NeIC - CodeRefinery 2

Project directive

The purpose of the project directive is to form the basis for a decision on starting the a second phase of the CodeRefinery project (DP1, see annex 1) and to define the boundaries for the preparation phase.

Background

Research heavily relies on software and too often research software is developed inefficiently, often without peer review, and without considering reproducibility and reusability. The CodeRefinery addresses these issues for the Nordic research communities with training workshops and infrastructure services.

The proposed project is a follow-up of the CodeRefinery project (phase 1), a 2-year project which started in October 2016 and is planned to conclude in September 2018. During phase 1, the project will have delivered 13 three-day workshops across the Nordics and 5 shorter events and taught over 400 students and researchers.

The project has informal ties to The Carpentries[1] and the Software Sustainability Institute[2] and can be seen as an extension of The Carpentries to more advanced programming skills and the beginning of the formation of a Research Software Engineering network in the Nordics. The CodeRefinery project has developed a very fruitful informal cooperation with Aalto Science-IT which resulted in two workshops and collaborative lesson development.

In phase 1, we learned that students and researchers who write software to produce or analyze data are in critical need of the proposed training portfolio on software management, and currently only NeIC has the scale, network, and expertise to fill this need in the Nordics until these workshops become part of university curriculum. Phase 1 of the project could already match this obvious need of the research community exceptionally well as shown by almost always fully booked workshops and the resulting surveys.

We are convinced that the project requires commitment to phase 2 right after or soon after phase 1 will conclude in order to keep the momentum of an established motivated team and to reach the necessary scale, visibility, and community involvement to make this initiative self-sustained.

Project idea

Picking up on the success of the previous CodeRefinery project, the second phase will contribute to enhancing the productivity of researchers in the Nordic region and in addition to the goals for phase 1 serve as a Nordic hub for research software engineers.

Expected benefit

The expected benefit is an increased competence in research software development which will lead to more efficient collaboration and produce more reproducible and reusable research software.

---

[1] https://carpentries.org
[2] https://www.software.ac.uk

1. Increased competence (continued from phase 1)

1.1 More competent scientists and researchers

Software development is a core part of research for many scientists and researchers in the Nordics. This project will contribute to an increased competence in software development tools and techniques among scientists and researchers.

1.2 More competent support staff

The preparation and delivery of CodeRefinery workshops and instructor workshops is a great learning opportunity also for the CodeRefinery team and indirectly also their colleagues.

2. Better science (continued from phase 1)

2.1 Reproducibility of research code

The expected benefit to the national e-infrastructure provider organizations and NeIC stakeholders will be a user community trained in best practices of modern software development with focus on reproducibility, collaboration, and peer review.

2.2 Reusable and extensible code

The training focus will not only be on reproducibility of published computational results but also on helping researchers creating reusable, modular, and extensible code where code modules and libraries can be composed to solve new problems according to FAIR principles (findable, accessible, interoperable, and reusable).

2.3 Collaborative learning and knowledge transfer

Peer review in programming and collaborative code development will be a core pillar of the training with the goal of not only improving code quality but also facilitating knowledge transfer within research groups and across research groups.

3. Strengthening Nordic communities and relations (new in phase 2)

3.1 Train-the-trainer program

To reach scale we will not only train the research community but also promote, train, and certify future trainers at yearly train-the-trainer workshops. At the same time we will improve competences and promote development of the project staff which will benefit the infrastructure providers.

3.2 Symbiosis with Software Carpentry, Data Carpentry, HPC Carpentry workshops, and PRACE advanced training centers

While The Carpentries provide often the introductory steps for researchers into programming, scripting, and version control, CodeRefinery can be positioned as the second step, focusing on collaboration and reproducibility, in particularly targeting researchers who already develop code. This project will actively foster collaboration and common projects with The Carpentries to catalyze Carpentry workshops, recruit instructors from the Carpentries ecosystem, and to enable knowledge and lesson transfer in both directions by contributing lesson material to the Carpentries "contributed lessons" and proactively approaching

Carpentries instructors to teach at CodeRefinery workshops. The national project partners will benefit by having access to customized instructor training with certification (15 persons each year), administrative support for 6 coordinated Software Carpentry workshops and generally more Carpentry workshops catalyzed by the instructor training and the coordination assistance. The advertising of CodeRefinery workshops will be coordinated with related courses and workshops organized at PRACE advanced training centers such as CSC IT Center for Science.

