Capability Maturity Models as a Means to Standardize Sustainable Development Goals Indicators Data Production

Ignacio Marcovecchio[1,2], Mamello Thinyane[1], Elsa Estevez[2,3] and Pablo Fillottrani[2,4]

[1]*United Nations University, Institute on Computing and Society (UNU-CS), Macau*
[2]*Depto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Argentina*
[3]*Instituto de Ciencias e Ingeniería de la Computación, CONICET, Argentina*
[4]*Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina*
E-mail: ignacio@unu.edu; mamello@unu.edu; ece@cs.uns.edu.ar; prf@cs.uns.edu.ar

Abstract

Achieving the Sustainable Development Goals (SDGs) demands effective harnessing of the ensuing data revolution – the integration of new and traditional data to produce high-quality indicators that are detailed, timely, and actionable for multiple purposes and to a variety of users. The quality of these indicators, defined in terms of completeness, uniqueness, timeliness, validity, accuracy, and consistency, is crucial for their use in national level planning, monitoring and evaluation (PM&E) processes, for facilitating global monitoring of progress on the SDGs, and for enabling comparative evaluation between countries. The use of indicators for trans-national analyses and global-level decision making necessitates coordination, integration, and inter-operation between the various stakeholders within the global data ecosystem. Various instruments, including protocols, models, frameworks, specifications,

and standards are used widely to facilitate the coordination, integration, and interoperation across various global systems, such as telecommunication systems. In this paper, we posit that Capability Maturity Models (CMMs) can be an instrument and a mechanism towards not only ensuring the production of high-quality indicators data, but also for standardizing the key processes around the production of SDG indicators data, and for facilitating interoperation within the data ecosystem. This paper motivates for the adoption and mainstreaming of organizational CMMs within the SDGs activities. It also presents the preliminary formulation of a multidimensional prescriptive CMM to assess and articulate pathways towards the maturity of organizations within national data ecosystems and, therefore, the effective monitoring of the progress on the SDG targets through the production of high-quality indicators data. Furthermore, the paper provides recommendations towards addressing the challenges within the increasingly data-driven domain of social indicators monitoring.

1 Introduction

In September 2015, leaders of 193 countries agreed on 17 Global Goals for Sustainable Development which set off a worldwide call to protect the planet and ensure peace and prosperity for all people by the year 2030. These goals, known as the SDGs, define the global development agenda and present aspirational objectives that must balance the three pillars of sustainable development: social inclusion, economic development, and environmental sustainability. The SDGs build on the success of the Millennium Development Goals (MDGs), a set of time-bound and quantified targets agreed in September of 2000 during the UN Millennium Declaration [1]. In particular, the SDGs prioritize areas not considered before such as climate change, economic inequality, innovation, sustainable consumption, peace, and justice [2]. The 17 goals aim at reaching 169 targets, which will be monitored and evaluated through 230 indicators. The United Nations Statistical Commission (UNSC) [3] is the body within the UN system responsible for the development of a global indicator framework for monitoring the progress towards the achievement of the SDGs. At the national level, National Statistics Offices (NSOs) typically have the custodianship for the production of relevant indicators and for coordinating activities within the National Statistical Systems (NSS).

A crucial component of the SDG agenda is the monitoring of progress towards the achievement of the targets, as well as the development of suitable technology tools and platforms to support the activities of different stakeholders [4]. It is expected that the monitoring of the SDG indicators will demand further efforts to produce reliable and high-quality data that is inclusive and ensures the principle of 'leave no one behind' [5]. However, there are deep-rooted capacity challenges for many countries in measuring progress on the proposed SDGs [6]. The capacity of key players in the data ecosystem, including governments, institutions, and individuals, also needs to be enhanced to be able to deliver and take advantage of this data. There is, therefore, a universal imperative to ensure that countries have effective NSS, capable of measuring and producing high-quality statistics in line with global standards and expectations [5].

High-quality data is critical for transforming the SDG indicators into useful tools for national level PM&E processes, for facilitating global monitoring of progress on the SDGs, and for enabling comparative analysis between countries. Indicators data that accurately reflect the progress, monitor the allocation of resources, inform policy making, and assess the impacts of policy and programs, are fundamental for accountability and monitoring of the 2030 Agenda for Sustainable Development. Beyond the use of data for informing national-level monitoring and analyses, indicators data is used for informing progress on the SDGs at the global level and for facilitating analyses at that level and scale. The use of indicators for trans-national analyses and global-level decision making necessitates coordination, integration, and inter-operation between the various stakeholders within the global data ecosystem. One of the key aspects of the analysis that is undertaken at the global level is to perform comparative ranking and clustering of different countries based on various aspects associated with their progress and achievement towards the SDGs. At this level, there is a need for *coordinating* the processing of indicators between countries, the requirement to *integrate* indicators data from various actors and sources, and to *interoperate* with diverse and heterogeneous elements within the data ecosystem. These complex interactions and data dependencies within the indicators data ecosystem can be harnessed through standardization processes towards ensuring a more effective and efficient use of indicators data.

In view of the inherent and increasing complexity of the national and the global data ecosystems, this research adopts an organization thinking approach to explore the potential interventions towards improving the capacity of organizations within the national data ecosystem to be more effective in

producing high-quality data and therefore in monitoring the SDG indicators. In particular, the research explores the opportunity for standardizing aspects of the production of the SDG indicators through the formulation of a CMM. This paper is organized as follows – having provided an introduction to the research in Section 1, Section 2 discusses the unfolding data revolution especially in the context of social indicators monitoring for SDGs. Section 3 presents an extensive review of the current initiatives on improving the quality of statistical data. Section 4 motivates for the use of CMMs within the SDG indicators framework for improved quality of SDG indicators data. This is followed by a presentation of a multidimensional prescriptive CMM in Section 5. Sections 6 and 7 provide recommendations and a conclusion to the paper respectively.

2 The Data Ecosystem Evolution

The volume of data in the world is increasing exponentially. One estimate is that 90% of the data in the world has been created in the last two years [5]. The volume and types of data available nowadays have increased exponentially due to the evolution of technology and its impact on the social practices and behaviors. All stakeholders within the global data ecosystem, including governments, companies, academia, and civil society, are needing to adapt to this new reality and need to be prepared to continue adapting to a world that produces more and more data (i.e., volume), generated at a faster speed (i.e., velocity), and coming from new sources (i.e., variety). This evolution of the data ecosystem is being fuelled by the new reality that has been defined as the data revolution.

