

#108705 CodeRefinery (sustainability phase)

Application Info

Applicaton ID 108705
Call NelC - Invitation for e-Infrastructure Collaboration Development project
Applicant Radovan Bast

GENERAL PROJECT INFO

Development project title CodeRefinery (sustainability phase)

Acronym

Time span

From date 01.11.2021

To date 31.10.2024

CONSORTIUM

Consortium partners

Add the consortium partners

Institution	Country	Contact person	Position	E-mail
Aalto University School of Science (Science-IT project)	Finland	Richard Darst	Staff Scientist	richard.darst@aalto.fi
CSC - IT Center for Science	Finland	Pauliina Somerkoski	Development Manager	pauliina.somerkoski@csc.fi
DeiC	Denmark	Kristoffer Laigaard Nielbo	Director of CHCAA	kln@cas.au.dk
Digital Scholarship Center, University of Oslo	Norway	Matthew Good	DSC team lead	matthew.good@ub.uio.no
EuroCC National Competence Center Sweden	Sweden	Thor Wikfeldt	ENCCS Training Coordinator	kjartan.thor.wikfeldt@it.uu.se
UNINETT Sigma2	Norway	Vigdis Guldseth	Senior Advisor	vigdis.guldseth@uninett.no
Swedish National Infrastructure for Computing	Sweden	Hans Karlsson	SNIC Director	hans.karlsson@it.uu.se
Type 1 Consortium for Interactive HPC	Denmark	Kristoffer Laigaard Nielbo	Director of CHCAA	kln@cas.au.dk
USIT: University Center for Information Technology, University o	Norway	Gard Thomassen	Assistant Director	g.o.s.thomassen@usit.uio.no

Contact person for the consortium

First name Radovan
Last name Bast
Position Senior Engineer, current PM
Institution UiT The Arctic University of Norway
Country Norway
Mobile +47 96752679
E-mail radovan.bast@uit.no

DEVELOPMENT PROJECT

Public summary

Short description of the proposed collaboration

The current emphasis on open science is caused by the crisis of reproducibility and inability to make full use of existing research. Reproducibility comes with significant technical challenges that affect researchers' day-to-day work, and over the past five years CodeRefinery has addressed them by teaching key tools and practices. CodeRefinery has shown that this training works in international collaboration, and the sustainability phase will ensure a long-lasting effect.

CodeRefinery (coderefinery.org) supports researchers by advancing the FAIRness of software and development practices so that they can collaboratively develop, review, discuss, test, share, and reuse their codes. Between 2016 and 2020, about 1500 researchers from the Nordic countries have been trained. The long-term impact of CodeRefinery workshops, measured through post-workshop surveys (3-6 months later) shows that, for example, about 60% of our learners have improved their usage of version control. Yet, the tools we teach are rarely part of academic curricula and the resulting waste of resources and results to society is massive. The need for training in research software engineering will only increase and become more diverse. Partner organizations have recognized the value of collaboration instead of re-inventing courses and are committing substantial in-kind contributions. Moreover, raising the skills of researchers will catalyze the use of national e-infrastructure resources, both hard (facilities) and soft (other advanced training). For the continued collaboration and development of the community effort to be successful and sustainable, it is important that the collaboration is supported by dedicated coordination during this transition phase.

CodeRefinery (sustainability phase) will ensure the long-term success of the lessons developed during previous project phases by funding a central coordination of in-kind contributions, thus i) coordinating and delivering a sufficient number of accessible high-quality training events contributing to level-up research software development skills at reduced costs for each participating institution, ii) maintaining and developing further an inclusive community for software- and skill development and fostering the co-creation and co-maintenance of FAIR learning materials, videos, and other resources, iii) and increasing the number of instructors trained with pedagogy suitable for teaching technology from novice to intermediate level learners.

Development project plan

Development project plan 1 attachment ([coderefinery-proposal.pdf](#))

Budget and resources

Budget and time commitment in person months 1 attachment ([budget.pdf](#))

Please give a description of and justification for the needed resources

This project will operate for a duration of 3 years with the following composition:

- 1 full-time equivalent (FTE)/ year or 12 person months (PM)/ year funded by NeIC
- On average 3.8 FTE/year or 46 PM/year contributed in-kind

The average cost for 1 FTE is estimated to be 1.38 MNOK (based on the estimate that 1 PM costs 115 kNOK).

Activity A (Management and Coordination) will be fully funded by NeIC, in contrast to Activities B (Operations, exception: GitLab service operation) and C (Sustainability and Governance), which are expected to be funded in-kind and carried by the partners (see also the letters of commitment). To make this distinction in funding clearer, we have split the “Worktime” table in two: one listing the NeIC-funded person months, and one summarizing the in-kind person months.

As we expect to onboard new organizations each year and to grow, this is reflected in the “Worktime: in-kind contribution” table with a progression of 3 new organizations entering the consortium.

