be used and disclosed in different contexts.
2. How would you describe the organisation's adherence to policy/guidance?	The current policies/guidance/sanctions do not adequately address the misuse of non-traditional/big data.	We are developing a Policy or guidance for the misuse of non-traditional/big data.	We have a Policy or guidance in place but there are no sanctions in law for violations of these.	We have a Policy or guidance in place and there are sanctions in law for violations of these.
3. How would you describe the legal considerations for the disclosure of Data at your organisation?	There are no legal constraints for the disclosure of identifiable information; we may disclose identifiable information for any purpose.	We may disclose identifiable information for statistical purposes.	We may disclose identifiable information for statistical purposes only under certain controlled circumstances (e.g. for a considered 'safe' person, or for use within a 'safe' setting).	We are prohibited by law from disclosing identifiable data to anyone for any purpose and may only release non-identifiable datasets.

4. How would you describe your Partnership & Agreements practices?	We have limited or no statutory data acquisition powers and therefore any data acquisition is done through mutual agreement and in accordance with other national laws, e.g. Data Protection.	There is a permissive gateway for third parties to share data with the organisation which we use in order to acquire data. Sharing agreements are recorded and linked to existing non-traditional and/or big datasets.	We have statutory powers to require some third parties to share data with the organisation.	We have statutory powers to require all non-traditional and/or big data providers (private, public, international) to share data with the organisation.
5. How would you describe your Intellectual Property & Copy rights of Big Data?	The organisation is discussing or considering the adherence to IP & Copy Right requirements for non-traditional and/or big data.	There is a clear statement in the communication channels regarding IP & Copy Right of non-traditional and/or big data.	The organisation has registered and adheres to Local Laws of IP & Copy rights that includes non-traditional and/or big data.	The organisation has registered and adheres to International Laws of IP & Copy rights that include non-traditional and/or big data.

B. Information Technology Infrastructure Dimension

The questions below will allow you to reflect on your IT infrastructure and resources vis-à-vis big data utilization. We recommend that you answer the questions with the IT Department to ensure accuracy and strategic buy-in. Whether the storage is hosted internally or externally does not affect maturity level.

Questions	Answers by Level			
	Pre-Foundation	**Foundation**	**Practitioner**	**Expert**
1. How would you describe the computing power and data hosting at your organisation?	Big data projects do not have dedicated infrastructure, but rather use existing resources. Big data computing power and storage capacity is severely affected and limited due to lack of dedicated resources.	Big data projects may have their own infrastructure, but it may not scale well or may constrain the scope of projects. Some specialized big-data infrastructure is used (e.g. a Hadoop or Spark cluster), but it is not tightly integrated with normal operations. There are backups, but they may not be tested regularly.	There is dedicated big-data IT infrastructure that does not limit the size or complexity of current big data operations. Deployment may require manual intervention. Big data is stored and analyzed on specialized infrastructure, with developed disaster-recovery.	The dedicated big-data IT infrastructure is designed to scale with increasingly complex projects and big data is an integral part of IT operations. Infrastructure deployment and computer resource provisioning can be automated.
2. How would you describe the processes and tools used in big data projects at your organisation?	Processes and tools used in big data projects are created fresh each time on non-advanced office packages. There is no specific big-data deployment process, or an existing data warehouse is used. All configuration is performed manually.	The tools used are often adjusted only for the purpose of the big data project/task. The processes and tools that are reused, are often extensive configuration requirements or versions. Infrastructure status may require manual checks. May have deployment code stored online.	A development process is in place to prevent the need to create new processes and tools for new projects, and they are generally designed for scalability and reuse. Infrastructure status is monitored and measured.	The organisation possesses a wide range of tools that can be used for new tasks. Code, data configuration, and processes are designed for reusability, as part of a mature software-development process. Infrastructure teams manage system configuration in version control. All configuration is performed through code and data pipelines.
3. How would you describe the security of the data	The data import/export processes used for big data do not integrate security	The data import/export processes may be modified to increase security and/or	Big data (or its derived datasets) is accessed and/or imported and exported in a	Security and confidentiality are integral to the big-data strategy, and the organization

