



# SCB - download metrics notes

Amanda Infeld  
August 24, 2021

---

## Overall

### Data processing

- Removed downloads with *bytessent* = 1
- As of 8/17/21: comprised ~8% of downloads

### Downloads

Defined as: unique daily dataset downloads by remoteip (unless otherwise stated)

- Only includes downloads for datasets currently on the cellxgene platform

### Users

- Proxied by *remoteip*
- Unable to identify downloads from internal CZI employees, other than myself
  - We do not believe CZI employees are downloading datasets frequently through the course of their work

## Metrics

### Most downloaded datasets

Normalized

- Value:  $[\# \text{ of downloads} / \# \text{ of days on the platform since MAX(upload date AND April 23, 2021)}] * 100$
- Downloads didn't start getting tracked until April 23, 2021

### Downloads by dataset characteristic

Normalized

- First we account for datasets that have been on the platform longer than others
  - Value =  $[\# \text{ of downloads} / \# \text{ of days on the platform since MAX(upload date AND April 23, 2021)}] * 100$
- Second, we account for more datasets with cells of a certain category than others
  - value / # of datasets of that category