

CodeRefinery (phase 2) project plan

Version history:

- v1.3 (2021-10-07, "NeIC matching" for 2021 in Table 3.1 corrected from 23 to 22 PMs)
- v1.2 (2021-02-10, moved 7 PMs from ETAIS and DeiC to Aalto Scientific Computing) -
- v1.1 (modified ETAIS contribution from 7 PMs to 6 PMs)
- v1.0 (approved 2019-08-29 [by the steering group](#))

1 - Idea and objective

Picking up on the success of the phase 1 CodeRefinery project ([report](#)), 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.

The objective is to render this project largely self-sustained after phase 2 (after October 2021). This will depend on developing institutional recognition for the project, which should be feasible as it is the universities and the research institutes who benefit from Coderefinery. We will elaborate on this point under **2 - Tasks**.

For further background, objective, and benefit analysis we refer to the [project directive](#).

2 - Tasks

2.1 - Reaching sustainability

A completely self-financed organization is not realistic and not our goal but our aim is to arrive at a project which can continue using 0.5 - 1 FTE/year for coordination, funded either by NeIC or by engaged national or institutional partners.

After phase 2, travel and accommodation for workshops will be financed by the hosting institution, similarly to the Carpentries model. In contrast to the Carpentries model, however, we will not ask the hosting institution for a workshop fee.

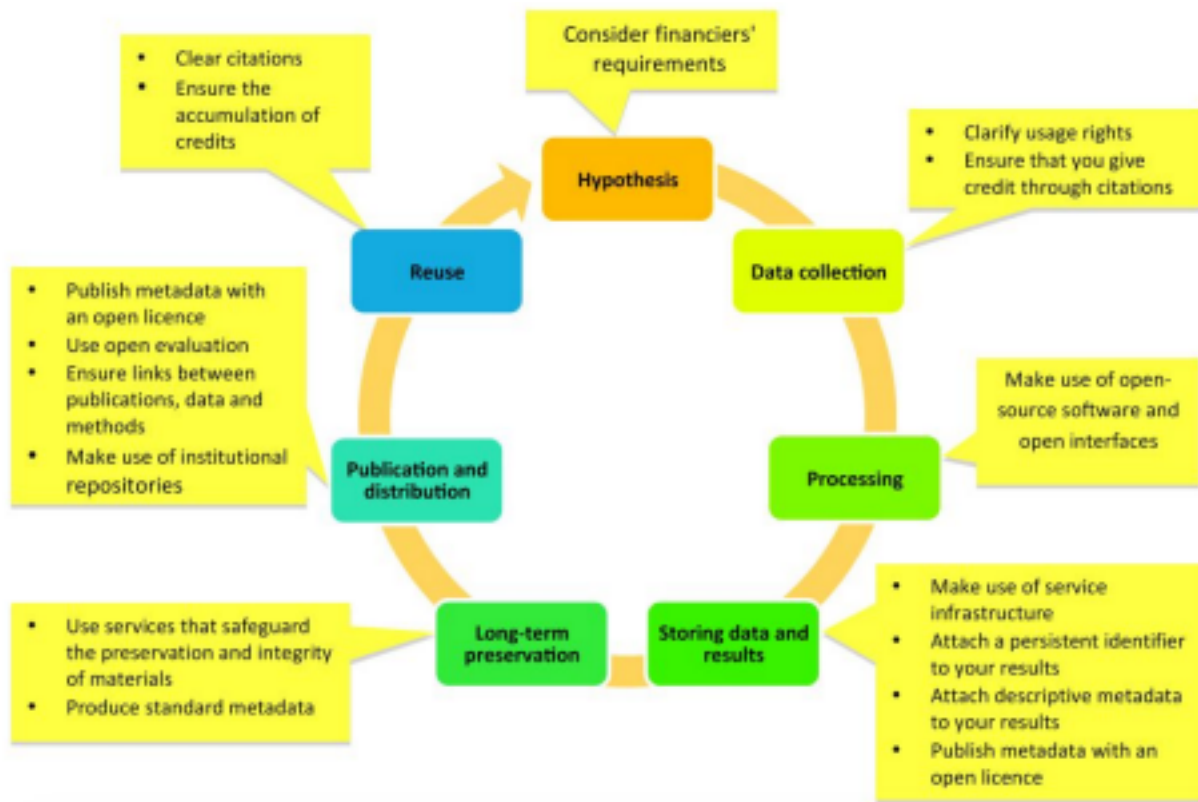
The lesson material will be maintained by the community and volunteers but employers will be encouraged to allow instructors across nordics to participate in-kind in 1-2 workshops/year. The added bonus for the participating institutions will be improved networking across borders and learning opportunity both in teaching and in the actual collaborative tools. Workshops will be arranged with two local and one travelling instructor, again to encourage networking and

