



Statistical Open Source Software

HLG-MOS Workshop 2023

Geneva, 21-22 November 2023

Barteld Braaksma

Background



A lot of activity related to open source

- BSTN sessions on Open Source (© Kate Burnett)
- CES session on Open Source
- ESS Taskforce on Open Source
- SAS-to-R conversion using GenAI
- .Stat presentation (this workshop)

Time for a more structured approach?

Overall Goal of SOSS project



The purpose of the Statistical Open Source Software project is:

- to develop a better common understanding
- of the pros and cons, do's and don'ts
- of moving to further and more comprehensive use
- of open source software
- as a cornerstone for official statistics production,
- based on concrete experiences
- through use cases of broad interest

Scoping (top down)



What generic aspects to consider?

- Organisation of maintenance, support and training
- Legal aspects and liabilities/responsibilities
- Licensing models and fair distribution of costs
- Community building
- Communication
- Engagement with e.g. academia and private sector

Scoping (bottom up)



What use cases to consider?

- Input domain: Rapid survey systems, smart surveys, ...
- Throughput domain: Awesome list
- Output domain: .Stat, PxWeb, Demetra, tau-Argus, ...
- Use GSBPM for structuring?

Include open source software developed by others?

- What about e.g. Pandas, PySyft, Spark, TensorFlow, ... ?
- Proposal to leave out for now (open to debate!)

Work Packages



(very preliminary!!!)

- **WP0. Scoping and ambition level**
 - Agreed scope and ambition
- **WP1. Generic aspects**
 - Guidelines, principles, frameworks, ...
- **WP2. Use cases**
 - Focus on codevelopment and community building
- **WP3. Management, synergies and communication**
 - Communication plan

Further Considerations



- **Relations/synergies may be possible with**
 - Carpentries and GenAI projects
 - Modernisation Committees
 - Thematic workshops/steering committees
- **Leave no one behind!**
 - Smaller offices have specific needs but limited resources
 - Maturity levels between (even within) institutes may differ
- **Learn from past experiences**
 - For example: what worked well (and what not) with CSPA?



Thank you