Moving towards open-source technologies

Strategic and managerial perspectives

Marleen Verbruggen

Conference of European Statisticians
22-23 June 2023, Geneva

Contributions from:

1. The Journey to Using R – Experience of Central Statistics Office Ireland
2. Free and Open-Source Software at Statistics Netherlands
3. DAPLA – A cloud-based statistical production system and its implications for Statistics Norway
4. Open-Source - Note by Statistics Poland
5. Transforming statistical workflows to use open-source technology at the UK Office for National Statistics

Open-source: different goals and angles

Ambitions:

- Demand for more and faster statistics, optimize production processes, use of open-source as part of the solution (NO, NL)
- Need to identify an ideal analytics environment for the longer term (IE)
- Need for reproducible research, transparency, openness (PL, UK)

Different angles:

- Policy around open-source technologies, (inter)nationally (NL, PO)
- Change management (IE, NO, UK)

UK: Transforming statistical workflows to use open-source technology

Scope and goals:

- UK govt. committed to open-source solutions: 'Open-source, unless..'
- Benefits: transparency, reproducibility, efficiency (re-use) and good software engineering

Approach:

- Strategic: software is output in itself
- The RAP approach (reproducible analytical pipelines): "RAP MVP", "RAP concepts principles and practices", "RAP transformation", "RAP workflows", "RAP deployment model". "RAP support team"
- RAP quality demands: minimalize handwork, audit trail, version management, open code, documentation

Main challenges:

- Competency development
- Culture, commitment
- Tooling

Observations:

- The challenges mentioned in the different papers show overlap, challenges are mainly non-technical
- Can RAP approach be developed as a more generic concept, being of value for other countries?

Norway: DAPLA – A cloud-based statistical production system

Scope and goals:

- Need for more / faster statistics, including being able to quickly respond when crises occur to stay relevant
- Use of open-source as part of the solution

Approach:

- Building own Data Platform ("DAPLA") in the public cloud (Google). Services: security, storage, data sharing, pseudonymisation, and to produce statistics with Python and R
- Move all data to the public cloud and to re-create all statistical production processes from scratch

Challenges:

- Competency development
- Time and Resource constraints
- Managing costs
- Agile approach
- Common functionality

Observations:

- Many NSI's face similar challenges and similar approaches. Opportunities for sharing best practices?

Ireland: The Journey to Using R

Scope and goals:

- Transformation from SAS to R in statistical production
- In 2027 all scripts must be available in R

Approach:

- A Technology-led initiative
- Wider potential to improve quality and methodologies through code refactoring and rewriting
- High-level plan focused on: communication, training, software environment, migration, standards and methodology
- Use of early adopters in implementation

Challenges:

- Commitment at senior management level and resistance to change by the statisticians
- Competency development

Observations:

- Interesting development: exploring ChatGPT (Davinci model) to translate SAS code into R code
- Opportunity: explore the potential of ChatGPT in international context; exchange of lessons learned

Poland: Open-Source

Scope and goals:

- "Being Open" Manifesto: adapt to digital world in an innovative, safe, and affordable way with Open-Source software, Open Data, Open Algorithms, Open Access, and Open Knowledge
- Benefits of open-source policy: improved quality, cost benefits, opportunities for public-private cooperation
- Open-source in many areas: technical infrastructure, programming, data processing and analysis, data sharing etc.

Approach:

- Competency building: Data Science Academy and open-source community
- Strategy for process of selecting software (identifying needs), testing and implementation of open-source solutions
- Stimulate participation of the community to influence the development of the software

Challenges:

- Security and continuity issues: testing, software updates, continuity of built libraries

Observations:

- How does the Polish "Being open manifesto" relate to similar initiatives? Is there one unifying approach possible?"

Netherlands: Free and Open-Source Software

Scope and goals:

- Policy of Stat. Netherlands: choice for R, Python, "awesome list of official statistics software"
- International contributions: "*Awesome list of official statistics software*" and: *ESS principles on OSS*

Approach:

- Formal introduction 2010, first bottom up. Since 2020 strategic approach. Policy comprises 4 areas: Production and Operations, Building Tailored Solutions, Use for R&D, and Contributing to FOSS
- Production and Operations, same selection process as for Commercial off-the-shelf (COTS) software
- Contributing to FOSS, all statistical software newly developed, to be applied generically, preferably FOSS
- Active community, also internationally
- International collaboration to the *7 ESS principles on FOSS*: 1. OSS by default 2. Work in the open 3. Improve and give back 4. Think generic statistical building blocks 5. Test, package and document 6. Choose permissive license 7. Promote

Observations:

- Can 7 principles be adopted as cornerstone for further development of open-source community for official statistics?
- How to stimulate and facilitate competency building and work 'in the open', internationally, at all levels of maturity?

Conclusions

- The benefits of open-source solutions: improved quality, transparency, reproducibility, flexible statistical production processes, possibilities for (inter)national cooperation, cost benefits, adoption of common standards and services

- Open-source initiatives often technically driven (at first)

- Challenges are mainly non-technical: change management issues, commitment management, competency building, resources

- Many best practices available regarding strategy / policy and change management / implementation

- Open-source solutions: strategic issue involving the whole organization: **How can we make the journey easier?**

Points for reflection and discussion

1. *The strategy and policy perspective:* How to bring principles (e.g. 7 principles), best practices together? How to take into account different levels of maturity?

2. *The change management perspective:* How to bring lessons learned and best practices together? Non-technical community: Use of OSS?

3. *The competency perspective:* How to stimulate and facilitate competency building and working 'in the open'?

4. *The legal perspective.* Is the legislation underlying our work prepared for the OSS developments that we want? Any burdens or improvements possible? Should FOSS principles be added to the code of practice?