# Proposal to the NeIC Strategic Committee to support phase 4 of CodeRefinery (2025 - 2028)

Authors: Radovan Bast (radovan.bast@uit.no) and Samantha Wittke (samantha.wittke@csc.fi) with contributions from the CodeRefinery team and steering group

Version of this document: 2025-01-24T10:00



## Summary of what we are asking for

This is a **continuing collaboration** between 9 organizations. Partner organizations provide project staff in-kind with a total volume of 130—160 person months over the 3-year project period. We request that costs for meetings and conference outreach (as well as minor operational costs) are covered by NeIC.

The project period is expected to be **from 2025-03-01 to 2028-02-29**.

## Background

CodeRefinery is a **community-driven initiative designed to empower students, researchers, staff, and research software engineers** with the skills and tools needed to develop high-quality, efficient, reusable, and reproducible research software. By focusing on practical training in areas such as version control, documentation, testing, sharing, and reproducible workflows, we bridge the gap between research and software engineering. Our mission is to foster collaboration, improve research output, and promote

open-source principles across academic disciplines. Open science is an important goal for universities and research institutions, and CodeRefinery is a key component in working towards this goal.

Since its beginning in 2016, CodeRefinery has contributed to enhancing the software development capabilities of the academic community in the Nordics and beyond. Over the past eight years, we have developed a comprehensive, openly maintained curriculum of lessons and courses that has been taught at over 40 workshops. **We reach circa 500 learners each year at all career stages and across all disciplines** and with a diverse range of programming expertise. We achieve this with the help of around 30 instructors, organizers, and facilitators.

CodeRefinery has cultivated a vibrant, collaborative community of instructors, learners, helpers, and partner organizations. Together, **we have built a robust cross-border training network** and a foundation for long-term improvements in research software development.

The need to train the academic community in software development skills has not decreased over the course of the last eight years. On the contrary, we see a growing demand and **we anticipate that this demand will continue to increase significantly** with the ongoing digital transformation of research and the omnipresence of machine learning (ML) and artificial intelligence (AI). Researchers will need even more training on how to deploy, maintain, and share ML and AI models, how to transfer them from their laptops to high-performance computing (HPC) environments and AI factories, and how to make them reproducible and reusable for the broader research community. However, we will continue our training focus on coding skills for researchers applying AI/ML, not AI/ML itself.

Proposals need to offer not only a foundation and stability but also innovation. In this project the innovation is to constantly **update and improve how we teach**. We have developed novel teaching methods to reach more learners, to be able to on-board more instructors, and to offer a learning path not only for learners but also for instructors. Our approach to training has inspired other training networks to follow our methods and to join us and teach together.

With this proposal, we are reaching out to NeIC and its partner organizations. These organizations value training and digital skill development as part of their mission. With the support of our project partners, we can ensure that CodeRefinery continues to enhance the capabilities of researchers and software engineers, ultimately advancing the quality and reproducibility of academic research. We want to offer quality training in the spirit of a **training cooperative** with our partners that would not be possible for any of the project partners alone. We all need to provide training in digital skills, but we can do it better together.

## Teaching solid foundations in the age of artificial intelligence

Since 2022, the field of artificial intelligence and large-language models have transformed education, research, and the funding landscape. So-called AI Factories which are set to provide an ecosystem for AI research by combining compute capabilities, data, and support, are being established in many countries and are attracting significant funding and attention.

In this project we will continue to focus on what we do well: provide training in robust research software engineering tools and practices for researchers and staff supporting researchers. But we will work on connecting our material better to new communities, user groups, and support staff. In our existing curriculum we can adapt examples and add exercises about how to version and share data and models, in addition to the existing material on how to version and share code and dependencies. We expect that we will be an important partner in the training of staff about how to train researchers in computational tools and how to develop new material.

## CodeRefinery and research data management

FAIR (findable, accessible, interoperable, reusable) data practices and research data management (RDM) are relatively well established across the Nordics. To those who focus on the software side of open science in their work and teaching, it can feel like RDM is a decade ahead in terms of maturity, acceptance, and the existing network reach.