Justification for the cost calculations:

- Workshop/meetings organized by CodeRefinery: making it possible for staff to attend two events per year (own workshops and hackathons, as well as NeIC All-Hands Meeting; 10 kNOK per event and person).
- Attending conferences: Making it possible for staff to attend one international conference per year (25 kNOK per event and person).
- GitLab hosting: Estimate based on current pricing of DeiC-hosted service (50 kDKK/year). Our ambition is to develop a sustainability plan before 2024 so that the service can be financed by participating organizations instead of NeIC after 2023.
- Web domain and promotion: Domain name registration, stickers, posters, and stands.
- Carpentries membership: 5 kUSD/year for platinum membership with discount for co-funding the regional coordinator.
- Collaborative tools: Chat platform and tools for collaborative notes and documents such as <https://hackmd.io/>, and a helpdesk platform (currently we use free plans). Our estimate is based on 50 NOK per user per month and service (2 services, 20 users).
- Registering organization: Registering the organization “CodeRefinery”: registration and legal fees.
- In-kind travel/conference budget: Event organization (tasks B-1 and B-2) will require that also in-kind staff receive support to travel and attend our events (team meetings and hackathons) as well as NeIC All-Hands meetings and NeIC conferences.

Services and tools

Services and tools 1 attachment (services-tools.pdf)

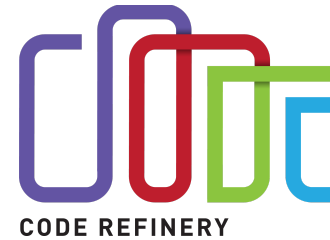
LETTERS OF COMMITMENT

Letters of Commitment 17 attachments (aalborg.pdf, aalto.pdf, aarhus.pdf, carpentries.pdf, csc.pdf, deic.pdf, dsc.pdf, encs.pdf, newcastle.pdf, sigma2.pdf, sintef.pdf, snic.pdf, ssi.pdf, tud1.pdf, tud2.pdf, uit.pdf, usit.pdf)

CodeRefinery (Sustainability Phase)

Project partners:

- Aalto University School of Science
- CSC – IT Center for Science (Finland)
- Danish e-Infrastructure Cooperation
- Digital Scholarship Center (University of Oslo)
- EuroCC National Competence Center Sweden
- Swedish National Infrastructure for Computing
- Type 1 Consortium for Interactive HPC (Denmark)
- UNINETT Sigma2 (Norway)
- University Center for Information Technology (University of Oslo)



1. Significance

Across all scientific disciplines, there is currently a great emphasis on reproducible research and FAIR data so that research can have the greatest benefit to society. Underlying these principles is usually custom-written research software and workflows. This research software is the key to reproducibility and data management, but most researchers do not have training beyond basic programming (if that), leading to a crisis of reproducibility: it is difficult to ensure that software is reusable by others, or even correct in the hands of the original author. This shift towards research enabled by software requires a profound and continuous transformation of research practices. A major challenge for the research community is the pressure to gain and maintain the necessary software skills in very limited time while both technologies and best practice evolve rapidly. These tools and practices are almost never part of a traditional academic curriculum and learning of foundational coding, data science skills, and associated best software practices in isolation, without a support structure is inefficient and costly, both for the individual researcher and the research community as whole. Over the last 5 years we have built up a curriculum and a training network to address this challenge and trained about 1500 researchers (see [list of our past workshops](#)).

National and international e-infrastructure and high-performance computing providers, organizations, and funding agencies have recognized the need for more training and have launched efforts such as PRACE, EuroCC, and CASTIEL. These offer a great opportunity for researchers to improve their competence, but in general serve advanced users focusing on specific tools. Unfortunately, these projects will not fill the increasing gap in data science and coding skills between novice and intermediate users among researchers in all academic disciplines. We are convinced that developing and supporting these skills by offering basic (Carpentries) and intermediate (CodeRefinery) courses to future users and developers of high-performance workflows is essential, because FAIR research data and reproducible research directly come from good practice and software. Basic- to intermediate skills are also required for effective use of advanced facilities. The learning process, from novice to a competent practitioner in coding, needs appropriate pedagogy. However, it would be wasteful to try to develop sufficient trainer resources independently at each institution. The same applies to lesson development and maintenance, since material does not have to be always site-specific. Collaborating would not only save costs but also open up opportunities for courses which would only materialize by combining expertise from different organizations or countries.

We observe that the trend points away from local computing centers towards national (e.g. Sigma2/Metacenter) and international compute resources (e.g. EuroHPC/LUMI) with distributed support staff. This centralization of resources and decentralization of users and staff forces organizations to rethink training: organizations will have to collaborate on training and stay close to the users, and CodeRefinery will help combine and coordinate these efforts.

