Clio Infra Website:
Creation, Use and Maintenance.

Michail Moatsos*

May 9, 2017

Abstract

Everything one needs to know for how the new Clio Infra website is created and how to recreate it from scratch if necessary. All the requirements are spelled out, and all materials and software used are open source. The overall approach uses R-coded scripts to fill-in manually made html templates, and produce the static pages for the web server. R code is fed with the material (both data and metadata) available on the Clio Infra dataverse repository. The html layout is manually made using Bootstrap. The demo front end is reachable here. Pending tickets for website and service improvement are available on Github

*michalis.moatsos@iisg.nl, IISG

Contents

1 Introduction

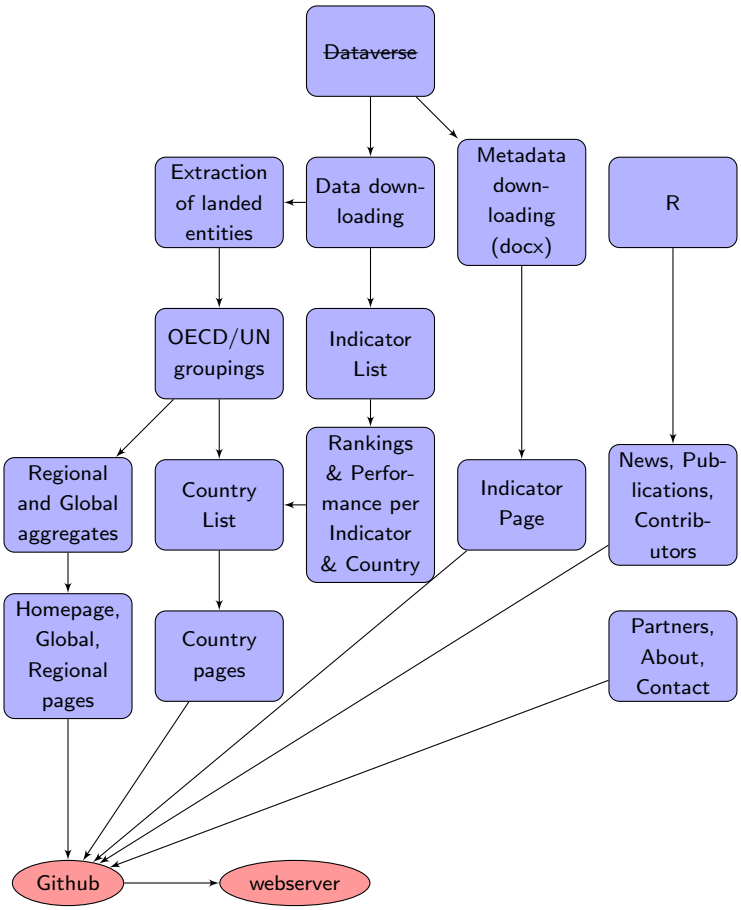
1.1 Built using Bootstrap

The website is developed using Bootstrap version 3.3.7. However, some of the css files have been modified for additional functionalities (such as the back to the top button). Thus, one needs to use the Bootstrap files as they are found on the demo server for smooth operation. For the exact templates developed and used see section 2.

1.2 Workflow

Figure 1 shows the workflow of website creation via the various R scripts and the available data. The Dataverse block is strikedthrough to underlie the fact that in the current version those files need to be fetched manually.

Figure 1: Workflow for an R generated website



1.3 Overview of items generated by the scripts

1. Indicators per country in separate xlsx with long format, created by running CountryHTML_JSON.R, having filename structure:
   CountryName_IndicatorName_TerritorialRef_BorderStart_2012_CCode_XXX.xlsx
   for example: Zambia_LabourersRealWage_TerritorialRef_1964_2012_CCode_894.xlsx
   Exported at folder: /IndicatorsPerCountry.

2. All indicators for a given country in one xlsx with wide and long format, created by running CountryHTML_JSON.R, having filename structure:
   CountryName_AllIndicatorsAvailable_TerritorialRef_BorderStart_2012_CCode _XXX.xlsx
   for example: BosniaandHerzegovina_AllIndicatorsAvailable_TerritorialRef_1992_2012_CCode _70.xlsx
   Exported at folder: /CountryData.

3. (a) A "broad" xlsx per indicator using all "2012 border" countries of the Clio Infra dataset, and (b) a "compact" xlsx per indicator with rows only for countries with available data. Both are created by using IndicatorHTML.R, with both wide and long formats.
   Exported at folder: /data.

4. One html page for each indicator
   Exported at folder: /IndicatorPagesWithMenus.

5. One html page for each country
   Exported at folder: /CountryPagesWithMenus.

6. index.html, indexOECD.html, Partners.html, News_Publications.html located at the same folder as the R scripts.

7. The html files that actually need to be uploaded on the server are located in the folder "ToUpload".

8. An xlsx file containing all indicators for historical entities at "DataAtHistoricalBorders.xlsx".
   Exported at folder: /data.