The concept of data revolution was coined in 2013 in the report of the High-Level Panel of Eminent Persons on the post-2015 Development Agenda [7] and it is defined as "an explosion in the volume of data, the speed in which data is 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 perception data" [5, p. 6].

Applying a data revolution perspective to the SDGs involves the integration of new data (e.g., crowd-sourced data, citizen-generated data) with traditional data (e.g., census data, household surveys) to produce high-quality information that is more detailed, timely and relevant for many purposes and users, especially to foster and monitor sustainable development [5]. Traditional statistics entities must therefore not only engage with new data sources but also

with new technologies and data analysis tools. The role of technology within the SDG indicators monitoring framework and towards the achievement of the SDGs is paramount, and this is highlighted and supported by the articulation of technology as an explicit Means of Implementation (MoI) under SDG17. Further, fuelling the evolution and the need for modernization of the statistics production systems is the fact of the large number of SDG indicators for which novel and innovative data sources, methodologies and technologies are needed [4].

NSOs, who are the traditional custodians of national indicators data, remain central to the government efforts to harness the data revolution for sustainable development. To fill this role, however, they need to be able to adapt to the constant changes, they need to embrace efficient production processes, incorporate new data sources, and ensure that the data cycle matches the decision cycle. However, many NSOs lack sufficient capacity and funding, and remain vulnerable to political and interest group influence. Data quality should be protected and improved by strengthening NSOs, and ensuring they are functionally autonomous, independent of sector ministries and political influence. Their transparency and accountability must be improved, including their direct communication with the public they serve [5]. The role of the NSOs within the data ecosystem, in particular with regards to the SDG indicator data production, is also in a flux. While NSOs have traditionally played a central custodianship role with regards to the production of indicators data, an alternative trajectory would see NSOs role being more of a facilitator and a coordinator within the data ecosystem, responsible for shaping, curating, and integrating various elements of indicators data [8].

The data revolution not only presents opportunities and positive prospects for the monitoring of SDG indicators, it also presents challenges and raises new associated risks. One of the main challenges for monitoring the SDGs is to minimize the risks and maximize the opportunities that come from the data revolution for sustainable development. One of the key challenges and risks is that of accentuating the data divide (i.e. the gap between those who have ready access data and information, and those who do not). The current paradox in sustainable development indicators data production is that those countries that are presumably in most need of high-quality indicators data are the same countries that might be least able to produce the high-quality data [9]. Furthermore, the current articulation of the role of data and technology towards supporting the SDGs does not address nor challenge the underlying structural inequalities between the developed countries and the developing countries, and between governments and citizens. Inequalities in the access and use

of indicators data must be tackled to reduce the gap between data-rich and data-poor countries. Another risk and challenge of the data revolution (and in part of the evolving data ecosystem) is that which is associated with the proliferating heterogeneity and variety (e.g., a variety of new data sources, new actors and players). The underlying phenomenon of increasing diversity and heterogeneity within the indicator data ecosystem is not necessarily negative and it can, in fact, be seen as positive; however, it introduces a need for increased coordination, integration and interoperation between the various stakeholders and elements within the data ecosystem. It also amplifies the need for standardization across the various elements and flows within the data ecosystem. Further significant challenges that need to be considered include:

- There are data and knowledge gaps – new science, technology and innovation (among others) are needed to fill such gaps.
- There is not enough high-quality data – many countries cannot rely on their data because it is outdated, incomplete, or it simply does not represent the reality accurately.
- Lots of data that is unused or are unusable – many countries still have data that is of insufficient quality to be used to make informed decisions, for governments to be accountable or to fostering innovation.

A key role of the UN and other international organizations is to set up principles and standards, and to harness actions that encapsulate common norms and best practices. Mobilizing the data revolution for achieving sustainable development urgently requires actions such a raising awareness, improving capacity, setting standards, and building on existing initiatives in various domains. In particular, initiatives built over previous foundations should consider the data production ecosystem to understand the multi-stakeholder engagement issues related to data sharing, ownership, risks, and responsibilities. Such initiatives are indispensable to enable data to play its essential role in the implementation of the development agenda.

The Independent Expert Advisory Group on a Data Revolution for Sustainable Development calls for "international and regional organizations to work with other stakeholders to set and enforce common standards for data collection, production, anonymization, sharing and use to ensure that new data flows are safely and ethically transformed into global public goods, and maintain a system of quality control and audit for all systems and all data producers and users" [5, p. 18]. Towards this aim, efforts must be made to support countries in empowering their NSS to be able to adequately respond to the realities of evolving data ecosystem, and to produce and use high-quality indicators data in ways that advance national and global aspirational development goals.

3 Standards and Related Instruments for Indicators Data

In order to serve sustainable and inclusive development, social indicators and statistics should be of high-quality, timely, easily accessible, reliable, and sufficiently disaggregated. Data disaggregation, in particular, is key to achieving the principle of "leaving no one behind" [10]. The importance of the role of the national statistics entities in the production of official statistics for the monitoring and implementation of the development agenda, and the importance of high-quality statistics have been well articulated in the literature [11]. Further, there has been extensive work with the domains of social indicators and statistics to develop instruments to facilitate production of high-quality social indicators.

From an extensive literature review, a set of initiatives aimed at improving the functioning and the results generated by the national statistics entities have been identified – these comprise standards, models, frameworks, processes and programs, enterprise architectures, and readiness studies (Figure 1).

