CGIAR

Platform for Big Data in Agriculture

Findability of gender datasets

Marcelo Tyszler and Ewen Le Borgne

Findability of gender datasets

Marcelo Tyszler and Ewen Le Borgne

About the authors

Marcelo Tyszler[1] and Ewen Le Borgne[1]

[1] Royal Tropical Institute KIT, Netherlands

About the CGIAR Platform for Big Data in Agriculture

The CGIAR Platform for Big data in Agriculture aims at using big data to solve agricultural development problems faster, better and at greater scale. Data has become a valuable global commodity, but it is much more than simply information: in expert hands, it is intelligence.

Already, analysts are finding ways to turn big data — the immense stocks of information collected in computers worldwide — into an invaluable resource for planning and decision-making. It is helping accelerate the development of robust responses to some of the most pressing challenges of our time: climate change/variability, food insecurity and malnutrition, and environmental degradation. The smart and effective use of data will be one of the most important tools for achieving the United Nations' Sustainable Development Goals. Big data represents an unprecedented opportunity to find new ways of reducing hunger and poverty, by applying data-driven solutions to ongoing research for development impact.

About CoP_SED

The Community of Practice on socio-economic data (SED-CoP), led by CIMMYT, aims at bringing together CGIAR centers, academia, not-for-profit research and development organizations and private sector partners willing to tackle major issues related to socio-economic data.

The community works together on strategies to make the data interoperable, in order to enhance the impact and the use of CGIAR-related socio-economic data for partners in development.

This space can be used as a discussion area, share and request relevant information and contribute towards building the community as a whole.

Citation

Marcelo Tyszler and Ewen Le Borgne. 2019. *Findability of gender data sets*. Community of Practice on Socio-economic Data report COPSED-2019-002. CGIAR Platform for Big Data in Agriculture.

License

Creative Commons Attribution License 4.0 International
https://creativecommons.org/licenses/by/4.0/legalcode

Foreword

Late 2018, the CGIAR gender platform through the Royal Tropical Institute KIT received a mini-grant to do a scoping study on the findability of gendered data sets within the CGIAR. The current report presents the findings of this work.

Gender is a very important body of work of CGIAR and it has a lot of potential to increase the richness of the data sets investigated through the Big Data Platform. There are also very interesting challenges inherent to gender data and its tendency to mix qualitative and quantitative research methods.

The Gender Community wants to take a much more active role in the Big Data space to help shape up how gender data can be better analysed and reveal new insights, but also in the process to engage with other socio-economic scientists and even biophysical scientists that may not know how to best engage with gender research and data.

Besides the findings, the R code used for the analysis has been made available through the GitHub repository of the Community of Practice on Socio-economic data. The CoP is dedicated to generating global public goods through open access publications, open data and open source models tolls and scripts.

December, 2019

Gideon Kruseman, community of practice on socio-economic data coordinator

Table of Contents

About the CGIAR Platform for Big Data in Agriculture iii

About CoP_SED iii

Citation iii

License iii

Foreword iv

Table of Contents v

1. Introduction 1

2. Approach 2

3. Initial exploration 3

3.1. Comparing direct search requests 3

3.2. Keywords 4

3.3. Initial insights 5

4. Meta data fields or keywords? 6

5. Broader dataset harvesting 7

References 8

Annexes 9

Annex 1: R-script for the comparison search 9

Annex 2: R-script for the harvesting of gender datasets 14

1. Introduction

Many CGIAR gender scientists have a qualitative research background, including researchers in the agri-food systems research programs. This suggests untapped potential in terms of using qualitative gender data collected through Focus Groups Discussions (FGD), Key Informant Interviews (KII) and other forms of Participatory Action Research (PAR).

However, these datasets vary widely on how data is structured and the detail of documentation provided. By facilitating the tagging, findability and accessibility of quantitative and qualitative gender data we hope to facilitate mixed methods research by providing opportunities for both qualitative and quantitative researchers to exchange insights and create a stronger dialogue.

This is in line with the Big Data Platform approach to use metadata fields to characterize quantitative data. So far, however information related to 'gender' data has been limited.

This pathway proposes the identification of a set of smart 'gender metadata fields' that could ensure that all CGIAR data sets properly assess gender dimensions, which would boost the visibility of gender research.

As a complementary pathway, support to the exchange of gender-focused data collection tools, such as FGD exercises and KII questions, increasing the potential for comparability.

In this seed project we focused on the general findability of gender data / other research data containing strong gender components.