We see data and software as connected: Data is often produced, analyzed, and visualized by software. It makes sense to at least know about good software practices when teaching and supporting RDM, and vice versa.

In this project we wish to help closing the gap between training in RDM and research software engineering (RSE). It will require that we adapt our material to be useful to RDM support staff. We plan to do that by providing summaries for the episodes and lessons that we teach. Here it will be less about hands-on practice and more about the basics and awareness of tools and techniques to send the researchers towards the right direction (e.g. to RSEs). It will be very beneficial that the CodeRefinery network is well connected to RSEs in the Nordics. We expect these summaries to also be useful to give group leaders and managers an overview of modern tools and practices for FAIR software development to motivate them to send their students and staff to our workshops.

## Planned activities for 2025–2028

2025:

- Online CodeRefinery workshop (multi-day)
- Kick-off meeting (lunch to lunch)
- Train-the-trainer workshop (2 half-days, online)
- Python for scientific computing workshop (3 half-days, online)

2026:

- Trial CodeRefinery workshop in massive open online course (MOOC) format where some lessons are pre-recorded and live sessions focus on Q&A and discussions (continuous, online)
- Online CodeRefinery workshop (multi-day)
- Writing retreat for lesson development (3-4 days, in-person)
- Workshop at international conference (half-day or full day)
- Python for scientific computing workshop (3 half-days, online)

2027:

- Online CodeRefinery workshop (multi-day)
- Workshop at international conference (half-day or full day)
- Train-the-trainer workshop (possibly at international conference)
- Python for scientific computing workshop (3 half-days, online)
- Workshop on reproducible machine learning workflows (2-3 half-days, online, this is a new course to be developed)

2028:

- Hand-over of activities to the continuation of phase 4
- Closing event for phase 4 (in-person)

## Benefits for partner organizations

**Continuous access to training:** These workshops empower researchers and enhance organizational capacity in digital and computational skills. By being a partner, organizations can influence scheduling and make sure that the training is visible to their students, researchers, and staff.

**Strategic input on curriculum development:** Partners have the opportunity to shape the evolution of present and new course materials, providing an opportunity that training aligns with the specific needs of their researchers and institutional priorities.

**Participation in collaborative training initiatives:** Partnering organizations gain early access and priority in event scheduling to collaborative training events, including train-the-trainer programs, which help build internal capacity for delivering impactful training programs.

**Increased visibility and networking opportunities:** Workshops and events co-hosted with CodeRefinery are promoted to an extensive network of academic institutions, researchers, and trainers, offering partners visibility and opportunities for collaboration on a European to even global scale.

**Representation in the steering group:** Partner organizations receive a seat in the CodeRefinery steering group, providing them with a direct role in governance and strategic decision-making, ensuring their voice is heard in shaping the future of the project.

**Recognition and public acknowledgment:** Contributing partners are acknowledged in every workshop they support and prominently highlighted on CodeRefinery's website and social media channels, showcasing their commitment to advancing research software development.

## Collaboration model

CodeRefinery will continue as a decentralized network with the following structure:

**In-kind contributions:** Partners provide staff time funded by their own organizations to support training and workshop activities. Partner organizations retain full administrative control over their staff but are expected to allocate time for collaborative efforts that benefit the broader CodeRefinery network. The concept is to allocate staff time to work on training efforts that the organization would like to do anyway but collaboratively with other partners and in a broader context.

**Workshops:** CodeRefinery will maintain its popular workshops at minimum once a year.

**Specialized training:** In addition to core workshops, we encourage partners to collaborate on specialized sessions (e.g. data visualization, containerization, reproducible workflows, data management).

**Outreach and engagement:** Partners help the project to promote CodeRefinery workshops to their local and institutional audiences. We see this as a win-win situation so that both the partner organization and CodeRefinery benefits from having better visibility and reach.

## Governance and sustainability plan

NeIC will administrate the collaboration and appoint a steering group chair and a project manager. Collaboration partners will sign a collaboration agreement with NeIC (for reference, here is the collaboration agreement from phase 3 of the project).

