

Birth of an Ivy+ Web Archive: National Statistical Offices and Central Banks

**North American Coordinating Council
on Japanese Library Resources
November 15, 2018
Princeton University**

Joann Donatiello
Population Research Librarian
Stokes Library
jdonatie@princeton.edu



BIGGEST CHALLENGE:

Changes in how governments publish statistical information

RESULT:

Major gaps in the collection

TO THE RESCUE ...

Ivy Plus Web Collecting/Archiving Program!

Ivy+ Libraries interested in building curated,
thematic collections of freely available, but
at-risk, web content

Brainstorming tool – “half-formed and not-yet-conceptualized
ideas or an idea that is ready to go, but lacking collaborators”

Your name:	Your institution:	Your contact information:	Your idea:	Additional notes:	Interested folks and their contact information:	Proposal submitted on:
Joann Donatiello	Princeton University	jdonatie@princeton.edu/ 609-258-1377	<p>Collect individual country statistical agency web sites that provide valuable summary population data, such as census results, migration figures, and vital statistics. Often, the sites also include figures about the labor force, housing, agriculture, infrastructure, and economics (such as GDP, trade data, etc.) as well. Typically, there is one statistical agency in each country (about 200) that provides access to the majority of statistics about that country. In the past, these agencies produced print publications that were available for purchase and collected by research libraries. For many countries, this is no longer the case as they are increasingly providing this information on the web in the form of interactive sites, excel spreadsheets or PDFs of one or two pages, or links to numerous reports that are too labor intensive to collect systematically. These web pages contain valuable information that is used by library patrons in a multitude of disciplines. Although they are hosted by governments, the sites are at risk for the long term; they are impacted by politics, economic conditions, technological capacity, etc. and, at times, information once available becomes inaccessible. To the best of my knowledge, they are not systematically captured and preserved at the agency level, let alone cataloged with corresponding metadata, and much of this data will be lost to future generations of researchers.</p> <p>Here are some examples of the census information that is available on these sites:</p> <p>http://statsmauritius.govmu.org/English/CensusandSurveys/Pages/census/Census-2000.aspx</p> <p>http://armstat.am/en/?nid=12</p> <p>http://www.geostat.ge/index.php?action=page&p_id=152&lang=eng%20/</p> <p>http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Census?OpenDocument&ref=topBar</p> <p>http://www.cbs.gov.il/census/census/pnimi_sub_page_e.html?id_topic=11&id_subtopic=2</p> <p>As a population research librarian, I collect this type of information regularly and would be able to provide recommendations of sites. I am seeking the endorsement of librarians in other institutions.</p>	Critical project for many disciplines	Bobray Bordelon bordelon@princeton.edu (This project would benefit researchers across the disciplines. With data and statistics disappearing from both developing and developed country websites and the Internet Time Machine not able to capture interactive pages, a systematic approach is needed.	2/14/2018
Gwyneth Crowley	Yale University	gwyneth.crowley@yale.edu	Archiving economic and banking statistics for countries in the Global South that are at a high political risk	We receive many of these statistics in print but some sporadically. It would be a more complete set from the countries' website.	Bobray Bordelon bordelon@princeton.edu	2/14/2018

SCOPE: National Statistical Offices and Central Banks

- The world (about 240 countries)
- No more than two sites for each country
(sometimes only one)

Metadata – What to include?