3.3 Symbiosis with local units

We will not just provide services and training ourselves, but (where desire exists) partner local universities to bring our products to their community as a default service. Aalto Science IT will be our partner when organizing workshops at Aalto/Espoo/Helsinki. Aalto Science IT has committed to organize two workshops a year and contribute one instructor to each of these workshops.

3.4 Implementation of a Research Software Engineers (RSE) community in the Nordics

CodeRefinery will serve as a hub for a growing Research Software Engineers community in the Nordics. This effort will be catalyzed by contributing to a survey of the RSE landscape, providing an open chat platform, and by organizing yearly workshops. One goal of the phase 2 project will be to organize a Nordic conference for RSEs.

Basis

- Mid-term report: http://cicero.xyz/v2/remark/github/coderefinery/report/master/talk.md/
- SG group meeting minutes recommending a follow-up: https://wiki.neic.no/wiki/CodeRefinery-SG-meeting-2017-12-05#4._Planning_the_project_follow-up
- Lesson material: http://coderefinery.org/lessons/
- Long-term impact survey data: https://github.com/coderefinery/post-workshop-survey/blob/master/survey_analysis.ipynb
- List of past workshops: http://coderefinery.org/workshops/
- Feedback of past workshops (available upon request)

Contact persons

Name	Email	Role
Hans Eide	hans.eide@uninett.no	SG member phase 1 (Sigma2)
Lene Krøl Andersen	lene.krol.andersen@deic.dk	SG member phase 1 (DeiC)
Kristoffer Laigaard Nielbo	knielbo@sdu.dk	SG member phase 1 (DeiC)
Michaela Barth	caela@kth.se	project owner phase 1
Rossen Apostolov	rossen@kth.se	SG member phase 1 (SNIC)
Pauliina Somerkoski	pauliina.somerkoski@csc.fi	SG member phase 1 (CSC)

Richard Darst	richard.darst@aalto.fi	Aalto Science IT http://science-it.aalto.fi
Radek Lonka	radek.lonka@ntnu.no	Research software engineer at NTNU Trondheim
Anne Claire Fouilloux	a.c.fouilloux@geo.uio.no	Software Carpentry instructor trainer
Lex Nederbragt	lex.nederbragt@ibv.uio.no	Software Carpentry Steering Committee member
Radovan Bast	radovan.bast@uit.no	project manager phase 1
Bjørn Lindi	bjorn.lindi@ntnu.no	team member phase 1 (Sigma2)
Sabry Razick	sabry.razick@usit.uio.no	team member phase 1 (Sigma2)
Thor Wikfeldt	kthw@kth.se	team member phase 1 (SNIC)
Jyry Suvilehto	jyry.suvilehto@csc.fi	team member phase 1 (CSC)
Sri Harsha Vathsavayi	sriharsha.vathsavayi@csc.fi	team member phase 1 (CSC)
Hedi Peterson	hedi.peterson@ut.ee	ELIXIR Estonia coordinator (UTARTU)
Juhan Ernits	juhan.ernits@ttu.ee	Programme Director (Software Engineering), Tallinn University

Timeframe and estimates for the preparations

Commitment up to	Date	Estimated effort	Estimated expenditures
DP2: preparation phase	2018-08-01	0.5 person month	
DP3: approval of project plan	2018-10-01	0.5 person month	Either cost of one steering group meeting or 0 if video meeting (apart from in-kind contribution of work time).

Project goals

Result goals and quantitative targets

The goals of the project remain the goals of the phase 1 (providing researchers with infrastructure and training in the necessary tools and techniques to create sustainable, modular, reusable, and reproducible

software). Phase 2 will sharpen the project focus, increase visibility and scale, and integrate better with the community. A new goal of phase 2 is to reach sustainability after phase 2 is concluded.