WARNING: if at any point the xlsx files fail to export in the r scripts replace write.xls with write.xls2. Do note that most of them have been already replaced.

2 Requirements in tools and data

The website is created by using a set of templates and some scripts written in R. The templates are written in html and make use of the css functionalities of Bootstrap.

Those HTML templates are:

- IndexTemplate.html
- IndicatorsTemplate.html
- CountryTemplateJSON.html and CountryTemplateNoMoreVisualsJSON.html
- NewsPublicationsTemplate.html
- PartnersTemplate.html

The R scripts are (in strict order of execution):

- ReadData.R
- IndicatorHTML.R
- CountryHTML_JSON.R
- HistoricalBordersData.R
- AddMenusToIndicatorAndHome.R
- AddCountryMenu.R (only run this after running AddMenusToIndicatorAndHome.R)
- Contributors.R this needs to run only once to create ContributorsList.xlsx
- NewsAndPublications.R
- PartnersPage.R
- FooterSubstitution.R
- some supportive scripts that only need to be called from the other scripts are (*you don't need to call them directly*): IndicatorHTMLSubstitutions.R, LongToClio.R, citations.R, MakeTheGGplot.R, f_IndicatorMenu.R, and f_IndicatorMenu2.R.

Some main texts and external images used in the process are:

- AboutClioInfra.txt
- logo of various participating entities (found in images folder under www)

To run those scripts you need to install the following R libraries: *readxl, xlsx, ggplot2, jsonlite, tidyr, RefManageR.*

Files related to Bootstrap that need to be copied to the server www folder are all placed in the www folder of the Github repository.

Data files that are required (not included in the dataverse repository, but included in the Github repository) are:

- "CIA-Factbook-Countries with notes for their independence status.xls", constructed with information found in: `https://www.cia.gov/library/publications/the-world-factbook/fields/2088.html#af`
- "statcan - countries list and codes from statistics canada.xls", constructed with information found in: `http://www.statcan.gc.ca/eng/subjects/standard/sccai/2011/scountry-desc`
- "UN Countries or areas, codes and abbreviations.xls", constructed with information found in: `http://unstats.un.org/unsd/methods/m49/m49alpha.htm`
- "oecdregions.csv" file containing the OECD split of the world intro regions. Provided by Auke Rijpma.

3 Processing

The data files are not fetched automatically from the dataverse repository in the current version of the R scripts. This will be the case in future versions. Thus the data (and metadata) files need to be manually placed on the proper folder so that R scripts can locate them.

3.1 Metadata files

Until the metadata files on dataverse are updated with new ones, the pre-edited metadata files found on this Github repository must be fed to the R scripts instead.

Because the R script requires the metadata in txt format follow the instructions found in ConvertDOCStoTXT.xls to convert the docx files to txt. For convenience the txt files are also provided in the Github repository.

3.2 The website layouts

On the same folder level as the R scripts all the html layouts (listed in section 1) need to be placed.

3.3 Exporting data

This takes place in

`CountryHTML_JSON.R`

and can be found at the comment:

```
### Now exporting indicator and country files
```

Change the value of ExportData flag to export the files or not. If no change takes place in the data, then to save time considerably set this variable to FALSE.

4 Other pages

4.1 News & Publications page

This page is completely manual. When new items are added then they have to be written in html code within the News_Publications.html page. This could be slightly improved in later versions, but the idea is not to recreate a cms functionality with R.

4.2 Partners page

Some preprocessing to extract the list of contributors is necessary. This is done by Contributors.R script. This needs to run only once to create ContributorsList.xlsx. After that only calling the PartnersPage.R is enough to create the html page. Then of course you need to run the FooterSubstitution.R script to properly set the footer.

5 Adding new countries and indicators

This is to explain how to update the files and scripts when one needs to add an indicator and countries.

First step is to place the new data and metadata files in the proper folders and, in case of a new indicator, convert the metadata doc file that the creator of the dataset has filled to a txt file. To convert the doc file to txt file see ConvertDOCStoTXT.xls for the commands to use in Ubuntu. Also make sure that the data are given in the proper layout (see file ClioLayout.xlsx), otherwise convert them in the proper layout. If this is not possible then additional R programming in necessary to digest the new format.

If the metadata are not given in the format required (see current txt files for the format required) then the process will not conclude. Thus, some manual editing of the metadata file to achieve conformity might be necessary. Currently the template I'm using is slightly different than the one used before. Thus I had to downgrade the current template to achieve conformity with the old one. This means that at some point I need to update the procedure that will allow for the new template to be used.

Second step is that the *ReadData.R* script needs to be run to read the new data in, and then follow and execute the remaining scipts in order of appearance in the R script list of section 2. Before running this script some modifications are necessary. Go to section starting with the comment: "ADD WebName and WebCategory" and add those two variables for the new dataset you want to include. Example:

```
 ClioMetaData$WebName[which(ClioMetaData$title=="Composite Measure of
Wellbeing")] <- "Composite Measure of Wellbeing"
ClioMetaData$WebCategory[which(ClioMetaData$title=="Composite Measure of
Wellbeing")] <- "Demography"
```

For ease of use keep the same WebName as it is given in the original xlsx top cell. If not problems with naming consistency arise among the various tables in the scripts.

