

Making Data Portals work for SDGs: A view on deployment, design and technology

Geoffrey Greenwell, John Kinyua,
Thilo Klein, Rajiv Ranjan



PARIS21

Partnership in Statistics for
Development in the 21st Century

Discussion Paper No. 8
April 2016



Disclaimer

The opinions expressed in this paper are those of the author and should not be attributed to the PARIS21 partnership or its members. The opinions expressed and arguments employed are those of the authors. Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which PARIS21 works.



Making Data Portals work for SDGs: A view on deployment, design and technology^{*}

Geoffrey Greenwell, John Kinyua, Thilo Klein, Rajiv Ranjan

^{*} We are grateful for comments from colleagues at the PARIS21 Secretariat. In particular we would like to thank Trevor Fletcher, Johannes Jütting, El Iza Mohamedou, Leslie Rae, Marcel Raecker, Gay Tejada and Koffi Zougbede. We would also like to thank representatives from the National Statistical Offices of Ethiopia, Kenya, Rwanda, Tanzania and Uganda.

ABSTRACT

The adoption of the Millennium Development Goals some 15 years ago has led to a surge and proliferation of “data portals” in developing countries. The majority were put in place by international agencies for monitoring purposes as well as to help National Statistical Offices to improve the dissemination of data to a broader public. Based on an analysis of several data portals in a variety of countries, this paper offers lessons on how the deployment, design and technology considerations can be improved as we enter the implementation phase of the Sustainable Development Goals – which will lead to a further push for data portals. The paper finds that while there have been many good intentions to make these portals available to countries, when it comes to the actual implementation the outcomes are rather mixed. One particular problem is the set-up of data portals with overlapping functionalities and their lack of integration – particularly in the most aid-dependent countries. This results in (i) a duplication of workload for already resource-constrained NSOs who have to maintain several portals and update information manually, (ii) confusion for users who consult the various portals with often conflicting results, and (iii) overall high costs for demonstrably low usage of these portals. The paper concludes with recommendations for international agencies and NSOs when deploying new data portals.

KEYWORDS: Data portals, National Statistics Offices, Dissemination, Data Revolution, Sustainable Development Goals

Contents

1. Introduction	4
2. Data portal landscape	8
3. Operating environment	10
3.1. Shift in data dissemination practice.....	10
3.2. Vertical interoperability: Leveraging data portals for reporting.....	11
3.3. Horizontal interoperability: When data portals become burdens	13
4. Financial sustainability	15
4.1. Country demand for portals	15
4.2. Estimated cost of portals	15
4.3. Implementation incentives	16
5. Design and technology considerations	18
6. Recommendations	20
6.1. Reconsider software design.....	20
6.2. Promote standards for integrated reporting.....	21
6.3. Prepare for phased progress.....	21
6.4. Agree on co-ordinated solutions	22
7. Conclusion.....	23
References	24
Annex 1: Review of Data Portals in selected East African countries	26

1. Introduction

In 2014, the UN Secretary General formed an Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) and tasked members to make concrete recommendations on how to catalyse a Data Revolution for sustainable development. The conclusions of this IEAG were published in a report that highlighted the persistent problems of data gaps and inequality. The paper called on governments and other development partners to enable data “to play its full role in the realization of sustainable development by closing these key gaps and allowing greater access and use of data” to inform policy processes (UN Data Revolution Group, 2014, p. 5). The IEAG paper cites examples of how data availability can help analysts identify problems and create an environment for creative change. The benefits of more data are enumerated and include possibilities of targeted interventions to fight malaria, predict and measure price changes due to food supply shocks, target benefits to the most needy and improve agricultural productivity.

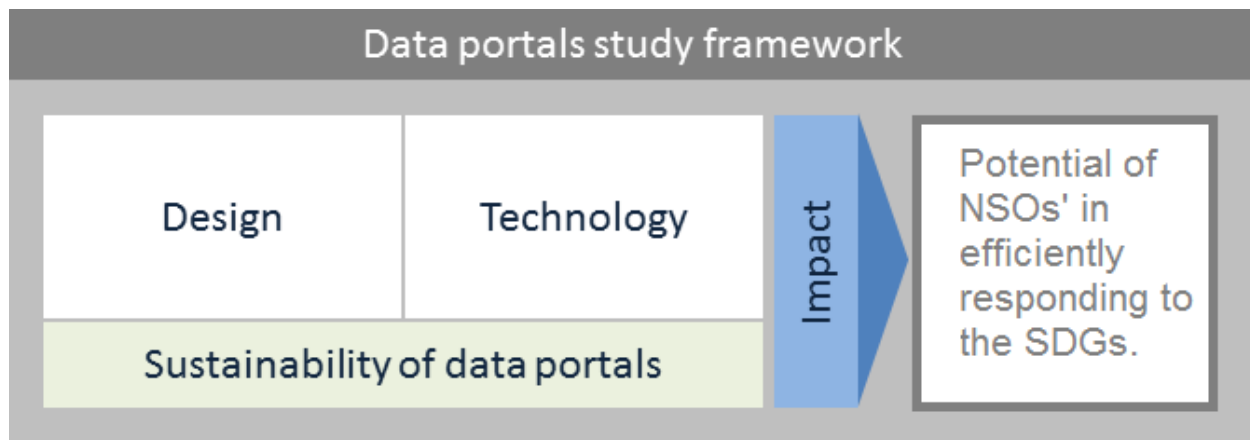
In this context, technological innovations that facilitate the dissemination and delivery of data faster, more efficiently and at low levels of disaggregation are perceived as a means to advance sustainable development and effectively respond to the calls of *leaving no one behind* (Goldstein, 2015). In fact, over the 15 years of monitoring the Millennium Development Goals (MDGs), Information Technology enabled tools have played a significant role in addressing many of the challenges in data dissemination faced by National Statistics Offices (NSOs) (UNDG, 2014 and Stockins, 2013, p. 41) – the primary institutions responsible for producing official statistics at the country level. One of the tools categories used by the NSOs for the purpose of dissemination is the data portal. For the purposes of this paper, a data portal is defined as an NSO’s specific adaptation of generic, stand-alone, web-based and interactive data, and metadata platform, each with a dedicated database system modelled for a specific data type e.g. microdata, aggregate data or geo-spatial data.