The CodeRefinery (sustainability phase) project will ensure the long-term success of the lessons of CodeRefinery by funding central coordination of in-kind contributions, thus i) coordinating and delivering a sufficient number of accessible high-quality training events contributing to level-up research software development skills at reduced costs for each participating institution, ii) maintaining and developing further

an inclusive community for software development skill development and fostering the co-creation and co-maintenance of FAIR learning materials, videos, and other resources, iii) and increasing the number of instructors trained with pedagogy suitable for teaching technology from novice to intermediate level learners.

Expected benefits:

1. More researchers have the prerequisite skills to benefit from national research computing, PRACE, EuroHPC and Euro national competence centre (ENCC) offerings.
2. Increasing the number and diversity of codes being able to access national, pre-exascale and exascale systems.
3. Supporting diversity and improving equality of researchers writing and publishing research software.
4. More researchers are sharing their software and data according to FAIR principles.
5. More software developed and published in the Nordics with an increased participation in EOSC and EuroHPC joint undertaking calls (with the possibility of joint Nordic participation).
6. Lower costs of training for participating organizations with a significant increase of the quality (both of the training material and the delivery).
7. More instructors in the Nordics capable of co-creating, co-maintaining and co-delivering high-quality training on foundational coding, data science skills and associated best software practices.
8. Instructor training benefits not only future CodeRefinery instructors but also instructors and future workshops given within ENCC and given on a national scope.
9. Benefits above are more sustainable and outlive the 3-year project period.

1.1 Research community

The **innovation** of this project is the creation of a virtual collaborative space for teaching together across Nordic countries, a place to meet, discuss, and facilitate training development and maintenance to keep the high standard and inclusive community for co-learning.

Traditionally the development of research software has often had a strong link to scientific computing (quantum chemistry, theoretical condensed matter physics, computational astrophysics, bioinformatics and biostatistics, oceanography, meteorology, and others). Recently however, also other academic disciplines in humanities and social sciences use computational research as a key technique and research software as a key tool. Thus the **relevant research community** is growing and becoming more diverse.

The **benefits for the research community**: This project benefits the entire spectrum: from users to developers, from novice users to intermediate/advanced users. Because not only developing research software but also using software requires tools and skills to combine existing tools, pre- and post-process, analyze, and visualize data and to share the relevant artifacts with the research community in a reproducible way. And we also need to connect and support experts who know “everything” but can use our network and tools to teach their knowledge to others.

The **acceleration of scientific output** thanks to this project will be more reusable and reproducible software which will accelerate reuse of the scripts, programs, workflows, and pipelines in the same way as reusable data or papers enable *reuse* and combined works in future studies. CodeRefinery will provide researchers with essential skill sets to develop, improve, and share their codes and to reuse and combine codes written by their peers.

All **Nordic countries provide relevant and related services** in the form of training offerings. The goal of this project is to combine these. For more detail please see Section 1.2 (National and international context).

1.2 National and international context

The CodeRefinery project has created interfaces to a number of national and international training efforts (detailed below) and our ambition is not to replace their training portfolio or to compete, but rather to provide basic groundwork which is not in scope of any other training initiative. In addition to serving as a bridge between these efforts and between different levels (from novice to advanced), we can contribute

instructor training to these initiatives to collaboratively improve the quality of training events and to minimize the gap between these offerings. We are also convinced that NeIC is a suitable platform for this coordination and pooling of resources and competencies.

National e-infrastructure providers (CSC, DeiC, Sigma2, SNIC)

All national e-Infrastructure providers organize training events for their user communities, mainly motivated to improve resource usage and increase the productivity of computational research. These courses range from basic Linux courses to advanced programming and code optimization courses. The offerings range from local courses to centralized courses to multi-center collaborations, depending on the organizational structure of the provider. As we have mentioned previously, the gradual centralization of computing and storage resources as well as the decentralization of users and support staff create challenges and opportunities. Our project can help connect these efforts and we can help to avoid reinventing similar courses and reinventing best practices for online training. This will free up resources to increase course offerings for all partners and thus make also the Nordic e-Infrastructure collaboration more visible and tangible.

The Carpentries

The Carpentries is a global organization dedicated to teaching fundamental data science and coding skills in a fair and inclusive manner. The Carpentries workshops on core curricula give foundational skills to proceed to CodeRefinery workshops. In addition, CodeRefinery and the Carpentries have similar teaching philosophy and style. Therefore, CodeRefinery and the Carpentries complement and mutually benefit each other to create an efficient learning flow to develop researchers' skills further, as well as developing instructor resources that are trained in special pedagogy. With this background, NeIC has been a Carpentries platinum member and in 2019 the CodeRefinery project started to co-fund a Carpentries Regional Coordinator position for the Nordic/Baltic region which has catalyzed new workshops, collaborations, and recruiting the Carpentries instructor trainers to grow the capacity of the Carpentries training. From the experiences in the past two years, we learned that our coordination work is needed to connect instructors and to make the workshops happen. With the current project proposal we wish to further strengthen the coordination of Carpentries and CodeRefinery workshops.

