There were several driving use cases that the DL2021 were to benefit. Some of them were related to traditional HPC capacity needs, while others were new needs from the research community. In general, there were needs to manage and utilise data in a more versatile manner than before. New user groups, like governmental research institutes and increasing amount of users from universities of applied sciences added to the demand. The infrastructure and related services is to benefit the whole research and higher education community in their work, resulting benefits for society as well.

2.1 Demand for data management and computing infrastructure

One of the goals for the programme was to respond to the new needs for services from the research community and new research fields. The programme responded to the need by not only renewing the centralised data management and computing infrastructure operated by CSC, but also by developing capabilities to use the services also in fields that have not traditionally used supercomputers in their work. Furthermore, the services are now offered more widely than before, to the governmental research institutes and also for education use. This change in the usage policy and the supportive actions affected to the growth in the number of users, that has been over 40% (Figure 2.1).
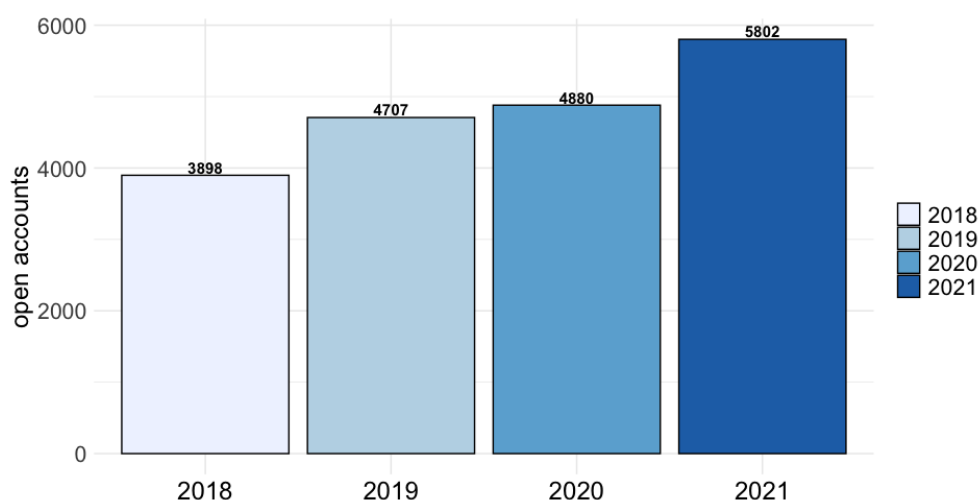

Figure 2.1. Number of users of the data management and computing infrastructure 2018–2021. The figure includes other services besides DL2021 related as well (e.g. Fairdata services).

Data source: CSC user statistics.

Puhti and Allas were opened for use in the fall 2019, Mahti a year later in the fall 2020. There is a growing trend in the number of projects that are using the services (Figure 2.2). The number of projects using Mahti is lower than for Puhti, which is due to the nature of the services. Mahti is primarily designed for medium to large scale simulations that require a significant amount of computing capacity while Puhti serves a wide range of use cases from data analysis to medium scale simulations. About two fifths of the resources CSC has provided in 2021 has been used with Mahti.
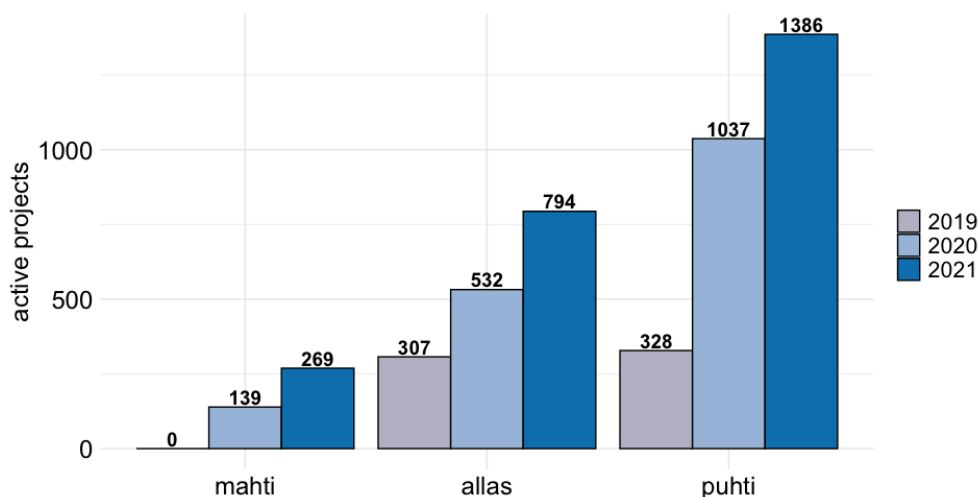
Services respond to the needs

