

daten.berlin.de Usage Statistics

This dataset contains usage statistics (page impressions and page visits) for the Berlin Open Data Portal <https://daten.berlin.de>. Statistics are collected per month, both for the domain as such, and for all datasets (pages below /datensaetze).

Statistics are given in both CSV (split over two files) and JSON (one combined file).

Until 2019-12-31, usage statistics were collected with our internal *BerlinOnline Site Statistics* (BOSS) tool. As of 2020-01-01 we have stopped using BOSS on the Berlin Open Data Portal, and replaced it with Webtrekk. While BOSS and Webtrekk provide the same metrics, the actual results differ. *Page visits* is always an approximation, in that the statistics software has some kind of heuristics to determine which *page impressions* should be considered part of the same visit. The heuristics for BOSS and Webtrekk are slightly different, so unsurprisingly the numbers for page visits differ as well. However, the numbers for *page impressions* differ as well. This can e.g. be attributed to the fact that BOSS was based on the actual server logs, while Webtrekk is triggered by the client via JavaScript. Some clients choose to block JavaScript, so requests by those clients are not counted.

daten_berlin_de.domain_stats.csv

Download here: [daten_berlin_de.domain_stats.csv](#)

Domain-wide statistics. One row per month, columns for page impressions and page visits.

```
month,impressions,visits
2018-05,28563,10436
2018-04,24779,9282
2018-03,27728,10548
...
```

daten_berlin_de.page_stats.datensaetze.csv

Download here: [daten_berlin_de.page_stats.datensaetze.csv](#)

Per-dataset statistics. One row per dataset, two columns per month (page impressions and page visits).

```
page,2013-04-01 pi,2013-04-01 pv, ... ,2018-05-01 pi,2018-05-01 pv
liste-der-h%C3%A4ufigen-vornamen-2017,,, ... ,279,246
alkis-berlin-amtliches-liegenschaftskatasterinformationssystem,,, ... ,211,185
...
```

daten_berlin_de.stats.json

Download here: [daten_berlin_de.stats.json.tgz](#) (compressed)

The structure of the data is as follows:

- `/source`- From which source system the statistics were generated. One of ["Webtrekk", "Boss"].
- `/timestamp` - when these usage statistics were generated
- `/stats/site_uri` - domain of the data portal
- `/stats/earliest` - first month for which domain-wide statistics have been collected
- `/stats/latest` - last month for which domain-wide statistics have been collected
- `/stats/totals` - domain-wide statistics
- `/stats/totals/{MONTH}/impressions` - domain-wide page impressions during MONTH
- `/stats/totals/{MONTH}/visits` - domain-wide page visits during MONTH
- `/stats/pages/datensaetze` - statistics for individual datasets
- `/stats/pages/datensaetze/page_uri` - parent page for all datasets
- `/stats/pages/datensaetze/earliest` - first month for which dataset-specific statistics have been collected
- `/stats/pages/datensaetze/latest` - last month for which dataset-specific statistics have been collected
- `/stats/pages/datensaetze/sub_page_counts/{MONTH}/{DATASET}/impressions` - page impressions recorded for DATASET during MONTH
- `/stats/pages/datensaetze/sub_page_counts/{MONTH}/{DATASET}/visits` - page visits recorded for DATASET during MONTH

```
{
  "timestamp": "2018-06-12 14:29:47 +0200",
  "stats": {
    "site_uri": "daten.berlin.de",
    "earliest": "2011-09-01",
    "latest": "2018-05-01",
    "totals": {
      "2018-05": {
        "impressions": 28563,
        "visits": 10436
      },
      ...
      "2011-09": {
        "impressions": 54454,
        "visits": 23765
      }
    },
    "pages": {
```

```

"datensaetze": {
  "page_uri": "daten.berlin.de%2Fdatensaetze",
  "earliest": "2013-04-01",
  "latest": "2018-05-01",
  "sub_page_counts": {
    "2018-05": {
      "liste-der-h%C3%A4ufigen-vornamen-2017": {
        "impressions": 279,
        "visits": 246
      },
      "alkis-berlin-amtliches-liegenschaftskatasterinformationssystem": {
        "impressions": 211,
        "visits": 185
      },
      ...
    }
  }
}

```

License

All software in this repository is published under the MIT License. All data in this repository (in particular the `.csv` and `.json` files) is published under CC BY 3.0 DE.

Dataset URL: <https://daten.berlin.de/datensaetze/zugriffsstatistik-datenberlin.de>

2019, Knud Möller, BerlinOnline Stadtportal GmbH & Co. KG

Last changed: 2020-01-21