In addition to this document we have drafted a governance charter for the project which we expect to be approved in February 2025, before the start of phase 4. The latter document outlines the decision process and the community-facing aspect of the collaboration. The charter also defines what the project actually contains in terms of lessons, data, communication channels, and accounts, and clarifies responsibilities and ownership.

With the in-kind model we hope to make it easier for new organizations to join the project, also possibly only for an event or only short-term. This will require us to communicate the value and benefits. Going forward, we will however also seek national and European funding and the governance charter gives the steering group the possibility to form a consortium applying for external funding on behalf of the CodeRefinery collaboration.

We are also in contact with several other organizations and projects, e.g. The Carpentries, Digital Research Academy, and Open Life Science. These organizations work in the same space to integrate our results with other existing efforts through for example learning paths and instructor/helper exchange. We plan to strengthen our existing connections to research infrastructures like Elixir, European Open Science Cloud (EOSC), and EuroCC. Multiple partners are already involved in these initiatives and common goals and potential collaborations are identified.

## Risks

**Key persons leaving the project** or their organization which partners in this project: It will take time to on-board new staff that is interested in contributing to the overall direction of the project in addition to teaching, lesson development, and facilitating events.

- Mitigation: Giving staff recognition for their work and the possibility for career development. An important part of this will be generous travel support to attend relevant conferences and meetings and being able to engage with other projects in this space.

Lack of commitment to an **in-kind model**: The in-kind model risks putting the CodeRefinery project work at a lower priority compared to externally funded projects. Project team members are typically required to report their work time within their line management/home institutions.

- Mitigation: Support from local management in prioritizing project work. At places where team members are required to report individual hours, organizations need to offer them a cost object (project code) to give them the possibility to allocate a few work hours to this project. We will also seek external funding during phase 4 of the project which should help increasing the priority compared to other local projects.

## Project partners

**Aalto University, Finland**

- Contact: Mikko Hakala (mikko.hakala@aalto.fi)
- Short description: Aalto University is the premier technical university in Finland. As part of providing its own computing resources (and supporting the use of other resources, such as LUMI), the School of Science maintains a strong, open, practical scientific computing teaching program.
- Groups in the organization that will be involved: Science-IT
- Past involvement: Instructors and helpers, outreach, streaming and recording, transition to online teaching.
- Involvement in phase 4: Continuing the level of contribution as in phase 3.

**CSC- IT Center for Science, Finland**

- Contact: Atte Sillanpää (atte.sillanpaa@csc.fi)
- Short description: CSC is a company entrusted with special state assignment and owned by the state of Finland (70% of the shares) and Finnish higher education institutions (30% of the shares). CSC builds digital solutions for data management, scientific computing, and education and helps researchers, learners, and companies to efficiently use these resources. CSC hosts EuroHPC-JU Supercomputer LUMI and has been chosen to set up LUMI-AI and the connected AI factory in the coming years.
- Groups in the organization that will be involved: CSC pools staff and resources from the Advanced Computing Facility and Data Management Services and will allocate one project manager and 1-3 instructors and helpers from this pool.

- Past involvement: Community management and registration coordination, instructor and helpers, outreach.
- Involvement in phase 4: Provide project/community manager, provide helpers and potentially instructors, link to CSC training and outreach.

**NRIS/Sigma2, Norway**

- Contact: Vigdis Guldseth (vigdis.guldseth@sigma2.no)
- Short description: The Norwegian research infrastructure services (NRIS) is a collaboration between Sigma2 and the universities of Bergen (UiB), Oslo (UiO), Tromsø (UiT The Arctic University of Norway) and NTNU, to provide national supercomputing and data storage services in Norway. These services are operated by NRIS and coordinated and managed by Sigma2. Sigma2 provides the national e-infrastructure for computational science in Norway.
- Groups in this organization that will be involved: NRIS pools staff and resources from 4 universities and will allocate 3 instructors from this pool.
- Past involvement: Project management, instructors and helpers, collaboration with the NRIS training coordinator, outreach.
- Involvement in phase 4: Provide instructors and helpers, link to NRIS training, outreach.