The objective of this paper is to look at the sustainability of the current use of data portals and, more specifically, at the design and technological considerations which enable the NSO to deploy a data portal. It evaluates the ability of these portals to help position countries to be more effective in reporting on the Sustainable Development Goals (SDGs). For related work on the review of data portals, the interested reader is referred to two working papers by the World Bank (2014a, 2014b) and a handbook published by UNECA (2011).

The paper employs both quantitative and qualitative research designs. Quantitative research was used to review the NSOs' current data dissemination practices which included a review of the data portals. The qualitative research identified the decision process followed by the NSOs leading to the establishment of the data portals and their contextual suitability. The paper also draws from the Generic Statistical Business Process Model (GSBPM), which provides a useful framework for conceptualizing dissemination processes according to an organic and integrated process.¹ Figure 1 provides a conceptual reference and schematic for describing the approach of the paper. The regional focus is on data portals of NSOs in selected sub-Saharan African countries because this is where the problem of co-existence of data portals is severe (see Table 1 for the number of portals per country). In particular five countries in the region (Ethiopia, Kenya, Rwanda, Tanzania and Uganda) were selected for this study as a great deal of emphasis is placed on developing statistical systems in these countries and there is a large demand for development data by the international community in the region (UN, 2015, p. 10).

Figure 1: Conceptual outline of portal evaluation process

¹ For further information on the GSBPM see: <http://www1.unece.org/stat/platform/display/GSBPM/I.Introduction#I.Introduction-Toc375051192>. The GSBPM Page 7 deals specifically with dissemination products and could be used by NSOs as a conceptual tool.



The analysis in the paper has identified four key recommendations for NSOs and international agencies:

1. **Reconsider software design.**

- Follow an *integrated solutions framework* in portal design. The way forward will be modular solutions with easy integration interfaces, such as APIs, that communicate to different user groups and allow for better end-to-end integration.
- Have a clear understanding and ownership of NSOs' *business processes*. Following the Generic Statistical Business Process Model (GSBPM) facilitates the comparison of statistical processes and allows NSOs to better integrate systems that cross different functions in the organisation.
- Ensure a *user-centred design approach*. This paper finds very little demand or use of existing portals by local users and policy makers. Greater attention should be placed on monitoring the use of portals and promoting them through targeted national user forums.

2. **Promote standards for integrated reporting.** Promote the horizontal exchange of data within the National Statistical System (NSS) by leveraging metadata models, such as SDMX, to establish a coherent framework for transmitting data and metadata.

3. **Prepare for phased progress.** Harmonise data portal instances in the NSS in the short run while preparing for fundamental changes in the long run. As technological requirements change and evolve the focus shifts to leveraging existing technology,

openness and the community of data-users to transform the way NSOs work on other technologies. These long-term changes should be undertaken by a co-ordinating body.

4. **Agree on co-ordinated solutions.** Funding agencies need to broker better designed solutions. This requires improving technical awareness of the challenges that NSOs are facing, helping to generate collaboration between international agencies and regional bodies with competing interests, and applying technology that leads to integration.

The remainder of the paper is organised as follows. Section 2 presents the landscape of data portals reviewed. Section 3 characterises the operating environment in which data portals are run. Section 4 analyses demand, cost and implementation incentives for data portals while Section 5 considers the design and technology behind data portals. Section 6 lays out the recommendations and Section 7 concludes.

2. Data portal landscape

The new development agenda, as defined by the global community and country commitment to the SDGs, provides a new opportunity to address many persistent development problems. The SDGs have laid out a broad political and technical canvas where a transformative agenda is envisaged in an attempt to respond effectively to the growing and complex development challenges of the world. Additionally, national open data initiatives are taking root and demand for open data from civil society organisations is becoming more prevalent. Initiatives such as the Open Government Partnership, Open Data Watch, the International Open Data Charter (IODC) and the Open Data Institute (ODI) are assuring that the Open Data agenda is kept at the forefront of national statistics². Therefore, ensuring the availability and accessibility of usable data by NSOs is emerging as a key theme (PARIS21, 2015a).

In a development context, the NSO is a key player in guaranteeing data availability and accessibility given its position at the centre of the NSS and role as the key producer of official statistics (Kindornay, 2015). Furthermore, there exist crucial opportunities for development if the means of data management and storage are modernised. This is more critical for countries with scarce resources where statistical modernisation is still seminal and where data portals are often the only means to retrieve development data. The Handbook on Major Statistical Data Management Platforms by UNECA (2011) is an important resource for NSOs and can provide guidance during the decision-making process to help NSOs select the appropriate platform for managing and disseminating statistical data to their users. Another useful resource is the World Bank (2014) technical assessment of open data platforms for NSOs. Both papers highlight many of the portals that were reviewed in this study.