Priorities:

- Issuing agency/title
- Country name
- Native language version
- Subjects (Library of Congress)

Major heading (LC)	More detailed terms	Definition
Banks and Banking, Central--Country name --Statistics		Use for Central banks
Country name - Economic conditions - Statistics	Macroeconomic indicators	balance of payments, business indicators, capital expenditures, Gross Domestic Product (GDP), trade, imports/exports, currency , wage/price index, cost of living, consumer prices (CPI) , prosperity indexes, standard of living, price indices
	Financial indicators	lending, securities, wealth, taxation
Crime -- Country name -- Statistics		violence, domestic violence, prisons, prisoners, drugs, criminals, corrections, convictions, arrests, police,
Arts -- Country name - Statistics		events, venues, facilities, attendance, expenditures
Sports - Country name - Statistics		events, venues, facilities, attendance, expenditures
Country name - Religion - Statistics		affiliation, attendance
Education - Country name -- statistics		attendance, graduation, facilities , pre-school, primary school, secondary school, university, literacy, training, drop-outs
Housing - Country name -- Statistics		homeless, owned/rented, availability, costs/prices, financing, occupancy, density
Country name -- Population - Statistics	population counts (not a specific census)	
	immigration/migration - statistics	
	population forecasting/population growth	predictions about what the population will be in the future
	life tables	
	life expectancy	
	population distribution by age	
	ethnicity/ancestry - statistics	
Country name -- Census, (Year)		
Country name -- Statistics, Vital		births, deaths, mortality, natality, marriages, nuptiality, divorces,
Country name -- Environmental conditions - Statistics		water, land use, waste, air quality, pollution, climate change, development, conservation, energy use
Transportation - Country name - Statistics		motor vehicles, public transportation, roads, bridges, tunnels, commuting time, number of trips, mobility, road accidents
Country name -- Statistics, Medical		obesity alcoholism, diseases, epidemics, infections, smoking, addictions, causes of death, injuries, abortion, childbirth,

QUESTIONS for the GROUP

- How will you prioritize the collection, especially considering that one country may be more vulnerable than another? Do you have a list of countries that you'd like to start with?
- What kind of metadata do you envision for this collection? Prefer metadata be applied at the website-level (which is how our collections are typically handled), or at a more granular level (i.e. report)?
- Because creators are government agencies, and generally the information is public, is it okay to not send notification letters before collecting these websites?

RESPONSES

Prioritize on:

- Political and technological stability
- Format of the content/unavailable elsewhere
- Begin with 10 sites each

Metadata:

- Collect at the website or agency level, not at the report level

Notification letters:

- Concerns about international copyright laws - “take down” offer

Questions for Ivy+ Librarian:

- What does the technology permit (i.e. what formats will it be able to capture and provide access to in the future)?
- Frequency and depth of the crawling?
- Where should we store all the project documents?

REPLY

- Archive-It technology can easily capture uploaded documents (.pdf, .doc, .xls, etc.), embedded videos (YouTube, Vimeo), images, streaming audio, etc. Struggles with more dynamic content, like databases (and search boxes) that require user interaction, and anything that's Flash-based / JavaScript heavy. There are workarounds for these issues.
- Typically crawl semiannually, more frequently possible
- Crawling drills down every level of a site unless she tells it not to
- Ivy Plus Libraries operates under a 'no questions asked' take-down policy (haven't had any yet)
- Store on Google Drive