EuroHPC-JU and EuroCC National Competence Centres

EuroCC national competence centres (NCCs), funded in part by the EuroHPC Joint Undertaking and in part by national funding programs, are in some countries incorporated into existing national providers (e.g. CSC and Sigma2) while run as independent centres in others (ENCCS). NCCs focus on advanced training for high-performance computing (HPC), high-performance data analytics (HPDA) and artificial intelligence (AI). This training will however not cover basic and intermediate skills needed before one reaches the level of advanced HPC/HPDA/AI training. The CodeRefinery training material can fill this gap and thus perfectly complement NCC training portfolios. Nordic NCCs are important collaboration partners for CodeRefinery and there are already synergies and collaborations between CodeRefinery and the Swedish NCC (ENCCS), with the latter adopting best practices in teaching and lesson development from CodeRefinery and in turn contributing to CodeRefinery events.

We expect to:

- Share best practices for open teaching and lesson development to NCCs in the Nordics.
- Assist Nordic NCCs in adopting best practices and proven methods for online teaching to reach a wider audience regardless of physical location.
- Provide a hub for shared teaching across countries by contributing to NCC training, inviting NCC instructors to CodeRefinery workshops and aggregating training material from NCCs on the CodeRefinery web page.
- Provide instructor and workshop organizer training based on the CodeRefinery system to Nordic NCC instructors - both to help them improve their HPC/HPDA/AI training but also to get NCCs to provide CodeRefinery workshops to their target audience.
- Recruit more NCC nodes outside the Nordics as CodeRefinery partners - this will be facilitated by the close interaction of NCCs and the EuroCC mentoring and twinning system.

[EOSC & EOSC-Nordic](#)

CodeRefinery project will be useful for The European Open Science Cloud (EOSC) and EOSC-Nordic in two ways: Skill-up of researchers as users and contributors of EOSC services and provision of know-how to training initiated by EOSC.

EOSC aims to provide a large infrastructure to support and develop open science and open innovation in Europe and beyond. In this context, EOSC-Nordic aims at making sure that researchers from Nordic and Baltic countries can effectively benefit from EOSC and work together across countries and scientific disciplines. The benefit of EOSC can only be realized when a bottom-up approach (getting the requirements/needs directly from the research communities) is combined with a top-down one (harmonisation at policy and service provisioning across Nordic & Baltic countries). Applied knowledge in best software practices will allow researchers to get their voice heard and play a meaningful role in the choice of services and tools to be deployed in EOSC.

Once operational in EOSC, these services will need to be anchored in training to become more used and more usable. For the training we can help with our know-how on training coordination and instructor training to bridge the gap between what services exist and what students and researchers really need.

[Nordic Research Software Engineers](#)

The community of Nordic Research Software Engineers (RSEs) is a relatively young but growing community of researchers, engineers, and technical staff, interested in research software engineering, the career development of RSEs, interested in improving diversity and inclusivity, and also in exchanging solutions and collaborating on multi-disciplinary projects. There is a significant overlap and there are close connections between the persons driving the Nordic RSE community and the CodeRefinery community and staff. In this project phase we will keep and strengthen these connections and collaborate on event organization and dissemination. We will work towards making RSE opportunities more visible to CodeRefinery stakeholders and learners and to make CodeRefinery training events more visible to the RSE community.

1.3 The Nordics and NeIC

The need for training in research software development does not stop at organizational or national borders and the Nordic collaboration has good size and critical mass to solve this together. The **added value** is to combine efforts instead of every country trying to solve this on their own. This will reduce pressure and work for individual organizations and **amplify** the training portfolio and network of available trainers. The successful scaling to a large number of participants within individual CodeRefinery workshops given online during 2020 has demonstrated the capability to host pan-Nordic workshops that reach a large audience, thus providing a high value at a low cost per participant.

The NeIC priorities are to strengthen both deep Nordic and wide international collaborations and to improve sustainability for research and societal value. This project will **contribute to the goals of NeIC** by strengthening the collaboration among instructors and training efforts in the Nordics, and by addressing and improving the sustainability of research software helping researchers making their software more open and more FAIR, thus increasing its societal value. We will also work on strengthening collaborations with other NeIC projects through training and with our GitLab service which provides a platform for hosting software and tools (the NeIC projects NT1 and E3DDS are users of this service).

Together with other NeIC projects, the CodeRefinery team could take on a role as consultants for developing and hosting workshops on research software and infrastructure. In addition to connecting trainers and the research community with training offerings, we can also help connecting the community to services by making the services more visible, more accessible, and by improving their resource usage. Over time we wish to grow this project beyond the Nordics and to open up the collaboration to a wide international network.