**NTNU, Norway**

- Contact: Bjørn Lindi (bjorn.lindi@ntnu.no)
- Short description: NTNU (Norwegian University of Science and Technology) is Norway's largest university. Founded in 1996 through the merger of several institutions, it's particularly well-known for its strong engineering and technology programs, though it offers a comprehensive range of subjects across sciences, social sciences, humanities, and arts.
- Groups in this organization that will be involved: The HPC-group in NTNUs IT-division will be part of CodeRefinery.
- Past involvement: Instructors and helpers, collaboration with NRIS, some outreach.
- Involvement in phase 4: Provide instructors and helpers, link to NRIS training, outreach.

**USIT/University of Oslo, Norway**

- Contact: Sabry Razick (sabryr@uio.no)
- Short description: The University of Oslo is a leading European university. UiO's advantage is high quality throughout the organization and long-term basic research across the breadth of disciplines and activities. Our academic breadth and depth gives UiO a unique starting point for contributing to sustainable solutions that require both new technology and a better understanding of the interactions between people, nature, society and technological solutions.
- Groups in this organization that will be involved: Division for Research, Dissemination and Education. The training work package of Norwegian Ai Cloud (NAIC) project will allocate necessary funding for the contribution.
- Involvement in phase 4: Provide instructors and helpers, link to NRIS training, outreach, and using Coderefinary training material in other courses and projects.

**CHC/DeiC, Denmark**

- Contact: Kristoffer L. Nielbo (kln@cas.au.dk)
- Short description: The Center for Humanities Computing (CHC) at Aarhus University is a multifaceted research and development unit within the Faculty of Arts that bridges the gap between humanities and technology. It focuses on cutting-edge research in areas such as language technology, artificial intelligence, and cultural dynamics, while also providing essential services and support for computational and digital humanities projects. The CHC offers data management, high-performance computing access, and project support, collaborating with partners to develop and maintain apps for UCloud, a GDPR-compliant HPC platform. As a research liaison, the center advocates for technological solutions tailored to humanities researchers and offers various engagement opportunities, from consultancy to full partnerships, fostering interdisciplinary collaboration and advancing the integration of technology in humanities scholarship.
- Groups in this organization that will be involved: CHC
- Past involvement: Support in online teaching, DeiC is operating the GitLab service over many years now (we are in the process of handing this service over to DeiC).
- Involvement in phase 4: Provide instructors and helpers, lesson development.

**ENCCS, Sweden**

- Contact: Thor Wikfeldt (thor.wikfeldt@ri.se)
- Short description: The EuroCC National Competence Center Sweden (ENCCS) provides high-performance computing training and support for industry, academia, and public administration.
- Group(s) in this organization that will be involved: The ENCCS group itself.
- Past involvement: In the previous phase we have co-organized a number of events which were either led by ENCCS or CodeRefinery or Aalto Science-IT or PDC Stockholm, on Python, containers, GPU programming, and build tools.
- Involvement in phase 4: Provide instructors and helpers, assist in organizing and hosting of workshops, outreach, and using CodeRefinery training material in other courses. We will collaborate on moving some of our lessons to asynchronous format where lessons are pre-recorded.

**NAISS, Sweden**

- Contact: Joachim Hein (joachim.hein@naiss.se)
- Short description: The National Academic Infrastructure for Supercomputing in Sweden is operating high performing computing resources, large storage and sensitive data solutions, and training for higher education institutions in Sweden. NAISS is the largest and most widely used research infrastructure in Sweden.
- Groups in this organization that will be involved: The participating organization is NAISS. NAISS' CodeRefinery activity will become a part of NAISS training. NAISS will appoint staff for its CodeRefinery activities. CodeRefinery will have access to the expertise within NAISS training. A number of persons who have been involved in CodeRefinery for many years have now joined

NAISS as part of branch agreement. These staff members are expected to deliver a significant contribution of NAISS' CodeRefinery effort.