SUBMITTER	Country Name	Website	Stat office or Bank	Reasons Chosen	Formats	LC Subject Heading	Site name - Language 1	Site Name - Language 2
BB, MB	Afghanistan	dab.gov.af/	Bank	political instability	1. Documents – Mainly PDF; Some Excel 2. Did not find any real statistics other than those imbedded in reports	Banks and Banking, Central--Afghanistan--Statistics	بانک مرکزی افغانستان	Da Afghanistan Bank
JA	Afghanistan	http://cso.gov.af/en	Statistical Office	Political instability / ongoing conflict		Afghanistan -- Population -- Statistics Afghanistan -- Economic Conditions -- Statistics Afghanistan -- Statistics, Vital	داحصاي مرکزى اداره	Central Statistics Organization
JD	Algeria	http://www.ons.dz/	Statistical Office	numerous one or two page files that are difficult to collect by any other means	Documents: PDFs, Excel files	Algeria -- Economic conditions – Statistics Education – Algeria -- Statistics Housing – Algeria -- Statistics Algeria - Population – Statistics Algeria – Statistics, Medical Medical care – Algeria – Statistics Public health – Algeria -- Statistics Wages – Algeria -- Statistics Labor Supply – Algeria – Statistics Industries – Algeria -- Statistics Social Services – Algeria -- Statistics	الديوان الوطني للإحصائيات	Office National des Statistiques
JD	Armenia	http://www.armstat.am/am/ (Not all data available in English, although English option is available)	Statistical Office	numerous, short PDFs/difficult to collect by any other means/ instability of site	PDFs, HTML, Excel files	Armenia (Republic) -- Economic conditions -- Statistics Industries - Armenia (republic) - Statistics Armenia (Republic)-Census, 2011 Armenia (republic)- Census, 2001 Armenia (Republic)- Population-Statistics Labor supply-Armenia (Republic)-Statistics Arts -- Armenia (Republic) - Statistics Armenia (Republic)—Statistics, Vital Education – Armenia (Republic) -- Statistics Crime – Armenia (Republic) -- Statistics Medical care – Armenia (Republic) -- Statistics Environmental conditions - Statistics Social services – Armenia (Republic) - Statistics Armenia (Republic)-- Statistics, Medical	Հայաստանի Հանրապետության վիճակագրական կոմիտեն (?)	National Statistical Service of the Republic of Armenia
MB	Bangladesh	https://www.bb.org.bd/index.php	Central Bank	Environmental vulnerability; dynamic nature of historical information on site	HTML	Banks and Banking, Central--Bangladesh--Statistics	Central Bank of Bangladesh	

National Statistical Offices and Central Banks Web Archive

Collected by: Ivy Plus Libraries

Archived since: Jul, 2018

Description: The National Statistics Offices and Central Banks Web Archive is a collection developed by librarians at Dartmouth College, the University of Pennsylvania, and Princeton University in partnership (as Ivy Plus Libraries) with Brown University, Cornell University, Duke University, Johns Hopkins University, Harvard University, and Yale University. The collection is curated by James Adams (Dartmouth), Marcella Barnhart (Penn), Bobray Bordelon (Princeton), Gwyneth Crowley (Yale), and Joann Donatiello (Princeton). The Archive aims to preserve country statistical office and central bank websites that often contain ephemeral statistical data on topics such as economic conditions, population, education, housing, environmental conditions, transportation, health, social services, vital statistics, and more.

Subject: Government, Statistics, Banks and banking

Creator: Ivy Plus Libraries

Narrow Your Results

Group

Sort By: **Count** | (A-Z)Banks (14)
Statistical Offices (17)

Subject

Sort By: **Count** | (A-Z)Population (17)
Economic conditions (16)
Banks and Banking (14)
Education (14)
Industries (11)

More ▼

Creator

Sort By: **Count** | (A-Z)Afghānistān Bānk (1)
Algeria. Dīwān al-Waṭanī lil-Iḥṣā'īyāt (1)
Banco Central de Venezuela (1)
Banco de España (1)
Bank al-Markazī al-'Irāqī (1)

More ▼

Language

Sort By: **Count** | (A-Z)English (23)
Arabic (9)
Spanish (5)
French (4)
Ukrainian (2)

More ▼

Country

Sort By: **Count** | (A-Z)Haiti (2)
Iraq (2)
Syria (2)
Ukraine (2)

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Enter search terms here

Search

Clear

Sites

Search Page Text

Page 1 of 1 (31 Total Results)

Sort By: **Title (A-Z)** | Title (Z-A) | URL (A-Z) | **URL (Z-A)**

Title: بانک مرکزی افغانستان, Da Afghanistan Bank

URL: <http://dab.gov.af/>

Captured once on Jul 23, 2018

Subject: Banks and Banking

Group: Banks

Creator: Afghānistān Bānk

Language: Persian, English

Country: Afghanistan

Title: دائرة الإحصاءات العامة, Department of Statistics

URL: <http://dosweb.dos.gov.jo/>

Captured 2 times between Aug 23, 2018 and Aug 23, 2018

Subject: 1994 Census, 2004 Census, 2007 Census, 2015 Census, Transportation, Labor supply, Tourism, Industries, Social services, Medical care, Environmental conditions, Crime, Population, Education, Economic conditions, Vital statistics