1.4 Services and tools

Extrapolating from the first two phases, we can estimate reaching **ca. 300 new learners every year** and we believe we can really induce a cultural change. This [overview page](#) and this [more detailed page](#) summarize

attendance statistics from all our previous workshops. However, attendance numbers are only half the story: we have managed to engage dozens of [volunteer instructors and helpers](#) who made large-scale workshops possible. Our [post-workshop survey results](#) and letters of support from various organizations (see Section 5 and attachments), encourage us to continue and grow.

We can document the need for the training services also with two paid workshops that we delivered in 2020 and the limiting factor does not seem to be interest to attend or to pay but rather resources to prepare and deliver.

Beyond traditional teaching, the promise of the Internet is that one can reach everyone in the world. CodeRefinery has been active in online teaching, fully living up to the philosophy "as open as possible, as closed as necessary". We are equipped so that every course we do can have a public, online component (Twitch live streaming, YouTube recordings) while maintaining small exercise rooms for small-group mentoring. This strategy will be expanded, documented, and used to redefine the way teaching is delivered.

The GitLab service has currently ca. 700 active users but a significant portion of these have been created for student course projects. We can expect 50-100 new users per year who use this service for their research.

We constantly strive to **improve the quality** of our courses by tracking all issues, listening to feedback, and by lowering technical and organizational barriers to experiment with new ideas and by striving for a good balance between review of changes but also encouraging changes. We see indications for an increasing uptake of our lesson material from other communities: in 2019 we have changed the license of our material to CC-BY to make reuse easier and this was a response to a community request. We also see an increased number of pull requests (change proposals) to our material repositories.

To keep up the high quality of the material we also need to give instructors opportunities to apply these tools and techniques in their own work. An open organization and a large network of contributors and helpers is also an important feedback loop.

2. Activities and benefit realization plan

Activity A: Management and Coordination (For the whole period)

Partners: DSC (lead), Sigma2/UiT

Activity A will be fully funded by NeIC, in contrast to Activities B (exception: GitLab service operation) and C, which are expected to be funded in-kind and carried by the partners. The motivation for this distinction is that the project requires one central contact and coordination hub which at this stage cannot rely on distributed in-kind efforts only. One of the main goals of this activity will be to onboard new organizations joining the collaboration network.

Two main areas of tasks and deliveries have been identified:

A-1 Coordination of CodeRefinery and the Carpentries workshops and events

- Continuation of partnership with the Carpentries by having a Regional Coordinator in place all the time to serve as the "front face" of The Carpentries in the Nordics and CodeRefinery, including participation in the Carpentries Member Organization Council and updating CodeRefinery partners.
- Disseminating Carpentries workshop and instructor training opportunities to eligible institutions in Nordics, coordinating 3 Carpentries workshops, and managing 15 instructor-training trainees per year.
- Coordinating CodeRefinery workshops and events including on-demand ones (also see B-1).

A-2 Community building and outreach

- Accepting and onboarding the 9 initial (based on letters of intent) in the first year and 1 new organization per year for the rest of the project period to join the collaboration network.

- Coordinating 3 specialized/external workshops/events in the Nordics per year upon request by CodeRefinery partners.
- Developing a standard partner workshop guide within the second year so that any partner workshop can be seamlessly shared with others.
- Marketing events of common interest via appropriate channels.
- Developing and implementing a rewarding scheme for CodeRefinery community members (instructors, helpers, and contributors) in the first year.
- Maintaining and further developing a manual/handbook on workshop organization on the CodeRefinery website for the distributed-work era.
- Publishing anonymized pre- and post-workshop survey results on the CodeRefinery website annually.
- Running an annual survey to collect inputs and feedback from partners.

Activity B: Operations (For the whole period)

Partners: All partners involved

In collaboration with the partners and academic community in general, CodeRefinery will continue to give high-quality workshops (and provide the Gitlab service), with a new emphasis on reach and diversity of offerings.

The standard CodeRefinery workshop is well-developed and can be successfully delivered without trouble either in-person or online. This will continue, but we will place a new focus on community-driven workshop delivery (without as much need for centralized staff) to facilitate on-demand workshops and to turn our Nordic network into a collaboration platform for other workshops.

In addition, CodeRefinery has provided a GitLab code repository hosting service for Nordic research software since 2017 and the service hosts research code projects for currently 700 users and 1300 projects. This service is popular because the corresponding data resides in the Nordics (the service is run by DeiC) and for some groups using GitHub or GitLab (as provided by the GitLab company or self-hosted) are not viable alternatives.

Four main areas of tasks and deliveries have been identified:

B-1: Delivering Carpentries and CodeRefinery workshops

- Delivery of high-quality workshops in a variety of formats: Online, in-person, hackathons, community meetups. Our focus is on online workshops, but we support in-person where there is local interest. We aim at training over 300 learners per year in 4 CodeRefinery workshops and supporting 5 partner workshops per year (pilot phase in the first year). Note that unlike in pre-COVID time, our online CodeRefinery workshops can accommodate 100 participants, with similar effectiveness as previous small in-person workshops with typical size of 30 participants.
- Maintenance and improvement of workshop infrastructure, including web page template, learner guides, helper onboarding, and related infrastructure.
- Developing a standard partner workshop guide to support shared workshops between partners, to take full advantage of online formats.
- Discussing and improving methods for delivering workshops, in particular online and open teaching formats.

