

UNECE HLG-MOS Open-Source Software Charter for Official Statistics

Table of contents

Opening Statement	1
Principles (under process of CES endorsement)	2
1. OSS by default	2
2. Work in the open	2
3. Improve and give back	3
4. Think generic statistical building blocks	4
5. Test, package and document	4
6. Choose permissive	5
7. Promote	5

! Important

The text herein is currently under draft and should by no means be considered finalised. Any references to endorsement by the UNECE HLG-MOS or any other body are proposed and not valid as of the time of writing.

Opening Statement

UNECE HLG-MOS recognizes open source software as essential for modern statistical production, fostering transparency in methodology and promoting international collaboration in developing and supporting the production of official statistics. Open source solutions enable NSOs to develop, validate, and share statistical methods while ensuring reproducibility of official statistics. By adopting and developing open source tools, NSOs can establish flexible, cost-effective statistical infrastructures that adapt to emerging data sources and methodological innovations. The transparent nature of open source software allows for peer review of statistical procedures, strengthening the credibility of official statistics and facilitating methodological harmonisation across countries. NSOs shall prioritise open source solutions for statistical processing, particularly for data collection, analysis, and dissemination tools, ensuring long-term sustainability and knowledge retention within statistical

systems. Through open source adoption and development, NSOs can build robust statistical capabilities while promoting international statistical standards and best practices. Sharing statistical code and tools between NSOs creates a collaborative ecosystem that accelerates innovation in official statistics while reducing duplicate development efforts. This approach ensures efficient use of public resources while maintaining the independence and scientific integrity of national statistical systems. Open source statistical software enables NSOs to respond rapidly to emerging data needs while building trust through transparency in statistical production.

The UNECE HLG-MOS therefore endorses using the following Principles on Open Source Software in official statistics, both the production of software, and the adoption of software for statistical production.

Principles (under process of CES endorsement)

1. OSS by default

Statement

In the production of official statistics we prefer the use of open source software solutions over closed software solutions. Moreover we share our software solutions as open source.

Rationale

This principle adds to core values in official statistics such as being transparent and independent in how we make statistics and striving for high quality and reproducibility. Using and sharing open source software increases the transparency on how we work, it avoids black boxes in the implementation of official statistics.

Implications

This means that when implementing, redesigning or creating new processes, open source software solutions have preference. Only when no viable open source solutions exist it is possible to derogate from the default OSS option. The same holds sharing: sharing as open source is the default, but it is possible to derogate from this in justified cases. For NSIs this means that the methods used in the production of official statistics are not just described, the code used to actually apply the method is shared as OSS. For International Organisations this means openness on how international aggregates are computed via OSS solutions.

2. Work in the open

Statement

We start our projects in the open from the beginning and clearly mark maturity status.

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Rationale

Many projects have the intention to publish results as open source but have difficulties deciding on the best moment to do this. It might feel uncomfortable to put early ideas and rough implementation sketches on-line, but on the other hand sharing it too late prevents others from providing valuable comments and ideas or volunteering to work together on the project. To circumvent this dilemma we start working in the open right from the beginning wherever possible and clearly mark and update our projects development phase over time.

Implications

This means that it is recommended and accepted to start development projects in the public domain. We clearly show development status, which may vary from pre-alpha to stable and proven by showing a public roadmap, public source code repository, a public backlog of features, issues, bugs etc.

3. Improve and give back

Statement

We rather improve existing open source solutions than decide to create new solutions and we give our improvements back to the respective open source community.

Rationale

There are cases where existing open source solutions do not exactly cover the functionality needed in official statistics. The quickest way to cope with this is to copy a solution, adapt it and use it. However improvements implemented in the original solution will not be merged into the copy and our improvements made to the copy will not be visible in a wider context. Therefore we strive to give back our improvements to the open source community as change requests or suggestions even if it takes additional resources to do that. In the end this is an investment in the effectiveness and efficiency of the official statistics community as a whole.

Implications

This means that statistical organisations actively search for solutions that can be re-used instead of creating new solutions. Even if a solution does not exactly fit the needed functionality, it is examined how it could be improved, keeping the intended functionality in mind or even widening it. This also holds for partial solutions such as code snippets and (machine learning) models that could be valuable for others. The changes or extensions are tested, documented, and given back to the respective community to decide on possible integration into their solution.