Among the **frameworks** for data or statistics, the following can be highlighted:

- *National Statistics Quality Framework* – based on the European Statistical System dimensions of quality (as laid out in the National Statistics Code of Practice Protocol on Quality Management), aims to improve the quality of data collected, compiled and disseminated through enhancing the organization's processes and management [12].
- *Frameworks for National Statistics* – define the status and governance framework for official statistics. For example, the one developed by the UK Statistics Authority [13] focuses on economy and society.
- *Statistics Quality Frameworks (SQF)* – set forth main quality principles and elements guiding the production of statistics. An example is The European Central Bank Statistics Quality Framework [14].
- *Monitoring and Evaluation Frameworks* – aim at identifying trends, measuring changes and capturing knowledge to improve programs' performance and increased transparency. For example, the SDG Fund Secretariat [15] has established a Monitoring and Evaluation framework with key indicators that allows to obtain a comprehensive overview of the contribution to sustainable development.
- *Process Quality Frameworks* – the framework for process quality in national statistical institutes [16] proposes a structured framework for the quality of the statistical processes used to produce official statistics.
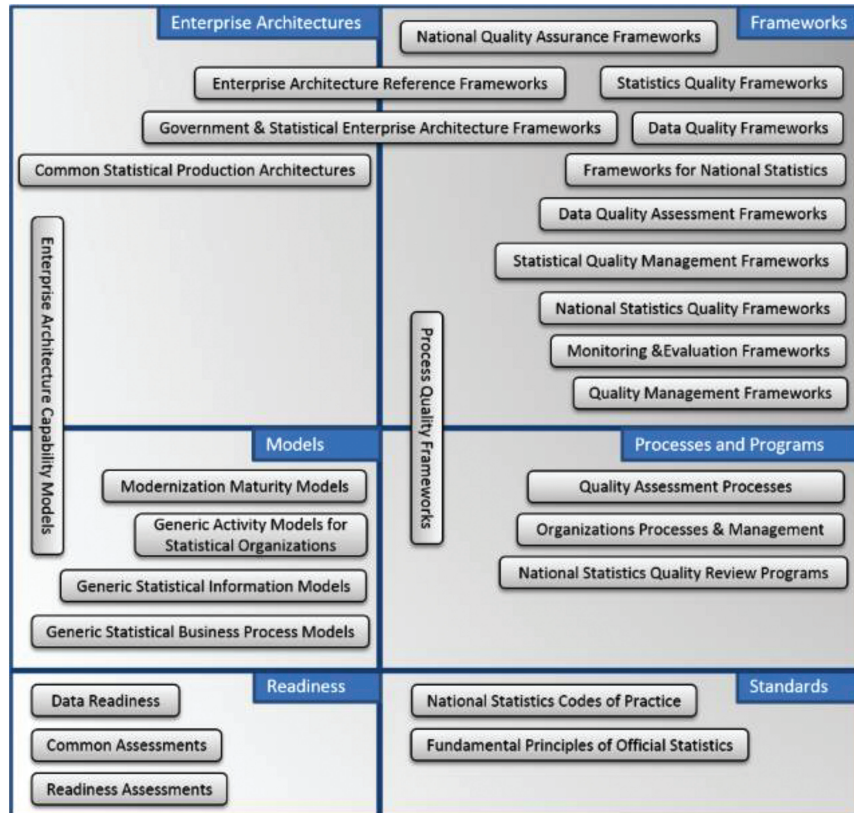
Figure 1   Initiatives for Improving Quality in Statistics Generation.

- *Quality Management Frameworks* – for example, the one implemented in the Central Statistics Office in Ireland [17] is an extensive and long-term program of activities aiming at ensuring that statistical production meets the highest standards as regards quality and efficiency.
- *Quality Frameworks* – provide a systematic mechanism for ongoing identification and resolution of quality problems and increased transparency to the processes used to assure quality. An example is the Quality Framework and Guidelines for Economic Co-operation and Development (OECD) Statistical Activities, developed by the OECD in 2012 [18].
- *Data Quality Assessment Framework* – evaluates the data quality of statistics. For example, the International Monetary Fund created a data quality assessment framework [19] for comprehensive assessments of

countries' data quality. It defines five dimensions and it covers institutional environments, statistical processes, and characteristics of the statistical products.

- *Statistical Quality Management Framework* – aims at setting out clearly and succinctly an organization's commitment to quality in respect of particular statistical outputs, and to describe the steps that it will take to meet its quality aims [20].
- *National Quality Assurance Framework (NQAF)* – developed by the Global Expert group under the UNSC is a tool comprising a template, guidelines, a glossary of quality-related terms, and an inventory of quality references aiming to provide the general structure within which countries can formulate and operationalize national quality frameworks of their own or further enhance existing ones [21].

Enterprise Architectures (EA) are formal descriptions of the structure and function of organizational components, the relationships between such components as well as the principles and recommendations for their creation and development over time [22]. Some EA applications to official statistics include:

- *Enterprise Architecture Reference Frameworks (EARF)* – aim at helping countries (in particular, EU member states) with the production of statistics that respond more quickly and cost-effectively to new statistical business needs [23].
- *Common Statistical Production Architecture (CSPA)* – provides support for the whole span of statistical production process and gives a framework for collaborating and sharing effectively [24].

Koskimäki and Koskinen [25] discuss Statistical Enterprise Architectures as tools for modernizing the NSS by identifying the gaps and overlaps between CSPA and EARF from the point of view of the National Statistics Institutes.

Readiness studies analyze the conditions in a country, city or sector to see if data initiatives are likely to be successful and, at the same time, they seek out suitable areas and identify challenges that may exist when implementing such policies [26]. Some readiness studies in the domain include:

- *Readiness Assessments* – are used to determine the existing environment and the preparedness for change. UNDP has developed a prototype tool – the Rapid Integrated Assessment (RIA) – to support countries in assessing their readiness for SDG implementation. RIA reviews the current national development plans and relevant sector strategies, and

provides an indicative overview of the level of alignment with the SDG targets.

- *Common Assessments* – useful for assessing and promoting common approaches towards objectives involving multiple stakeholders. The Common Country Assessment (CCA) prepared by UNDP informs the design of UN policies and programs at the country level based on the review of context-specific data that correspond to the SDGs and targets of the 2030 Agenda [27]. The CCA assists in identifying links among goals and targets in order to effectively determine mutually reinforcing priorities and catalytic opportunities for implementation of the new agenda as a whole.
- *Data Readiness* – a tool to assess an organization's ability to produce and report data. In [28], a design-reality gap model is applied for the assessment of big-data-for-development readiness, barriers and risks. This kind of tools could similarly be applied to assess readiness for monitoring the progress towards the achievement of the SDGs.

Processes and **standards**. A statistical process is defined as the collection, processing, compilation and dissemination of statistics for the same area and with the same periodicity [29]. A statistical standard provides a comprehensive set of guidelines for surveys and administrative sources collecting information on a particular topic [30]. The following are some processes and standards for statistics:

- *Quality Assessment Process* – their purpose is to define the steps to process data in such a way that quality is preserved. The quality assessment process for Big Data developed by the OECD [31] presents a data quality assessment process which includes a dynamic feedback mechanism to adapt to the characteristics of big data, and define the tasks that should be conducted at early stages to improve quality.
- *Codes of Practice (CoP)* – the European Statistics Code of Practice aims to ensure that statistics produced are not only relevant, timely and accurate but also comply with principles of professional independence, impartiality and objectivity [16]. Similarly, the UK National Statistics Code of Practice sets out conditions and procedures which govern access to data, including access to data for research purposes, and appropriate actions for unauthorized data disclosure [32].

