A vision for supporting research software sustainability across Europe and beyond

Authors:

Bezaye Tesfaye	
Carina Haupt	https://orcid.org/0000-0001-6447-1379
Carlos Martinez-Ortiz	https://orcid.org/0000-0001-5565-7577
Cees Hof	https://orcid.org/0000-0001-9624-2431
Jan van der West	
Jose Benito Gonzalez Lopez	https://orcid.org/0000-0002-0816-7126
Martin Hammitzsch	
Neil Chue Hong	https://orcid.org/0000-0002-8876-7606
Pedro Príncipe	https://orcid.org/0000-0002-8588-4196
Radovan Bast	https://orcid.org/0000-0002-7658-1847
Stephan Druskat	https://orcid.org/0000-0003-4925-7248
Yvonne Vermonden	

Research software is an essential output in different research domains and needs to be reproducible and extendable throughout future research. Achieving this requires shared principles and know-how among researchers and research institutes. The European Software Sustainability Infrastructure (EuSSI) consortium aims towards identifying key elements and providing common grounds to achieve such a goal in European level. Partners in the EuSSI consortium are all involved - in one way or another - in research software and share a common interest in improving the general quality of research software.This includes various aspects such as the availability of software in the future and the ability to reuse software beyond the project for which it was created. We generally refer to this as *Software Sustainability*.

Although consortium members already collaborate with each other, we still approach software sustainability in different ways. Coordination at a European level will help to align our approaches and provide a unified set of best practices of research software (in alignment with the emerging FAIR principles for research software).

Sharing best practices in research software development across Europe will promote reuse of research software across European organisations. This will help to accelerate scientific discoveries across Europe and help reduce intrinsic costs incurred when research software is created from each project, without building on previously existing software.

The main ideas in EuSSI provide a basis to establish the requirements across the European software sustainability landscape for best practices in developing research software. The ideas cover the following main lines:

- Research Software Guidelines and Policies
- Software quality verification mechanisms (Software Seal of Approval)
- Services and infrastructure

- Trainings
- Community engagement and consultancy support

Research Software Guidelines and Policies

We aim at developing a framework containing a common ground of baselines for software sustainability in research. These baselines should include governing policies that can be used throughout the implementation of a research software and be followed by all researchers and research institutes at European level. The policies will implement FAIR principles for research softwares as in FAIR is being implemented for data. In order to define policies for software sustainability, it's also important to have a list of guidelines that go hand in hand.

Different organisations across Europe already have some form of guidelines at institutional or national level. Therefore it is important to review already implemented policies, guidelines, and best practices by different organizations in the EU and internationally. This is required (1) to understand what has been adopted by different institutes to make research software FAIR, (2) to differentiate variants among existing guidelines, as well as to find out their commonalities, (3) compare results from surveys and publications, and (4) to understand the gaps in existing policies and guidelines related to findings.

Specific Challenge:

Develop European agreed guidelines, policies and best practices for development of research software

Scope:

The creation of template documents containing guidelines, policies and best practices for development of research software. Such template guidelines could be extended and to the specific needs of particular institutions or communities. These documents will specify:

- Guidelines which define what constitutes good research software in terms of quality and sustainability,
- Recommendations for policies which can support the creation of high quality research software and support sustaining it in the long term.
- Approved best practices and practical recommendations for developing good quality research software.

Expected Impact:

At organization level:

- What will be the impact/added value for your organization?

    - Individual organisations joining this effort are likely to have their own guidelines for software development. By joining this effort they ensure that their guidelines are inline with the most up-to-date guidelines in Europe.
- What will be the impact/added value for the partners?
    - Partners will benefit from the knowledge provided by each participating organisation.

At system level:

- What will be the impact/added value for the region/ Europe?
    - The European community will benefit from the existence of guidelines which can ensure a comparable level of software quality across all European organisations developing and using research software
- Is there a specific societal challenge which will be tackled?
    - Good quality in research software ensures scientific results obtained using software are reproducible, addressing the broader issue of scientific reproducibility crisis.

Activities:

What will be the main activities in the project?

- A workshop (or series of workshops) to create documents containing guidelines, policies and best practices for development of research software. Concrete outputs of these workshops would be:
    - Documents containing policies for FAIR software in research, aka policies for FAIR software in EU.
    - Documents containing a set of guidelines that need to be applied to make sure sw produced in research are under a common umbrella of policies, aka guidelines for FAIR software in the EU.
    - Documents containing a set of best practices that all researchers need to follow in order to implement defined guidelines, aka best practices for FAIR software in the EU.

