



WORLD STATISTICS
CONGRESS

2025

THE HAGUE



isi International
Statistical
Institute

IPS 748

Free and Open Source Software at Statistics Netherlands: a review and current challenges

Mark van der Loo, Olav ten Bosch
Statistics Netherlands, Leiden University
October 8, 2025 14:00

A BRIEF REVIEW

A brief history



Period	Milestone
≤ 2009	R used only in research; first R package published
2010	R adopted as formal tool, first FOSS policy
2012	Python as formal tool
2014	Git as formal tool (succeeding SVN)
2017	CBS starts awesomelist for OS software
2018	CBS hosts uRos2018
2019/2020	New FOSS policy, Software Development Guidelines
2023	ESS FOSS principles
2024	IT starts getting interested in FOSS as users (K8S)
2025	FOSS Principles Adopted at UNECE Conference of European Statisticians



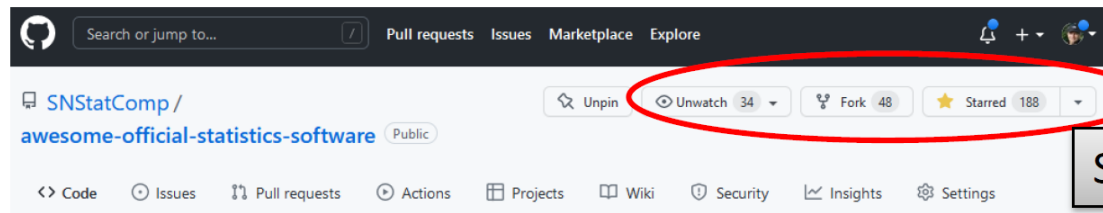
What is the awesome list?

Curated list of software for
official statistics



awesome

www.awesomeofficialstatistics.org



Social interactions, watch

Awesome official statistics software

An awesome list of open source software for official statistics

An item on this list is awesome because it is

1. free, open source, and available for download and
2. used in the production of official statistics by at least one institute or provides access to official statistics.

We prefer software that is easy to install and use, has at least one stable version, and is actively maintained. [Contributions](#) welcome.

License

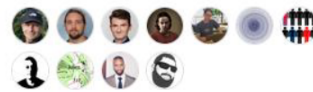


This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Criteria

Open license

Contributors 17



+ 6 contributors

Working together

Contributions

Awesome contributions are welcome, here are ways to do it:

- The GitHub way: send us a [pull request](#) to add directly to this list.
- Add an item to the [issue tracker](#) issue tracker. (you need a GH account)
- Send an e-mail to [mark dot vanderloo at gmail dot com](#) or [olav dot tenbosch at gmail dot com](#) or tweet [@olavtenbosch](#) or [@markvdloo](#)

Design frame and sample (GSBPM 2.1)

- CRAN 1.5-4 – a year ago license GPL (>= 2)

R package [SamplingStrata](#). Optimal Stratification of Sampling Frames for Multipurpose Sampling Surveys.

- CRAN 1.0.5 – 7 months ago license EUPL

R package [R2BEAT](#). Multistage Sampling Allocation and PS

Design variable descriptions (GSBPM 2.2)

- GitLab no releases found last commit november license MIT License

Excel [SDMX_Matrix_Generator](#). Excel-based visual SDMX and

up Statistical disclosure control (GSBPM 6.4)

- GitHub v5.1.7b3 last commit march license EUPL-1.2

Java and C++ application [Mu-ARGUS](#). Tool to cr

- GitHub v4.2.4.2 license EUPL-1.2

Java C++ Fortran and Delphi application [T-ARGUS](#)

- CRAN 5.7.6 – 2 months ago license GPL-2

R package [sdcMicro](#). Disclosure control for statist

- CRAN 0.32.6 – 4 months ago

R package [sdcTable](#). Disclosure control for tabul

Sampling (GSBPM 4.1)

- CRAN 2.10 – a month ago license GPL (>= 2)

R package [sampling](#). Several algorithms for drawing survey samples, including a variety of unequal probability sampling designs (high entropy, systematic, Rao-Sampford, etc.), and calibrating design weights.

- CRAN 4.0 – 4 years ago license GPL (>= 2)

R package [surveyplanning](#). Tools for sample survey planning, including sample size calculation, estimation of expected precision for the estimates of totals, and calculation of optimal sample size allocation.

- CRAN 1.4.2 – 6 days ago license GPL-3

R package [PracTools](#). Functions and datasets related to Valliant, Dever, and Kreuter (2018 2nd ed), [Practical Tools for Designing and Weighting Survey Samples](#).