Key outputs from this small grant study are:

a. An initial inventory of CGIAR gender datasets

b. A list of metadata fields that support gender researchers in accessing data

A simple but key take-away is that gender researchers should make more systematic and sharper use of keywords when describing their datasets. Simply adding the keyword 'gender' in the dataset description makes a difference. And there are other keywords that also contribute to indicate the gender focus of a piece of research. Any dataset containing 'sex-disaggregated' data should also indicate so in the keywords.

2. Approach

The approach is highlighted in the following table.

Project objective / result	Activity	Milestone (new end-state, once activity is done)
Inventory of gender data sets	- Contact CGIAR gender community about data sets to investigate - Compile list of data sets - Compile alternative data set lists (for the future)	Clear list of gender data sets used is created. Overview of potential other data sets to explore in the future is created
List of metadata fields to classify data sets is created – and applied (at least in testing)	- Review existing data and metadata recorded - Analyse gaps and opportunities - Draft a recommended list of research metadata fields - Review with gender community the validity of the draft list - Finalise the list	
Metadata suggested are stored in a findable and searchable portal such as CeRES	Make a structure file with metadata mapping	GARDIAN or alternative portal contains metadata suggested
The process of making gender data more comparable is documented to help other socio-economic purposes	- Continually keep track of the process (of problem definition, community mobilisation, engagement, analysis of data sets, sense-making and decisions about which meta data to favour and why etc.) - Write the report	A short process report contains a short overview of the process, key considerations, lessons and recommendations for any team willing to do something similar.

3. Initial exploration

3.1. Comparing direct search requests

CGIAR Collaborative Platform for Gender Platform sent a request for CGIAR gender researchers to provide links to the publicly accessible datasets they knew that comprised a gender component.

An initial inventory was developed on the basis of the inputs provided, in particular by researchers from the International Food Policy Research Institute (IFPRI) and CGIAR Research program on Climate Change, Agriculture and Food Security (CCAFS). We call this the *reference list*. Since most of the datasets provided by CCAFS were actually hosted at the IFPRI Dataverse[1], this reference list was compared with a direct search on the IFPRI Dataverse. The objective of this search is to investigate what a researcher would find if he would simply go to IFPRI Dataverse and search for gender datasets without knowing the reference list. The comparison of both lists provides insight on the findability of gender datasets and how it can be improved. The search was automated by an R-script (see it as annex 1).

We searched the IFPRI Dataverse for datasets using the terms 'gender', 'women' and 'female' and tracked overlaps in search results compared with the reference list. A search query returns a hit if the term is found in the dataset name, author, affiliation, description or keywords. For this purpose, the relevant fields are name, description and keywords.

The table below shows the summary of the exercise:

Keyword	'Gender'	'Women'	'Female'
Hits	57	61	27
Hits also on reference list	47	44	11
Items from reference list not found	14	17	50

This means for example that when we searched for 'gender', we found 57 datasets, 47 being present in the reference list. This means that 10 datasets were found by the search but not provided in the reference list, and that 14 datasets provided in the reference list were not found, because the search words do not appear in the description of the dataset.

Next, we checked how often each dataset provided in the reference list was found. The table below summarizes this:

[1] Dataverse is open source web application to share, preserve, cite, and explore research data. Data hosted in Dataverse can be properly cited and has a persistent identifier, i.e. a fixed link. Dataverse is widely used by the CGIAR centers to host their open data. Some centers use the Harvard Dataverse or have their own instance. An alternative for Dataverse also used by the CGIAR centers is CKAN.

Found	Count
Found in no search	6
Found in 1 of the searches	17
Found in 2 of the searches	29
Found in 3 of the searches	9

This means, for example, that six (6) datasets provided in the reference list were found by none of the search queries, and nine (9) were found in all search queries. The majority was found in 2 out of 3 search queries.

We still need to understand if the datasets found by the searches but not provided in the reference list are indeed gender datasets. Determining this, however, would require contacting the authors of each study, which is not feasible in the scope of this grant. For those studies provided but not found, the simple solution is to add the keyword 'gender' and preferably connecting it to a controlled vocabulary, such as AGROVOC. Controlled vocabularies provide smart ways to organize knowledge, translations, minimize spelling errors and facilitate retrieval.

3.2. Keywords

It was interesting to see the keywords listed by each of the datasets included in the reference list. The keywords were a mix of terms pointing to different topics such as crop, geographical region, methodology, etc. Some are linked to controlled vocabularies like AGROVOC, some not.