Traditionally CSC has offered especially high-performance computing services and there are research fields and research areas in which the availability of such services is vital. As data intensive research is emerging and good data management is becoming ever more important, services for data management has been one of the focus points of development within the programme and more widely at CSC. Besides computing and data management services, CSC also offer a wide range of softwares and databases for the users. A lot of effort has been put on supporting new fields and new types of users to fully benefit the use of the services offered by CSC, both by offering them training but also by developing CSC's user support and guidance. The services related to computing, data management and storage, softwares and databases and customer support and guidance were considered important by majority of the respondents in the user survey conducted in 2021 (Figure 2.3). Training was considered important by about two fifths of the respondents. For those who have attended trainings the experience has been good and the trainings have offered new skills and competences (more on this in section 4).
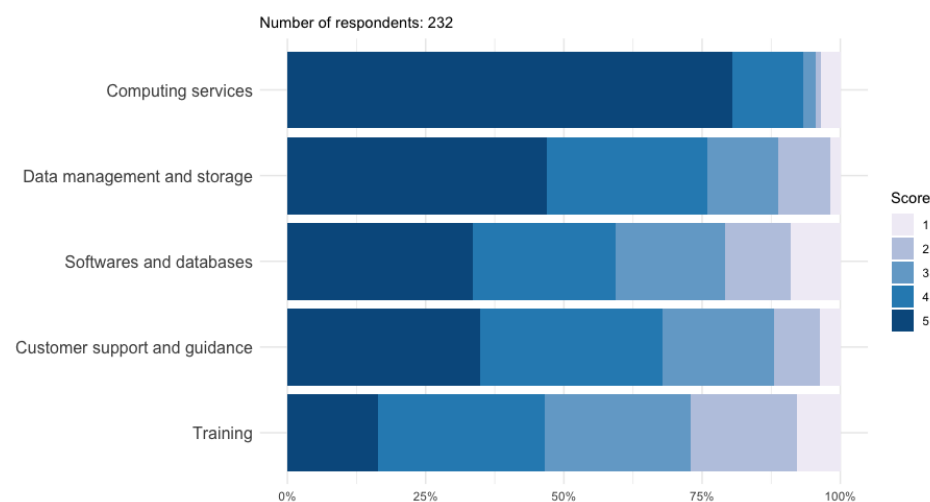

Figure 2.3. Importance of different services offered by CSC for the users' tasks.

We asked. How important are the following types of CSC's services for your work or studies? 1= unimportant, 2= fairly unimportant, 3= neither important nor unimportant, 4= fairly important, 5= very important, IDK= I don't know. IDK answers are not included in the figure.

Data source: Survey for users of CSC's data management and computing services 2021.
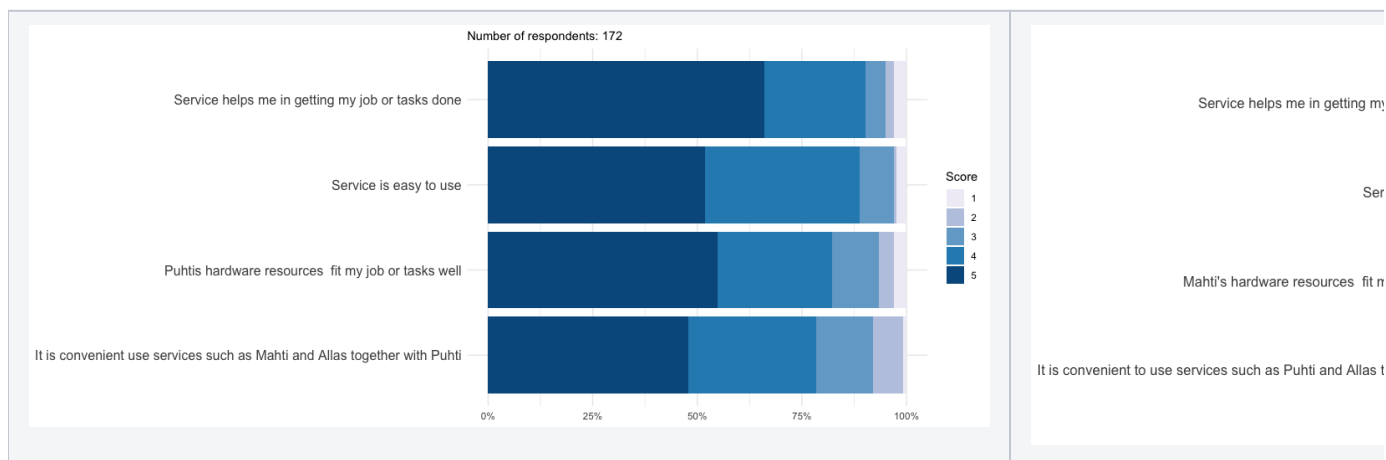

Figure 2.4. User experience of Puhti, Mahti and Allas.