experience sharing while also saving costs. Initially, this will only be possible in a few nodes such as Oslo, Stockholm, and Aalto, but by growing the community of local instructors this will become easier with time in other places, as well.

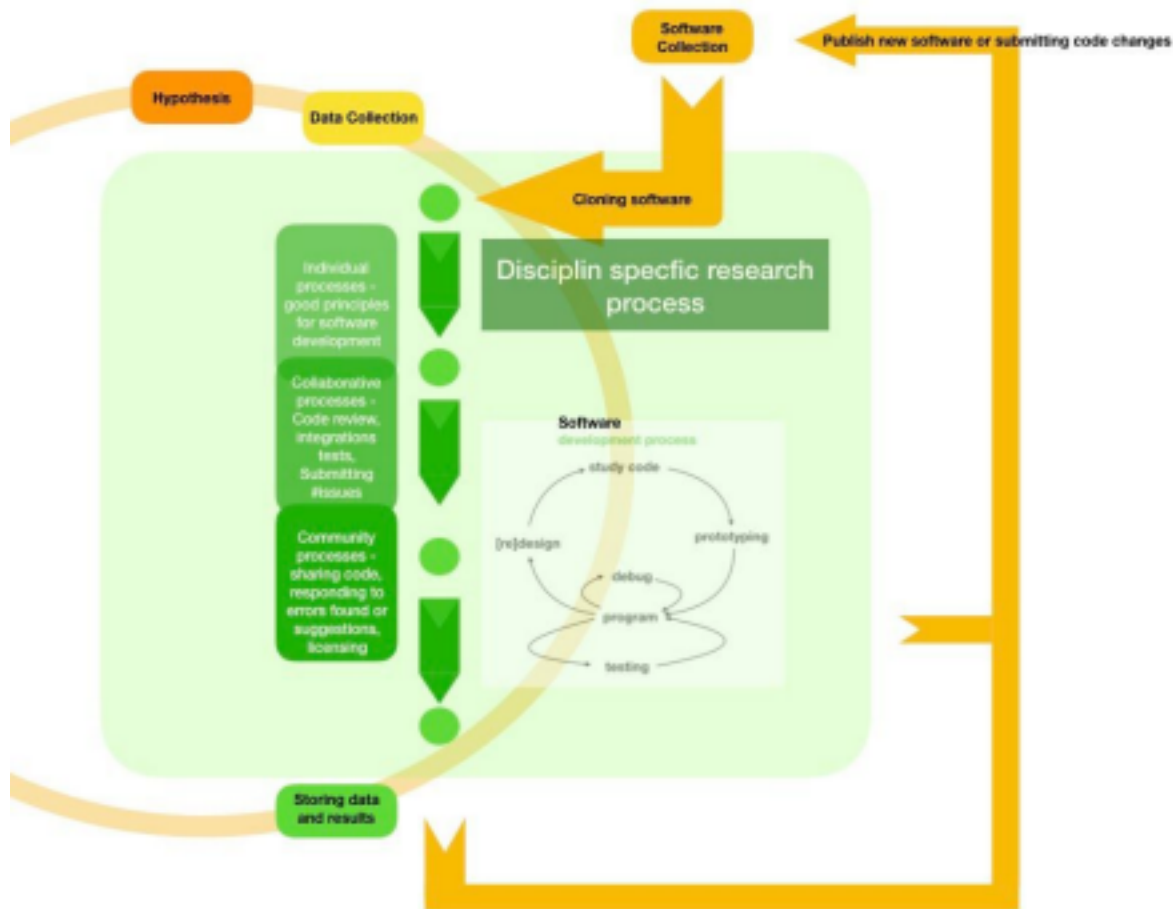
Convincing local e-Science initiatives to support CodeRefinery

CodeRefinery builds necessary competence at academic institutions. This is very fundamental competence needed in an Open Science environment. If we look at the Open Science lifecycle, CodeRefinery provides training in skills needed for interacting with and making use of Open Science resources.