From the 61 datasets provided in the reference list, we generated a list of 217 unique keywords. We considered that eight (8) of these keywords suggested gender research content, some very clearly, some less so.

These keywords are:

Keywords	Count
Empowerment	6
gender	37
gender analysis	2
Intrahousehold dinamics	1
WEAI	5
Women	31
Women Empowerment in Agriculture Index (WEAI)	2
women's empowerment in agriculture index	3

Not surprisingly, 'gender' is the most common keyword. There are some keywords which are not consistently used and therefore not really unique, such as the three versions of WEAI.

However, this search shows very few details as to what, about gender, is studied in each project. Studies in other fields (i.e. nutrition) seem to have much more granularity in the description, such as: nutrient intake, nutrition policy, micronutrient deficiencies, etc.

Interestingly, the AGROVOC keyword '*sex-disaggregated data*' is not used by any of the datasets investigated.

3.3. Initial insights

Outcomes of the initial investigation of the metadata reveals that the keyword 'gender' is a very strong predictor in finding a gender dataset. This is not surprising, as: 1) this inventory is not properly published anywhere, therefore making it available increases findability and visibility and 2) 'gender' is not precise enough a keyword to find all relevant gender-focused datasets. An immediate and simple recommendation to all gender researchers would be to systematically use the keyword 'gender' (or 'gender research' if preferable) when describing their gender research datasets when posted on (open) data repositories.

We also found out that 'sex-disaggregated data' was not among the keywords used to describe those datasets. We also recommend including such a keyword (where applicable, naturally) to the dataset description

4. Meta data fields or keywords?

Researchers have had various discussions in the community of practice on Socio-Economic Data (SED-CoP) from the CGIAR Platform for Big Data in Agriculture on this topic. One of their conclusions was that enriching keywords is doable and desirable, while altering meta-data schemes is not. Their advice is therefore to compile a good list of important gender-specific keywords.

Further discussions from the Data Harmonization workshop (in Rome, December 2018) have made it clear that there is a need to make a difference between (Van Wijk et al, 2019):

- *Data made available by gender researchers*: these require gender-related keywords that should be consistently used.
    - This can be study-specific but there should come from an agreed list of gender research specific keywords

- *Data made available by any researcher:* these would still benefit from keywords that is informative to gender researchers in assessing the usability of a dataset for gender research
    - The general advice here is be to have minimal requirements (e.g. sex-disaggregated data specification on the unit of observation), which can unlock extremely powerful potential for interoperability and combination of datasets.

5. Broader dataset harvesting

Based on the initial insights, we also performed a broader automated search on centers' and CGIAR Research Programs' (CRPs) repositories, using 'gender' as search term. Find the code (R-script) for the automated search in the annex 2.

The searches were done on:

- Harvard Dataverse:
  - Africa Rice
  - Bioversity
  - CIAT
  - IFPRI (which includes A4NH and PIM)
  - ICRAF (which includes WLE)
  - IRRI
  - WorldFish
  - *CCAFS (CRP)*
  - *FTA (CRP)*
- Own Dataverse:
  - CIFOR
  - CIMMYT
  - CIP
  - ICARDA
  - ICRISAT
- Own CKAN:
  - IITA
  - ILRI

The results were filtered to remove duplicates. This list resulted in 186 (potential) gender datasets.

We subsequently also retrieved in a similar way datasets with the keyword 'gender' from *gardian.bigdata.org*. The list resulted in 170 (potential) gender datasets. From the 186 datasets found from the search above, 33 were not found in the results from the search at *gardian*. From the 170 datasets extracted from *gardian*, 17 were not found in the previous list with 186 datasets. Most of these 17 datasets where ICARDA datasets not found via a direct search to *data.mel.cgiar.org*, ICARDA listed Dataverse.

We merged the combined list with the list of 14 IFPRI/CCFAS datasets provided but not found by the 'gender' searches. The final list contains 214 (potential) gender datasets, publicly available via the centers' data repositories, including title, permanent link, and data server source. The list is provided as supplementary material.