Table 1 presents the platforms reviewed in this study broken down by geographical region. A detailed summary of these portals (in the study area) is available in Annex 1.

² See: <http://www.opengovpartnership.org/> and <http://www.opendatawatch.com/>

Table 1: Types and number of data portals established by region

	No. of Countries	Country STAT	DevInfo	IMIS	NADA	OECD. STAT	Prognoz (Open Data Portal) & Knoema	Other portals	Total	Portals per country
Asia	31	9	22	2	9	5	0	6	53	1.71
Eastern Europe	15	1	9	0	0	4	0	8	22	1.47
Latin America	20	1	17	14	8	3	0	2	45	2.25
North America	2	0	0	0	0	1	0	4	5	2.50
Northern Africa & Middle East	21	7	14	0	4	1	8	1	35	1.67
Pacific	11	0	2	2	1	1	0	4	10	0.91
Sub-Saharan Africa	46	32	41	4	29	1	46	5	158	3.43
Caribbean	14	1	8	3	2	0	0	1	14	1.00
Western Europe	27	0	3	0	0	19	0	20	42	1.56
Total	187	51	116	25	53	35	54	50	384	2.05

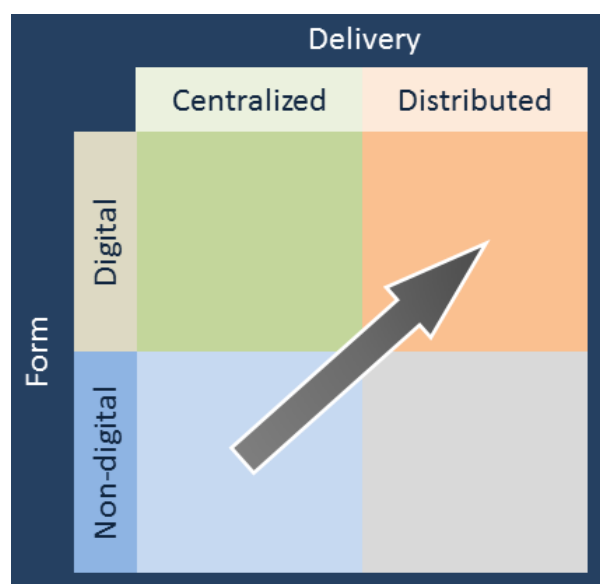
3. Operating environment

3.1. Shift in data dissemination practice

NSOs have been willing to adopt and utilise various data platforms for a number of reasons. Notably, there has been a trend resulting in greater reliance on digital dissemination formats for data. The data dissemination landscape is quickly becoming digital and data are gradually being dispensed in machine-readable formats (in addition to data in PDFs) using the Internet. Figure 2 provides a schematic that depicts the directional vector of dissemination from non-digital, centralised dissemination to distributed, digital means to disseminate data. The following examples explain the four quadrants of Figure 2:

1. Digital/Distributed: Primarily the use of the Internet, used by platforms such as the NADA, to disseminate data by different statistical producers within the NSO.
2. Digital/Centralised: This is largely microdata and PDF type documents (publications) that are made available on CDs/DVDs (optical media) from the NSO.
3. Non-digital/Centralised (traditional): The distribution of printed material at the NSO. Dissemination is usually done by formal request or a physical visit to the NSO and is usually in the form of statistical tables.
4. Non-digital/Distributed: Different statistical producers within the NSS disseminate printed data.

Figure 2: Schematic on data dissemination formats used by NSOs



Most statistical agencies aspire to move from the traditional model (non-digital, centralised) towards a distributed, modernised digital system. Yet many retain legacies in traditional data dissemination, despite digital options, and have mixed systems. Because of the inherent inertia in adapting to new systems, personnel at the NSOs often struggle with the need to retain traditional systems and respond to the new dissemination paradigms depending upon the demands and the habits of the user. In addition to resistance to change from within NSOs, resistance is often caused by external agents and sector statisticians or policy makers that, in many cases, have less support for capacity development than NSOs.

3.2. Vertical interoperability: Leveraging data portals for reporting

The impetus behind the development and implementation of the data portals is the monitoring agencies' need for data and assisting NSOs to disseminate data. Not only does the portal itself challenge technical capacities at NSOs, but there is the reporting obligation that an NSO has as part of the international community. For many NSOs, reporting (which is done in parallel to dissemination and electronically) requires data to be shared directly with specific entities – mostly international agencies – in a predetermined format while dissemination comprises the proactive distribution or transmission of data to the public-at-large, including national policy makers. In the case of the sample study of countries, regular reporting includes the NSO

requirements to report to: the IMF, World Bank, EAC, COMESA, FAO and ILO, each of which has different reporting standards. Some organisations may not be aware of the accessibility of information and data on the portals and so place additional burdens on the reporting agency to report which often creates scenarios of duplicate reporting.