- CRAN 0.3.0 – 9 months ago license MIT + file LICENSE

R package [prnsamplr](#). Coordinated stratified sampling using permanent random numbers (PRN's). Supports simple random sampling and probability proportional to size sampling, and includes a function for transforming

Data integration and record linkage (GSBPM 5.1)

- CRAN 0.3.4 – 5 months ago license GPL-3

R package [reclin2](#). Functions to assist in performing pairs, comparing records, em-algorithm for estimation, also be used for pre- and post-processing for machine

- CRAN 0.4-12.4 – a year ago license GPL (>= 2)

R package [RecordLinkage](#). Implementation of the Fellegi-Sun

- CRAN 1.4.1 – 2 years ago license GPL (>= 2)

R package [StatMatch](#). Statistical Matching or Data F

- CRAN 0.6.1 – 24 days ago license GPL (>= 3)

R package [fastLink](#). Implements a Fellegi-Sunter procedure and the inclusion of auxiliary information. [Documentation](#)

- CRAN 0.9.12 – 13 days ago license GPL-3

R packages [stringdist](#). Approximate string matching (Levenshtein, Hamming, Levenshtein, optimal string alignment), q-gram (Jaro, Jaro-Winkler). An implementation of soundex

- CRAN 0.1.6 – 4 years ago license MIT + file LICENSE

Over 30 software packages, giving access to > 60 data providers
majority are R-packages

Access to official statistics (GSBPM 7.4)

- CRAN 0.6-3 – 7 months ago license GPL (>= 2)

R package [rsdmx](#). Access to data or metadata from statistical organisations through SDMX. The package contains a list of SDMX access points of various national and international

- CRAN 0.3.1 – 7 months ago license GPL-3

R package [readsdmx](#). Read SDMX into dataframes from local SDMX-ML file or from OECD.

- GitHub v2.14.0 last commit last wednesday license Apache-2.0

Python [sdmx](#). Python library that implements SDMX 2.1 to explore data from SDMX data and metadata and convert it into Pandas objects.

- CRAN 0.4.3 – 7 months ago license MIT + file LICENSE

R package [rjstat](#). Read and write data sets in the JSON-stat format.

- PyPI v2.4.0 license Apache License 2.0

Python [pyjstat](#). Read and write JSON-stat.

- GitHub v0.2.8 last commit march 2023 license MIT

Java application [json-stat.java](#). Read and write JSON-stat. By Statistics Norway

- CRAN 0.2.5 – 2 years ago license CC0

R package [oecd](#). Search and Extract Data from the OECD

- CRAN 0.8.21 – 7 months ago license BSD_2_clause + file LICENSE

R package [sorvi](#). Finnish Open Government Data Toolkit

- CRAN 4.0.0 – 3 months ago license BSD_2_clause + file LICENSE

R package [eurostat](#). Tools to download data from the Eurostat database together with manipulation utilities.

- CRAN 0.22.5 – 3 months ago license EUPL

R package [restatapi](#). Search and retrieve data from Eurostat database, by Euro