How the partners can jointly address the expected impact:

Each partner will be able to contribute their own expertise in the topic towards achieving standardised European guidelines. Each partner will contribute the perspective and knowledge from local initiatives in their area which could be adapted and adopted by others.

Consortium

1. GFZ
2. UEDIN

3. NLeSC
4. OpenAIRE
5. German Aerospace Center (DLR)

Discussion to join: Research groups at universities, research institutes and other institutions involved in development of research software.

Software Seal of Approval

The existence of guidelines research software guidelines ultimately aims at improving the quality of research software. The existence of research software quality metrics which assess the compliance with those guidelines provides objective means to measure such quality, and provides the opportunity for improving specific aspects of the software.

Quality metrics for research software must take into account various aspects of the research software's context, and establish a hierarchy of classes of software:

- the type of software, ranging from small analysis scripts (notebooks) used to process data for a single study, to generic software libraries with high potential for reuse;
- the field of application and the common practices in that research community;
- the time frame in which the software is expected to continue to be relevant.

We aim to develop an internationally supported Software Seal of Approval (SSoA) comparable to the Data Seal of Approval, but directed to the software development cycle. Different types of seal may result, depending on the sustainability requirements of the individual code (e.g. short lived vs long term preservation software). Existing comparable approaches are already emerging, therefore we aim at building the SSoA on existing tools.

Specific Challenge:

There are three main challenges in the development of a software seal of approval:

- The development of quality metrics appropriate to different types of research software. These metrics will allow developers to benchmark their software and understand what they need to do to improve on it.
- The development of tools which can automatically determine the level of compliance of a specific piece of software with aforementioned quality metrics.
- The development of accepted community processes which facilitate trusted application of the metrics and tools, enabling others to understand and trust the use of, and compliance to, the SSoA.

Scope:

This project will focus on the design and develop an internationally agreed research software quality metrics which will support internationally supported Software Seal of Approval (SSoA) comparable to the Data Seal of Approval, but directed to the software development cycle.

Expected Impact:

At organization level:

- What will be the impact/added value for your organization?
    - Participating organisations will gain the ability to demonstrate that the software they are producing meets an agreed upon level of quality.
- What will be the impact/added value for the partners?
    - Partners will benefit from having a common standard for evaluating the quality of research software.

At system level:

- What will be the impact/added value for the region/ Europe?
    - European parties will benefit from having an agreed upon standard for evaluating software quality. This will facilitate research software reuse across the European union, ensuring software meets certain minimum requirements which will facilitate its reuse (e.g. documentation, dependencies, docker images, etc.).
- Is there a specific societal challenge which will be tackled?
    - The availability of a SSoA will bring the research software policies and guidelines mentioned before into practice.

Activities:

What will be the main activities in the project?

- Development of a proof-of-concept software implementation capable of evaluating compliance with the research quality metrics. The following are some considerations which should be taken into account while developing a SSoA:
    - Tools developed should integrate with ease with platforms generally used during the software development process, such as Github and Gitlab.
    - Tools developed should provide a familiar integration process, such as those provided by TravisCI, Github Actions and SonarQube.
    - Tools should provide transparent information on how quality metrics are calculated. For example, when a piece of software is given a score, it should be clear to the developer why this score was achieved and what would be required in order to improve such score.
- Promote adoption of SSoA. In order to have a meaningful impact in the research software landscape, the SSoA requires a wide level of adoption. Only then will

researchers benefit from knowing the level of quality of a piece of software before they attempt to use it.

How the partners can jointly address the expected impact:

Two main forms of contribution are necessary in order to achieve this ambition:

1. Partners can contribute their expertise and engineering capacity in developing a prototype SSoA.
2. Partners can adopt and promote the adoption of the SSoA in research software projects in which they are involved.

Consortium

1. NLeSC
2. UEDIN
3. GFZ

Discussion to join: Research groups at universities, research institutes and other institutions involved in development of research software, policy makers.

Services and infrastructure

Sustainable software is supported by component-based platforms which are interoperable to provide support. A European Software Sustainability Infrastructure should be capable of supporting the development, maintenance, publishing, validation / verification, reuse, preservation and archival of software, no matter the size of the team developing it.