Further evidence of the lack of awareness by users of dissemination portals comes from national policy makers and analysts. NSO management has, as their primary obligation, the responsibility to respond to demands from national policy makers. International reporting, therefore, can become less of a priority than responding to a minister and demands within the NSS. Inter-agency reporting is often conducted informally through collegial “drop-ins” by NSS agents to the NSO as they search for data to fulfil internal demands. New skills need to be acquired both by the user (regionally and nationally) and the producer to proactively use the data portals. The process of building capacity becomes more complicated with increasing choice. As the number of electronic data dissemination platforms increases, and options multiply, NSOs may find that they are unable to cope or rather, in an effort to manage the demands, may assign maintenance to different technical staff and task them in an ad hoc fashion to manage the portal. This is not necessarily part of a well-considered dissemination plan owned by the NSO and can catalyse reporting stove pipes within the NSO. Different staff assigned to reporting may not communicate with each other, unless part of a designated reporting team, which may result in reporting statistical results in a non-harmonised fashion. Many NSOs still do not have data dissemination units that would be in charge of all dissemination related matters.

Unfortunately, data portals are not designed to support reporting as they do dissemination. Easing the reporting burden through the data portals seems to have been an afterthought in portal design. Most global reporting is still conducted using spreadsheets and electronic mail. Adapting data portals to facilitate reporting is certainly an area for improvement and would enhance and increase the utility of data portals by NSOs.

A good example of how to streamline systems is the CountryData platform developed by the United Nations Statistics Division (UNSD), see Box 1.

Box 1: UNSD-DFID project

UNSD-DFID project -- Improving the Collation, Availability and Dissemination of National Development Indicators, including MDGs

The project *Improving the Collation, Availability and Dissemination of National Development Indicators, including Millennium Development Goals* is supported by the UK's Department for International Development (DfID). This project used SDMX and a specific Data Structure Definition (DSD) to enable NSOs to deliver information to the UNSD's CountryData platform automatically. There is no data portal that requires installation in the country. Rather it is fed from an existing portal. This eases the reporting burden and also helps to introduce standards in metadata exchange. The NSO in Rwanda, for example, provided information to the UNSD using the database from its DevInfo portal and the inbuilt SDMX registry to share its MDG data.

This UNSD-DfID project demonstrates the importance of data portals in reducing reporting burdens and improving the quality of reporting of development indicators. The ideal would be to have one data portal reporting data and metadata to different agencies using machine-to-machine communication.

3.3. Horizontal interoperability: When data portals become burdens

Data portals can become real burdens for NSOs and create management overheads on several counts that are often not addressed. First, data portals that are not automated (i.e., not linked to the data production systems and requiring manual data entry) place a heavy burden on NSO staff to manually upload the data. In most cases, the databases of the data portals require a stage of manual data of data that are not dynamically linked. This creates a burden in terms of data entry but also requires sufficient conscientiousness to assure that the data are maintained at regular intervals and accurately uploaded. The manual process of data entry has many disadvantages in terms of speed, accuracy, comprehensiveness and cost (mainly human resource related) and increases the risk of duplicate entries due to failure in quality assurance

of the NSO. Failure to seamlessly connect data digitally at its source (i.e. born digital data) with the dissemination outlets (data portals) raises serious questions about the information architecture. This would require a source portal where other portals could communicate machine-to-machine. Portals should be able to talk to each other as the developers of the portals are largely limited to international organisations that have ample opportunity to co-ordinate using available discussion fora. This is a topic which will hopefully be taken up by the Global Partnership on Sustainable Development Data (GPSDD).

Second, inconsistent reporting is evident between data portals. An indicator can be reported differently in different data portals in the same NSO. In Uganda, for example, the indicator for population size reported for the year 2010 differed in three data portals: CountrySTAT, ODP, and IMIS: the number reported on these portals was: 31 784 600, 30 719 810, and 28 087 685 respectively. Methodologies change and metadata systems are often not well developed to provide clear information on how the indicator was computed. This burden may in part be relieved if the data portals were synchronised with each other (e.g. DevInfo could exchange with the Prognoz portal).

Finally, with limited screen real estate on an NSO website, design issues are not properly taken into consideration. Maintaining a website and maintaining the data portal are two different tasks. NSO home pages appear more as votive sites acknowledging international organisations as they brandish logos of stakeholders and their links to the portal. The home page is often crowded out and real statistical content is lost in the logos.

4. Financial sustainability

This section analyses the country demand for data portals and contrasts this with cost estimates of portal implementation and maintenance. It then explores the drivers behind the observed inundation of portals in some of the world's poorest countries and its implications for the existing portals' financial sustainability.

4.1. Country demand for portals

As previously mentioned, the use of data portals by national and regional officials is limited. Indeed, it is questionable whether the data portal is even demanded by local officials and whether they are the appropriate instrument for encouraging the national use of data. In a survey undertaken by this study, a review of user behaviour on the NSO website was conducted using Google Analytics. On a given month, there were 54 056 visits to the NSO website. Most of the visitors remained on the main web page. The highest number of clicks leaving the main page was to public job offers (6.5% of site visitors). With regard to data demand, 4.37% of site visitors downloaded discrete publications in PDF format as their primary source for data. An evaluation of the traffic using the NSO website to access the portals showed that 0.002% visited the country's DevInfo website; 0.0015% visited the Open Data Portal and 0.001% visited the NADA system. At these current levels of use, this extremely low demand simply cannot justify the investment in data portals and the imposed technical demand and burden on NSOs.

