

Seminar Component

Name of author: Mahesh Kumar Subedi, PhD

Organization: Central Bureau of Statistics (CBS), Kathmandu, Nepal

Contact address: Agriculture and Livestock Census and Survey section, Central Bureau of Statistics, Kathmandu, Nepal

Contact phone: Tel. +977-1-4245947 (office); +977-9860905752 (mobile)

Email: mk.subedi@gmail.com

Title of Paper:

The National Data Archive System – cataloging surveys for SDG indicators

Abstract

The National Data Archive (NADA) is a web-based survey cataloging system that serves as a portal for researchers to browse, search, compare, access, and download relevant census or survey information. The online catalogue of surveys provides users with access to rich metadata as well as microdata in compliance with dissemination policy of the data producer. This paper is based on a direct observation of existing NADAs published in the national statistical offices (NSO) websites of developing countries, particularly in Asia and the Pacific region. The paper presents the architecture of NADA functionality, takes stock of national surveys and censuses published online in NADA and provides a list of possible socio-economic sustainable development goal (SDG) indicators that could be drawn from the survey repository. The microdata of national surveys and censuses are considered a prime source of data for in-depth research and data integration for producing disaggregated indicators of national priority and international development indicators like the SDGs. The findings show that several SDG indicators can be obtained from the existing surveys and censuses in NADA and concludes that the NADA is one platform that can potentially be considered by countries as a repository of data for accessing data required for estimation of SDG indicators.

Keywords: NADA, survey, census, dissemination, NSO, data portal, indicators, SDGs

Contents

Abstract	1
Introduction	2
Objective	3
Methodology	3
National Data Archive	3
Architecture and functionality	3
List of surveys and censuses	4
SDG indicators and data sources	8
Conclusion	11
References	12

I. INTRODUCTION

Official statistics are considered as public goods and it is a duty of the related government agencies to make it available for public use. The first point of the Fundamental Principles of the Official Statistics also states that the official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information. Generally, the statistics produced by the various agencies of the National Statistics System (NSS) are managed systematically in a standard format for further use by policy makers and researchers. The process of an organized collection, storing and dissemination of statistical data is defined as database management. The databases are mainly built for online compiling, storing, processing, archiving and disseminating data for public use. The databases are also termed as "Data Portals".

The data portal could be multisectoral or for single topic. A common characteristic of the various data portals is that they provide data dashboards on various topics for public access and display indicators in tables and charts (WB 2014). For example, the DevInfo developed by United Nations Development Group (UNDG) for monitoring human development indicators under MDGs; CountrySTAT of FAO for disseminating food and agriculture statistics are some good examples of Data Portals in country and international level. In such, the national data archive (NADA) is an online cataloguing and dissemination system for survey and census microdata and metadata which allows data users to browse, search, access, and download relevant survey and census materials including microdata in compliance with the data dissemination policy of the producer. It is developed by [the International Household Survey Networks \(IHSN\)](#) in 2006 with the main

objectives of promoting open access to and use of data for monitoring the MDGs and it has been implemented by the World Bank and PARIS21 in countries.

With the technical support of the World Bank and the PARIS21, the NADA has been installed in the websites of NSOs of more than 65 countries and a 4000 and above surveys and censuses micro and metadata information are displayed online for public use¹. The IHSN itself has published more than 6600 studies online for the users². It has provision of providing survey microdata online to users based on the data dissemination policy of the host organization.

In this backdrop the paper is based on a direct observation of existing NADAs published in the NSO websites of developing countries, particularly in Asia and the Pacific region. The paper presents the architecture of NADA functionality, takes stock of national surveys and censuses published online in NADA and provides a list of possible socio-economic SDG indicators that could be drawn from the repository.

II. OBJECTIVES

The objective of this paper is to portray the overview and the key features of the National Data Archive (NADA) developed by the IHSN; to take stock of surveys and censuses published in the NSO websites of Asia and the Pacific region, and to provide a list of possible socio-economic SDG indicators under Tier I and Tier II that could be drawn from the survey and census repository of the NADAs.