Quantitative targets (key performance indicators):
- 12 carpentry instructors trained each year (ca. 3 per participating country)
- 3-8 CodeRefinery workshops organized per year (depending on the staff resources)
- 1 train-the-trainer workshop per year
- 1 research software engineers (RSE) workshop/meet-up per year (collocate with train-the-trainer workshop)
- Organization of 1 RSE conference

1. Sharpen the project's focus on training and knowledge

Reduce organisational overhead by employing a training coordinator

The position that the phase 1 was missing the most was administrative support in outreach and coordinating workshops and instructors. The training coordinator (0.5 FTE) would have the following tasks:
- Engage with NeIC Ratatosk Training Coordinator
- Engage and market to local universities
- Be the contact point for workshop requests
- Scheduling of instructors and helpers for workshops
- Room reservations
- Advertising of workshops
- Travel reimbursement coordination
- Registration for the workshop
- Surveys: pre- and post-workshop
- Sending out course certificates (which do not include ECTS credits)
- Outreach via Twitter
- Coordinate blog-posts
- Newsletter (alternatively using Twitter)

Also we can seek inspiration from https://software-carpentry.org/workshops/operations/.

Registration procedure and administration of participants is streamlined

Managing registrations for workshops can be non-trivial due to room limitations, cancellations, waiting lists, workshop preparation. The CodeRefinery team relies on a set of tools to manage workshops but for phase 2 these routines not only have to be GDPR-compliant, they also have to be smooth and decentralized.

Infrastructure services are delegated using service level agreements

Infrastructure services such as the source code repository hosting are not deliverables anymore, bound to FTE resources, but are delegated using service level agreements which are transparent to the users of the services. This way we engage partners within the Nordic e-Infrastructure Collaboration with the benefit that these services do not bind project staff directly and that support personnel operating these services are not pulled into an administrative overhead of reporting and engaging in meetings which may not be relevant to their work. The steering group of the phase 2 project will have the responsibility to evaluate the use of the provided services.

A goal of this phase of the project would be to hand over not only the operation of the infrastructure services but also the financing to national infrastructure providers for long-term maintenance.

In contrast to phase 1, the service level agreements must also contain support to avoid the situation that support requests are directed to CodeRefinery staff (or manager) who has little direct technical control over

the services to answer support requests. The challenge will be to render the support uniform from the user perspective.

Decentralize support structure

The support structure of phase 1 was very much centralized around the project manager, even though the project introduced a support request ticketing system. The dependency on the PM is unsustainable both for the project and for the PM and phase 2 needs a clear separation of responsibilities with respect to questions and support requests. Not every single request should need to go through the PM in contrast to phase 1.

2. Visibility and accessibility

Increase visibility

After two years of project work the project has reached the visibility that institutions and research groups approach the project and request workshops. One result goal will be to increase visibility by marketing the project not only to local support staff at universities, but also by directly approaching managers and group leaders. For this the project will prepare a short blog post and presentation specifically targeting managers and group leaders and motivating why they should consider sending their staff and students to CodeRefinery workshops. The training coordinator will take an active role in reaching out over Twitter with at least weekly posts. We will publish a call for workshops where institutions can apply.

Federated access for infrastructure services

For services to reach scale they need to become part of teaching at university courses and for this often the current lack of federated authentication is a barrier that the phase 2 of the project will aim to overcome. The project will provide Feide/eduGain/Haka access to services and advertise the services to teaching staff. Since this task will not be core to the project, we will consider to achieve this in cooperation with the Glenna project.

3. Larger scale and reach

The community need for training is estimated to be tenfold compared to the present offer. To make the biggest impact towards a systemic change in research software development, provided the funding envelope, the phase 2 of the project will actively seek, encourage, and involve volunteer instructors participating in-kind (albeit travel expenses reimbursed) and encourage satellite/spin-off workshops reusing the workshop material.