There are also **models** to represent information, activities, capabilities, business processes, and modernization of statistical organizations. Examples of such models are:

- *Generic Statistical Information Model (GSIM)* – a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics [33]. It describes the information objects and flow within the statistical business process.
- *Generic Statistical Business Process Model (GSBPM)* – describes and defines the set of business processes needed to produce official statistics [34]. It covers all the activities undertaken by producers of official statistics – at both national and international levels – which result in data outputs. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources.
- *Generic Activity Models for Statistical Organizations (GAMSO)* – describes and defines the activities that take place within a typical statistical organization. It extends and complements GSBPM by adding additional activities needed to support statistical production. It is useful to assess the readiness of organizations to implement different aspects of modernization.
- *Modernization Maturity Models (MMM)* – self-evaluation tools to assess the level of organizational maturity against a set of pre-defined criteria. The United Nations Economic Commission for Europe (UNECE) defined a MMM that considers multiple aspects of maturity and distinct dimensions in the context of modernization [35]. The model defines maturity levels allowing identifying the organizational maturity, which can be compared between organizations, and between statistical domains/business units within an organization.

4 Statistical Processes Standardization and CMMs

Most of the existing work aimed at supporting the NSOs and the stakeholders within the NSS – including some of those discussed in Section 3 above, typically focus on assessing and improving the quality of the statistics and the indicators data produced. As such, they tend to embrace a data-centric and an output-focused approach wherein, for example, the quality of the indicators is conceptualized in largely data-centric terms (e.g., validity, accuracy, completeness). In this research, we posit that it is important to primarily consider the quality of the processes that are employed and engaged towards the production of the indicators, as these processes eventuate the

final output (i.e., the indicators). Thus, while it is important to achieve coordination, integration and interoperability at the SDG indicators level, it is paramount that standardization is considered at the process and function level, towards informing not only the assessment of capability and the maturity of NSOs, but also towards prescribing pathways towards mature operations and processes.

To be able to monitor progress, make governments accountable, and advance sustainable development, having mature organizations that are able to fulfil the rapidly changing demand for high-quality information is utterly important. It is also imperative for the improvement of the capability of the national data ecosystem that frameworks, models, and standards are formulated to support the adoption of best practices for improving the production of the SDG indicators.

The capability of national data ecosystems, focusing particularly on organizational capacity, can be improved through the formulation of a new multidimensional prescriptive CMM that would enable the assessment of the capability (i.e. as far as, collecting, analysing, processing, and reporting data about the SDGs.) of the entities responsible for reporting on the progress of the SDGs at the national level (i.e., the NSOs). Maturity reflects a level of organizational development which can be used to determine the capability of organizations to perform certain activities. Maturity models are an important tool to assess the quality and effectiveness of processes. Evaluating maturity became popular with the introduction of the CMM for software defined by the Software Engineering Institute at Carnegie Mellon University [36]. Maturity models can be used to identify organizational strengths and weaknesses, and as tools for benchmarking [37].

Prescriptive models surpass descriptive ones since they are good not only for assessing the here-and-now (also known as the "as-is" situation) but also to indicate the way to improve the level maturity by enabling organizations to develop a roadmap for improvement [38]. Organizations applying these types of models benefit from the ability to measure and assess their capabilities at a given point in time and to have guidelines on improvement measures.

Some of the statistics data quality initiatives (as discussed in Section 3) stand to make a contribution to improving the quality of the data produced by the NSOs. However, it remains that none of these initiatives is specifically aimed at improving the quality of the data generated for the monitoring of the SDGs, and at assessing the capability maturity of such entities and the processes they use to produce SDG statistics. The closest initiative would be the MMM as it can be used to identify the maturity of statistical organizations

and it helps them to modernize the way they produce official statistics. Nevertheless, the most critical difference with the model presented in this paper relies on the focus: while the CMM targets specifically the process that informs the progress towards the SDGs, the MMM focuses on the approach followed by statistical organizations to modernize the way they produce official statistics as a whole. The evolution of the model is also different; while the CMM is prescriptive, the MMM is descriptive. Defined as a "self-evaluation tool to assess the level of organizational maturity against a set of pre-defined criteria" [35, p. 1], the MMM is complemented by a roadmap where the guidelines to reach higher levels of organizational maturity are defined. The CCA by UNDP can also be a useful input to the CMM since it holds the potential for ensuring that the support provided by UN agencies as a whole in a country is coherent and complementary, drawing from each agency's expertise, resources, and mandate. Other existing efforts, like GAMSO and CSPA, could also inform the CMM.

5 Capability Maturity Model for SDG Indicators Monitoring

The CMM developed in this research is multidimensional and focuses on the statistical production of data to inform the SDG indicators. The high-level elements of the model are a set of dimensions (i.e., people, processes, technology, data), phases of the process(es) for data production (i.e., collection, processing, uptake, impact), and the levels that describe the maturity of the capability to produce such SDG data (i.e., basic, supported, managed, proficient). The description of the high-level elements along with the rationale and justification for their definition is described below, followed by the resulting model.

5.1 Dimensions

Since the end of the 20th century, organizations have been described in terms of their people, processes, and technology [39]. These three pillars are known as the "golden triangle" and have been used to describe not only organizations but also processes, projects, systems, and frameworks. However, due to the rapidly changing environment and the different nature of organizations, numerous other dimensions have been proposed to better describe the specifics of the context and the situation of the various organizations. Some of these further dimensions include data, information, strategy, measurement, and organizational culture.