4.2. Estimated cost of portals

Given that high demand for the data portals by local policy makers and academics is not yet evident, the study undertook an evaluation of the investment in data portals over the last 10 years. Estimates for the costs in developing these data portals over this period indicate that over USD 70 million have likely been invested in developing the products and implementing them, including the provision of training. Most of these investments come from international agencies and make a substantial part of USD 450 million in development assistance to statistics annually (PARIS21, 2015b). With an average investment of USD 10 million per system, given the apparent low country demand for the data they supply, the investment in various portals can

be questioned. Countries should be able to eventually bear the costs of the data portal in their national budgets. Since most of these products are intended to provide lower income countries with better data management, we assume that the cost would be borne by the 77 IDA countries. If these solutions were to be covered through annual licensing fees, then each country would have to pay a fee of approximately USD 13 000 per portal per year. With some countries hosting up to five data portals, this translates to an annual cost of USD 65 000 that would have to be borne by the country in order to sustain the portal development process. This is clearly an unsustainable business model.

Estimates for what a country would be willing to pay for a licensing fee are difficult to obtain. In the survey of NSOs, most of the respondents agreed that managing one data portal was sufficient. To estimate the market in the 77 IDA countries, if we assume that the country is willing to manage one data portal and therefore assume the licensing fees of USD 13 000 then the sustainable market for data portals is about USD 1 million per year. However, it is not likely that governments with constrained budgets would be willing to pay this amount. What is certain is that international monitoring agencies cannot finance their development efforts in this market. So, for the moment, it is extremely likely that products will continue to be highly subsidised by external funding and grants. But greater effort can be made to bring down the collective price tag through more collaboration. Examples of such collaboration exist with the IMF and the AIH now collaborating and integrating IMF standards into their ODH initiative. This should continue and broaden.

4.3. Implementation incentives

With very little country demand for portals and substantial costs of implementation and maintenance, it is natural to ask why we currently observe an inundation of portals in some of the world's poorest countries.

With the primary demand for data coming from international monitoring agencies and often not from local policy makers, implementation and training will continue to be funded by the agencies that develop the portal. Soft incentives are provided to countries through training which make the product more appealing. The marginal utility of additional portals implies that

each subsequent portal will tend to encounter more resistance by the country. In order to introduce its solution (i.e. the platform for their data portal), countries are often provided with training incentives, such as participating in regional training events where travel and per diems are offered. These are very attractive to lower paid civil servants seeking to augment their salaries with training allowances. Additionally, in order to spread out the benefit, different data portals can be managed by different people in an NSO therefore increasing the costs of implementation. It is possible for the portals to become vehicles for individual technical persons to gain from their introduction. Personnel training for data portals should follow a more sustainable model where training needs and programmes are assessed by personnel managers and become part of an institutional process rather than opportunistic training offered by international agencies. This means that an NSO should budget for the costs of the training based on institutional priorities and preferences, separating the licensing fees and the training budget and providing for these in annual budgets as reflected in National Strategies for the Development of Statistics (NSDS)³.

³ The NSDS process “is expected to provide a country with a strategy for developing statistical capacity across the entire national statistical system (NSS)” (see: <http://www.paris21.org/NSDS>)

5. Design and technology considerations

Despite the cost of developing the various solutions and the burden they may present to the NSO, the emergence of data portals has proven to be of immense help in bringing the utility and efficiency of online data delivery mechanisms into the spotlight. The DfID-UNSD project mentioned earlier is an example where automating a data portal avoids duplication of work and rationalises the reporting obligations of the NSO. The project has also had a positive impact on improving data quality because of the reuse and extra visibility. As has been demonstrated, these have been heavily financed by international monitoring organisations. The demand for data will most certainly grow as NSOs prepare to respond to the increase in demand for data due to the 2030 Agenda and the Sustainable Development Goals (SDGs). The SDGs will demand greater frequency in reporting, a greater level of disaggregation and greater sub-national relevance, and hence, the importance of the data portal will grow. These platforms and their effective integration need to undergo a new generation of development anticipating more competition and product development and a more conscientious approach primarily by funding agencies. Furthermore, data portals should be developed to include reporting mechanisms so that the reporting burdens of NSOs can be reduced.

NSOs operate according to generic business processes (e.g. GSBPM). The tools and technologies that they use to facilitate these processes, including the use of data portals, should be designed to support the day-to-day operation of the NSOs. Data analysis and data dissemination are two examples of the processes undertaken by the NSO. The NSO may produce tables in standard software such as SPSS, Stata or R. The interface between the analytic and dissemination stages of the business process is now permeable as analytic results can interface with dissemination platforms and avoid manual data entry. Data portals which fail to enhance the productivity of the NSO also fail in the development process. International agencies should provide integrated solutions that are part of a holistic business approach.

It may be argued that, to address the technical peculiarities of different data types such as microdata, geo-spatial data (shape-files) and aggregate data (time-series), different data portals

are needed. However, in many NSOs, multiplicity of data portals, even within the same category of data type, have served only to create redundancies and confusion for the users. Driven by overlapping grouping of statistics such as, 'Census data', 'MDG data' and in some cases 'Sectoral data' (e.g. data on food and agriculture) etc. various data portals, often with conflicting results, have been implemented.

6. Recommendations

Efforts required to implement the data revolution in the context of SDGs at the country level imply addressing the role of NSOs in data dissemination. Data portals sit appropriately in this discourse leveraging the latest advances in technology to deliver on the expectations of data seekers. Driving the shift from manual to digital in data dissemination practices at NSOs data portals have, however, warranted critical examination as the opportunities offered by technological innovations resulting in efficiency gains at NSOs are yet to be fully exploited. In spite of significant investments in data portals by the international agencies, the unsustainable model of data portal deployments poses challenges requiring considerations in both technology and policy realms.