More targeted and invited workshops

Phase 2 will put more focus on quality and preparation of workshops in order to align better with available resources. Instead of "cold-calling" universities to host workshops, the project will put emphasis on invited workshops since it simplifies coordination and lifts a significant amount of administrative effort as well as financial load. In addition to traditional three-day workshops the project will offer more targeted 1- and 2-day workshops on special topics.

Larger instructor pool

This goal has two aspects: allowing small FTE-contributions from project partners and thus enlarging the instructor pool, as well as, providing infrastructure for guest instructors to participate in teaching. To establish a larger instructor pool the project will organize workshops for instructors where anybody who is interested can join. In order to accommodate guest instructors this the project needs to offer a simple reimbursement strategy to make it easy for volunteers to become instructors at workshops. The added benefit will be not only more workshops without additional FTE-engagement for the partners, but also networking and knowledge transfer across organizations. The project will encourage satellite workshops that build on top

of existing CodeRefinery teaching material and also seek partnership with local partners closer to the research community such as the libraries.

Train-the-trainer concept to reach scale

In order to reach scale the goal of the project needs to be to train future trainers and establish a certification scheme. For this we will ask former participants to become helpers in future workshops nearby. Helpers will be encouraged to participate in the Software Carpentry instructor training and CodeRefinery will require that all instructors take the Software Carpentry instructor training. The project will establish workshops for instructors where they can meet, discuss, and improve material and skills.

Scale and visibility with online material

In addition to training workshops which only have a limited scale, the project will develop and market super short best practice guides and develop brief (5 minute) screencasts demonstrating best practices in action.

4. Community and integration

Foster integration with The Carpentries

The goal of the project phase will be that lesson material will be contributed to https://software-carpentry.org/lessons/. CodeRefinery will invite Software Carpentry instructors to teach at CodeRefinery workshops and also encourage CodeRefinery staff to contribute to Carpentries workshops with the aim to recruit trainers in both directions and to participate in the Software Carpentry and Data Carpentry instructor training.

Catalyzing a Nordic Research Engineers community

CodeRefinery will serve as a hub for a growing Research Software Engineers community in the Nordics. We will provide an open chat platform for stakeholders, team, and users/learners to meet, interact, and discuss. This community will not be forced but grow out of the present good communication culture within the team and close stakeholders organically. For this we plan to open the chatroom from currently Slack and invite-only to a more open platform where anybody can join. Since most staff will only contribute part-time to the project we will consider establishing office hours where persons are present in the chat room to discuss and answer questions. The project will also consider to offer lightweight post-workshop mentoring without having the goal/focus on community building. The open chat could also become the principal entry point for support requests which are currently 1-1 and require a paid subscription.

Establish a Centre of Excellence on Software Sustainability in the Nordic Countries

Similar to the Software Sustainability Institute in UK but with a larger range of actions, including co-design activities (as suggested in section 6.3 "Programming Environment" of the ETP4HPC Strategic Research Agenda) and strong coordination between national, European and international activities.

Organizing a Nordic conference around the RSE community

The project will involve people with RSE-type positions and organize the first Nordic conference with this focus which will catalyze new collaboration and the Nordic RSE community.

Time goals

Since this is a follow-up project, the team can start delivering workshops right from project start provided there are no large changes in staff. The latter point requires that there is no significant break between phase 1 and phase 2. The goal will be to organize a Nordic RSE conference within the first year of the project. Best

practice guides will be drafted at the beginning of the project. Mid-term and final report time-line will be standard and will be detailed in the project plan.

Cost goals

See "Financing" section below.

Project objective priority

Top priority is to educate the research community in modern collaborative software development workflows with emphasis on reproducibility and peer review in software development. The duration of the project is planned to be 3 years. Since this project would be a continuation of a successful 2-year project, the time and cost boundaries are well understood and calibrated.

Result	Time	Cost
0.6	0.3	0.1

Financing

Duration

Phase 1 of the project was well received among the research community with essentially always full attendance of workshops. For phase 2 we request support for a **project duration of 3 years** to assure continuity and also to provide enough time to develop a self-sustained project by transferring the financing of services to national infrastructure providers and by transferring lesson material and lesson development to The Carpentries community.