This research centres on the data for the SDG indicators, since only robust and reliable data for evidence-based decision-making enables the development of implementation strategies and proper allocation of resources. The hypothesis underlying this work postulates that the more mature a statistical organization is, the higher the quality of the data it can produce and, therefore, the better the decisions that can be made. Although data quality is commonly determined by the fitness for use or fitness for purpose, among statistical organizations the definition by the ISO 9000:2015 standard [40] – "the degree to which a set of inherent characteristics fulfils requirements" – is the most commonly used [41]. This definition is also utilized in the Statistical Data and Metadata Exchange (SDMX) Common Vocabulary [42] and in the NQAF Glossary [43]. However, the concept of quality of statistical data is considered to be multidimensional as no single measure of data quality is considered to be comprehensive enough [42]. There is no unique set of dimensions to describe data quality, but some of the most recurrent ones include relevance, timeliness, accuracy, accessibility, and coherence. Beyond the quality of the data, the SDGs demand a set of key principles to be respected when data is involved [5]. Such principles are disaggregation, timeliness, transparency and openness, usability and curation, protection and privacy, governance and independence, resources and capacity, and rights.

Such is the importance of SDG indicators data for the accomplishment of the 2030 Agenda for Sustainable Development that the dimensions measured by the CMM are people, processes, technology, and data. The *data* dimension considers the sources that recognize unheard voices so nobody remains uncounted or is left behind, and ensures the rights that every individual deserves. It does also consider the ethics in the manipulation of data that guarantee protection and privacy to all human beings, as well as promotes openness, exchange and sharing of data. The *people* dimension covers the resources and capacity principles examining both, the individual capabilities (e.g. education, training, skills) as well as the organizational capabilities (e.g. culture, policy, strategy). The *processes* dimension ponders on the methodological aspects, the existence and utilization of standards, guidelines, good and best practices, and the overall quality management processes. Finally, the *technology* dimension – besides covering resource aspects – analyses the supporting ICT infrastructure and the tools, platforms, systems, and services available for the production of statistical data for the SDG indicators. With these four dimensions, all the principles mandated by the SDGs are covered, as shown in Figure 2.
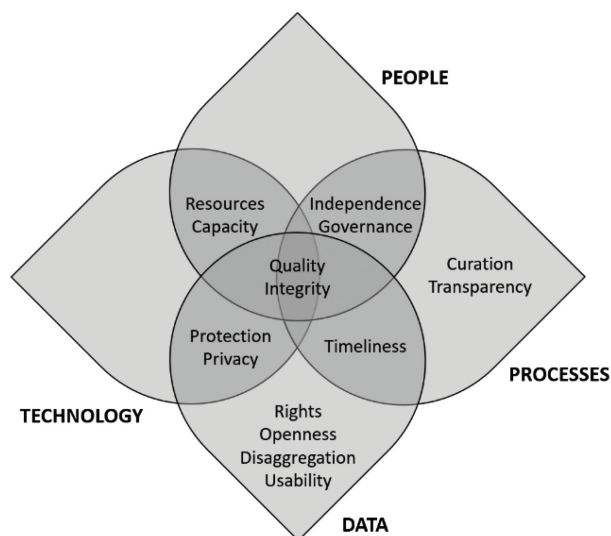
Figure 2 CMM Dimensions and Data Key Principles.

5.2 Levels

Maturity models generally include a sequence of levels that define a path from the lowest (initial) to the highest (ultimate) state to maturity [44]. The most popular way of evaluating maturity is through a five-point Likert scale where the higher the level, the higher the level of maturity [38]. Within one statistical organization, though, there can be different maturity levels depending on the statistical domain or part of the organization [45]. Nonetheless, the model presented in this research focuses only on one domain – the production of data to inform the SDG indicators, and only on the processes for producing data for the SDG indicators.

In this model, the levels define the maturity of the statistical organization to carry out the different phases of the production of SDG indicators data. Four levels have been defined to categorize the maturity of the capability to carry out such processes:

- *Basic* – the capability maturity of the organization is low, and the confidence in its results is little due to the fact that the data produced might provide an inaccurate, partial or incomplete assessment of the reality.

- *Supported* – the capability maturity of the organization is intermediate and the results promise an acceptable degree of quality and reliability, although they might not accurately and completely illustrate the reality.
- M*anaged* – the capability maturity of the organization is high, offering a complete and accurate description of the national state.
- *Proficient* – the capability maturity of the organization is very high, providing a precise and highly reliable reflection of the national reality.

Considering the SDG indicators data production as an atomic process, the levels of maturity can be disaggregated by the dimensions of the model, as shown in Table 1.

Table 1   CMM Maturity Levels

	Basic	Supported	Managed	Proficient
People	Individuals are not well qualified and prepared to cope with the changing demands. The organization lacks a clear strategy and culture, and it is susceptible to external pressures.	Individuals possess the required skills but lack the proper knowledge and training required. Cohesion among roles is missing. The organization is well administered but is vulnerable to internal and external impacts.	Individuals are well equipped to perform and deliver. They have a sense of belonging and team spirit. The organization is properly managed and has strong policies in place.	Individuals are fully committed and well prepared to perform their activities. The organization is strong and independent from external influences.
Processes	Processes are not clear, non-existent, or do not contribute to the business.	There are certain processes to guide the execution of activities but they are not well integrated.	Processes are well managed and contribute to the proper execution of activities.	There are clear, well defined and carefully followed processes across the whole lifecycle.

Table 1 Continued

	Basic	Supported	Managed	Proficient
Technology	The infrastructure is insufficient, outdated or poorly managed, becoming an obstacle for certain activities. There are not enough systems or services to support the production of data.	The infrastructure satisfies the requirements but does not add value to the business. There are systems and services to support the simple tasks but not to facilitate the complex ones.	The infrastructure permits the effective performance of processes and activities. Systems and platforms are integrated and well managed, proactively responding to the business needs.	The infrastructure is appropriate and well managed, boosting the possibilities for improvement and innovation. The systems and services fully meet the needs and leverage the performance.
Data	Most of the data quality dimensions are not met resulting in low-quality results. Data is poorly handled permitting infringements to the key data principles.	Some quality dimensions are met but some of the key dimensions are not. Results quality cannot be guaranteed. Data is acceptably handled and some key principles are respected.	Most quality dimensions are met, including all the key dimensions. The quality of the results is suitable for decision-making. Data is correctly managed and most of the key principles are protected.	All the quality dimensions are respected and the results are expected to be of high-quality. Data is accurately managed ensuring the fulfilment of all the key principles.

However, the SDG indicators data production is the result of complex processes. Although there is not a single deterministic process for statistical organizations, nor would there be a realistic motivation for having such a universal process, the general phases of the statistics lifecycle can be identified.