(figure taken from <https://www.fosteropenscience.eu/content/what-open-science-introduction>)



Zooming in on 'Processing', we see that to make use of 'Data Collection', software is very much a part of the picture, either by use of existing software or by writing new software.



The discipline-specific research process is accompanied by a generic software process part. CodeRefinery is concerned with establishing the software process at the individual level. Through the workshops, participants are trained in better individual practices, how to collaborate on software development, and how to interact with a software community. These are necessary skills which are not covered well in current academic programs.

If especially the universities pool some of their resources together, their local e-Science initiatives can embrace and sustain CodeRefinery as a program which makes the local e-Science resource better at covering a wider set of skills. CodeRefinery's ambition is to get local e-Science initiatives to allocate 0.25 FTE each for community events, lesson development, and workshop delivery. The arguments for this level of institutional support are as follows:

- It is the institution's PhD-students and researchers attending a workshop who benefit from the project
- The “field” (research software engineering) is in a state of flux (cf. project Jupyter). Participating in a network/project makes it possible to learn from each other. • Allowing staff to participate in workshops as instructors attracts staff who will attract workshops to the local institution.
- Their employees become experts and share the knowledge inside their institutions/groups as well.

- By having local e-Science initiatives cooperating on providing these necessary skills, good Code of Conduct principles already established in the Open Software community can be more easily transferred to the scientific field.

To nudge universities and e-Science initiatives towards contributing to CodeRefinery workshops, we will develop material for a 1 hour seminar directed towards university and research leaders with the purpose of showcasing CodeRefinery success stories and highlighting the value brought to researchers by participating in and contributing to workshops.

For the infrastructure service (GitLab repository hosting; currently 50k NOK/year) we will work towards anchoring the service within the EOSC Nordic project and in the long term find a financing model where participating countries can contribute a share which roughly reflects the usage of the service.

2.2 - Building a community

Building a community will be an essential requirement to reach sustainability.

In order to motivate institutions and departments to participate, we will need to make our work more visible to the departments. For this we will need to highlight on our website what the benefits are for the departments and partners to participate.

We will need to actively seek out contacts and encourage them to contribute to workshops as instructors or helpers. To simplify onboarding, instructor training material specific to teaching CodeRefinery workshops should be developed. To reach scale, we need more instructor trainers, and in fact many of the project staff should become instructor training instructors. It will also be important to recruit more existing Carpentry instructors to teach CodeRefinery workshops, and for this the Carpentries membership will be key. These activities should happen early in the project.

We will co-finance half a position for a Carpentries Regional Coordinator for Nordics & Baltics, for the two final years of the project phase 2. This position will not only unify coordination of Carpentries workshops and instructor training across Nordics and Baltics, but also provide a better link between CodeRefinery and the Carpentries and help improve administrative routines and workshop planning also for CodeRefinery workshops.

To recruit new volunteer instructors and partners, we will add a module at the end of all workshops about community building, where we invite to participate in instructor training, join our chat, etc.

We will also use the results of the Nordic RSE survey: these are contacts to engage in RSE meetups and the RSE conference which we will organize in 2020.

We will organize yearly hackathons where Nordic RSEs can meet and collaborate on short term projects which can spawn longer term collaborations and networks.

We will also need more media coverage in various university and research newsletters. This will raise awareness about the project and demonstrate to university leaders that CodeRefinery workshops deliver valuable competence development to researchers, which in turn will make it easier for project staff and volunteers to motivate to their university/department leaders to allocate funds for hosting workshops.

Events and tasks

2019

- Establish an open chat: anybody can join, make it easy to get involved and to watch/unwatch.
- 2-day train-the-trainer and lesson development workshop in 2019.
- 2-day hackathon in Autumn 2019 to bring the Nordic RSE community together. - “Bring your own code” events, can be in hackathon format.
- Publish a 10-points checklist towards more FAIR code.
- Train the trainer material is published as instructor notes for individual lessons. - Coordinate Carpentries instructor training for 15 persons, we want these to be strategic for the CodeRefinery project.
- Find one enthusiastic person per university/node.
- Document how volunteers can get involved.
- Map workshops/places that have used the CodeRefinery material.
- Create a guide on “how to run a workshop” and “what to do if you want to get started/contribute” based on existing guides (e.g. BioExcel: <https://github.com/blauc/workshop-orga/>).
- Website will list core team and partners. We will add a guide on “what to do to get listed as a partner”.
- Create an onboarding document for new members.
- Aggregate contact points by respecting GDPR where we ask the contact whether it is OK if we track the contact.
- Print and distribute stickers.