- Past involvement: Early Swedish involvement was mostly through two groups. PDC Center for High Performance Computing was among the initial partners of CodeRefinery in 2016 and has been a partner since. Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) joined CodeRefinery in 2020. NAISS has supported CodeRefinery since spring 2023 by providing teachers and helpers for CodeRefinery events as part of NAISS training. Staff based in Sweden has over the years contributed to lesson development, teaching, and hosting of individual events, onsite as well as online.
- Involvement in phase 4: Provide teachers for upcoming courses, contribute to the development of new teaching material, assist in organizing and hosting of workshops, and take a lead role in organizing selected events.

**University of Iceland**

- Contact: Ebba Þóra Hvannberg (ebba@hi.is)
- Short description: The University of Iceland (UICE) is a leading European university. UICE is a progressive educational and scientific institution, renowned in the global scientific community for its research. It is a state university.
- Groups in this organization that will be involved: Department of Computer Science and the EuroCC National Competence Center Iceland.
- Past involvement: This is the first time Iceland is participating in CodeRefinery as a partner.
- Involvement in phase 4: Provide instructors and helpers, assist in organizing and hosting of workshops, outreach, and using CodeRefinery training material in other courses and projects.

# Personnel contributions (in-kind)

All contributions are in-kind and listed in person-months (PM).

| | 2025 | 2026 | 2027 | 2028 | Total |
|---|---|---|---|---|---|
| Aalto | 5—10 | 6—12 | 6—12 | 1—2 | 18—36 |
| CSC (coordination) | 3—5 | 4—6 | 4—6 | 1 | 12—18 |
| CSC | 5—6 | 6—8 | 6—8 | 1—2 | 18—24 |
| NRIS/Sigma2 | 5 | 6 | 6 | 1 | 18 |
| NTNU | 2 | 2.5 | 2.5 | 0.5 | 7.5 |
| USIT/UiO | 4 | 4 | 0 | 0 | 8 |
| CHC/DeiC | 5[a)] | 6[a)] | 6[a)] | 1[a)] | 18[a)] |
| ENCCS | 1.5 | 2[b)] | 2[b)] | 0.5[b)] | 6[b)] |

| | | | | | |
|---|---|---|---|---|---|
| NAISS | 5 | 6 | 6 | 1 | 18 |
| University of Iceland | 2 | 2 | 2 | 0.5 | 6.5 |
| Total in-kind PMs | 37.5—45.5 | 44.5—54.5 | 40.5—50.5 | 7.5—9.5 | 130—160 |

Notes:

a) The actual commitment might be larger than indicated. This depends on a pending grant application.
b) The current funding phase for ENCCS ends 2025. Contributions after 2025 rely on follow-up funding. In-kind support will be provided on a best-effort basis.

## Travel budget and other costs (covered by NeIC)

All values are in EUR. **These costs will be invoiced to NeIC**.

| | 2025 | 2026 | 2027 | 2028 | Total |
|---|---|---|---|---|---|
| Allocated NeIC Travel budget | 2000 | 3000 | 3000 | 1000 | 9000 |
| Required conference/outreach budget [a)c)] | 5000 | 7000 | 7000 | 2000 | 21000 |
| Required costs for team meetings [b)c)] | 4000 | 8000 | 8000 | 4000 | 24000 |
| Web domain | 40 | 40 | 40 | 40 | 160 |
| Newsletter and support desk | 400 | 500 | 500 | 100 | 1500 |
| Promotion/marketing | 200 | 300 | 300 | 100 | 900 |

Notes:

a) This is to give each team member the possibility to travel once to an international conference during the 3-year project period to promote the project. This would enable team members to be able to talk about the project, provide in-person training, and bring in new ideas, perspectives, and collaboration opportunities back into the project. This is essential if our goal is sustainability and growth.
b) In order to maintain the collaboration we believe that yearly team meetings are necessary. Our calculation assumes a 3-day meeting in the Nordics once per year for 8-10 participants. This is

now even more important since we need to find a different meeting venue after the previously yearly NeIC all-hands meetings have been discontinued.

c) We cannot leave it up to the participating organizations to cover travel to conferences and meetings in addition to work time. Leaving this up to organizations to decide this later puts the project at risk of not being able to do outreach and continue our success story.