References

van Wijk M., Alvarez C., Anupama G., Arnaud E., Azzarri C., Burra D., Caracciolo F., Coomes D., Garbero A., Gotor E., Heckert J., Johnson N., Kim S., Miro B., Muliro J., Shikuku K., Tyszler M., Valdivia M., Viviani S., Vrolijk H., Kruseman G. (2019) Towards a core approach for cross-sectional farm household survey data collection: a tiered setup for quantifying key farm and livelihood indicators. *Community of Practice on Socio-economic Data report COPSED-2019-001*. CGIAR Platform for Big Data in Agriculture. https://sites.google.com/view/copsed-shared-space/resources

Annexes

The R script can also be found through the GitHub site of the Community of Practice on socio-economic Data: https://github.com/GideonKruseman/CoP_SED_Rcode

Annex 1: R-script for the comparison search

```
#<%REGION File header%>
#================================================================================
# File      : GenderMeetsBigData_001_comparison_search.R
# Author    : Tyszler, Marcelo <m.tyszler@kit.nl>
# Version   : 1.0
# Date      : July 15, 2019
# Changed   : October, 7 2019 9:35:56 AM
# Changed by: Gideon Kruseman (g.kruseman@cgiar.org)
# Published by: Community of Practice on Socio-economic Data, CGIAR Platform for Big Data in
Agriculture
# Remarks   :
#   This code was developed under the mini-grant provided by Community of Practice on Socio-
economic Data,
#   CGIAR Platform for Big Data in Agriculture to the CGIAR Gender Platform to identify gender
related data sets
#   within the CGIAR.
#
#   This is an R-script for the comparison search using the identified gender data sets at IFPRI
#
# License:
# Creative Commons Attribution License 4.0 International
# https://creativecommons.org/licenses/by/4.0/legalcode
#
# Citation:
# Tyszler, M. 2019 Comparison search: use case IFPRI gender data sets.
GenderMeetsBigData_001_comparison search.R
#================================================================================
#<%/REGION File header%>
library(dataverse)
library(WriteXLS)
Sys.setenv("DATAVERSE_SERVER" = "dataverse.harvard.edu")


## Supporting functions
# this function returns a list of keywords based on a dataset (from dataverse) id
keywords <-function(id) {
```

```
  # load dataset metadata
  ds<-dataset_metadata(id)
  content<-ds$fields$value


  # search for keywords list
  for (i in 1:length(content)){
    sub_content <- content[[i]]
    if ("keywordValue" %in% colnames(sub_content)) {
      list_of_kw <- paste(sort(sub_content$keywordValue$value), collapse = ",")
      return(list_of_kw)
    }
  }
}


# create a data frae with handles provided by IFPRI researchers
ifpri<-data.frame(c("hdl:1902.1/17954",
                    "hdl:1902.1/11189",
                    "hdl:1902.1/11180",
                    "hdl:1902.1/17753",
                    "hdl:1902.1/17801",
                    "hdl:1902.1/17608",
                    "hdl:1902.1/17408",
                    "hdl:1902.1/17357",
                    "hdl:1902.1/17531",
                    "hdl:1902.1/19058",
                    "hdl:1902.1/17606",
                    "hdl:1902.1/17082",
                    "hdl:1902.1/17045",
                    "hdl:1902.1/15580",
                    "hdl:1902.1/15640",
                    "hdl:1902.1/17079",
                    "hdl:1902.1/15646",
                    "hdl:1902.1/17988&version=6.1",
                    "hdl:1902.1/19160",
                    "hdl:1902.1/19236",
