# setwd("~/git/urv-sdgs-tracker")

setwd("C:/Users/Lluis Salvat/git/urv-sdgs-tracker")

library(readr)

library(dplyr)

library(jsonlite)

# Load and setup main docs ----

degree\_programs\_list <- read\_delim("data/degree\_programs\_list.csv",

delim = ";", escape\_double = FALSE, locale = locale(encoding = "WINDOWS-1252"),

trim\_ws = TRUE)

course\_details\_list\_guido <- read\_csv("data/course\_details\_list\_guido.csv")

course\_details\_list\_docnet <- read\_csv("data/course\_details\_list\_docnet.csv")

colnames(course\_details\_list\_docnet) <- c("web\_scraper\_order", "degree\_url",

"course\_delivery\_mode", "course\_code", "course\_name","course\_url\_first\_join",

"course\_period", "course\_type", "credits", "year")

#Load and replace the first course\_url for iframe\_url in Docnet courses

docnet\_iframes <- read\_delim("data/docnet-iframe.csv",

delim = ";", escape\_double = FALSE, locale = locale(encoding = "WINDOWS-1252"),

trim\_ws = TRUE)

docnet\_iframes\_subset <- docnet\_iframes[, c("web-scraper-start-url", "iframes")]

colnames(docnet\_iframes\_subset) <- c("course\_url\_first\_join","iframes")

course\_details\_list\_docnet <- merge(course\_details\_list\_docnet, docnet\_iframes\_subset, by = "course\_url\_first\_join", all = TRUE)

#Edit course\_details\_list\_docnet and course\_details\_list\_guido to rbind

colnames(course\_details\_list\_docnet) <- c("course\_url\_first\_join", "web\_scraper\_order", "degree\_url",

"course\_delivery\_mode", "course\_code", "course\_name",

"course\_period", "course\_type", "credits", "year","course\_url")

course\_details\_list\_docnet <- course\_details\_list\_docnet[, c("web\_scraper\_order", "degree\_url",

"course\_url", "course\_code", "course\_name", "course\_delivery\_mode",

"course\_period", "course\_type", "credits", "year")]

colnames(course\_details\_list\_guido) <- c("web\_scraper\_order", "degree\_url",

"course\_name", "course\_delivery\_mode", "course\_url",

"course\_code", "course\_period", "course\_type", "credits", "year")

course\_details\_list\_guido <- course\_details\_list\_guido[, c("web\_scraper\_order", "degree\_url",

"course\_url", "course\_code", "course\_name", "course\_delivery\_mode",

"course\_period", "course\_type", "credits", "year")]

colnames(degree\_programs\_list) <- c("web\_scraper\_order", "faculty\_school\_url",

"degree\_name", "degree\_url")

course\_details\_list <- rbind(course\_details\_list\_guido, course\_details\_list\_docnet)

rm(course\_details\_list\_guido,course\_details\_list\_docnet,docnet\_iframes,docnet\_iframes\_subset)

# Load and setup coordinators ----

guido\_course\_coordinators <- read\_csv("data/guido-course-coordinators.csv")

docnet\_course\_coordinators <- read\_csv("data/docnet-course-coordinators.csv")

colnames(guido\_course\_coordinators) <- c("web\_scraper\_order", "course\_url", "coordinators","coordinators\_url")

colnames(docnet\_course\_coordinators) <- c("web\_scraper\_order", "course\_url", "coordinators")

guido\_course\_coordinators <- guido\_course\_coordinators[, c("web\_scraper\_order", "course\_url", "coordinators")]

docnet\_course\_coordinators <- docnet\_course\_coordinators[, c("web\_scraper\_order", "course\_url", "coordinators")]

course\_coordinators <- rbind(guido\_course\_coordinators, docnet\_course\_coordinators)

rm(guido\_course\_coordinators,docnet\_course\_coordinators)

# Load and setup professors ----

guido\_course\_professors <- read\_csv("data/guido-course-professors.csv")

docnet\_course\_professors <- read\_csv("data/docnet-course-professors.csv")

colnames(guido\_course\_professors) <- c("web\_scraper\_order", "course\_url", "professors", "professors\_url")

colnames(docnet\_course\_professors) <- c("web\_scraper\_order", "course\_url", "professors")

guido\_course\_professors <- guido\_course\_professors[, c("web\_scraper\_order", "course\_url", "professors")]

docnet\_course\_professors <- docnet\_course\_professors[, c("web\_scraper\_order", "course\_url", "professors")]

course\_professors <- rbind(guido\_course\_professors, docnet\_course\_professors)

rm(guido\_course\_professors,docnet\_course\_professors)