import/export processes at your organisation?	measures specific to big data concerns.	confidentiality of data, but it may be ad hoc or inconsistent.	systematic fashion that ensures the security and confidentiality of data.	is continuously looking for ways to improve it.
4. How would you describe the accessibility to big data and related applications?	Only IT staff can access the data and applications, or analysts access them on an ad hoc basis.	Analysts and application users either are unsure how to get the data or access the applications, or each has their own method for accessing the data they need.	Analysts need to use multiple processes and/or need to make a service request in order to access the data.	Analysts have straightforward access to the data and use it as a part of their normal operations.
5. How would you describe the data governance practices for big data at your organisation?	Any control over data publication, retention, or adherence to contractual limitations on data use is performed manually.	Any control over data publication, retention, or adherence to contractual limitations on data use requires manual configuration of settings or case-by-case management.	Control over data publication, retention, or adherence to contractual limitations on data use may have some automation or external controls, or automation of policies. Reproducibility may therefore be limited.	Data publication and sharing is confirmed independently prior to publication, while ensuring control of data. There is robust automation of policies on deriving data, ability to set individualized policies on datasets, and dissemination criteria. It is clear how or whether results can be replicated.

C. Human Resources Dimension

The questions below will allow you to reflect on your Human Resources practices with regards to big data/data science posts and skills. We recommend that you answer the questions with the Human Resources Department to ensure accuracy and strategic buy-in.

Questions	Answers by Level			
	Pre-Foundation	**Foundation**	**Practitioner**	**Expert**
1. How would you describe the awareness of non-traditional/big data sources within your organisation?	There is currently little awareness of non-traditional/big data sources at the organisation, although senior managers are discussing it more frequently.	Awareness of non-traditional/big data sources is growing within the organisation, especially through IT practitioners, analysts and their senior managers.	Most of the organisation understands how non-traditional/big data analytics can benefit outputs.	The culture for the use of non-traditional/big data sources is well established and adopted as a normal consideration when undertaking research projects at the organisation.
2. How would you describe your organisational structure with regards to non-traditional/big data/data science posts?	The organisation is starting to consider and discuss what is needed in terms of non-traditional/big data/data science positions at the organisation.	A plan for non-traditional/big data/data science positions is being developed by the organisation and is planning where posts should be placed and how to fill them.	Non-traditional/big data/data science positions are present within the organisation and feature in the regular organisational business planning process.	Posts are well established and succession plans are in place for more senior positions. A big data/data science job family is defined within the organisation.
3. How would you describe current big data/data science skills within the organisation?	Big data/data science skills are present *ad hoc* within the organisation.	Some analysts are developing new skills for the extraction, transformation, linkage, and analysis of big data. Some may also be developing knowledge of big data programming tools such as R and/or Python.	Big data analytics and data science techniques are well established. Project findings, methodologies and code are readily shared across the organisation for knowledge exchange.	Big data/data science skills and techniques are very well established. Research papers, methodologies and findings are presented nationally and internationally via journals and conferences.
4. How would you describe the acquisition of in-house/outsourced big data/data science expertise for project delivery?	The organisation is discussing the possibilities for acquiring big data/data science expertise for project delivery.	The organisation is drawing mainly on external resources to aid big data/data science project delivery. Internal staff are encouraged to participate to aid knowledge exchange.	The organisation draws on both internal and external expertise for big data/data science project delivery. Knowledge exchange to internal staff is fully enabled through well managed	The organisation uses mainly internal resources for the delivery of big data/data science projects. Internal team collaboration is commonplace, aiding knowledge exchange and the development/growth of

			collaborative projects and research.	internal staff. External collaborations continue, with two-way knowledge exchange well established.
5. How would you describe the availability of big data/data science training courses at the organisation?	There are currently no training opportunities at the organisation to develop employees in the field of big data/data science.	The organisation commissions others to provide training for employees in the field of data science/big data.	The organisation is building internal training resources for internal delivery. Training may be delivered by those who have already acquired new skills at the organisation, by an external provider, or by a dedicated data science trainer.	The organisation has established the resources and materials for the internal delivery of training. Formal programmes of assessed training are established via emerging collaborations with universities, coaching and learning centres, or others. The organisation is capable of offering big data/data science training, consultancy, guidance and mentorship externally to other organisations.
6. How would you describe your big data/data science recruitment processes?	The organisation is discussing the processes required for big data/data science recruitment.	Big data/data science recruitment processes are being established by the organisation that include a simple assessment of technical competencies.	Big data/data science recruitment processes are established by the organisation that include a good assessment of technical competencies against an appropriate Competency Framework.	Well-defined recruitment processes are in place that use robust assessments of technical competencies against an appropriate Competency Framework, using approaches such as a presentation of project case studies or assignments. The organisation may offer advice and templates to others who are seeking advice on how to recruit for data science posts.