III. METHODOLOGY

This paper is based on a qualitative research design to review the existing NADAs published in the NSO websites of developing countries, particularly in Asia and the Pacific region. An online observation and review of technical features and functionalities of NADA was conducted. It was taken the stock of national surveys and censuses published online in the country NADAs. And, a list of socio-economic SDG indicators is drawn from the Global SDG portal and their data sources are matched with existing survey and censuses in the NADA repository.

IV. NATIONAL DATA ARCHIVE (NADA)

i. Architecture and functionality

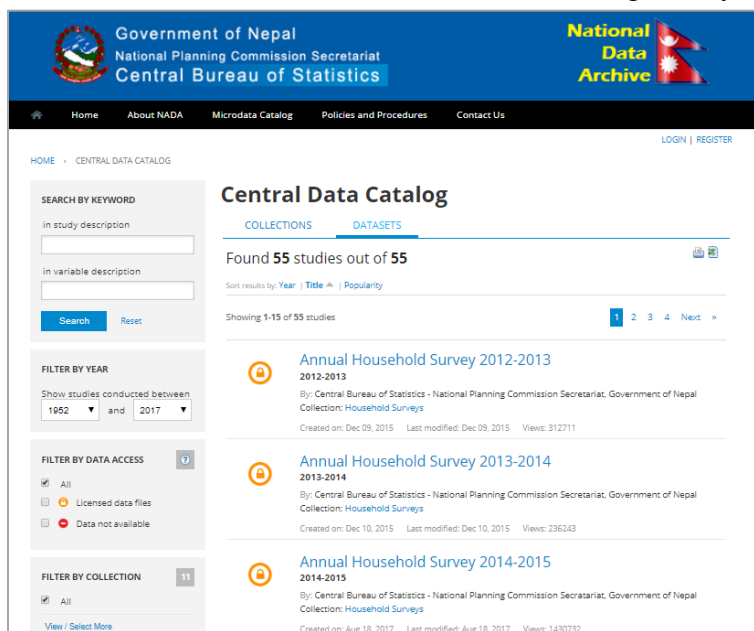
The National Data Archive (NADA) is a web-based survey cataloging system that serves as a portal for researchers to browse, search, compare, access, and download relevant census or survey information. It is developed by International Household Survey Network (IHSN) and established in NSO's website by the Partnership in Statistics for Development in the 21st

¹ <http://adp.ihsn.org/survey-catalogs>,

² <http://catalog.ihsn.org/index.php/catalog>

Century (PARIS21) under Accelerated Data Program (ADP). The portal was established with a noble objective of promoting open access to micro and metadata by establishing national and regional survey data repositories to make existing survey data more accessible to users as well as for monitoring national priority agenda and international development indicators like MDGs and SDGs.

The application is used by several national, regional, and international organizations including NSOs. It uses the Data Documentation Initiative (DDI), a XML- based international metadata standard. The DDI is an international standard for describing surveys, questionnaires, statistical data files in the social, behavioral, economic, and health sciences. Microdata Catalog consists of a number of data collections and documentations such as questionnaire, technical guidebook, report, etc classified into social and economic data collection. It can be browsed and searched centrally or individually as per type of data collection, year of data, type of access, or based on keywords in the survey catalog.



The NADA catalog has four types of data access based on the sensitivity of their content and their inherent disclosure risks. They are: 1) Public use files: the data file has low risk of disclosing personal information and the microdata is downloadable through an online process, 2) Licensed files: The Licensed files are semi-anonymized (direct identifiers are removed) data files characterized by a low disclosure risk. The datasets are available to public use, but it needs a signed agreement between the organization and the trusted users, 3) Data Enclave: data is available on-site in controlled environments, and 4) Data not available: data is not allowed online for download. However, the severity of data access is based on as per the data dissemination policy of the organization.

ii. List of studies (surveys and censuses)

The NADA is used by NSOs of more than 60 developing countries in Africa, Asia and the Pacific, Latin America and the Caribbean and over 4000 national survey and census microdata and metadata information are published for public use. In Asia and the Pacific region about 27 countries have been using the NADAs for publishing more than 1550 national surveys and

censuses (Table 1). Among them microdata of about 817 surveys and censuses are available to researchers once they fulfill the data access conditions set forth in a formal microdata access agreement of the agency.