5.3 Phases

As described in Section 3, there are different tools to support the production of national statistics. Some of them guide, and sometimes prescribe, the sequence of activities NSOs must undertake to produce statistical data. Among them, one of the most widely adopted is the GSBPM, which is currently used by more than 50 statistical organizations and defines and describes the statistical processes these organizations follow. GSBPM was extended and complemented by GAMSO, which incorporated activities that were not thoroughly defined in GSBPM and were considered as overarching processes instead. Although they are defined as non-linear models (i.e. the sub-processes do not have to be followed in a strict order) they present the following limitations associated with the demands of the SDGs and the evolving role of the statistical organizations within the evolving indicators data ecosystem.

- *Responsiveness* – the role that the NSOs play in the monitoring and reporting of indicators for the SDGs varies from country to country and it depends, typically, on existing national legislation and policy mechanisms. While processes like the one prescribed by GSBPM can adapt well to countries with centralized models [8] – where the NSO is a coordinator of all SDG statistical reporting, they do not respond well when decentralized models – where the responsibility for the production, development and dissemination of official statistics is dispersed over many agencies or line ministries and the role of the NSOs is mainly of coordination – are in place.
- *Completeness* – processes like the one given by GSBPM focus mostly on the production of statistical data – from the specifications of needs to the evaluation – but do not consider the full lifecycle of data, ignoring the uptake and impact of the data produced. In order to benefit the human beings, who are at the center of sustainable development, the way the data is consumed and the impact data has in the society are utmost important. Additionally, there should be constant feedback from the stakeholders to the producers of data.
- *Specificity* – all of the instruments identified in the review of the literature cover the production of official statistics in general, but do not consider the particularities demanded by the 2030 Agenda for Sustainable Development. Some of the drivers of the data for the SDGs – like the premise on "leaving no one behind", the focus on certain social groups (mainly on children and youth), and the need for data disaggregated by several

social criterial – demand tailor-made processes that emphasize on the particularities of this public data for the public good.

- *Update* – in the current world of the data – where the volume, sources, and types of data as well as the technology evolve rapidly – traditional processes do not manage well with this evolution. Traditional processes should abandon slow practices and be more agile and dynamic to better cope with the constantly evolving pace set by the data revolution and maximize the opportunities arising from the data revolution for sustainable development.

To address some of these limitations, the Open Data Watch developed a Data Value Chain that targets the needs for gender data of the SDGs but that can be applied to all types of data in the context of the SDGs [46]. This value chain describes the evolution of data from collection to analysis, dissemination, and the final impact of data on decision making. The data value chain presents four major stages: collection, publication, uptake, and impact; which are further separated into twelve steps. Despite its focus on data for the SDGs, such stages do not fully represent the phases that the CMM anticipates for the statistical process of data production for the SDG indicators.

Based on the strengths and limitations identified in the different instruments, the following four general phases that articulate the full data lifecycle were defined:

- *Collection* – comprises the activities involved to collect data which goes from identification and specifications of needs to the actual data acquisition, through the design and construction of tools.
- *Processing* – includes the activities needed to process the data collected and validate the results (analyse, evaluate) until they reach a state where they can be published.
- *Uptake* – contains the activities carried out to promote the usage and consumption of the results achieved, including connecting data to users, incentivizing users to incorporate data into the decision-making process, and influencing them to value data.
- *Impact* – encompasses the activities performed to use data to make an impact and create change. This phase includes activities as actually using the data for decision making, influencing decisions based on the data and reusing or combining the data to create more knowledge.

The resulting CMM is the outcome of cross-cutting the phases with the dimensions and segmenting them according to the maturity levels.
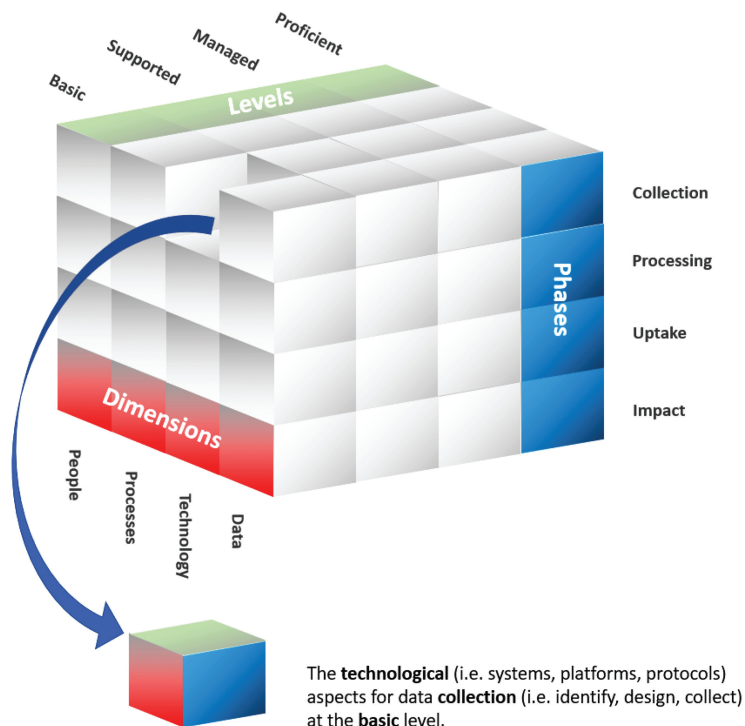
Figure 3   High-Level View of the CMM for the SDG Indicators Data.

Each block in Figure 3, representing the intersection of the three high-level elements of the model, describes the level of maturity of a statistical organization to perform a given phase of data production for the SDG indicators, analysed from a certain organizational dimension.

6 Discussion and Recommendations

There is a clear need for reliable information within the international statistical community, and there have been a number of efforts to ensure quality and accuracy of data. However, the process is long and technology is changing rapidly, directly affecting the lives of human beings and in turn, the data they produce. Therefore, reliability must be safeguarded by robust and mature organizations which are independent of their employees, and the current and future administrations. To this end, NSS must be empowered to quickly and easily adapt to the new reality of data.

International organizations play an important role in supporting countries in being able to produce reliable and efficient indicators, and in providing them with adequate tools for achieving so. For instance, UNECE is making great contributions with the development of GAMSO, GSBPM, GSIM, CSPA. There is space, however, for other organizations to also make a contribution.

Every country, regardless of their advancement and level of development, can benefit from the CMM. While developed countries tend to lead the way and have more resources for improvement and innovation, developing countries can benefit greatly from the efforts and experience gained from those leading the way. The achievement of the global development agenda is not a competition among countries and it depends on every member to be able to achieve its goals and targets. One of the beliefs and principles of the SDGs states that "The United Nations member states work together with a high level of cooperation to improve the circumstances of all people in the world, and place them at the core of future development" [47, p. 1].