## Our achievements in the last 3 years

Below we summarize our achievements during the last phase. We have also recently published a blog post in collaboration with partners and supporters celebrating eight years of CodeRefinery.

**Sustained momentum and vision:** Even after eight years since project start, **CodeRefinery continues to grow**, innovate, and inspire. We transitioned to fully online workshops and embraced the **streamed format**. The streamed format encouraged inclusivity by lowering logistical barriers like travel costs. Through refinement of the streaming setup we can often make the **recordings** available the same day, which enables learners to re-watch or catch-up and continue the workshop without missing out. We learned so much along the way. But as the tools evolve, so needs our workshop and the format.

**Reaching more learners:** Over the last 3 years, we held 6 large streamed online **workshops with an increasing number of participants** and multiple smaller related workshops as collaborative efforts (workshop statistics). We are currently piloting and developing **"bring your own code"** sessions, about a month after the workshop, where participants can meet instructors and helpers to discuss and ask questions about applying what they have learned in the workshop to their own work.

**Instructor development:** We host a community of skilled instructors who can effectively deliver our workshops and contribute to our mission. Our approach to instructor development includes:

- **Train-the-trainer events** provide new and aspiring instructors with the skills and confidence to teach CodeRefinery workshops, fostering a supportive and knowledgeable teaching community.
- **Co-Instructor onboarding and mentoring** opportunities for new instructors by pairing them with experienced co-instructors, enabling hands-on learning and collaborative teaching experiences.
- **Collaborative lesson development:** Instructors actively participate in developing and refining lesson materials, to keep the curriculum up-to-date, relevant, and reflective of diverse perspectives.

Through these initiatives, we were able to create an inclusive environment where instructors can grow, share expertise, and contribute to advancing our materials.

**Community engagement:** We have a vibrant supportive community of learners and contributors from diverse disciplines and career stages. We held several **Open House** events on topics like the workshop format and advanced Git curriculum. The "**Bring Your Own Classroom**" initiative encouraged local groups to participate collectively in workshops, fostering collaborative learning environments (blog post). We also started the **CodeRefinery ambassador program**: The program is meant for people who like what we

do and would like to help spread the word to their community. An ambassador does not have to be an active CodeRefinery instructor, helper or even community member. It is highly appreciated by the community that we send out occasional summaries of everything that happens in our chat as "**chat digest"** (archive) to people that are too busy to follow everything going on. Our **Zulip chat** is now home to almost 500 people. The workshop format and several lesson materials got an overhaul by the team during our **in-person writing retreat**. Meeting in-person in beautiful Tromsø boosted the team spirit and we got to know each other better.

**Outreach activities and recognition within the RSE community:** We have been actively spreading the word via **talks and posters** at conferences (overview), invitation to workshops, podcasts, and panels. A community manager was hired. We increased **social media outreach and engagement** for events and to celebrate achievements and share blog posts. The name CodeRefinery became a recognizable brand in the RSE community.

**Resource access and reuse:** We are committed to making our lesson materials as accessible, reusable, and impactful as possible. To achieve this, we are actively working on applying FAIR principles to our resources, including making them citable. This ongoing work is summarized in a blog post.
Here are some of the steps we are taking to improve resource access and reuse:

- **Enhanced lesson descriptions:** By providing detailed descriptions for each lesson (overview), we aim to make it easier for users to discover and identify materials relevant to their needs, increasing their overall impact.
- **Extensive manuals:** Our comprehensive manuals guide both instructors and learners, with the goal to keep our resources accessible, and easy to implement and adapt for different audiences.
- **Tracking engagement:** We monitor usage metrics, including recording watcher numbers and lesson page visits, to understand how our materials are being used and to inform future improvements.

**Positive feedback and impact stories:** We have collected a wealth of feedback (mainly positive, but also constructive negative) from our daily workshop notes as well as long term impact of our workshops (long term survey results).

We worked out a **governance charter** (draft to be approved) for the project to define the roles and responsibilities of different actors involved in the project. The goal is to increase flexibility and lower the barrier for new partners.