There is no denying that the use of data portals to disseminate data obtained from statistical activities of NSOs is becoming common practice. It has widened the reach of data dissemination and extended data usability options. However, the proliferation of such platforms has not only stretched the capacities of the NSOs but their context oblivious deployment has also contributed to creating redundancies and confusion among data seekers and users.

It is required therefore, that appropriate solutions are established at the intersection of technology and policy to meet the challenges being faced by NSOs in the midst of growing data demand.

6.1. Reconsider software design

Integrated solutions framework: The data-workflows in NSOs are mainly driven by the types of data that they store, requiring distinct considerations regarding dissemination. Therefore, attempts to gain efficiencies through monolithic solutions may prove to be unattainable. Instead, interactive modular solutions that communicate to different user groups are likely to be a more realistic model for end-to-end integration. These modular solutions with easy integration interfaces, such as the APIs, could be a way forward.

Process orientation: NSOs should have a clear understanding and ownership of their business processes. Clearly articulating different steps in the production of official statistics, the GSBPM

provides a good starting point. This approach facilitates the definition and description of statistical processes in a coherent way to compare and benchmark processes within NSOs and supports NSOs in making better decisions, especially, with regard to systems that cut across different functions in the organisation.

User-centred design: The needs of users should be at the heart of the design process whether platforms are built or obtained and customised, ensuring that the use of the system is simple and straightforward. An early and continual focus on users and their tasks, by systematically capturing the needs, is critical for having context aware software solutions. It has been demonstrated that there is very little demand or use of the portals by local users and policy makers. Greater attention should be placed on monitoring the use of portals and promoting them through targeted national user forums. The viability of a product should be based on tangible evidence that portals are contributing to the public discourse on policy and advancing the development agenda by national decision makers.

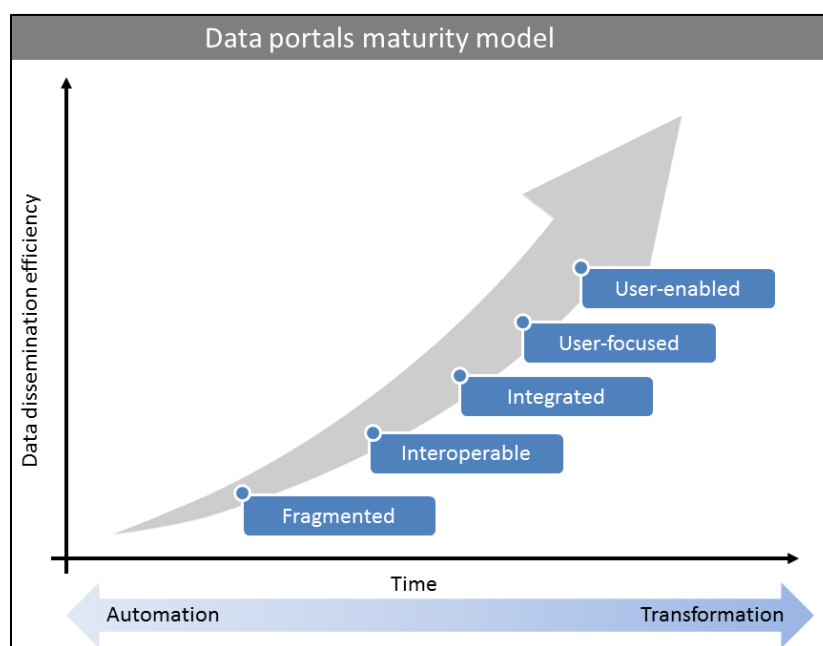
6.2. Promote standards for integrated reporting

Promoting the horizontal exchange of data within the NSS using the model undertaken by the UNSD-DfID project might also prove productive. This leverages metadata models such as SDMX and provides a more coherent framework for transmitting data and metadata.

6.3. Prepare for phased progress

It will be necessary to harmonise data portal instances in the NSOs in the short run while preparing for fundamental changes in the long run as technological requirements change and evolve. Automation and transformation are the part of the development spectrum in which, as the data dissemination matures, the focus shifts from technology on one end to leveraging the same technology, openness and community (data-users and consumers) to transform the way NSOs works on the other (Figure 3). These long-term changes should be undertaken by a co-ordinating body. A particular effort should be made for better co-ordination between the various international agencies and regional bodies introducing tools. This could be proactively pursued within the SDG context and find a forum at the UNSC or the Global Partnership on Sustainable Development Data.

Figure 3: Schematic on data portals maturity model



6.4. Agree on co-ordinated solutions

Solutions available to the NSOs will also likely increase over the short-term and will be available based on funding. Funding agencies should become more technically aware of the challenges that NSOs are facing and help in forming greater collaboration between apparently competing goals that lead to fragmentation and applying technology that leads to integration. The role of funding agencies in brokering better designed solutions is paramount in this context.

7. Conclusion

Data portals will gain renewed attention as a tool to help monitor the progress we are making towards the achievement of the SDGs. While the cost of the deployment of the data portals is borne by international agencies there are quite a few hidden and indirect costs that have to be borne by already very resource strapped NSOs – in particular when it comes to the sustainability of the data portal in the long run. Any data portal should therefore be part of a broader, well-articulated strategy. This strategy should ensure that IT solutions are fully budgeted for through the national budget process, integrated into the capacity development plans of the NSO instead of ad hoc implementations, targeted to local policy makers and researchers, relevant to civil society or the country context, and that they show evidence of their utility by regular and critical demand analysis of dissemination products.