The model proposed in this paper is targeted to the SDGs in particular, and to the social indicators for the public good in general. Practices and solutions taken from the private sector have to be analyzed and adapted carefully since their priorities and goals are different. As an example, while developmental indicators pay attention to inclusion (no one should be invisible) and respect for the privacy of individuals and their communities, the private sector solutions may have other priorities.

Other efforts for monitoring social indicators exist and can be taken advantage of. For instance, big investments have been made in improving data for monitoring and accountability for the MDGs. Similarly, UN member states have been reporting for over ten years on human rights in compliance with the Universal Periodic Reviews. All such efforts (and in particular, their results) have to be standardized and considered to develop the synergies that they can facilitate.

Data has to include everyone and has to be useful for everyone. The trend shows that businesses and governments are increasingly relying on big data and the associated analytics. While businesses use big data to inform business decisions and strategy, governments use big data to provide better service delivery and citizen engagement [48]. Initiatives on small data (in which data, instead of being aggregated is processed at the same unit as it was sampled [49]) are also important to make sure nobody is left out. The model proposed in this paper integrates both, the big and the small data approaches to promote inclusiveness.

7 Conclusions

This paper advocates the achievement of the sustainable development agenda through interventions towards improving the capabilities of the entities within the national data ecosystem responsible for monitoring its progress. By the adoption of an organization thinking approach, this research motivates for the adoption and mainstreaming of CMMs within the SDG activities.

The main contributions of this paper are: a thorough definition and the problematization of a space where research and actions are urgently needed, the formulation of a multidimensional prescriptive CMM to assess and improve the maturity of organizations within national data ecosystems, and a set of recommendations towards addressing the challenges within the increasingly data-driven domain of social indicators monitoring. Furthermore, and aiming at reaching globally accepted standards, an extensive review that describes the landscape of the current initiatives on improving social statistics was also presented. This contribution can be helpful for informing statistics institutions of the domain of tools and platforms available. Gaps and overlaps were also identified, and the lack of integration among these efforts, leading to a poor utilization of current and previous investments, was highlighted.

References

[1] "UN Millennium Project | About the MDGs." [Online]. Available: http://www.unmillenniumproject.org/goals/. [Accessed: 31-Jan-2017].

[2] "Sustainable Development Goals | UNDP." [Online]. Available: http://www.undp.org/content/undp/en/home/sustainable-development-goals.html. [Accessed: 31-Jan-2017].

[3] "UNSD – United Nations Statistical Commission." [Online]. Available: http://unstats.un.org/unsd/statcom. [Accessed: 31-Jan-2017].

[4] Thinyane, M. (2016). "Small Data for SDGs Community-Level Action and Indicators Monitoring."

[5] Independent Expert Advisory Group on a Data Revolution for Sustainable Development, "A World that Counts: Mobilising the Data Revolution for Sustainable Development," 2014.

[6] "Metrics & Indicators – Business for 2030." [Online]. Available: http://www.businessfor2030.org/metrics-indicators/. [Accessed: 31-Jan-2017].

[7] High-Level Panel of Eminent Persons on the Post-2015 Development Agenda, "A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development," 2013.

[8] United Nations Economic Commission for Europe (UNECE), "National Mechanisms for Providing Data on Global SDG Indicators," 2018.

[9] Moorosi, N., Thinyane, M., and Marivate, V. (2017). A Critical and Systemic Consideration of Data for Sustainable Development in Africa. In *International Conference on Social Implications of Computers in Developing Countries* (pp. 232–241). Springer, Cham.

[10] Sustainable Development Solutions Network, "Leaving No One Behind: Disaggregating Indicators for the SDGs," 2015.

[11] United Nations, "Fundamental Principles of Official Statistics," 2014.

[12] Robinson, H., and Obuwa, D. (2006). Quality assurance of new methods in National Accounts. *Econ. Trends*, 629, 14–19.

[13] Office for National Statistics United Kingdom, "Framework for National Statistics," 2000.

[14] European Central Bank, "ECB Statistics Quality Framework (SQF)," 2008.

[15] "Monitoring and evaluation | Sustainable Development Goals Fund." [Online]. Available: http://www.sdgfund.org/monitoring-and-evaluation. [Accessed: 02-Mar-2017].

[16] Brancato, G., D'Assisi Barbalace, F., Signore, M., and Simeoni, G. (2017). Introducing a framework for process quality in National Statistical Institutes. *Stat. J. IAOS*, 33(2), 441–446.

[17] Portillo, S., and Moore, K. (2016). "A systematic approach to quality: the development and implementation of a quality management framework in the Central Statistics Office, Ireland," in *European Conference on Quality in Official Statistics*, 1–12.

[18] OECD, "Quality Framework and Guidelines for OECD Statistical Activities," 2012.

[19] International Monetary Fund Statistics Department, (2010). "IMF's Data Quality Assessment Framework," in *Conference on Data Quality for International Organizations*, 1–11.

[20] Dahlgaard-Park, S. M. (2015). "Total Quality Management (TQM)," *SAGE Encycl. Qual. Serv. Econ.*, 808–812.

[21] United Nations Statistics Division, "National Quality Assurance Framework." [Online]. Available: https://unstats.un.org/unsd/dnss/qualitynqaf/nqaf.aspx. [Accessed: 24-Apr-2018].

[22] Dygaszewicz, J., and Szafranski, B. (2016). "Introducing Enterprise Architecture Framework in Statistics Poland," *Comput. Sci. Math. Model.*, 3, 23–32.

[23] Eurostat, "ESS EA Reference Framework," 2015.

[24] Lalor, T., and Gregory, A. (2015). "Common Statistical Production Architecture," in *5th Annual European DDI User Conference*, 1–50.

[25] Koskimäki, T., and Koskinen, V. (2016). "Governmental and Statistical Enterprise Architectures as Tools for Modernizing the National Statistical System," in *European Conference on Quality in Official Statistics*, 1–10.

[26] Elena, S., Aquilino, N., and Pichón Riviére, A. (2014). *Emerging Impacts in Open Data in the Judiciary Branches in Argentina, Chile and Uruguay*, 1–48.

[27] United Nations Development Group, "Common Country Assessment," 2002.

