

R-skript i kvalitetsarbetet med DiVA

DiVA-mötet 2017-10-25

Greta Linder, Södertörns högskola

Bakgrund

Projektet Demeter

Behov av en mer effektiv och säker arbetsprocess för våra bibliometriska analyser.

Kompetenshöjande då det gäller datahantering.
Möjligt att koppla ihop flera källor.

Pågår under 2017: Camilla Hertil Lindelöw (bibliometriker),
Greta Linder (DiVA-specialist),
Jonas Öberg (systemutvecklare)

Vi jobbar med

R Studio

- Gratis program
- Installerat på en virtuell server hos Sunet, för att vi skall kunna göra uppdateringar själva
- Programmeringsspråket R

GitHub

- Södertörns högskolebibliotek
<https://github.com/SodertornHB>
- För att kunna dela skript med andra
- Strukturen under arbete...

Exempel: Skript för städning

Läs in en csv-fil och få ut en lista som underlättar arbetet

- Epub ahead of print
- Manuskript i avhandlingar
- ISSN-validering
- Poster med *Reviewed = false*, trots att de inte ligger i granskningen
- Kapitel som saknar både start- och slutsida

Skriptet kan se ut så här

```
diva <- read_csv(file = "diva_csvall2.csv")  
ahead_of_print <- diva %>%  
  subset(Status == "aheadofprint") %>%  
  select(PID, DOI) %>%  
  transmute(PID, link = str_c("https://doi.org/", DOI))
```

Leveransformat

Det går att skriva resultatet till csv- eller tsv-fil.

I R Markdown blandas den egna texten med skript och genererar en rapport som html, pdf eller word.

Epub ahead of print

Greta Linder

2017-09-29

Sammanställning av artiklar som i DiVA är registrerade som Epub ahead of print. Behöver kontrolleras om de fått placering i voyrn och nummer.

PID	link
1117907	https://doi.org/10.1016/j.indmarman.2017.06.002
1106422	https://doi.org/10.1017/S2045796017000233
1082801	https://doi.org/10.1080/17512786.2017.1299032
1088818	https://doi.org/10.1080/00036846.2017.1310999
1135112	https://doi.org/10.1007/s10803-017-3245-7
1138673	https://doi.org/10.1177/1555412017710616
936151	https://doi.org/10.1080/13691457.2016.1255928
899311	https://doi.org/10.1080/13504622.2017.1320704
1141601	https://doi.org/10.1016/j.jad.2017.09.013
1135721	https://doi.org/10.1007/s40926-017-0061-2
1053021	https://doi.org/10.1016/j.destud.2017.07.002



This organ



This repository

Search



This repository

Search

Pull requests

Issues

Marketplace

Explore



Repositories

Search repository

common

automated scripts,

R Updated 4 da

diva

R Updated 5 da

exploration

analyses for explor

R Updated 5 da

SodertornHB / diva

Code

Issues 0

No description, website, or top

Add topics

11 commits

Branch: master

New pull req

GretaL Anpassad till tidyverse

data_quality

.gitignore

README.md

diva.Rproj

README.md

diva

Scripts to use with dat

DiVA is a finding tool a
research institutions in

Här samlar vi skript so

SodertornHB / diva

Code

Issues 0

Pull requests 0

Projects 0

Wiki

Insights

Settings

Branch: master

diva / data_quality /

Create new file

Upload files

Find file

History

GretaL Anpassad till tidyverse

Latest commit d2dccc8b 5 days ago

README.md

Anpassad till tidyverse

5 days ago

ahead_of_print.Rmd

Anpassad till tidyverse

5 days ago

cleanup_diva.R

Mindre uppdateringar

5 days ago

validator_issn.R

Anpassad till tidyverse

5 days ago

validator_issn.Rmd

Anpassad till tidyverse

5 days ago

README.md

data_quality

Skript som sammanställer poster som kan behöva åtgärdas i DiVA.

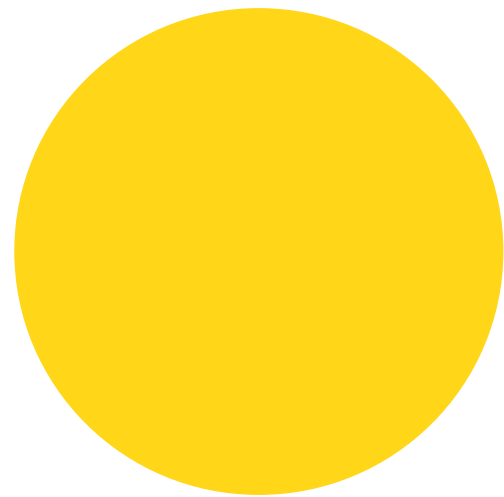
cleanup_diva

Innehåller: Ahead of print (även som RMarkdown), Ej granskningsmärkta, Manuskript i sammanläggningsavhandlingar, Sidor saknas i kapitel

Hur lär vi oss R?

- Finns olika tutorials på nätet, till exempel:
DataCamp: www.datacamp.com
R Markdown rmarkdown.rstudio.com/
- Vi läser:
Wickham, H. & Grolemund, G. (2016) *R for Data Science: Import, Tidy, Transform, Visulaize, and Model Data*. Sebastopol, CA, O'Reilly.
<http://r4ds.had.co.nz/>

Övningarna i boken är bra, men jag tycker det är lättare att förstå då jag får jobba med DiVA-data.



SÖDERTÖRNS HÖGSKOLA | STOCKHOLM
sh.se