B-2: Lesson development and maintenance

- Collaborative development of new lesson material based on input from participating organizations.
- Making lesson material findable, accessible, and citable (with a digital object identifier) and making sure that all contributors are credited for their work (following “[Ten simple rules for making training materials FAIR](#)”).
- Developing and maintaining a scheme to collect usage metrics for lessons.

- Adopting or co-maintaining lessons from partner organizations by giving them a “home” (organizational repository), and open-sourcing them.
- Developing a guideline to decide which lessons can be adopted and taken over.

B-3: Instructor training and development

- Maintaining CodeRefinery instructor workshop training material.
- Delivery of annual instructor training workshops, and other instructor/helper training workshops to different audiences.
- Continuous improvement in workshop delivery among partners.

B-4: Operation and support of the GitLab service

- Making definition of and publishing a GitLab service level agreement.
- Operation of this service (including backup, security patches, and upgrades).
- Provision of a help desk and user support for account creation and activation.
- Publishing GitLab usage statistics on the CodeRefinery website.

Activity C: Sustainability and Governance

Partners: Aalto University School of Science (lead), all partners involved

Our ultimate goal of this project is to decouple CodeRefinery management and coordination from NeIC project funding, thus ensuring a self-sustaining organization. Through the first two project phases, we have convinced ourselves that **centralized project coordination** will be needed for a distributed training project like ours, also in future: Cross-border and cross-organization training activities will not run on its own “for free”, even if free in-kind contributions exist for the bulk of the work. Our current plan is to develop a **membership and sponsorship model** where member organizations will co-fund the necessary coordination work which we estimate to require 1 FTE (likely shared by 2-3 persons).

At the end of the project period we wish to be fiscally able to receive sponsorship/membership fees, have a mechanism to receive payments for workshops, have established the project as an organization and as a brand, and be an organization which other Nordic research- and educational organizations support and are partners of. For this it will be essential that we make the benefits for organizations to join us evident, and that we make the **onboarding mechanism for member organizations** easy.

Two main areas of tasks and deliveries have been identified (expected time for delivery given in brackets):

C-1: Establishment of CodeRefinery as an organization

- Investigation of incorporation strategies in various Nordic countries (year 1, Q1).
- Organizing in-person or online consultation events to collect feedback on the CodeRefinery business model from CodeRefinery stakeholders and partners (year 1, Q2-4).
- Determining and publishing a governance model and by-laws, including board roles and responsibilities, and appropriate organizational policies and structures (year 2, Q2).
- Organizing a board election within the first 2 years of the project (year 2, Q4).
- Developing and publishing a mission and strategic plan for the organization (year 3, Q2).
- Registering CodeRefinery as an organization that can accept payment for training requests outside of the partner network (year 3, Q4).

C-2: Business model development and implementation

- Developing a business model for CodeRefinery and its financing plan in collaboration with CodeRefinery partners, and publishing it on the CodeRefinery website (year 2, Q2).
- Defining a membership model where organizations can join as sponsors or as contributing partners (with one-time donations or annual fees) (year 2, Q4).
- Creating a structure and pricing model to be able to accept payment for training (year 3, Q2).
- Implementing the chosen CodeRefinery business model by starting membership fees for CodeRefinery association (year 3, Q4).

- Developing a financing plan for the GitLab service and anchor it to one of the participating national infrastructure providers within the first two years of the project (year 2, Q4).

3. Resources and viability

3.1 Sustainability plan

Our sustainability plan is detailed in Activity C. The fact that we are requesting NeIC funding for only 1 FTE and that a number of organizations are prepared to participate in-kind can be taken as an indication that we are on a good track towards this goal. We believe that we will be able to convince also some of our in-kind partners who know the project and its value for their organization and their users or customers to support the project coordination in future.

Concerning the GitLab service, we plan to anchor this service within one of the participating national e-infrastructure providers and to develop a financing model where national providers and organizations from benefiting countries or organizations contribute a share of the costs for operation, maintenance, and support. For this it will be vital to highlight usage statistics and feedback from user communities and their research impact facilitated with this service.

3.2 Current team and viability

CodeRefinery consists of a broad team of staff and volunteers. At this point, there are roughly three types of team members: a) paid NeIC staff, b) paid staff at other organizations contributing to CodeRefinery on a voluntary and in-kind basis as a part of their job, and c) researchers who contribute to CodeRefinery on a voluntary basis.