- CRAN 2.1.4 – 5 years ago license GPL-3

UNECE OSS Principles



1. OSS by default

2. Work in the open

3. Improve and give back

4. Think general statistical building blocks

5. Test, package and document

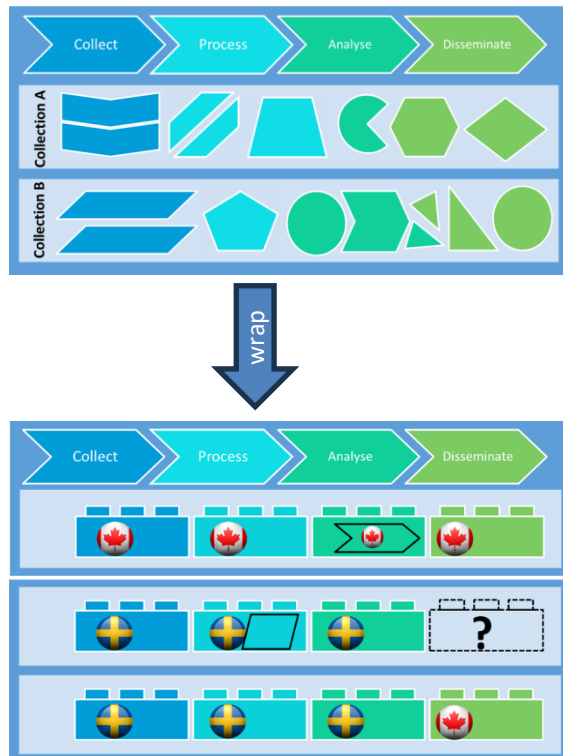
6. Choose permissive

7. Promote



CHALLENGES

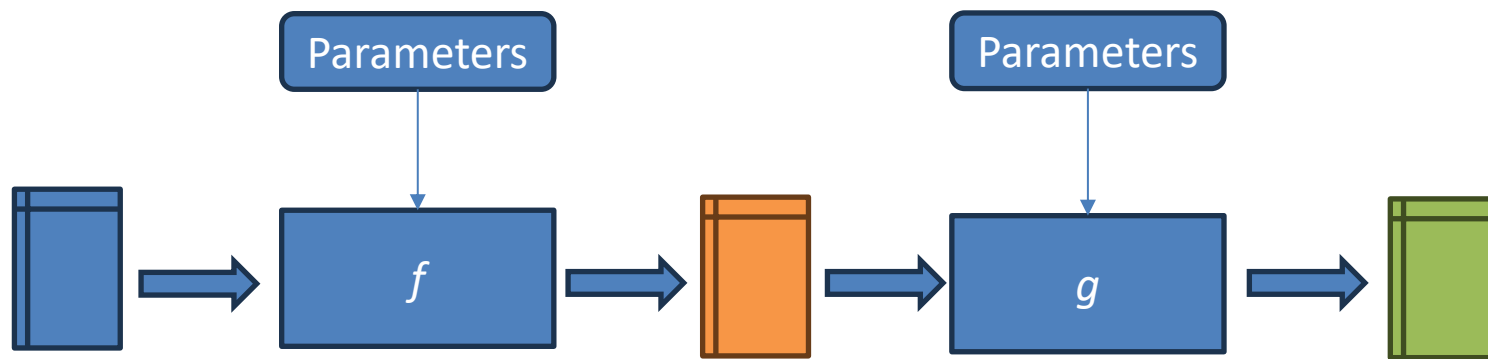
Composability



Is Wrapped
Software
Composable?