Table 1: Studies in NADA by country in Asia and the Pacific region

#	Country	Studies (survey and census) by data access type			Total studies
		Public use files	Licensed files	Data not available	
1	Bhutan	1	3	3	7
2	Cambodia	3	30		33
3	Fiji		1		1
4	India	142			142
5	Indonesia		318		318
6	Jordan		1	3	4
7	Lao PDR		2	20	22
8	Mongolia	34		3	37
9	Nepal		21	34	55
10	Philippines		126	57	183
11	Sri Lanka		21	367	388
12	Viet Nam			28	28
13	West bank and Gaza		105	124	229
14	Pacific island countries (14)		9	101	110
	Total	180	637	740	1557

The NADAs of 27 countries in Asia and the Pacific region accommodate household and establishment surveys and censuses in 19 different broad sectors and contain 1557 studies in the national portals. The NADAs of some countries also contain few administrative records related to custom, price and CRVS data. However, the main purpose of the NADA is to publish the household and establishment surveys and censuses. A list of key surveys and censuses published in NSO's NADAs of the Asia and the Pacific region are listed in Table 2 below as follows.

Table 2: Major studies (survey and census) published in NADAs of Asia and the Pacific countries

Name of the study (survey and census)			
1	Income, expenditure and living conditions statistics	13	Agriculture statistics
i	Living standards survey (LSMS)	i	Agriculture census
ii	Living conditions survey	ii	Land and livestock holdings survey
iii	Socio-economic survey	iii	Livestock household survey
iv	Household consumption expenditure survey	iv	Aquaculture and fishery survey
v	Household income and expenditure survey	v	Forestry and plant cultivation survey
vi	Survey on household expenditure on service and durable goods	vi	Forestry survey
vii	Cost of living survey	vii	Horticultural cultivation survey
viii	Poverty survey/ annual poverty indicator survey	viii	Floriculture survey
ix	Expenditure on health/education survey	ix	Corns farm cost survey
2	Labour force and employment statistics	x	Commercial poultry survey
i	Labour force survey	xi	Vegetable crops survey
ii	Child labour survey	xii	Commercial livestock and poultry survey
iii	Employment and unemployment survey	xiii	Cost of production of paddy/ tea survey
iv	Labour turnover survey	xiv	Survey of small tea holdings
v	Work condition survey	xv	Rural agriculture and fishery census
vi	Annual employment survey	xvi	Farm expenditure survey
vii	Occupational wage survey	xvii	Situation analysis survey of agriculture households
viii	Job displacement monitoring survey	xviii	Survey of food demands and agricultural commodities
ix	Seasonal employer program impact evaluation survey	xix	Crops production survey
3	Health statistics	14	Economic and industrial statistics
i	Demographic health survey (DHS)	i	Economic census
ii	Multiple indicator cluster survey (MICS)	ii	Census of (small, medium and large) manufacturing establishments
iii	Social indicator sample survey	iii	Census of industry, trade and service
iv	Mortality survey	iv	Annual manufacturing survey
v	Fertility survey	v	Small and medium industries survey
vi	Health facility survey	vi	Annual survey of industries
vii	Survey of morbidity and health care	vii	Unorganized manufacturing enterprises survey
viii	Survey of health care	viii	Unorganized service sector survey
ix	Survey of maternity, child care and family planning	ix	Informal non-agricultural enterprises survey
x	Child development survey	x	Survey of business and industries