Staff budget

The project will be co-financed with 50% contributions from Denmark, Estonia, Finland, Norway, and Sweden.

Budget option 1 for growing the project

Delivering 8 workshops per year

Partner	Staff phase 1 (previous project)	Staff phase 2 (current proposal)
Denmark	0.5 FTE	0.5 FTE
Estonia		0.5 FTE
Finland	0.5 FTE	0.5 FTE
Norway	0.5 FTE	0.5 FTE
Sweden	0.5 FTE + 0.25 FTE (SGAS project)	0.5 FTE
Project management	0.5 FTE	0.5 FTE

Administrative support		0.5 FTE
sum	2.75 FTE	3.5 FTE

Theoretically, the administrative support could also provided "outside" of the project e.g. within NeIC's Ratatosk Training Programme.

Budget option 2 for keeping the project alive

Delivering 3 workshops per year

Partner	Staff phase 1 (previous project)	Staff phase 2 (current proposal)
Denmark	0.5 FTE	0.25 FTE
Estonia		0.25 FTE
Finland	0.5 FTE	0.25 FTE
Norway	0.5 FTE	0.25 FTE
Sweden	0.5 FTE + 0.25 FTE (SGAS project)	0.25 FTE
Project management	0.5 FTE	0.25 FTE
Administrative support		
sum	2.75 FTE	1.5 FTE

Theoretically, the administrative support could also provided "outside" of the project e.g. within NeIC's Ratatosk Training Programme.

Travel and other budget

Operation of the source code repository: 50k NOK/ year (unchanged).

Workshops and meetings: 80k - 160k NOK/ year (depending on staff budget).

Conference budget: 30k - 60k NOK/ year (depending on staff budget).

Other costs (domain name, registration pages, promotion, support line): 3k NOK/ year.

Conference organization: 50k NOK.

Carpentries partner membership at the price of a gold membership and a 5000 $ discount (resulting in 10000 $ annually) including access to customized instructor training with certification (15 persons each year), administrative support for 6 coordinated Software Carpentry workshops and generally more Carpentry workshops catalyzed by the instructor training and the coordination assistance.

Other

Steering group

We propose to continue with the steering group which brings the phase 1 of the project to a conclusion (see "Contact points" above) plus a potential Estonian steering group member.

Known risks

Description of risk	Probability	Impact	Priority	Responses and responsible
High administrative and coordinating load to organize workshops in case the project does not employ coordinating staff.	high	high	high	Allocate resources for training coordination and administration.
Support for services and service maintenance are disconnected.	high	medium	low	When drafting service level agreements, allocate and plan for support. Responsible: PM
Insufficient in-kind contribution from partners.	medium	medium	medium	Dialogue with management to motivate and free up in-kind resources for teaching.

Opportunities

Please see the section "Expected benefit" above.

Annex 1 – Terminology

Decision points

During the life span of the project from startup to termination, a number of formal decisions must be made by the steering group. These fall into eight different types; which are numbered in the chronological order in which they are typically made.

DP1 – Decision point type 1; steering group decision to start the project, based on the project directive.

DP2 – Decision point type 2; steering group decision to continue, change or interrupt the project based on findings during the preparation phase. A project may have multiple DP2.

DP3 – Decision point type 3; steering group decision to approve the project plan developed during the preparation phase. Typically this is tied to a DP4 decision to start the execution phase.

DP4 – Decision point type 4; steering group decision to start the execution phase.

DP5 – Decision point type 5; steering group decision to continue, change or interrupt the project based on findings during the execution phase. A project may have multiple DP5.

DP6 – Decision point type 6; steering group decision to approve the result of a delivery, for example to end users. A project may have multiple DP6.

DP7 – Decision point type 7; steering group decision to transfer the responsibility for a delivery, typically to operations in a receiving organization.

DP8 – Decision point type 8; steering group decision to approve the final report and terminate the project.