2020

- 2-day train-the-trainer and lesson development workshop.
- Organize a Nordic RSE conference.

2021

- 2-day train-the-trainer and lesson development workshop.
- 2-day hackathon, in case the 2019 event was well received.
- Transition coordination.

2.3 - Workshop delivery

2019

- 7 workshops (we send 3 instructors)
- 2 spin-off workshops (we send 1 instructor)
- 1 instructor training, back to back with hackathon

2020

- 7 workshops (we send 3 instructors)
- 3 spin-off workshops (we send 1 instructor)
- 1 instructor training, back to back with hackathon

2021

- 6 workshops (we send 3 instructors)
- 6 spin-off workshops (we send 1 instructor)
- 1 instructor training, back to back with hackathon

This is a conservative target. The stretch target would be 9 staff-organized workshops in 2019 and only spin-off workshops in 2021.

2.4 - Lesson development

- Lessons issue triage.
- Archive old lessons (such as automation/devops).
- Consolidate repositories towards <https://github.com/coderefinery> (all repos needed during a workshop should be under CodeRefinery organization)
- Remove/clean-up attempts which we have abandoned.
- **Make it possible to host a spin-off workshop without us doing anything:** currently this is not possible (repositories require us to rewind them and activate them, the lessons are centralized, we need to **decentralize them**).
- Document how to contribute new lessons.
- Offer the possibility to use both GitHub as well as our GitLab service during workshops. - Create a “Git refresher” lesson to reduce the length of the “Introduction to Git” lesson by 50% - thus also closing the gap with the Software Carpentry lesson.
- Build all workshop lessons around a single example project (e.g. the word-count project) to create coherence and show how all tools and practices come together in a single project.
- Develop and teach 1-day and shorter events.
- Document how lessons can be combined into alternative workshop formats of different duration (1-3 days) and theme (focus on version control, reproducibility, or large-scale scientific software development).
- Develop a 2-day workshop schedule which can be combined with additional lessons (developed by local instructors) relevant to the local audience
- Develop presentation material for a 1 hour event for research, university leaders, and e-Science centers and organization such as the Swedish e-Science Research Centre (SeRC) to convey the benefits of funding workshops and in-kind instructor time

- Make our lessons more relevant also for humanities and social sciences by including the following themes:
 - Web scraping
 - Web APIs
 - Text mining
 - Building on top of existing Carpentries lessons
 - Working with regular expressions
 - Visual analytics, visualization, InfoVis
- Also discuss data- and software management in our workshops (this is not the focus of our lessons but we link to relevant material and invite experts to give a 1 hour session instead of developing a material from scratch):
 - Data management plan
 - Data storage
 - Data cleaning
 - Software management plan

If some topics are outside of the team expertise, we will check with other projects/programs, reuse and collaborate, rather than starting from scratch.

2.5 - Good practice guides and screencasts

Each team member develops one short guide and one screencast in 2019. These will be peer reviewed and further improved by others. The topics of the good practice guides are our lessons: <https://coderefinery.org/lessons/> - for each lesson we will try to distill its essence into a good practice guide.

2.6 - Operation of the GitLab repository hosting

This service is subcontracted to our partner at DeiC. However, it is the task of the CodeRefinery project to establish a sustainability plan for continued funding.

For this we will collect usage statistics and usage patterns and create an impact analysis and start a dialogue with national infrastructure providers and institutions to gradually scale in a financing model, starting already early 2021.

We will improve support level by connecting infrastructure staff with our support queue.

Our goal is to grow from 300 to 1000 projects. We will ask the “most interesting”/prominent projects to contribute a paragraph or two and we will publish these as blog posts. We will also

ask projects to use relevant citation tags when reporting their publications to tools such as <https://www.cristin.no>.