Third step requires running the IndicatorHTML.R script. Before running this script you need to upload the new indicator on dataverse to be able to get the citation info. After depositing the new indicator on dataverse you need to download the two types of available citations to the local citation folder[1]. Then to add the bib type of citation you need to either do that manually, or install "sudo apt install bibutils" on an Ubuntu system. To get the command for doing this simply paste the complete short description from dataverse to the third column in " /Clio Infra/Website/Citations.xls", and then export a static version of this in " /Clio Infra/Website/CitationsStatic.xls".[2] Further, you need to add the document filename of the new indicator in the " /Clio Infra/Website/IndicatorsListWithDocFiles.xlsx", and appropriately in " /Clio Infra/Website/IndicatorsGraphType.xlsx"; and in " /PhD/Clio Infra/Website/IndicPriorityList.xlsx" set the priority in selecting the variable for country visualizations.

Important for completing successfully the procedure: the consistent naming of the indicator in various locations is critical. The filename of the xlsx with the original data must end with "-historical.xslx". The prefix (e.g. CompositeWellbeingIndex) should be used as it is in the docx and the txt files. It is preferable that the prefix will also be used in the "ADD WebName and WebCategory" above. Very important: the WebName must be used in the Citations.xlsx list above.

Running ReadData.R will execute in about 2 minutes on a good laptop, and IndicatorHTML.R will do so in approximately 10 minutes (with all flags marked as True). CountryHTML_JSON.R takes about 35 minutes with all flags up.

In the section that begins with "# PARAMETERS:" in scropt CountryHTML_JSON.R you can set the parameters with which the data for the visualizations in the country HTML pages will be selected.

Note that when adding a historical entity a different process needs to be followed in script: AddMenusToIndicatorAndHome.R

To avoid reproducing the historical data files set CreateHistoricalDataFiles to false in

[1] .../Clio Infra/Website/Citations

[2] Note that when working with the citation files from the Composite Wellbeing Index the bibutils command did not work, so I produced the bib file manually so that I can proceed.

script: AddMenusToIndicatorAndHome.R

6 Uploading to the server

6.1 Folder correspondence table

Table 1: Where on server to upload which items from which local folder.

Local Folder	Server Folder (under /var/www)
ToUpload (no subdirectories)	/
ToUpload/CountryPagesWithMenus	/Countries
ToUpload/IndicatorPagesWithMenus	/Indicators
Country data	/docs
JSON	/json
IndicatorsPerCountry	/IndicatorsPerCountry
data	/data
html/graphs	/graphs
Citations	/Citations

Do note that the docx and xlsx files with the original data as deposited by the contributors must also be placed in the /docs folder of the server

6.2 Commands to upload to initial test server

```
ssh michalism@clio2.sandbox.socialhistoryservices.org

scp -r "/var/www/html/theme.css"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp theme.css /var/www/

scp -r "/home/michalis/PhD/Clio Infra/Website/CountryPagesWithMenus"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp CountryPagesWithMenus/* /var/www/Countries/
scp -r "/home/michalis/PhD/Clio Infra/Website/JSON"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp JSON/* /var/www/json/
scp -r "/home/michalis/PhD/Clio Infra/Website/IndicatorsPerCountry"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R IndicatorsPerCountry/ /var/www/

scp "/home/michalis/PhD/Clio Infra/Website/index.html"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp index.html /var/www/
```

```
scp -r "/home/michalis/PhD/Clio Infra/Website/html/graphs"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R graphs/ /var/www/
scp -r "/home/michalis/PhD/Clio Infra/Website/IndicatorPagesWithMenus"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R IndicatorPagesWithMenus/* /var/www/Indicators/
scp -r "/home/michalis/PhD/Clio Infra/Website/Citations"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R Citations/* /var/www/citations/
scp -r "/home/michalis/PhD/Clio Infra/Website/data"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R data/ /var/www/
scp -r "/home/michalis/PhD/Clio Infra/Website/CountryData"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R CountryData/* /var/www/docs/

scp "/var/www/html/index.html"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
scp -r "/var/www/html/images"
michalism@clio2.sandbox.socialhistoryservices.org:/home/michalism
sudo cp -R images/* /var/www/images/
```

7 Appendix

7.1 Functions and Scripts

7.1.1 LongToClio.R

This function converts long format to clio infra format

7.2 MadToClio.R

This function converts the format used in the Maddison project, to that used in the clio-infra website and dataverse. This function was written when adding the Composite Measure of Wellbeing, and it is written around it, thus it is not a generic script yet.

7.2.1 Contributors.R

Exports the list of authors with their homepage and affiliation to ContributorsList.xlsx. The correct path needs to be provided.