                    "hdl:1902.1/19237",
                    "hdl:1902.1/21266",
                    "doi:10.7910/DVN/27857",
                    "doi:10.7910/DVN/27704",
                    "doi:10.7910/DVN/27883",
                    "doi:10.7910/DVN/26930",
```

```
                    "doi:10.7910/DVN/28558",
                    "doi:10.7910/DVN/MUOX19",
                    "doi:10.7910/DVN/29015",
                    "doi:10.7910/DVN/DH1O3J",
                    "doi:10.7910/DVN/YW4WIT",
                    "doi:10.7910/DVN/AXGCHT",
                    "doi:10.7910/DVN/T9GGYA",
                    "doi:10.7910/DVN/RN40SP",
                    "doi:10.7910/DVN/0R5WTU",
                    "doi:10.7910/DVN/KUSXJR",
                    "doi:10.7910/DVN/BXSYEL",
                    "doi:10.7910/DVN/5CXCLX",
                    "doi:10.7910/DVN/ODARXH",
                    "doi:10.7910/DVN/DKURGR",
                    "doi:10.7910/DVN/LT631P",
                    "doi:10.7910/DVN/FO8WDU",
                    "doi:10.7910/DVN/PUK1P7",
                    "doi:10.7910/DVN/JJJBQ0",
                    "doi:10.7910/DVN/UP7WQ2",
                    "doi:10.7910/DVN/FSMCTQ",
                    "doi:10.7910/DVN/JWMCXY",
                    "doi:10.7910/DVN/LEP9KF",
                    "doi:10.7910/DVN/K5NSAF",
                    "doi:10.7910/DVN/GI0TEC",
                    "doi:10.7910/DVN/FOYZBL",
                    "doi:10.7910/DVN/AORZAU",
                    "doi:10.7910/DVN/XNAHHB",
                    "doi:10.7910/DVN/OWOETW",
                    "doi:10.7910/DVN/MP1KRD",
                    "doi:10.7910/DVN/DXMARV",
                    "doi:10.7910/DVN/BP23OB",
                    "doi:10.7910/DVN/VA2MER",
                    "doi:10.7910/DVN/JU7QP6",
                    "doi:10.7910/DVN/CBVLK5",
                    "doi:10.7910/DVN/RBW801"),
                  stringsAsFactors = FALSE
)


colnames(ifpri)<-"id"


## Search:
```

```
# retrieve datasets with the word "gender" in a metadata field and collect keywords
df1<-dataverse_search("gender", type = "dataset", subtree = "IFPRI", per_page = 100)
a1<-df1[c("name","global_id")]
a1$keywords<-""
for (i in 1:nrow(a1)) {
  print(a1$name[i])
  a1$keywords[i]<-keywords(a1$global_id[i])
}


#export
WriteXLS(a1, "temp1.xlsx", Encoding = "latin1")


# retrieve datasets with the word "women" in a metadata field and collect keywords
df2<-dataverse_search("women", type = "dataset", subtree = "IFPRI", per_page = 100)
a2<-df2[c("name","global_id")]
a2$keywords<-""
for (i in 1:nrow(a2)) {
  print(a2$name[i])
  a2$keywords[i]<-keywords(a2$global_id[i])
}



#export
WriteXLS(a2, "temp2.xlsx", Encoding = "latin1")


# retrieve datasets with the word "female" in a metadata field and collect keywords
df3<-dataverse_search("female", type = "dataset", subtree = "IFPRI", per_page = 100)
a3<-df3[c("name","global_id")]
a3$keywords<-""
for (i in 1:nrow(a3)) {
  print(a3$name[i])
  a3$keywords[i]<-keywords(a3$global_id[i])
}


#export
```

```
WriteXLS(a3, "temp3.xlsx", Encoding = "latin1")



###
# retrieve keyword of the pre-listed datasets
ifpri$keywords<-""
for (i in 1:nrow(ifpri)) {
  print(ifpri$id[i])
  ifpri$keywords[i]<-keywords(ifpri$id[i])
}


##
# using temp1.xlsx-temp4.xlsx it is possible to compare the findability of the datasets


#============================   End Of File   ================================
```