4. Think generic statistical building blocks

Statement

In our open source work we strive for re-usable generic functional building blocks that support well-defined methodology in statistical processes.

Rationale

Publishing source code as open source is not enough for effective re-use in the global official statistics community. It is necessary to think about the design of what is to be shared and identify generic statistical building blocks that can be used in different contexts. Therefore we design the software from the intended user point of view and in a way that it can be re-used in multiple statistical domains or organisations as possible. This helps maintain complex statistical processes and high-quality official statistics.

Implications

This means that monolithic applications are componentised as much as possible in generic configurable statistical building blocks. We put statistical functionality in code and make statistical expert knowledge configurable. We make these components as much as possible generic in time, across statistical domains and across statistical organisations. For individual NSIs this means that not just “its” statistical production process should be kept in mind when developing tools but also the possible wider applicability. International Organisations should actively encourage the development and sharing of generic OSS solutions within their domain of expertise.

5. Test, package and document

Statement

We test, package and document our open source software for easy re-use.

Rationale

Re-using generic statistical software in the official statistics community is not always easy due to differences in statistical processes, technological environments, and way of working. Testing our software on functionality and security and packaging our software with good documentation is of utter importance as it improves the chances of re-use. General purpose package management systems offer versioning and documentation facilities to exchange generic statistical software. The use of such packaging systems helps maintain complex statistical processes and guarantee high-quality official statistics.

Implications

This means that we invest in testing, security scans, packaging and documentation to enable re-use. Security patches are applied as soon as possible. Documentation is designed from the viewpoint of a statistical user, keeping it concise, understandable but also complete and covering at least the basic functionality and a complete API reference. Packaging is a key success criterion for open source projects. Larger projects should adopt modern approaches such as containerisation, automate as much as possible, and smaller projects can follow these practices. Every package is downloadable without registration, is installable with minimal efforts and has a minimal viable example that can be executed. Dependencies are managed and minimised as much as possible. Versioning is implemented according to the principles of the respective package exchange platform with a preference for semantic versioning. Security patches are implemented with priority. For individual NSI this means that published OSS software is maintained and updated according to the policies of the relevant platforms, e.g. CRAN. International Organisations should play an active role in knowledge exchange on test, package and documentation policies in their domain of expertise.

6. Choose permissive

Statement

We choose the most permissive OS licence possible for sharing our software.

Rationale

Re-using software is in the common interest of the official statistics community. Re-use of our software not only enhances efficiency but also improves the quality of the software since the larger user community can contribute to development and maintenance. To maximise re-use by others it is necessary to choose an OS licence that maximally allows re-use, and minimises conflicts with other licences. This is known as “permissive” (see for example here). When choosing the appropriate OS licence we strive for maximum re-use.

Implications

This means that when sharing software we opt for a permissive licence (e.g. Apache 2.0/MIT) over a “Copyleft” licence, taking into legal, organisational and societal considerations. Mandatory acknowledgement / attribution of sources and authors is a viable additional option.

7. Promote

Statement

We invest in promoting new developments or improvements on our open source software within the official statistics community and where applicable in a wider context.

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Rationale

Re-use of generic software is not going to happen if no one knows what can be re-used. On the other hand it is difficult to know beforehand what the value of our software is for others. The only way out is to communicate, even if we have no clue if it is usable in a wider context. We advertise our software in an honest, brief way, mentioning its core functionality. Let the public know our plans for new developments and improvements and be open for suggestions for improvements.

Implications

This means working together on communication facilities targeted at the open source community. A community-driven approach of sharing knowledge, possible OS building blocks and its application in the statistical production should be the preference instead of centrally maintained repositories. A centrally maintained repository of software tools can become outdated soon and collecting information from the community could be a big effort. Therefore, such a repository should be maintained by the whole community. For individual NSIs this means that they are active participants in the OSS community by participating in events, joining relevant only forums, etc. International Organisations should play an active role in the organisation of the statistical OSS community in their domain of expertise.