Currently there are gaps in the integration of services and infrastructure available. There is a need to integrate and extend existing services to support software sustainability workflows. Examples of concrete actions in this direction are:

- A meta-search engine for research software, building on CERN's pilot Asclepias service
- Improved integration between Zenodo / InvenioRDM and other repositories to support ingest (code repositories, digital repositories) and archival (Software Heritage)
- Services for automatically assessing quality metrics for various degrees of software sustainability (SSoA).
- A unified platform for developing and sharing training materials for research software skills and for finding instructors providing research software skills training, that leverages the Carpentries infrastructure and EOSC training material catalogues used by many of the consortium partners.

We aim to build on existing services and infrastructure, connecting and extending them to enable new workflows to support software sustainability.

Specific Challenge:

For the “services and infrastructure” aspects there are still many unknown factors and further details such as scope and specific activities remain to be defined.

Discussion to join: Universities, SMEs, Large Companies, clusters, policy makers (at least through letters of support), user association. Most projects should consist of balanced consortia between companies, research institutions and/or associations.

Trainings

It is known from previous experience of the project partners that the requirements, priorities and knowledge of different communities with respect to software sustainability differs considerably. Adherence to software quality guidelines requires specific expertise which is not broadly available in the current research landscape. To disseminate such expertise training materials and trainers are required. We will create training materials for developing good quality research software which can be adapted by the communities to meet their specific needs and train instructors with the expertise to teach such lessons.

A wide range of researchers and relevant actors require training to increase their knowledge, skills and attitude towards software sustainability. The audience for training events is not only the research community, but also the support staff, as well as participants from industry partners.

The lesson content should address multiple aspects: the preservation of software, reutilisation of software and the preservation of data. Lesson content should also discuss the technical tools relevant to each community, but also guidelines and policies discussed earlier. The lesson content should also be adaptable to the individual needs of different communities. Here we can build on already mature training material as well as documented experience and expertise consortium partners who have delivered similar workshops across many countries and across academic disciplines.

We consider recording the instruction part of the training and experiment with “flipped-classroom” approaches where we optimize for discussions, interaction, collaborative problem-solving, and case studies. The events will be exercise-driven and in “Carpentries style” of teaching where we can rely on experienced instructors within the project consortium.

These extended training events will not only address sustainability and archiving of software and data but also reproducibility (distributed version control, reproducible code/environments/workflows, containers, software citation, licensing, notebooks, and automated testing).

Specific Challenge:

The main challenges in providing training for developing research software are:

- To organise workshops and surveys for the relevant research communities and private sector to gather requirements of the training materials.
- The development of training materials to increase the knowledge, skills and attitude concerning software sustainability for the relevant research communities and private sector.
- To establish, maintain, and support a virtual forum to discuss questions and best practices and support the training efforts.

Scope:

We aim at gathering information from diverse communities, without the claim to be exhaustive. We do not aim to create a multi-dimensional matrix of communities, nor will it be feasible to approach all possible combinations, but will be pragmatic in our selection.

Expected Impact:

At organization level:

- What will be the impact/added value for your organization?
    - Participating organizations will benefit from having access to training materials which can help them in increasing their own institutional knowledge and that of other organisations in their network.
- What will be the impact/added value for the partners?
    - Partners will benefit from having a shared framework of training materials which will help allow them to learn from the expertise of other partners in the consortium and to standardize their knowledge on best practices for research software development.

At system level:

- What will be the impact/added value for the region/ Europe?
    - The availability of standardized training materials will facilitate the spread of a common knowledge base for all European organisations involved in research software development and reuse.
- Is there a specific societal challenge which will be tackled?
    - An European shared level of knowledge in research software development and reuse will facilitate collaboration and software reuse across European organisations.

Activities:

What will be the main activities in the project?

- Research/Development/Cooperation

Consortium

1. Sigma2
2. German Aerospace Center (DLR)
3. NLeSC

Discussion to join: Universities, SMEs, Large Companies, clusters, policy makers (at least through letters of support), user association. Most projects should consist of balanced consortia between companies, research institutions and/or associations.

Community engagement and consultancy support

Building software in accordance with best sustainability guidelines requires a certain level of expert knowledge. Researchers in specific domains are often not trained as research software engineers and sometimes they do not want or need to become an expert, but would rather collaborate with someone who already has the required expertise and who can support them, for example in selecting the most suitable technology to ensure that their software will be sustainable. Examples of topics where expert knowledge is required include: software preservation (archiving of code), keeping software operational (maintaining or emulating a platform for the software to run on), and sustaining the service (which also involves keeping the knowledge about the software and supporting its users).