Name of the study (survey and census)			
xi	Reproductive health survey	xi	Informal sector survey
xii	Household health survey	xii	Construction contractors survey
xiii	Population and AIDS indicators survey	xiii	Construction business survey
4	Disability statistics	xiv	Finance and insurance survey
i	Disability prevalence survey	xv	Service sector industries survey
ii	Survey of disabled persons	xvi	Integrated business survey
5	Population and demography	15	Price statistics
i	Population and housing census	i	Consumer price survey
ii	Census of population and dwelling	ii	Retail price survey
iii	Population, housing and establishment census	iii	Wholesale price survey
iv	Inter censal population survey	iv	Farm price survey
v	Housing conditions survey	v	Commodity flow survey
vi	Urban slums survey	vi	Commodity and outlet survey
vii	Demographic sample survey	16	Transport statistics
viii	Youth survey	i	Land transport survey
6	Migration statistics	ii	Transport and storage survey
i	Migration survey	iii	Transport survey of informal sector
ii	Survey of overseas migrants	17	Tourism statistics
iii	Household condition and migration survey	i	Domestic and outbound tourism survey
7	Education statistics	ii	Hotel guests survey
i	Census of government schools	iii	Survey of hotels and lodges
ii	Functional literacy, education and mass media survey	18	ICT statistics
iii	Survey of literacy and culture	i	Survey of information and communication technology
8	Crime and violence statistics	ii	Computer literacy survey
i	Survey of crisis impact early detection survey	iii	Business survey of ICT
ii	Victimization survey	iv	Communication and information survey
iii	Violence survey	v	Computer, internet and mobile survey
iv	Gender based violence survey		
9	Vital statistics	19	Energy and environment statistics
i	Vital statistics of births, deaths and marriages	i	Renewable natural resources census
10	Access to facility statistics	ii	Tropical cyclone impact assessment survey
i	Village infrastructure/facility survey	iii	Drinking water, sanitation, hygiene and housing condition survey
ii	Access to health service survey	iv	Environmental care behavior survey
11	Trade and custom statistics	v	National climate change impact survey
i	Custom statistics	vi	Household energy consumption survey
ii	Import and export statistics	vii	Environmental survey of health care centers

Name of the study (survey and census)			
iii	Internal trade survey	viii	Household environment survey
iv	Distributive trade	ix	Household energy survey
12	Time use survey	x	Health effect and pesticide use survey

V. SDG INDICATORS AND DATA SOURCES

The IAEG-SDGs has classified the SDG indicators into three tiers based on their level of methodological development and the availability of data at the global level. The recently updated tier classification contains 101 Tier I indicators, 91 Tier II indicators, 34 Tier III indicators and 6 indicators that have multiple tiers³. The Table 3 below presents some major SDG indicators under tier I and II indicators, related survey and census as data sources as mentioned in the Metadata repository of the SDG indicators⁴ and the availability of the study in countries' NADAs. Although the table below does not present an exhaustive list of indicators by different dimensions and geographical disaggregation, the 27 countries' NADAs contains related surveys and censuses as data sources for 23 Tier I and 13 Tier II indicators. A careful and comprehensive study of variables of the published survey and census and their links to the SDG indicator could provide more possibilities of positive linkages of the surveys as data sources for more indicators in the list in many dimensions.

Table 3: Major SDG indicators and related studies as data sources

Goal	Indicator	Tier	Data sources (related survey/ census)	Availability of the survey/ census in NADA (Yes/No)
Goal 1: End poverty in all its forms everywhere	1.1.1: Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)	I	1) Household income and expenditure surveys (HIES), 2) Living standards measurement surveys (LSMS) with employment modules, or 3) Labour force surveys (LFS) that collect information on household income	1. Yes 2. Yes 3. Yes
	1.2.1: Proportion of population living below the national poverty line, by sex and age	I	1) HIES, 2) LSMS with employment modules, 3) LFS that collect information on household income	1. Yes 2. Yes 3. Yes
	1.2.2: Proportion of men, women and children of all ages living in	II	1) HIES, 2) LSMS with employment modules,	1. Yes 2. Yes 3. Yes 4. Yes