Composable parametrisations

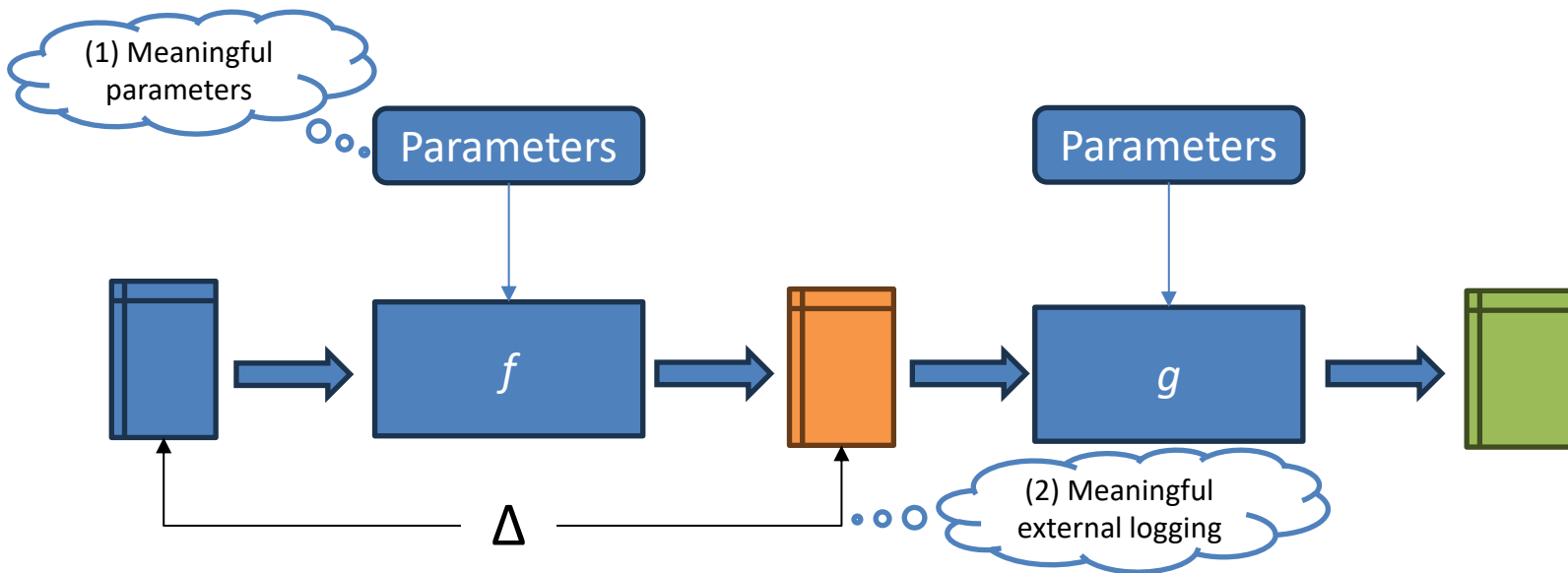


Updating parameters of f should not force one to update parameters of g

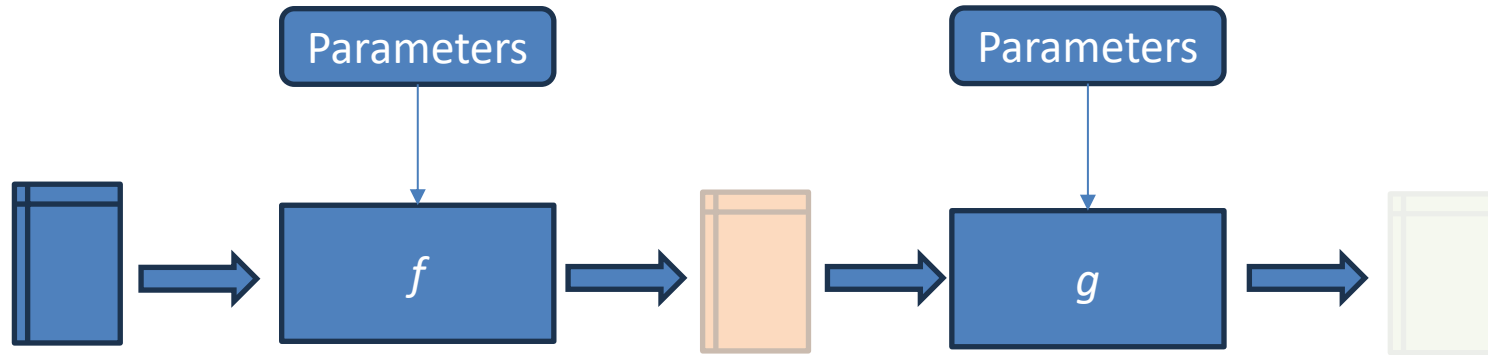


Granularity of Statistical Building Blocks

A statistical component should (1) be large enough such that subject matter experts recognize its function as a statistical task, but (2) small enough such that its action upon a dataset can be understood by comparing output with input.



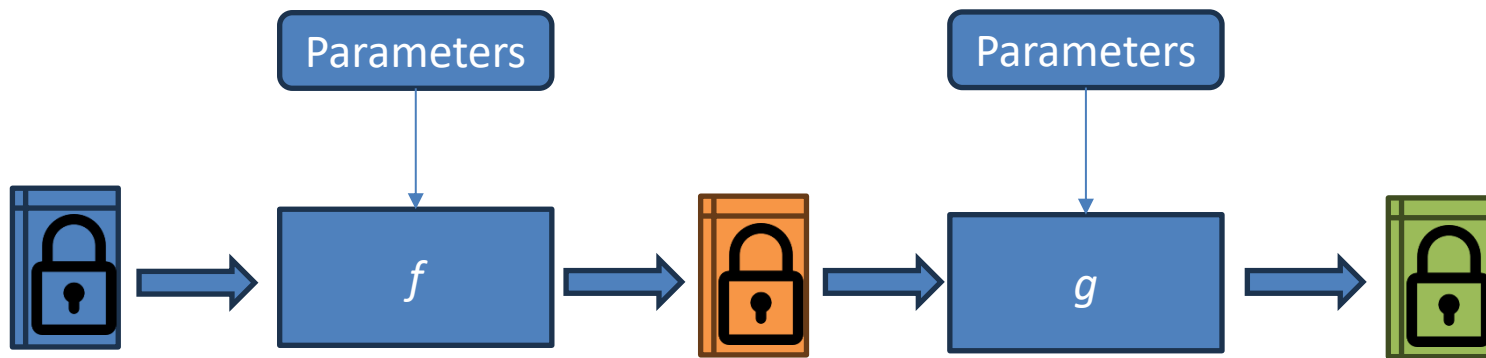
Composable Uncertainty



How do we gain insight into the uncertainty in results automatically when configuring a system?



Composable privacy



How do we ensure privacy protection in composed systems?



Cultural challenge

From Contract Handling to Community Membership



Summary

Statistical Open Source Software for Official Statistics: State of Play and Future Directions

Olav ten Bosch, o.tenbosch@cbs.nl, Statistics Netherlands

Mark P.J. van der Loo, mpi.vanderloo@cbs.nl, Statistics Netherlands and Leiden University

To appear in SJIAOS (2025)

1. We have come a long way
 1. Use & contribute since 2009
 2. FOSS and Software Development Policies
 3. Awesomelist
 4. Internationally supported Principles
2. There is still much to do
 1. Cultural challenges
 2. Challenges in composability



THANK YOU.