

Webscraping legal issues

NL

Olav ten Bosch

8 November 2016, ESSnet big data WP2, Rome



Statistics
Netherlands

Legal

- Dutch Statistics Law:
 - Enterprises have to provide data to SN on request
 - Scraping information from websites reduces response burden
 - SN uses data for official statistics only
- Dutch database law / Legislation on protecting IP rights
 - Commercial re-use of intellectual property is forbidden
 - SN will never publish or re-sell scraped content
 - web scraping by SN does not infringe the dutch database law
- Legislation on data protection:
 - Statistics Netherlands does only scrape public sources and processes data within Statistics Netherlands' safe environment, just as with other (privacy-sensitive) data internally
- Netiquette:
 - respect robots.txt
 - identify yourself (user-agent)
 - Provide information on website and inform website owners if a considerable amount of data is collected on a regular basis
 - do not overload servers, use some idle time between requests²