³ <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>

⁴ <https://unstats.un.org/sdgs/metadata/>

	poverty in all its dimensions according to national definitions		3) Demographic health survey (DHS) 4) Multiple indicator cluster survey (MICS)	
Goal 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture	2.1.1: Prevalence of undernourishment	I	1) Agriculture census/ survey, 2) HIES, 3) LSMS, 4) DHS, 5) Population and housing census (PHC), 6) Time use survey	1. Yes 2. Yes 3. Yes 4. Yes 5. Yes 6. Yes
	2.2.1: Prevalence of stunting among children under 5 years of age	I	1) MICS, 2) DHS, 3) LSMS	1. Yes 2. Yes 3. Yes
	2.2.2: Prevalence of malnutrition among children under 5 years of age, by type (wasting and overweight)	I	1) MICS, 2) DHS, 3) LSMS	1. Yes 2. Yes 3. Yes
Goal 3: Ensure healthy lives and promote well-being for all at all ages	3.1.1: Maternal mortality ratio	I	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
	3.1.2: Proportion of births attended by skilled health personnel	I	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
	3.2.1: Under-5 mortality rate	I	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
	3.2.2: Neonatal mortality rate	I	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
	3.7.1: Proportion of women of reproductive age (aged 15–49 years) who have their need for family planning satisfied with modern methods	I	1) MICS, 2) DHS, 3) Reproductive Health Surveys (RHS), 4) Contraceptive Prevalence Surveys (CPS)	1. Yes 2. Yes 3. Yes
	3.7.2: Adolescent birth rate (aged 10–14 years; aged 15–19 years) per 1,000 women in that age group	I	1) MICS, 2) DHS, 3) RHS	1. Yes 2. Yes 3. Yes
Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities	4.3.1: Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex	II	1) LFS 2) PHC	1. Yes 2. Yes
Goal 5: Achieve gender equality and	5.3.1: Proportion of women aged 20–24 years who were married or in a union before age 15 and before age 18	II	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes

empower all women and girls	5.4.1: Proportion of time spent on unpaid domestic and care work, by sex, age and location	II	1) Time use survey	1. Yes
Goal 6: Ensure availability and sustainable management of water and sanitation for all	6.1.1: Proportion of population using safely managed drinking water services	II	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
	6.2.1: Proportion of population using (a) safely managed sanitation services and (b) a hand-washing facility with soap and water	II	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
Goal 7: Ensure access to affordable, reliable, sustainable and modern energy for all	7.1.1: Proportion of population with access to electricity	I	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
	7.1.2: Proportion of population with primary reliance on clean fuels and technology	I	1) MICS, 2) DHS, 3) PHC	1. Yes 2. Yes 3. Yes
Goal 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	8.3.1: Proportion of informal employment in non-agriculture employment, by sex	II	1) LFS 2) LSMS	1. Yes 2. Yes
	8.5.2: Unemployment rate, by sex, age and persons with disabilities	I	1) LFS 2) LSMS	1. Yes 2. Yes
	8.6.1: Proportion of youth (aged 15–24 years) not in education, employment or training	I	1) LFS 2) LSMS	1. Yes 2. Yes
	8.7.1: Proportion and number of children aged 5–17 years engaged in child labour, by sex and age	II	1) LFS 2) LSMS	1. Yes 2. Yes
Goal 9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	9.2.1: Manufacturing value added as a proportion of GDP and per capita	I	1. CME	1. Yes 2. Yes
	9.2.2: Manufacturing employment as a proportion of total employment	I	1) CME 2) LFS 3) LSMS	1. Yes 2. Yes 3. Yes
Goal 10: Reduce inequality within and among countries	10.1.1: Growth rates of household expenditure or income per capita among the bottom 40 per cent of the population and the total population	II	1) LSMS 2) HIES	1. Yes 2. Yes
	10.2.1: Proportion of people living below 50 per cent of median income, by sex, age and persons with disabilities	II	1) LSMS, 2) HIES	1. Yes 2. Yes
Goal 11: Make cities and human settlements inclusive,	11.1.1: Proportion of urban population living in slums, informal settlements or inadequate housing	I	1) LSMS, 2) HIES 3) PHC	1. Yes 2. Yes 3. Yes

safe, resilient and sustainable	11.2.1: Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities	II	1) LSMS, 2) HIES 3) PHC	1. Yes 2. Yes 3. Yes
Goal 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development	14.4.1: Proportion of fish stocks within biologically sustainable levels	I	1) Aquaculture and fishery survey	1. Yes
Goal 15: Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests,...	15.1.1: Forest area as a proportion of total land area	I	1) Forestry survey	1. Yes
Goal 16: Promote peaceful and inclusive societies for sustainable development, ...	16.2.1: Proportion of children aged 1–17 years who experienced any physical punishment and/or psychological aggression by caregivers in the past month	II	1) DHS 2) MICS	1. Yes 2. Yes
	16.2.3: Proportion of young women and men aged 18–29 years who experienced sexual violence by age 18	II	1) DHS 2) MICS	1. Yes 2. Yes
	16.9.1: Proportion of children under 5 years of age whose births have been registered with a civil authority, by age	I	1) DHS 2) MICS 3) Vital statistics of births, deaths and marriages	1. Yes 2. Yes 3. Yes
Goal 17: Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development	17.3.2: Volume of remittances (in United States dollars) as a proportion of total GDP	I	1) LSMS 2) HIES 3) LFS	1. Yes 2. Yes 3. Yes
	17.8.1: Proportion of individuals using the Internet	I	1) MICS 2) DHS 3) LSMS 4) PHC	1. Yes 2. Yes 3. Yes

