2 Downloading the reports/spreadsheets from SciELO Analytics

The SciELO Project provides spreadsheets in a CSV format with the metadata of the articles stored/accessed in its database. These reports can be found at https://analytics.scielo.org/w/reports as ZIP packages that are monthly updated.

There's an easy way to download all the ZIP packages from that link on Linux or any environment with `wget` available: we just need to download that web page with a single "crawl" step, i.e., download all files that has some reference on that page. That can be done with:

```
wget -rcHl1 https://analytics.scielo.org/w/reports
```

It should create several directories, one for each host. The ZIP packages are in the `static.scielo.org/tabs` directory.

As of today, there are 22 links:

- https://static.scielo.org/tabs/tabs_arg.zip
- https://static.scielo.org/tabs/tabs_bol.zip
- https://static.scielo.org/tabs/tabs_bra.zip
- https://static.scielo.org/tabs/tabs_chl.zip
- https://static.scielo.org/tabs/tabs_col.zip
- https://static.scielo.org/tabs/tabs_cri.zip
- https://static.scielo.org/tabs/tabs_cub.zip
- https://static.scielo.org/tabs/tabs_ecu.zip
- https://static.scielo.org/tabs/tabs_esp.zip
- https://static.scielo.org/tabs/tabs_mex.zip
- https://static.scielo.org/tabs/tabs_network.zip
- https://static.scielo.org/tabs/tabs_per.zip
- https://static.scielo.org/tabs/tabs_prt.zip
- https://static.scielo.org/tabs/tabs_pry.zip
- https://static.scielo.org/tabs/tabs_psi.zip
- https://static.scielo.org/tabs/tabs_rve.zip
- https://static.scielo.org/tabs/tabs_rvt.zip
- https://static.scielo.org/tabs/tabs_spa.zip
- https://static.scielo.org/tabs/tabs_sss.zip
- https://static.scielo.org/tabs/tabs_sza.zip
- https://static.scielo.org/tabs/tabs_ury.zip
- https://static.scielo.org/tabs/tabs_ven.zip

The file names follows a `tabs_COLLECTION.zip` structure, that is, the file name suffix before the extension is usually the collection code. There are only 2 exceptions to this rule:

- `tabs_bra.zip`: `Brazil` collection, the first `SciELO` collection which has the `scl` code (legacy code).
- `tabs_network.zip`: All entries from all collection-specific reports together.

These ZIP packages have files with the following names:

- `accesses_by_journals.csv`
- `documents_affiliations.csv`
- `documents_altmetrics.csv`
- `documents_authors.csv`
- `documents_counts.csv`
- `documents_dates.csv`
- `documents_languages.csv`
- `documents_licenses.csv`
- `journals.csv`
- `journals_kbart.csv`
- `journals_status_changes.csv`

The CSV type is the name of the file without its extension, e.g. `documents_counts`.

The specs (in Portuguese) for all the CSV files can be found in http://docs.scielo.org/projects/scielo-processing/pt/latest/public_reports.html but as of today the English-only reader should rely on these notebooks.

For the remaining notebooks, the contents of every file `tabs_COLLECTION.zip` had been extracted on the `tabs_COLLECTION/` directory. On a Linux shell that could be done with this command:

```
for f in tabs_*.zip ; do unzip -d $(basename ${f%%.zip}) $f ; done
```