

## Scraping data for rented vacation home statistics

Statistics Finland is looking into web scraping as potential new source of data. Recently 11 scrapers were built using open source software for vacation home rental information. Scraping offers data that

1. are more accurate and up-to-date than administrative data. There are many real estates, identified in administrative data as being private vacation homes and private houses, that are actually rented out at least for part of the year. Furthermore, administrative data are updated very slowly in comparison to websites. Data collection is often done on monthly or yearly basis.
2. are unit-level. Unit level rental revenue tax returns data covers about 40 % of rental returns. Eight of the largest central booking agencies provide mostly aggregate occupancy and price related data.
3. have information not available in administrative data. This offers interesting possibilities in estimating and predicting prices.
4. can be linked. Most scraped data can be linked to administrative data on micro level by owner information, street address, or coordinates. However, some administrative data lack spatial information.
5. carries minimal perceived response burden.

Here, we discuss ethical, technical and analytical issues related to scraping, data linking, and visualization. Most web sites are well structured and produce good quality data that offer interesting possibilities to Official Statistics.

### Contact information of authors

Katja Löytynoja  
Statistics Finland  
+358 29 5513537  
[katja.loytynoja@stat.fi](mailto:katja.loytynoja@stat.fi)

Ossi Nurmi  
Statistics Finland  
+358 29 5512984  
[ossi.nurmi@stat.fi](mailto:ossi.nurmi@stat.fi)

Johannes Kolu  
Statistics Finland  
+358 29 5513686  
[johannes.kolu@stat.fi](mailto:johannes.kolu@stat.fi)

Matti Kokkonen  
Statistics Finland  
+358 29 5513770  
[matti.kokkonen@stat.fi](mailto:matti.kokkonen@stat.fi)