2.7 - Administration and reporting

- Organization of steering group meetings
- Mid-term report: April 2020 (will be a written document, not a presentation) -
Final report: October 2021
- Reporting to the project portfolio for the NeIC board
- Reporting to the NeIC annual report
- Reporting to the PM/PO group

We will track the project and tasks openly and publicly in [GitHub project boards](#). From this we will be able to automatically produce status updates and reports which can be filtered using labels. Status updates and reports will become transparent and almost effortless.

2.8 - 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. Since this project is 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

3 - Project cost estimate and resource needs For

more background we refer to the [project directive](#) and the [collaboration agreement](#).

3.1 - Staff costs

This project will operate on a budget of 3.444 FTE/year and a total of 10.333 FTE with the following composition:

All numbers are in person-months.

	2018	2019	2020	2021	total
CSC	0.75	6	2.25	0	9

DeiC	0.5	4	3	0.5	8
------	-----	---	---	-----	---

ETAIS					0
Aalto SC				7	7
Sigma2	1	8.5	2	1.5	13
SNIC	1	6	1	1	9
NeIC matching			24	22	46
NeIC project management	1	6	6	5	18
Carpentries Nordics/Baltics coordinator		2	6	4	12
Sum	4.25	32.5	44.25	41	122

3.2 - Travel and other budget

	2018	2019	2020	2021
GitLab service	9 kDKK	50 kDKK	50 kDKK	42 kDKK
Workshops/meetings	26 kNOK	155 kNOK	155 kNOK	129 kNOK
Attending conferences	10 kNOK	58 kNOK	58 kNOK	48 kNOK
Other (domain name, promotion)	0.5 kNOK	3 kNOK	3 kNOK	2.5 kNOK
Conference organization			50 kNOK	
Carpentries membership		84 kNOK	84 kNOK	84 kNOK
Sum	45.5 kNOK	350 kNOK	400 kNOK	305.5 kNOK

4 - Risks

We use the scale 1-3 (1: low, 2: medium, 3: high) and for overall risk we multiply the numbers for the probability and impact to form an overall risk.

- Travel budget for workshops and meetings is exhausted:

- Risk: 2
- Probability: 2
- Impact: 1
 - Responsibility: project manager to compare expenses and budget projections - Responses: we ask local organizers to cover not only room and catering but also travel - we can probe this model even earlier to improve sustainability

- Administrative overhead affects delivery of services:

- Risk: 4
- Probability: 2
- Impact: 2
- Responsibility: NelC and national infrastructure providers represented by the steering group.
- Responses: dialogue between project manager, steering group, and project owner

- No-shows at workshops:

- Risk: 1
- Probability: 1
- Impact: 1
- Responsibility: CodeRefinery team and local organizers
- Responses: local organizers should have access to list of participants so that they can do the communication; follow-up on registrations before the workshop starts

- Staff is not available or overcommitted:

- Risk: 9
- Probability: 3
- Impact: 3
- Responsibility: national infrastructure providers represented by the steering group.
- Responses: dialogue between project manager, steering group, and service providers

- Staff change jobs:

- Risk: 3
- Probability: 3
- Impact: 1 (every team member is important but we prepare the project for staff leaving the project to other projects and jobs)
- Responsibility: project management

- Responses: decentralize dependencies, emphasize documentation, rotate lessons
- **Carpentries membership is not used to the full:**
 - Risk: 2
 - Probability: 2
 - Impact: 1 (but for long term the effect is higher)
 - Responsibility: national training coordinators (reference group) and project management
 - Responses: close dialogue between training coordinators and project, active outreach
- **No financing for the repository hosting is found beyond the project end: -**
 - Risk: 6
 - Probability: 2
 - Impact: 3
 - Responsibility: project management and steering group
 - Responses: collect convincing usage statistics and engage national partners early to agree on a financing model, explore other partnerships such as EOSC - **Our current hosting partner DeIC ends the contract for the repository hosting:** - Risk: 2
 - Probability: 1
 - Impact: 2 (difficulty is not the migration but to secure commitment from another partner)
 - Responsibility: project management
 - Responses: maintain a close dialogue with the service provider and with possible alternatives
- **We do not manage to find a sustainability model for the material to continue being maintained and taught:**
 - Risk: 2
 - Probability: 2
 - Impact: 1 for the project but high for its future
 - Responsibility: project management and all stakeholders
 - Responses: grow a community of instructors and establish partnerships with universities and institutions. Ensure that our material is still available for independent use, even without a CodeRefinery organization. Donation of materials and experience to Software Carpentry.