Within the research sector, which is our primary target audience, we distinguish the following broad disciplinary sectors to engage:

- Natural sciences (including Geosciences) - engage via CERN, GFZ and UEDIN networks and relevant ESFRI Landmark projects, such as EMFL and SKA
- Life sciences - engage via bioinformatics communities and research infrastructures such as EATRIS, ELIXIR and BBMRI
- Engineering (especially Aerospace, Space, Traffic, and Energy) - engage via DLR
- Social sciences - engage via CESSDA, ESS and SHARE
- Humanities - engage via CLARIN, DARIAH, EHRI and related European projects
- Next, and partly overlapping with the humanities, is the cultural heritage sector of Libraries, Museums and Archives (the “GLAM” sector) - engage via national institutions and their international organisations, such as IFLA, OCLC, ICA and LIBER.
- The EOSC cluster projects ENVRI-FAIR, PaNOSC, ESCAPE, SSHOC and EOSC-Life

- Interdisciplinary - engage with national eScience and Data Science institutes and international organisations such as Plan-E (Platform of National eScience Centers in Europe)
- National and international communities and events of research software engineers (de-RSE, NL-RSE, Nordic-RSE, UK-RSE, international RSE leaders).

Although this is not an exhaustive list, it provides a starting point for engaging with relevant communities.

Specific Challenge:

Specific challenges pertaining to community engagement and consultancy support include:

- To establish and support points of contact with the EuSSI for domain and regional communities.
- To establish a research network for driving research of topics within the scope of the EuSSI, and organize knowledge exchange events.

Scope:

The creation of two schemes to support community building and engagement, and to establish a research scheme for research software sustainability research:

- The scheme to support community building and engagement (“ambassador scheme”) will establish a network of EuSSI Ambassadors and equip them with travel funding for networking and dissemination work. Candidates for the ambassador role are invited to apply for the role, which follows a fellowship model with an active phase for the duration of the project, and the right to reference this role beyond the project running time. THe scheme is targeted at research leaders such as professors and heads of institutes.
- The scheme to support research in research software sustainability topics will establish a network of PhD students who work on research software sustainability topics. Students who successfully apply for the scheme are equipped with travel funding to network and disseminate their PhD work, and are being invited to a workshop of funded students to discuss their work and network with their peers.

Expected Impact:

At organization level:

- What will be the impact/added value for your organization?
    - Participating organizations will gain expertise in cutting edge research software sustainability research specifically through the research scheme.
    - Participating organizations will be able to draw upon a network of research software sustainability ambassadors to further awareness of research software sustainability.
- What will be the impact/added value for the partners?

    - Partners will gain access to cutting edge research into research software sustainability through a network of EuSSI ambassadors through their delegates in the schemes.

At system level:

- Europe will have a network of ambassadors and research leaders furthering awareness of research software sustainability, driving policy change and compliance in this area, and performing knowledge exchange.
- Europe will be at the cutting edge of research software sustainability research through the research scheme and network.
- Awareness of the central role of sustainable research software for research will be increased throughout society.

Activities:

What will be the main activities in the project?

- Development and implementation of an ambassador scheme to create a network of EuSSI Ambassador that disseminate the project goals, increase awareness and drive policy change and compliance.
- Development and implementation of a research scheme in research software sustainability to create a network of PhD students to drive cutting edge research software sustainability research.
- Planning and organization of networking and knowledge exchange workshops for PhD students funded under the research scheme.

Consortium

1. DLR
2. NLeSc

Discussion to join: Universities, SMEs, Large Companies, clusters, policy makers (at least through letters of support), user association. Most projects should consist of balanced consortia between companies, research institutions and/or associations.

Potential funding ideas

For nearly all these ideas the new Digital Europe Programme could be very interesting. At the time of writing the programme is still being developed, but the main aims seem to fit very well with these ideas.

Additionally, a workshop (or series of workshops) in the direction of Research Software Guidelines and Policies could be organised as a Lorentz or Schloss Dagstuhl workshop.

One possibility to fund Trainings would be via an Erasmus+ Project. The project could focus on exchanging good practices and developing a common program. The Erasmus+ program will be renewed under Horizon Europe.