References

- Goldstein, J. (2015). The Data Revolution Has a Software Problem: Measuring the UN Sustainable Development Goals Requires Modern, Flexible Web and Mobile Applications. Blog. Accessed at https://medium.com/@african_minute/the-data-revolution-has-a-software-problem-ff5fa322b010#qwrikv81q
- IMF (2016). Using SDMX 2.0 Web Service. Technical Document, International Monetary Fund. Accessed at <http://datahelp.imf.org/knowledgebase/articles/667675-using-sdmx-2-0-web-service>
- Kindornay, S. (2015). Flashy innovation will not fuel the data revolution. Blog. Accessed at <http://www.scidev.net/global/data/opinion/flashy-innovation-fuel-data-revolution-post-2015.html>
- PARIS21 (2015a). A Road Map for a Country-led Data Revolution. Technical Report, Partnership in Statistics for Development in the 21st Century. Accessed at http://www.oecd-ilibrary.org/development/a-road-map-for-a-country-led-data-revolution_9789264234703-en
- PARIS21 (2015b). Partner Report on Support to Statistics. Technical Report, Partnership in Statistics for Development in the 21st Century. Accessed at <http://www.paris21.org/PRESS2015>
- Stockins, P. (2013). MDG data dissemination using online information systems; Good practices in monitoring and reporting on the Millennium Development Goals: National lessons from Latin America. Technical Report, UN ECLAC. Accessed at http://repositorio.cepal.org/bitstream/handle/11362/4787/LCL3564_en.pdf
- UN (2015). The Millennium Development Goals Report, 2015 (Summary). Technical Report, United Nations. Accessed at http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20Summary%20web_english.pdf
- UN Data Revolution Group (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development. Technical Report, UN Data Revolution Group. Accessed at <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>
- UNDG (2014). UNDG Meeting Decision Sheet. Technical Report, UNDG. Accessed at <https://undg.org/wp-content/uploads/2014/12/Item-5f-Dcision-sheet-DevInfo.pdf>
- UNECA (2011). The Handbook on Major Statistical Data Management Platforms. Technical Report, United Nations Economic Commission for Africa, African Centre for Statistics. Accessed at

http://ecastats.uneca.org/acsweb/Portals/20/Data%20Management/EGM%20HANDBOOK%202011/EDITED%20Handbook_SDMS_final_2-edited-clean.doc

World Bank (2014a). Technical Assessment of Open Data Platforms for National Statistical Organisations. Working Paper, The World Bank. Accessed at <http://documents.worldbank.org/curated/en/2014/10/20451797/technical-assessment-open-data-platforms-national-statistical-organisations>

World Bank (2014b). Open Data Challenges and Opportunities for National Statistical Offices. Working Paper, The World Bank. Accessed at <http://documents.worldbank.org/curated/en/2014/07/20467305/open-data-challenges-opportunities-national-statistical-offices>

Annex 1: Review of Data Portals in selected East African countries

The following platforms were reviewed during the initial phase of this study. This Annex provides a brief description of each platform, identified by the type of data they handle. Furthermore, examples of NSOs' data portals utilizing each type of platform are listed.

Table 2: List of reviewed data portals

	Ethiopia	Kenya	Rwanda	Tanzania	Uganda
NADA	Ethiopian National Data Archive (ENADA) / http://213.55.92.105/enada/index.php/catalog	Kenya National Data Archive (KeNADA) / http://statistics.knbs.or.ke/nada/	Microdata / http://microdata.statistics.gov.rw/	Tanzania National Data Archive (TNADA) / http://nbs.go.tz/tnada/index.php/catalog	Data Archives / http://www.ubos.org/unda/index.php/catalog
Redatam+SP/IMIS	IMIS / http://imis.csa.gov.et/imiseth/	IMIS / http://statistics.knbs.or.ke/binken/RpWebEngine.exe/Portal	N/A	N/A	IMIS / http://ugandadata.org/imis/
Prognoz	N/A (But exists independently as Data Portal Ethiopia / http://ethiopia.africadata.org/)	Kenya Data Portal / http://kenya.africadata.org/	Indicators / http://rwanda.africadata.org or http://indicators.statistics.gov.rw/	N/A (But exists independently as Data Portal Tanzania / http://tanzania.africadata.org/)	N/A (But exists independently as Uganda Statistical Data Portal / http://uganda.africadata.org/)
Knoema	N/A (But exists independently as Ethiopia Data Portal / http://ethiopia.opendataforafrica.org/)	N/A (But exists independently as Kenya Data Portal / http://kenya.opendataforafrica.org/)	N/A (But exists independently as Rwanda Data Portal / http://rwanda.opendataforafrica.org/)	Africa Information Highway / Open Data for Tanzania / http://tanzania.opendataforafrica.org/	Open Data Uganda / Open Data for Uganda / http://uganda.opendataforafrica.org/
DevInfo	EthioInfo / http://192.168.100.13/ethioinfo3/	KenInfo / http://statistics.knbs.or.ke/keninfo/	DevInfoRwanda / http://devinfo.statistics.gov.rw/ or http://devinfo.org/devinforwanda/	Tanzania Socio-Economic Database / http://devinfo.org/tanzania/	UgandaInfo / http://www.ugandadata.org/ugandainfo/
CensusInfo (adaptation of DevInfo)	N/A	N/A	N/A	Census Database / http://www.devinfo.org/CensusInfoTanzania	N/A
CountryStat	CountryStat / http://www.countrystat.org/home.aspx?c=ETH	N/A (But exists independently as http://www.countrystat.org/home.aspx?c=KEN)	CountrySTAT / http://www.countrystat.org/home.aspx?c=RWA	Food and Agriculture Statistics / http://countrystat.org/home.aspx?c=TZA	COUNTRYSTAT UGANDA http://www.countrystat.org/home.aspx?c=UGA