5 - Project organization and methods

5.1 - Project organization

The project is organized according the [PPS project method](#) with the following groups or roles:

- **Steering group:** The steering group takes the major decisions in the project. The steering group accepts the project plan, approves deliveries, and decides on transferal of

responsibility of the project results. The steering group also has the possibility to modify the project plan.

- **Project owner:** The project owner is responsible for the practical level project follow-up. The project owner is the chair of the steering group. The project owner will oversee that the project works towards the project idea and project objective.
- **Project leader:** The project leader prepares the project plan and is responsible for the day to day management of the project and of leading the project group. The project leader makes proposals for the steering group on decision points, such as acceptance of deliveries, change requests or any running business that needs to be decided.
- **Project group:** The project group is formed by the experts contracted for work in the project. The project group is distributed over the participating countries and several locations. Each project group member assumes responsibility for a certain task in order to level workload and to increase identity with the project.
- **Reference group:** The task of the reference group is to support the project by giving input requirements and advice to the project, by ensuring that the project outcomes are acceptable for the user communities, and by maintaining an active communication between the stakeholders and the project.

5.2 - Requirement dialogue

The requirements for the project are acquired from the project directive which has been submitted to and approved by the NeIC Board. The project requirements are supplemented and adjusted during the course of the project by the steering group. The experts working in the project will refine the requirements to a more technical level. Finally, the use cases to be implemented are a direct route to learn user requirements as they document a certain workflow that the end user wants to accomplish. The requirements will be regularly discussed at the project meetings and steering group meetings.

5.3 - Change management

Change requests which require a significant change in the project requirements and the project plan are documented in writing. The decision on approval of change requests will be made by the project leader with consultation to project owner or in a case of major change will be taken to the steering group for decision.

Small changes can be accommodated within the normal development cycle of the project. The project leader will follow the development of the project and if a significant factor affecting the project is emerging, may decide to re-prioritize the requirements accordingly.

5.4 - Internal information distribution and document management

- Discussion takes place on <https://coderefinery.zulipchat.com>.

- Weekly virtual coordination meetings within the project team are held over video. Everybody is welcome to join and listen in and minutes are stored and visible to all team members.
- Kick-off meeting at the beginning of the project.
- Project team meetings once every year.
- Steering group meetings every 4 months.
- Coordination meetings between the project owner and the project leader are scheduled when needed. Meeting minutes are stored in NeIC internal wiki pages.
- Only one private CodeRefinery staff channel on Zulipchat exists. Only internal contracts, participant lists, and the like will be kept private to paid staff.

5.5 - External information distribution and outreach

The project will strive for open and transparent communication. The main hub for communications is the project website: <https://coderefinery.org>. The main hub for task coordination and material output is GitHub: <https://github.com/coderefinery>.

- Project directive and project plan are publicly available online.
- Steering group meeting minutes are publicly available online after being approved by all meeting attendees.
- Public files available through <https://coderefinery.org> and <https://wiki.neic.no/wiki/CodeRefinery>.
- Tasks and progress can be tracked publically at <https://github.com/orgs/coderefinery/projects>.
- News and events are announced via <https://twitter.com/coderefine>.
- Discussion chat is open to everybody: <https://coderefinery.zulipchat.com> - Newsletters and blog posts on good practices will be published every second month on <https://coderefinery.org>.

5.6 - Administrative routines

Service contracts will be made for persons working in the CodeRefinery project. The contract will be made between the employing organization and NeIC. The service contract specifies the amount of working time committed for the project. Formally, the project leader decides on nominating persons to work for the project.

The project personnel should follow the practices of their home organization for time reporting. The CodeRefinery project does normally not require time reports. Instead, the results will be closely followed. However, time tracking is strongly encouraged for personal use to simplify time reporting within the home organizations.

Team members track the progress of work items and issues at <https://github.com/orgs/coderefinery/projects> which is publicly accessible.

NeIC will cover costs for travel which is necessary for the project work. Covering the travel costs need to be agreed beforehand with the project leader. The project leader manages the travel budget following the regulations of the home organisations.

Project leader reports expenses to the project owner upon request.

5.7 - Environment and economy

The project attempts to minimize environmental impact and costs by having online meetings instead of physical meetings whenever possible to minimize travel and costs.