- Paid NeIC staff and core team (as of March 2021): <https://coderefinery.org/about/who-we-are/>
- Speakers, instructors, hosts, helpers, and exercise leaders of past CodeRefinery workshops, as well as CodeRefinery events and shared events: <https://coderefinery.org/about/contributors/>

A primary goal of this project is to transition to fewer (a) and more (b) and (c) team members.

Core team applying to this NeIC Open Call 2021:

- Radovan Bast, University of Tromsø/ Serit
- Richard Darst, Aalto University
- Anne Fouilloux, University of Oslo
- Johan Hellsvik, PDC Center for High Performance Computing, KTH Stockholm, part of the Swedish National Infrastructure for Computing (SNIC)
- Diana Iusan, Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), part of the Swedish National Infrastructure for Computing (SNIC)
- Juho Lehtonen, CSC - IT center for science
- Sabry Razick, University of Oslo, University Center for Information Technology (USIT)
- Naoe Tatara, University of Oslo, Digital Scholarship Center (DSC)
- Thor Wikfeldt, The EuroCC National Competence Center Sweden
- Samantha Wittke, Finnish Geospatial Research Institute, Aalto University

With ca. 3 FTE/year we were able to deliver 58 workshops and events in the past 4 years. Thus, we believe that our projections are realistic and our proposal is feasible. But we are also aware of the risk of depending on key staff. To reduce this risk we will have to scale up our onboarding and mentoring processes.

4. Partners of the collaboration

Aalto University School of Science ([Science-IT, aka Aalto Scientific Computing](#))

Aalto Scientific Computing (ASC) provides HPC infrastructure, teaching, and hands-on support (Research Software Engineering) services for the *entire* Aalto University community. Rather than focusing on only one

aspect, ASC focuses on all three in order to enable research and independence of researchers. Aalto has been an in-kind partner in CodeRefinery since 2017, and became a full (co-funded) partner in 2021.

[CSC - IT Center for Science](#)

CSC - IT Center for Science Ltd. is a non-profit, state-owned company administered by the Ministry of Education and Culture of Finland. CSC maintains and develops the state-owned centralised IT infrastructure and uses it to provide nationwide IT services for research, libraries, archives, museums and culture as well as information, education and research management. CSC has been a project partner since 2016 and contributed both to training as well as by operating the GitLab repository hosting service before it moved to DeiC.

[DeiC - Danish e-Infrastructure Cooperation](#)

DeiC supports Danish research and research-based teaching through delivery of e-infrastructures (computing, storage and network). Organizationally, DeiC belongs to the Danish Agency for Science, Technology and Innovation. DeiC has been a project partner since 2017 and contributed both to training as well as by operating the GitLab repository hosting service.

[Digital Scholarship Center, University of Oslo](#)

Digital Scholarship Center (DSC) is a newly built center at University of Oslo (UiO) Library whose aim is to be a hub for courses and training in cross-disciplinary digital research skills within UiO. DSC has so far supported CodeRefinery and the Carpentries, and has been a project partner since October 2019 with a contribution of 0.5 FTE/ year for regional training coordination of both CodeRefinery and the Carpentries in the Nordic countries.

[EuroCC National Competence Center Sweden \(ENCCS\)](#)

The EuroCC National Competence Center Sweden is a joint initiative between the ten main Swedish research universities and RISE Research Institutes of Sweden. The center is hosted by Uppsala University on behalf of the consortium and includes the relevant competences at the other nodes. The initiative is funded by the EuroHPC JU, Swedish Research Council (Vetenskapsrådet) and the Swedish Innovation Agency (Vinnova). ENCCS provides advanced application support and training to researchers and engineers from academia, industry and the public sector.

[SNIC - Swedish National Infrastructure for Computing](#)

Swedish National Infrastructure for computing (SNIC) is a national research infrastructure that provides resources and user support for large scale computing and data storage to meet the needs of researchers from all scientific disciplines in Sweden. There are six universities that host and operate computing and storage resources for SNIC. The resources are made available through open application procedures so that the best Swedish research is supported. SNIC has been a project partner to CodeRefinery since the start of the project in 2016 and will continue to provide an in-kind contribution of 0.5 FTE for the entire duration of the project.

[Type 1 Consortium for Interactive High-Performance Computing \(Denmark\)](#)

This consortium consists of the University of Southern Denmark (SDU), Aalborg University (AAU) and the University of Aarhus (AU). The SDU YouGene cluster forms one part of the resources and the AAU CLAAUDIA's computer forms the other part. AU handles the support and helps new users get started. The focus of this collaboration is on interactive computing resources and easy access for new users offered through UCloud. The in-kind contribution from this consortium to the CodeRefinery project will be matched by DeiC by the same amount from 2022.

[UNINETT Sigma2](#)

Sigma2 manages the national e-infrastructure for large-scale data- and computational science in Norway and provides services for high-performance computing and data storage to individuals and groups involved in research and education at all Norwegian universities and colleges, and other publicly funded organizations and projects. Sigma2 has been a project partner since the start of the project in 2016 with a contribution of 0.5-0.75 FTE/ year (together with 0.5 FTE/ year project management).