# Load and setup description ----

guido\_course\_description <- read\_csv("data/guido-course-description.csv")

docnet\_course\_description <- read\_csv("data/docnet-course-description.csv")

colnames(guido\_course\_description) <- c("web\_scraper\_order", "course\_url", "description")

colnames(docnet\_course\_description) <- c("web\_scraper\_order", "course\_url", "description")

guido\_course\_description <- guido\_course\_description[, c("web\_scraper\_order", "course\_url", "description")]

docnet\_course\_description <- docnet\_course\_description[, c("web\_scraper\_order", "course\_url", "description")]

course\_description <- rbind(guido\_course\_description, docnet\_course\_description)

rm(guido\_course\_description, docnet\_course\_description)

# Load and setup contents ----

guido\_course\_contents <- read\_csv("data/guido-course-contents.csv")

docnet\_course\_contents <- read\_csv("data/docnet-course-contents.csv")

colnames(guido\_course\_contents) <- c("web\_scraper\_order", "course\_url", "contents")

colnames(docnet\_course\_contents) <- c("web\_scraper\_order", "course\_url", "contents")

guido\_course\_contents <- guido\_course\_contents[, c("web\_scraper\_order", "course\_url", "contents")]

docnet\_course\_contents <- docnet\_course\_contents[, c("web\_scraper\_order", "course\_url", "contents")]

course\_contents <- rbind(guido\_course\_contents, docnet\_course\_contents)

rm(guido\_course\_contents, docnet\_course\_contents)

# Load and setup learning-results ----

guido\_course\_learning\_results <- read\_csv("data/guido-course-learning-results.csv")

docnet\_course\_learning\_results <- read\_csv("data/docnet-course-learning-results.csv")

colnames(guido\_course\_learning\_results) <- c("web\_scraper\_order", "course\_url", "learning\_results")

colnames(docnet\_course\_learning\_results) <- c("web\_scraper\_order", "course\_url", "learning\_results")

guido\_course\_learning\_results <- guido\_course\_learning\_results[, c("web\_scraper\_order", "course\_url", "learning\_results")]

docnet\_course\_learning\_results <- docnet\_course\_learning\_results[, c("web\_scraper\_order", "course\_url", "learning\_results")]

course\_learning\_results <- rbind(guido\_course\_learning\_results, docnet\_course\_learning\_results)

rm(guido\_course\_learning\_results,docnet\_course\_learning\_results)

# Load and setup competences ----

docnet\_course\_competences <- read\_csv("data/docnet-course-competences.csv")

colnames(docnet\_course\_competences) <- c("web\_scraper\_order", "course\_url", "competences")

course\_competences <- docnet\_course\_competences

rm(docnet\_course\_competences)

# Load and setup bibliography ----

guido\_course\_bibliography <- read\_csv("data/guido-course-bibliography.csv")

docnet\_course\_bibliography <- read\_csv("data/docnet-course-bibliography.csv")

colnames(guido\_course\_bibliography) <- c("web\_scraper\_order", "course\_url", "reference")

colnames(docnet\_course\_bibliography) <- c("web\_scraper\_order", "course\_url", "reference")

guido\_course\_bibliography <- guido\_course\_bibliography[, c("web\_scraper\_order", "course\_url", "reference")]

docnet\_course\_bibliography <- docnet\_course\_bibliography[, c("web\_scraper\_order", "course\_url", "reference")]

course\_bibliography <- rbind(guido\_course\_bibliography, docnet\_course\_bibliography)

rm(guido\_course\_bibliography, docnet\_course\_bibliography)

# Join data----

# Remove the suffixes from degree\_url in both datasets so that degree\_url matches in both cases

degree\_programs\_list$degree\_url <- sub("/detall$", "", degree\_programs\_list$degree\_url)

course\_details\_list$degree\_url <- sub("/assignatures/gestionar$", "", course\_details\_list$degree\_url)

#Merge course details with additional information

course\_details <- course\_details\_list %>%

left\_join(course\_coordinators, by = "course\_url") %>%

left\_join(course\_professors, by = "course\_url") %>%

left\_join(course\_description, by = "course\_url") %>%

left\_join(course\_contents, by = "course\_url") %>%

left\_join(course\_learning\_results, by = "course\_url") %>%

left\_join(course\_competences, by = "course\_url") %>%

left\_join(course\_bibliography, by = "course\_url")

# Nest courses within each degree while keeping all degree information

final\_structure <- degree\_programs\_list %>%

left\_join(course\_details %>% group\_by(degree\_url) %>%

summarise(courses = list(cur\_data())), by = "degree\_url") %>%

mutate(courses = ifelse(is.na(courses), list(NULL), courses)) # Ensures degrees without courses still appear

# 3. Convert the final structure to JSON

json\_output <- toJSON(final\_structure, pretty = TRUE, auto\_unbox = TRUE, na = "null")

# 4. Save the JSON to a file

write(json\_output, "university\_data.json")