VI. CONCLUSION

NADA is an open source web-based cataloging system that serves as a data and metadata portal for household and establishment surveys and censuses, so that users can search, compare, apply for access, and download relevant dataset or information depending on the data access policy of the organization. The metadata in NADA created using Nesstar Publisher comply with the international standard for metadata, DDI (Data Documentation Initiative) and Dublin Core. The

DDI is an effort to create an international standard for describing data from the social, behavioral, and economic sciences. The microdata catalogue in NADA consists of numbers of data collections and documentations such as questionnaires, reports, manuals and code books of each survey classified into social and economic collection.

The study shows, with the technical support of the World Bank and the PARIS21, a 4000 and above national surveys and censuses are published online in NSO's NADA of more than 65 developing countries for public use. Among them, the NADA is available in 27 countries in Asia and the Pacific region with 1557 national surveys and censuses online for user access.

The household surveys and national censuses are considered important sources of data for many SDG indicators. In such, microdata of the national surveys and censuses which are available in country NADAs could provide ample opportunities to researchers for calculating more SDG indicators in many dimensions and geographical disaggregation levels as recommended by the SDG computational methods. A quick observation of the 27 country NADAs from Asia and the Pacific region shows that 23 Tier I and 13 Tier II indicators could easily be calculated from the microdata of published surveys and censuses in the NADAs. A rigorous and an appropriate use of available datasets from various surveys and censuses with advanced analytical methods could provide more opportunities to produce more socio-economic SDG indicators for more dimensions as specified in the SDG indicator framework.

In conclusion, the National Data Archive (NADA) with data and metadata information of national surveys and censuses could potentially serve as a source of data for many socio-economic SDG indicators. However, access to microdata of the surveys and censuses remains to be the biggest challenge. The constraints faced by each country vary significantly largely anchored on the varying level of statistical development as was observed in the 27 countries evaluated for this paper.

VII. REFERENCES

ADP, Survey catalogs (NADA). Web link: <http://adp.ihsn.org/survey-catalogs>

Greenwell, G., Kinyua, J., Klein, T. and Ranjan, R. (2014), Making Data Portals work for SDGs: A view on deployment, design and technology. Discussion paper no. 8, PARIS21, Web link: http://www.paris21.org/sites/default/files/Paper_on_Data_Portals%20wcover_WEB.pdf

Marrakech (2004), *The Marrakech Action Plan for Statistics: Better Data for Better Results An Action Plan for Improving Development Statistics*. Presented to the Second International Roundtable on Managing for Development Results Marrakech, Morocco February 4-5, 2004, Web link: <http://www.mfdr.org/documents/MarrakechActionPlanforStatistics.pdf>

PARIS21 (2011), *Statistics for Transparency, Accountability, and Results: A Busan Action Plan for Statistics*, Web link: https://www.paris21.org/sites/default/files/Busanactionplan_nov2011.pdf

UNECA (2011). The Handbook on Major Statistical Data Management Platforms. Technical Report, United Nations Economic Commission for Africa, African Centre for Statistics. Web link:

http://ecastats.uneca.org/acsweb/Portals/20/Data%20Management/EGM%20HANDBOOK%20011/EDITED%20Handbook_SDMS_final_2-edited-clean.doc

World Bank (2014), Technical Assessment of Open Data Platforms for National Statistical Organisations, Working paper, Web link:

<http://documents.worldbank.org/curated/en/744241468334210686/pdf/928100WP0Techn0Box385378B000PUBLIC0.pdf>

UN (2015), The Millennium Development Goals Report, 2015 (Summary). Technical Report, United Nations. Web link:

[http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20\(July%201\).pdf](http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf)

UN, DESA, Statistics Division, Sustainable Development Goals. Web link:

<https://unstats.un.org/sdgs/>