[University Center for Information Technology \(USIT\), University of Oslo](#)

USIT provides critical IT services to University of Oslo and at national level through projects like Services for Sensitive Data (TSD). USIT also participates in the Sigma2/Metacenter, which is the federation of four major Universities together with UNINETT Sigma2 providing high-performance computing and storage solutions to Norwegian researchers. USIT has been an important partner in CodeRefinery workshops in Oslo and beyond. This collaboration has helped users to improve their skills: both for using services provided by USIT and Sigma2/Metacenter, but also increasing the competence of researchers and students on software best practices in general. USIT has been a partner to CodeRefinery since 2017 via the Sigma2 partnership.

5. Additional information

Web presence:

- Website: <https://coderefinery.org/>
- GitHub organization: <https://github.com/coderefinery>
- Social media: [Twitter](#), [Twitch](#), [YouTube](#)

Project reports and documents from project phases 1 and 2:

- Phase 1: [Project directive](#), [collaboration agreement](#), [project plan](#), [report](#)
- Phase 2: [Project directive](#), [collaboration agreement](#), [project plan](#), [mid-term report](#), [mid-term presentation](#)

Articles and presentations: <https://coderefinery.org/about/reports/>

Letters of commitment^[a, b] (see attachments):

- Aalto University School of Science
- CSC – IT Center for Science (Finland)
- Danish e-Infrastructure Cooperation
- EuroCC National Competence Center Sweden
- Swedish National Infrastructure for Computing
- Type 1 Consortium for Interactive HPC (Denmark) represented by two letters:
 - CLAUDIA Research Data Services, Aalborg University
 - Center for Humanities Computing Aarhus (CHCAA), Aarhus University
- UNINETT Sigma2 (Norway)
- University of Oslo^[c] represented by two letters:
 - Digital Scholarship Center (DSC) at University of Oslo Library
 - University Center for Information Technology

Letters of support^[a, d] (see attachments):

- The Carpentries
- Research Data Services at TU Delft Library
- Research and education services group at UiT - The Arctic University of Norway
- Research Software Engineering, Newcastle University
- SINTEF
- Software Sustainability Institute

Notes:

[a] Letters were provided with “CodeRefinery (phase3)” as the project name, which was tentatively given.

[b] Listed letters show either in-kind or hosting contribution size.

[c] The 0.5 FTE/year in the letter from UB/DSC corresponds to two rows in the budget table: DSC and USIT

[d] Letters of support appear under “Letters of commitment” (limitation of the application portal).

Worktime (in person months): NeIC funding

	2021	2022	2023	2024	Total
Digital Scholarship Center, University of Oslo	1	6	6	5	18
Project manager (100% NeIC)	1	6	6	5	18
Total NeIC funded PMs	2	12	12	10	36

Worktime (in person months): In-kind contributions

	2021	2022	2023	2024	Total
Aalto University School of Science (Science-IT project)	1	6	6	5	18
CSC - IT Center for Science	0.5	3	3	2.5	9
DeiC		6	6	5	17
Digital Scholarship Center, University of Oslo	0.5	3	3	2.5	9
EuroCC National Competence Center Sweden (*)		3	3	3	9
UNINETT Sigma2	1	6	6	5	18
SNIC	1	6	6	5	18
Type 1 Consortium for Interactive HPC, Denmark	1	6	6	5	18
USIT, University of Oslo	1	3	3	3	10
Expected in-kind from new organization		2	2	2	6
Expected in-kind from new organization			2	2	4
Expected in-kind from new organization				2	2
Total in-kind PMs	6	44	46	42	138

(*) current ENCCS funding period ends 2022

Estimated average cost for 1 PM in NOK: 115 000

Cost calculations

	2021	2022	2023	2024	Total
Cost for NeIC funded person months	230 000	1 380 000	1 380 000	1 150 000	4 140 000
Workshops/meetings organized by CodeRefinery		40 000	40 000	40 000	120 000
Attending conferences		50 000	50 000	50 000	150 000
GitLab hosting		69 000	69 000		138 000
Web domain and promotion		3 000	3 000	3 000	9 000
Carpentries membership		43 000	43 000	43 000	129 000
Collaborative tools		2 000	2 000	2 000	6 000
Registering organization			2 000	2 000	4 000
Total cost NeIC funding (NOK)	230 000	1 587 000	1 589 000	1 290 000	4 696 000
In-kind person months	690 000	5 060 000	5 290 000	4 830 000	15 870 000
In-kind meetings/conference budget		330 000	345 000	315 000	990 000
In-kind funding (NOK)	690 000	5 390 000	5 635 000	5 145 000	16 860 000
Total budget for the project (NOK)	920 000	6 977 000	7 224 000	6 435 000	21 556 000