Group: Statistical Offices

Creator: Jordan. Dā'irat al-Iḥṣā'āt al-'Āmmah

Language: Arabic, English

Country: Jordan

Title: الجهاز المركزي للإحصاء الفلسطيني, Palestinian Central Bureau of Statistics

URL: <http://pcbs.gov.ps/>

Captured 6 times between Aug 28, 2018 and Aug 29, 2018

Subject: Medical care, Housing, Agriculture, Education, Economic conditions, Population, Vital statistics

Group: Statistical Offices

Creator: Dā'irat al-Iḥṣā' al-Markazīyah al-Filasṭīniyah

Language: Arabic, English

Country: Palestine

Spanish (3)
French (4)
Ukrainian (2)

More ▼

Country

Sort By: Count | (A-Z)

Haiti (2)
Iraq (2)
Syria (2)
Ukraine (2)
Venezuela (2)
Afghanistan (1)
Algeria (1)
Armenia (1)
Aruba (1)
Bahamas (1)
Bahrain (1)
Benin (1)
Burkina Faso (1)
Cuba (1)
Côte d'Ivoire, Mali (1)
Denmark (1)
Djibouti (1)
Equatorial Guinea (1)
Georgia (1)
Guinea Bissau (1)
Hungary (1)
Israel (1)
Jordan (1)
Myanmar (1)
Niger (1)
Palestine (1)
Senegal (1)
Somalia (1)
Spain (1)
Sudan (1)
Taiwan (1)
Togo (1)

Less ▲

Title: Հայաստանի Հանրապետության վիճակագրական կոմիտե, Статистический комитет Республики Армения, Statistical Committee of the Republic of Armenia

URL: <http://www.armstat.am/am/>

Captured once on Sep 17, 2018

Subject: 2001 Census, 2011 Census, Social services, Environmental conditions, Medical care, Crime, Education, Arts, Labor supply, Population, Industries, Economic conditions, Vital statistics
Group: Statistical Offices

Language: Armenian, Russian, English

Country: Armenia

Title: مصرف سورية المركزي, Banque Centrale de Syrie, Central Bank of Syria

URL: <http://www.banquecentrale.gov.sy/>

Captured once on Aug 3, 2018

Subject: Banks and Banking

Group: Banks

Creator: Maṣrif Sūrīyah al-Markazī

Language: French, Arabic, English

Country: Syria

Title: Banco Central de Venezuela

URL: <http://www.bcv.org.ve/>

Captured 5 times between Aug 3, 2018 and Aug 6, 2018

Subject: Banks and Banking

Group: Banks

Creator: Banco Central de Venezuela

Language: Spanish

Country: Venezuela

Title: Bank of the Republic of Haiti, Banque de la République d'Haïti

URL: <http://www.brh.net/>

Captured once on Jul 23, 2018

Subject: Banks and Banking

Group: Banks

Creator: Banque nationale de la Republique d'Haïti

Language: French

Country: Haiti

Title: Centrale Bank van Aruba

URL: <http://www.cbaruba.org/cba/home.do>

Captured 2 times between Aug 28, 2018 and Aug 28, 2018

CHALLENGES

- Mauritius - JavaScript errors
- Bangladesh – navigation errors
- Canada and Czech Republic National Banking sites - too large, working to block ads and unneeded content

MOVING FORWARD

- Add sites to spreadsheet as we find them
- Ivy+ librarian checks periodically
- Original member re-joined us
- 14 sites nominated since inception

National Statistical Offices and Central Banks Web Archive

<https://archive-it.org/collections/10637>

ありがとう

