Campus Luzern – A Cooperative Project of the University-Level Institutions in Lucerne

«*Web Scraping with R*»

Tutor	<u>Resul Umit</u>, University of Oslo
Organization	Campus Luzern
Language	English
ECTS-Points	(filled in by the administration)
Nos. of participants	max. 20
Content	This workshop will guide you in automating the process of extracting data from websites for analysis, in the software environment R. Some data are available in documents, ready to be downloaded from websites. Others are embedded into one or more webpages. This workshop will focus on the latter case, which poses specific challenges to researchers — in terms of extracting, cleaning, and tidying data. You will learn about how to deal with these challenges. To do so, we will work with a number of R packages — most importantly *rvest*, *tidytext*, and some of the packages in *tidyverse*.
Prerequisites/ Materials	The workshop assumes basic knowledge of R programming language. If you have no experience with R, but willing to familiarise yourself with it before the workshop, please contact Resul in advance to discuss your options.
Reading list	Steve Pittard, S. (2020). *Web Scraping with R*. Available at https://steviep42.github.io/webscraping/book/
Teaching method	This workshop is offered in English. It will take place online, via Zoom.
Charge	This Campus Luzern offering is directed at researchers, post-docs, doctoral students, people interested in starting a dissertation and employees of Lucerne universities and is free of charge for these persons. For persons not affiliated with Lucerne universities, the course fee is CHF 600/day. Certificates with the respective ECTS-Points are prepared on request (graduate.resources@campus-luzern.ch).