D. Application Dimension

The questions below will allow you to reflect on the application of big data/data science within your organisation. We recommend that you answer the questions with the Senior Management Team to ensure accuracy and strategic buy-in.

Questions	Answers by Level			
	Pre-Foundation	**Foundation**	**Practitioner**	**Expert**
1. Who carries out big data tasks/projects in the organisation?	Big data tasks/projects may not yet be happening, although Individual analysts within the organisation have started to become interested.	Ad hoc groups of analysts appointed to carry out a single big data task.	Permanent teams have been established to carry out specific tasks. They are activated when they are needed.	Units responsible for the implementation of big data tasks have a permanent place in the structure of the organization. They are likely to engage and/or collaborate with the rest of the organisation to help with the delivery of organisational objectives.
2. How are big data tasks/ projects carried out in the organisation?	Where big data tasks are undertaken, they may be very uncoordinated, initiated by individuals at the organisation out of own interest.	Big data tasks/projects are still uncoordinated, but they may now be requested (e.g. by a manager). They may feed into localized business needs or interests.	Big data tasks/ projects are strategically coordinated and effectively contribute to wider organisational objectives.	Tasks are performed systematically. The dedicated team is able to work collaboratively with other units within the organization, or with other departments or organisations to help others meet their organisational objectives.
3. How would you describe the procedures that have been developed in relation to acquiring, processing and sharing big data	The organisation has not developed procedures for obtaining, processing or sharing big data.	The procedures for acquiring, processing and sharing big data are created only for the needs of the team's tasks (they may be contractual, do not affect the functioning of the organization).	Within the organization, the procedures for handling big data are documented and are being applied.	The implementation of big data tasks within the organization is incorporated into the regular procedures.

in the organisation?				
4. How would you describe the implementation of big data solutions into the statistical production process at the organisation?	The solutions developed as part of the tasks are not implemented in statistical production.	Solutions developed by team of experts could be implemented into production to a limited extent.	The solutions developed by team of experts are implemented in the entire scope of statistical production.	The solutions developed as part of the task are implemented and the methods applied to other statistical operations.
5. How would you describe the process of integrating project results with the organisations resources?	Big Data resulting from project undertakings are not integrated with other data sources held at the organisation. The organisation has not developed methods and techniques for big data integration with other datasets.	Big Data sources are integrated with other data sources held at the organisation for the needs of individual tasks/projects. Data integration is temporary. New databases are redundant with limited metadata.	The integration of big data with other data sources is permanent. The developed methodological and technical solutions allow for constant data integration with other data sources held by the organisation. Databases have complete and consistent metadata.	Full data integration with other data sources held by the organisation takes place. Data is constantly monitored and adapted to emerging changes.
6. How are the results of the big data research projects disseminated?	The results are not shared outside of the organisation, but they may be disseminated to a narrow audience within the organisation. No documentation has been developed and made available as part of the big data tasks / projects carried out.	The results are shared outside of the organisation in the form of experimental statistics. Limited documentation is made available as part of the tasks / projects (e.g. methodological assumptions, user guide).	The results are published outside of the organisation in the form of single publications related to the topic. As part of the tasks / projects carried out, full documentation (e.g. technical - source codes, processing algorithms, information on data flows) is made available.	The results are published on a regular basis. They can be accompanied by additional analyses and interpretative comments by the authors. There are reports published concerning how progress is made as well as the planned implementation.
7. How is information from big data projects shared within or outside the organisation?	The internal exchange of knowledge and information about subcontracting activities does not exist.	Information and knowledge exchange are carried out only within the team and between team leaders.	Knowledge exchange takes place across the organisation, sharing project findings, methodologies and code.	Knowledge of big data tasks is widely spread throughout the organization. Research papers, methodologies and findings are presented nationally and internationally via blogs and/or conferences.