1. NADA (National Data Archive): **Microdata**

This is a web-based cataloguing system that serves as a platform for researchers to browse, search, compare, apply for access, and download relevant census or survey information. It was originally developed to support the establishment of national survey data archives. As of January 2016 there are a total of 84 active online catalogues used by a diverse and growing number of national, regional, and international organisations.⁴ NADA is a tool developed by the International Household Survey Network (IHSN) and has been implemented by the World Bank and PARIS21.

2. Redatam+SP/IMIS: **Hierarchical microdata**

Redatam+SP, the Integrated Multi-Sectoral Information System (Kenya) or the Integrated Management Information System (Ethiopia and Uganda) is a tool for disseminating large volumes of census and survey microdata. Unlike the NADA system, which is designed to provide files in analytic file format such as SPSS or STATA, IMIS has an integrated database that allows for dynamic querying of microdata and supports performing online cross-tabulations down to predetermined geographic levels. This portal is largely supported by the UNFPA.

3. Open Data Portal (or Prognoz): **Aggregate data**

The Open Data Portal (ODP) platform is not designed around microdata but rather delivers aggregate data and indicator data in a user friendly way. The ODP provides tools for visual discovery and advanced analytics utilizing aggregate data. The system's time series analysis functionality allows users to utilise a variety of advanced statistical functions. There is a dashboard designer component which facilitates the creation of customised dashboards. Data visualisations (including maps) can be delivered via email and exported to various file formats including Excel, PDF and HTML. The platform has been deployed in NSOs with support from the African Development Bank through its Africa Information Highway Program (AIH) (<http://africadata.org/>) and the application was developed by Prognoz (<http://www.prognoz.com/>). The ODP is a fully open data platform and distinguishes itself from

⁴ See the IHSN survey catalogue site at <http://adp.ihsn.org/survey-catalogs>

a country-based data portal that delivers official statistics. Both the ODP and country-based data portals in are implemented by the AIH in sub-Saharan Africa.

4. Knoema: **Aggregate Data**

As part of the activities of the African Information Highway Program (AIH), the African Development Bank, in partnership with the International Monetary Fund (IMF), worked on another platform to broadcast and simultaneously transmit data to international institutions such as the IMF. This is known as the Open Data Platform (as opposed to the Open Data Portal provided by Prognoz). After two years of collaboration with the IMF, this relatively new arrival promises to make distribution of data easier by gradually replacing the Prognoz platform. This new development will facilitate the management of the platforms by the AIH and was recently announced by the AfDB at an AIH conference in November 2015.

5. DevInfo: **Aggregate Data**

DevInfo offers aggregate data in tables, graphs and maps. These outputs can be represented as indicators, and organised by units of measurement, disaggregated by subgroups [gender (male/female), location (urban/rural), age groups and others], time periods, sources and geographic areas (national and sub-national levels). The Data Query Service allows for generating API calls, facilitating third-party applications to communicate directly with DevInfo to make queries and retrieve data in various formats⁵. API stands for Application Programme Interface. This is a versatile function that allows software to obtain data from other platforms and is web-based for further use. Data obtained through an API is mainly machine-readable. The multiplicity of data/results formats allows for wider choices in technology selection for further usage.

⁵ An example of an API to query Rwanda DevInfo:
<http://www.devinformatics.gov.rw/di7web/libraries/ws/REST/1/en/JSON/ALL/ALL/ALL/95c25365-a289-43f6-8723-a21ca4a0d9bb>. This information can then be used by the software that issues the call.

Furthermore, the use of standards in data and metadata exchange such as the Statistical Data and Metadata eXchange (SDMX) allow for greater interoperability. The SDMX Registry facilitates access to various web service functions. Data that conform to the structure are accessible to a variety of methods that allow the data to be exchanged (IMF, 2016). Adhering to these standards and data structures opens up data even more.

6. CountrySTAT: **Sector Aggregate Data**

CountrySTAT is a web-based information technology system for food and agriculture statistics at the national and subnational level. In practice, it acts as a one-stop centre which centralises and integrates data coming from various data sources (computed from sources such as agricultural surveys and censuses) and allows it to be harmonised according to international standards while ensuring data quality and reliability. This supports analysis, informed policy making and monitoring with the goal of eradicating extreme poverty and hunger. Through national and regional CountrySTAT projects, the FAO has formed partnerships with statistical offices and the ministries of agriculture, fisheries and forestry, among others, to introduce the system and build the national capacity to use it. In each country, the national government makes a substantial contribution to ensure its deployment and continued training and maintenance.⁶

⁶ It is important to distinguish the level of data which these portals serve. Microdata, or individual record level data derived from administrative systems, censuses or surveys are the basis for portals 1 and 2. The other data portals are effectively reporting aggregate data and require manual data entry into their databases of computed data.

PARIS21

Partnership in Statistics for
Development in the 21st Century

Discussion Paper No. 8
April 2016