[28] Gómez, L. F., and Heeks, R. (2016). "Measuring the Barriers to Big Data for Development: Design-Reality Gap Analysis in Colombia's Public Sector".

[29] Committee for the Coordination of Statistical Activities and Statistical Office of the European Communities, "Revised International Statistical Processes Assessment Checklist," 2009.

[30] Organisation for Economic Co-operation and Development, "OECD Glossary of Statistical Terms – Statistical standard Definition." [Online]. Available: http://stats.oecd.org/glossary/detail.asp?ID=4920. [Accessed: 26-Jun-2017].

[31] Giacalone, M., and Scippacercola, S. (2016). "Big Data: Issues and an Overview in Some Strategic Sectors," *J. Appl. Quant. Methods*, 11(3), 1–18.

[32] Office for National Statistics United Kingdom, "National Statistics Code of Practice: Statement of Principles," 2002.

[33] United Nations Department of Economic and Social Affairs Statistics Division, "Generic Statistical Information Model (GSIM): Statistical Classifications Model," 2015.

[34] United Nations Economic Commission for Europe Secretariat, "Generic Statistical Business Process Model: GSBPM," 2013.

[35] "Modernisation Maturity Model (MMM) – Roadmap for Implementing Modernstats Standards – UNECE Statistics Wikis." [Online]. Available: http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=12 9172266. [Accessed: 28-Feb-2017].

[36] Paulk, M. C., Curtis, B., Chrissis, M. B., and Weber, C. V. (1993). "Capability Maturity Model for Software, Version 1.1".

[37] Jugdev, K. and Thomas, J. (2002). "Project Management Maturity Models: The Silver Bullets of Competitive Advantage?," *Proj. Manag. J.*, 33(4), 4–14.

[38] De Bruin, T., Freeze, R., Kaulkarni, U., and Rosemann, M. (2005). "Understanding the Main Phases of Developing a Maturity Assessment Model," in *Australasian Conference on Information Systems (ACIS)*, 8–19.

[39] Simonin, T. (2009). "The Holistic Organization Model." p. 4.

[40] International Organization for Standardization, "ISO 9000:2015, Quality management systems — Fundamentals and vocabulary," 2015.

[41] United Nations Economic Commission for Europe (UNECE), "Quality Indicators for the Generic Statistical Business Process Model (GSBPM) – For Statistics derived from Surveys," 2016.

[42] Expert Group on National Quality Assurance Frameworks, "Guidelines for the Template for a Generic National Quality Assurance Framework (NQAF)," 2012.

[43] Expert Group on National Quality Assurance Frameworks, "National Quality Assurance Framework Glossary," 2012.

[44] Pöppelbuß, J., and Röglinger, M. (2011). "What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management," *Ecis*, p. Paper28.

[45] "Introduction to the Modernisation Maturity Model and its Roadmap – Roadmap for Implementing Modernstats Standards – UNECE Statistics Wikis." [Online]. Available: http://www1.unece.org/stat/platform/display/RMIMS/Introduction+to+the+Modernisation+Maturity+Model+and+its+Roadmap. [Accessed: 28-Feb-2017].

[46] Open Data Watch, "The Data Value Chain: Moving from Production to Impact," 1–8, 2017.

[47] "Sustainable Development Goals Beliefs and Principles Agora Portal." [Online]. Available: https://www.agora-parl.org/resources/aoe/sustainable-development-goals-beliefs-and-principles. [Accessed: 05-Jun-2017].

[48] Thinyane, M. (2017). "Investigating an Architectural Framework for Small Data Platforms," in *Proceedings of the 17th European Conference on Digital Government (ECDG 2017), Lisbon, Portugal,* 220–227.

[49] Best, M. (2015). "Small Data and Sustainable Development," in *International Conference on Communication/Culture and Sustainable Development Goals: Challenges for a new generation*, 1–6.

Biographies

Ignacio Marcovecchio is a Computer Scientist with almost fifteen years of experience in project and portfolio management, software development, data management, ICT administration, and multimedia development. His interests are software systems management, continuous improvement and quality assurance, and he has experience with several software development projects. Currently, he is a Senior Research Assistant at the United Nations University Institute in Computing and Society (UNU-CS), where he conducts research in the context of the Small Data Lab focusing in the improvement of quality of the SDG Indicators monitoring. Previously, Ignacio gained international experience from other UN and UNU organizations (UNU-IIST, UNW-DPC, UNU-FLORES) as well as some public and private institutions, all in the domain of research and education, where he played different roles.

Ignacio holds a Bachelor's Degree in Systems Engineering from the National University of Central Buenos Aires, and a Master Degree in Computer Science from the National University of the South in Argentina. He is currently pursuing a PhD in Computer Science at the same institute.

Mamello Thinyane (PhD, 2009, Rhodes University) is passionate about technology innovation and about seeing individuals and communities empowered to lead "their happy" lives. He works within the Small Data Lab at UNU-CS investigating the role of locally-relevant, citizen-generated data to empower individuals and community-level actors towards the Sustainable Development

Goals targets, as well as the role of this data within the larger social indicators data ecosystem.

Mamello is the Chairman of the board of the African Footprints of Hope Organization, an NGO that facilitates strategic multistakeholder engagements towards socio-economic development of communities in Southern Africa. He is also a Visiting Researcher at the Australian Centre of Cyber-Security at the University of New South Wales in Canberra.

Elsa Estevez is a Professor at the National University of the South and at the National University of La Plata and Independent Researcher at the National Scientific and Technical Research Council (CONICET), in Argentina. She was a Senior Academic Program Officer of the United Nations University contributing to digital government research and development. It is the President of the Steering Committee of the International Conference of Theory and Practice of Electronic Governance (ICEGOV) and Associate Editor of Government Information Quarterly (GIQ), Elsevier. She has many publications and has participated in numerous international events dedicated to issues of Digital Governance. Elsa has a PhD in Computer Science from the National University of the South, Argentina.

Pablo Fillottrani is a Professor at the National University of South (UNS) and Independent Researcher at the Commission of Scientific Research (CIC) of the Province of Buenos Aires. He is the Director of the Software Engineering

and Information Systems Laboratory (LISSI) at UNS. He is the Director of the degree programme of Engineering in Information Systems of UNS. He has numerous publications and extensive experience in building human capacity in Computer Science. Pablo has a Ph.D. in Computer Science from the National University of the South.