

Using an Agile Waterfall hybrid to manage a major Collaborative Computational Project.

Collaborative Computational Project Number 4 (CCP4) in Protein Crystallography was set up in 1979 to support collaboration between researchers working in structural biology, and to assemble a comprehensive collection of software to satisfy the computational requirements of the relevant UK groups.

Structural biology focuses on understanding large biological molecules like proteins and DNA, how they acquire the structures they have and how any structural alterations can affect their function.

Demand for computational support in these studies gave rise to the CCP4 program suite, which is distributed to academic and commercial users world-wide.

Taking a lead role

Scientist Eugene Krissinel, a member of the Science and Technology Facilities Council (STFC) Scientific Computing Department (SCD), has taken on the core lead of project managing the vast volumes of collaborative software development and its distribution for CCP4. In leading the core team in collaboration he says, "I am responsible for CCP4 infrastructure, software distribution, and everything which goes from CCP4 to users including some program development."

CCP4 is a well-known and respected open collaboration with a very good reputation. The project originated in the UK but the CCP4 team collaborates with a number of researchers from the Netherlands, Switzerland, Germany, the USA and other countries. There are a vast numbers of users – upward of 25,000.

The project now has a mature agile management style with an Executive Committee and two working groups. The Executive Committee includes 15 UK academics who oversee what the collaboration is doing. They meet every 2-3 months to assess progress and drive targets.

The two working groups have different remits. The first includes all UK Principal Investigators working in the structural biology field. They represent the communities which use the software being developed so they advise on what has been successful and what needs to be changed, and they consider what they will need to have in the future as science advances and goals change. They also elect the Executive Committee members.

The Agile Values

Individuals and Interactions	Over	Process and tools
Working Software	Over	Comprehensive documentation
Customer Collaboration	Over	Contract negotiation
Responding to change	Over	Following a plan

The second is more of a focus group which meets every quarter. It comprises 20-30 people from the practising research

community who oversee the CCP4 Educational Programme and give feedback on what has been done so far, what problems have been fixed, what the next steps will be and if they will meet the expectations of users. This process helps Eugene and his colleagues to manage the CCP4 Software Suite.

Challenges of the project

One of the first things that Eugene needed to address when he joined the CCP4 team in 2009 was the size issue, as the volume of software to be distributed was more than was manageable by the resources and technology of the time. The software suite had reached such a size that the ways of managing software were purely technical – from archiving, compilation, testing, to packaging and distribution – and this was taking all the effort from the core team. It was a considerable issue so his first goal was to suggest a more efficient way of handling the software.

“The Software suite grew very fast and now the size and complexity is comparable to Linux distribution, and is managed by only a handful of people.”

*Eugene Krissinel,
CCP4 core lead, SCD*



The team adopted technologies used by Linux maintainers, which enabled them to develop automatic software management pipelines, and introduce hot updates, so CCP4 is updated just like an operating system.

This is something Eugene designed, and it took about 3 years to actually implement to a stage where it was an established modus operandi for the team. “It took quite a sizeable development of new graphical installers, updaters and new pipelines,” he said. “Those pipelines are big because we have about 10 million lines of code.” With such a huge infrastructure, there is a lot to manage, and a way forward was to automate certain processes.

Eugene explains; “Regression testing of our software is an ongoing problem but now it’s completely automatic and happens every night.”

There is a great deal of communication and collaboration to achieve the mutual goal of the project. The CCP4 team links research community and developers, making sure that users’ feedback reaches program authors. It is a considerable size of code that needs distributing so this takes a lot of time and effort.

Management style

Eugene uses an agile style of management to organise the project. As part of this plan he uses a flow chart indicating which tasks have been completed or are still to be done, and how they are progressing along the timescale. Eugene used a coarse-grained plan, and tracking progress of tasks within projects is achieved through short regular group meetings. The usable outputs are shown and discussed with stakeholders on a continual basis, allowing the team to have a continuous stream of deliverables.

Eugene highlights the importance of good working relationships and mutual respect. His management style is to find out what people like to do and give them assignments that play to their strengths and match the project’s needs. The team has a diverse set of skills and interests and together they successfully deal with a wide variety of tasks; from scientific problems to very technical problems, things like managing websites, or mundane jobs to very creative jobs. For instance, one of the team members has developed skills in budget management. This is his strength and he enjoys doing it, so Eugene is happy to have his help in guiding the project finances. It is important to know exactly how much each staff post costs per year, how much is being received for industry licences, and how much will be spent each year. They also have an accurate prediction of what is planned to be spent on any new hardware or in holding training workshops for the crystallography schools they run worldwide over the course of the year. A formal report is made annually to the CCP4 Executive Committee.

As for identifying benefits in the project, this is done by monitoring updates for the software. If it is liked by the research community they will use the software and this will be shown in download stats and start up stats for the programs. This is collected only from academic users (not industrial users) and the information is completely anonymous.

The theory is that if academics are happy, then industry will listen. The more industry uses, the more sustainable the funding is for the project. The number of industrial licences is a crucial indicator for financial health. Currently, CCP4 sells on average 140 industry licences per year, and that number is growing.

Updates are generally released twice a month. This is dependent on updates from developers – once they have a sufficient number and tests are passed, the updates get pushed out.

Feedback is key and the CCP4 team has always been very strong on communicating with the community directly. They support the CCP4 'bulletin board', a mailing list which has about 8000 subscribers who post between 20 – 100 messages each day. So if something is not right, then it sparks discussions on the bulletin board. They also have a dedicated line for submitting bug reports, which are frequent and dealt with quickly. "If this line is completely silent I would personally worry because there are always bugs. If nobody is talking to us about them or thinks we can't be reached, that becomes a big problem," said Eugene.

The project is a great example of agile management because it's more focused on using regular direct discussion between the development team and the users, with 2 week continuous delivery slots. There is also a strong emphasis on stakeholder communication and reciprocal respect within the industry.

CCP4's success can be attributed in part to generous industry support. It basically has its roots in drug research and its industrial customers are all big pharma companies. By purchasing the software licences these

companies provide important funding to ensure the continuity of the project. Other funding comes from competitive grants, and STFC's Scientific Computing Department provides the overall setup and home for the project.

Improvement going forward

Despite its success, they are always trying to improve. Going forward Eugene would like to see a framework for supporting short-term activities as a much easier process. Currently the process to pay individuals, such as external students, post docs, invited experts and contingency workers involves extra administrative burden.

A little bit more autonomy in financial terms would benefit the project processes, especially in terms of purchasing hardware to be utilised on research. This can be slowed by the many channels necessary to make purchases.

The project, like so many others, has been impacted by Covid. Because of the pandemic, less spending has occurred, and the funds don't carry over to the following financial year automatically, something that Eugene, as core project lead, needs to take into consideration.

Rising to all of these challenges is something Eugene and his colleagues take in their stride. CCP4 is hugely successful – something that is borne out by its longevity, the ever-evolving software it produces and its growing community of users, and high demand from industry. Importantly, the first COVID-19 virus structures were solved using CCP4 Software. Taking the agile approach for managing the project has given it a further advantage of increasing the dialogue bandwidth between the development team and the users.

**Author: Layla Mirmalek,
SCD Project Management Office**

CCP4 is hosted and managed by STFC's Scientific Computing Department and funded by the Biotechnology and Biological Sciences Research Council, as well as via industrial licensing. STFC and BBSRC are both part of UK Research and Innovation.