Annex 2: R-script for the harvesting of gender datasets

```
#<%REGION File header%>
#================================================================================
# File      : GenderMeetsBigData_001_comparison_search.R
# Author    : Tyszler, Marcelo <m.tyszler@kit.nl>
# Version   : 1.0
# Date      : July 15, 2019
# Changed   : October, 7 2019 9:35:56 AM
# Changed by: Gideon Kruseman (g.kruseman@cgiar.org)
# Published by: Community of Practice on Socio-economic Data, CGIAR Platform for Big Data in
Agriculture
# Remarks   :
#   This code was developed under the mini-grant provided by Community of Practice on Socio-
economic Data,
#   CGIAR Platform for Big Data in Agriculture to the CGIAR Gender Platform to identify gender
related data sets
#   within the CGIAR.
#
#   This is an R-script for the gender datasets across the CGIAR data repositories
#
# License:
# Creative Commons Attribution License 4.0 International
# https://creativecommons.org/licenses/by/4.0/legalcode
#
# Citation:
# Tyszler, M. 2019 Dataset Harvester: searching for gender datasets at the CGIAR.
GenderMeetsBigData_002_dataset_harverster.R
#================================================================================

library(dataverse)
library(WriteXLS)
library(dplyr)
Sys.setenv("DATAVERSE_SERVER" = "dataverse.harvard.edu")


# Loop over dataserve repositories
for (subtree in c("IFPRI",
                  "AfricaRice",
                  "Bioversity",
                  "CIAT",
                  "RiceResearch",
```

```
                  "ICRAF",
                  "worldfish",
                  "CCAFSbaseline",
                  "crp6")) {

  print (subtree)
  # read datasets with the word "gender" in a metadata field
  df1<-dataverse_search("gender", type = "dataset", subtree = subtree, per_page = 100)

  # select
  selection <-df1[c("name","url")]
  selection$server <- "dataverse.harvard.edu"
  selection$subtree <- subtree

  #combine
  if (exists("complete_selection")) {
    complete_selection <- rbind(complete_selection, selection)
  } else {
    complete_selection <- selection
  }
}


# loop over Dataverse repositories which require an authentication key
for (source in c("data.cifor.org",
                 "data.cimmyt.org",
                 "data.cipotato.org",
                 "dataverse.icrisat.org",
                "data.mel.cgiar.org")) {

  print (source)
  # read datasets with the word "gender" in a metadata field
  if (source == "data.cipotato.org") {
    key = "USE_YOUR_OWN_KEY"
    df1<-dataverse_search("gender", type = "dataset", server = source, per_page = 100, key = key)

  } else  if (source == "data.cifor.org") {
    key = "USE_YOUR_OWN_KEY "
    df1<-dataverse_search("gender", type = "dataset", server = source, per_page = 100, key = key)

  } else  if (source == "dataverse.icrisat.org") {
    key = "USE_YOUR_OWN_KEY "
    #df1<-dataverse_search("gender", type = "dataset", server = source, per_page = 100, key = key)
```

```
    out <- jsonlite::fromJSON("http://dataverse.icrisat.org/api/search?q=gender&key=
USE_YOUR_OWN_KEY
&type=dataset&sort=name&order=asc&per_page=100&show_relevance=FALSE&show_facets=FALSE")
    n_total <- ngettext(out$data$total_count, "result", "results")
    message(sprintf(paste0("%s of %s ", n_total, " retrieved"),
                    out$data$count_in_response, out$data$total_count))

    df1<-out$data$items

  } else  if (source == "data.mel.cgiar.org") {
    #df1<-dataverse_search("gender", type = "dataset", server = source, per_page = 100, key = key)
    out <-
jsonlite::fromJSON("http://data.mel.cgiar.org/api/search?q=gender&type=dataset&sort=name&order=asc
&per_page=100&show_relevance=FALSE&show_facets=FALSE")
    n_total <- ngettext(out$data$total_count, "result", "results")
    message(sprintf(paste0("%s of %s ", n_total, " retrieved"),
                    out$data$count_in_response, out$data$total_count))

    df1<-out$data$items
  } else {
    df1<-dataverse_search("gender", type = "dataset", server = source, per_page = 100)
  }

  # select
  selection <-df1[c("name","url")]
  selection$server <- source
  selection$subtree <- ""

  #combine
  if (exists("complete_selection")) {
    complete_selection <- rbind(complete_selection, selection)
  } else {
    complete_selection <- selection
  }

}


## Cases where CKAN is used as data repository:
#ILRI
print ("data.ilri.org")
out <- jsonlite::fromJSON("http://data.ilri.org/portal/api/3/action/package_search?q=gender")
n_total <- ngettext(out$result$count, "result", "results")
```

```
message(sprintf(paste0("%s of %s ", n_total, " retrieved"),
                out$result$count,out$result$count))
selection <-out$result$results %>% select(title, url)
colnames(selection)<-c("name","url")
selection$server <- "data.ilri.org"
selection$subtree <- "CKAN"


#combine
if (exists("complete_selection")) {
  complete_selection <- rbind(complete_selection, selection)
} else {
  complete_selection <- selection
}


#IITA
print ("data.iita.org")
out <- jsonlite::fromJSON("http://data.iita.org/api/3/action/package_search?q=gender")
n_total <- ngettext(out$result$count, "result", "results")
message(sprintf(paste0("%s of %s ", n_total, " retrieved"),
                out$result$count,out$result$count))
selection <-out$result$results %>% select(title, identifier)
colnames(selection)<-c("name","url")
selection$server <- "data.iita.org"
selection$subtree <- "CKAN"


#combine
if (exists("complete_selection")) {
  complete_selection <- rbind(complete_selection, selection)
} else {
  complete_selection <- selection
}
# remove duplicates:
complete_selection<-distinct(complete_selection,url, .keep_all = TRUE)


#export
WriteXLS(complete_selection, "Gender_Inventory_Draft.xlsx", Encoding = "latin1")
```