Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederazium svizra

Swiss Confederation

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST
Federal Statistical Office FSO

Data Science
Competence Center DSCC

Lomas: The Hidden Data Oasis behind the Mist

The Lomas Service

Lomas aims to empower research, inform policy development and drive innovation across sectors, all while upholding the highest standards of data confidentiality.

It is a novel **open-source platform** designed to realize the potential of data held by public administrations, developed by the Data Science Competence Center (DSCC) of the Swiss Federal Statistical Office. It enables authorized users, such as approved researchers and government analysts, to **execute algorithms on datasets without directly accessing the data.**

Differential Privacy (DP)

DP is a framework for **releasing data products while controlling the risk of disclosure of personal data.**

Pure DP Mechanism

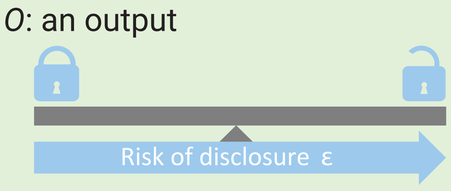
$$sens = \max |q(S) - q(S')|$$

$$Q(S) = q(S) + Lap\left(0, \frac{sens}{\varepsilon}\right)$$

Formal Guarantee

$$\frac{\Pr[M(S) \in O]}{\Pr[M(S') \in O]} \leq e^{\varepsilon}$$

- $q(S)$: result of query q on dataset S
- $Q(S)$: mechanism Q on dataset S
- S': dataset S plus or minus one record
- O: an output



Eyes-Off Analysis with Differential Privacy

Users **NEVER** see the private data. They can get DP protected results by sending requests to the server using one of the available DP libraries. To prepare their DP algorithm, they can retrieve metadata and dummy dataframes (randomly generated to simulate the sensitive data) through a request to the server.

To query a private dataset, a set of conditions must be met:
- The user is authorized to query the dataset and has enough remaining privacy-loss budget to make the query
- The query is executed by one of the available DP libraries:
  - **Smartnoise-SQL** for SQL-like queries
  - **Smartnoise-Synth** for generating synthetic datasets
  - **OpenDP** for summary statistics
  - **DiffPrivLib** for training Machine Learning models

Users can also perform other types of requests to get information about their privacy-loss budget and get the parameters and results of past queries.