We asked. Below there are some statements regarding this specific service. Please, assess them according to your experience.1= strongly disagree, 2= disagree, 3= neither agree or disagree, 4=agree, 5=strongly agree, IDK= I don't know. IDK answers are not included in the figures. There were IDK answers especially to the statement regarding using the services Puhti, Mahti and Allas together.

2.2 Research and education sector benefitting from the services

Research institutes

One of the goals for the programme was that the infrastructure and related services are offered for wider public research community. In the past, CSC's services were offered for research done in higher education institutes, and researchers in the governmental research institutes could use the services only as partners in the projects. In 2018 the services were opened for free for the governmental research institutes for academic research use for the same conditions than for higher education institutes. In 2017, before the services were opened, there were less than 100 users from governmental research institutes. Since 2018, the number of users from governmental research institutes has over doubled (Figure 2.5). Nowadays about 10% of the research personnel of governmental research institutes use CSC's services [1].
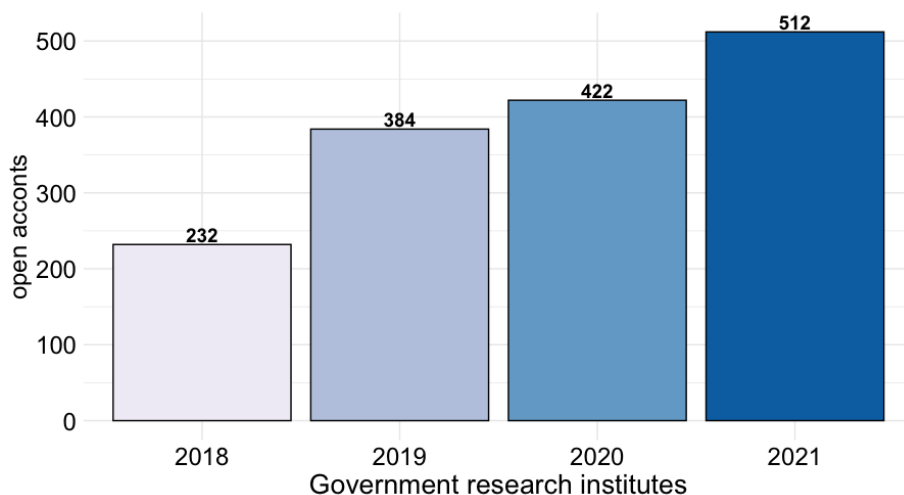

Figure 2.5. Number of users from governmental research institutes in 2018–2021.

Data source: CSC user statistics.

Being able to use a centralised national infrastructure rather than investing on own, local infrastructures or commercial services, has presumably mean savings to the governmental research institutes. For commissioned research and executing their primary tasks as a governmental official, an agreement of usage with CSC is needed. During the past few years, the interest of the governmental research institutes also towards solutions that CSC can offer in supporting them in executing their primary tasks have grown.

Finnish Meteorological Institute purchased computing capacity for their use, in the procurement of the DL2021. Puhti was extended with a dedicated cluster for FMI, that is integrated in the system. FMI covers the maintenance cost of their partition fully. In addition, FMI can use the DL2021 environment for research through normal resource allocation processes like other governmental research institutes as well.

The basic IT infrastructure for the governmental research institutes is offered by the Government IT Center Valtori. Opening the services for the governmental research institutes brought up questions related to user experience and service offerings by CSC and Valtori and possible overlap on those. The questions to solve were related to for example to end-to-end user experience from the terminal to the service, instructing in contacting the right user support in case of incidents or other user needs, solving the challenges related to the network and firewalls. Tripartite meetings were held between CSC, Valtori and the governmental research institutes in order to solve the questions.

Universities of applied sciences

The number of users from universities of applied sciences has increased and is now almost five times as much as it was in 2018 (Figure 2.6). One factor affecting this could be that since 2018 CSC's services can have been used for educational activities as well. About 10% of the research personnel of universities of applied sciences are using CSC's services [1].

Universities of applied sciences have used the CSC's services for a diverse set of applications. For example, Puhti has been used by Centria for researching, testing and developing security of industrial internet and wireless systems and also by HAMK for simulations of the reactions in the structures of industrial buildings in case of fire.
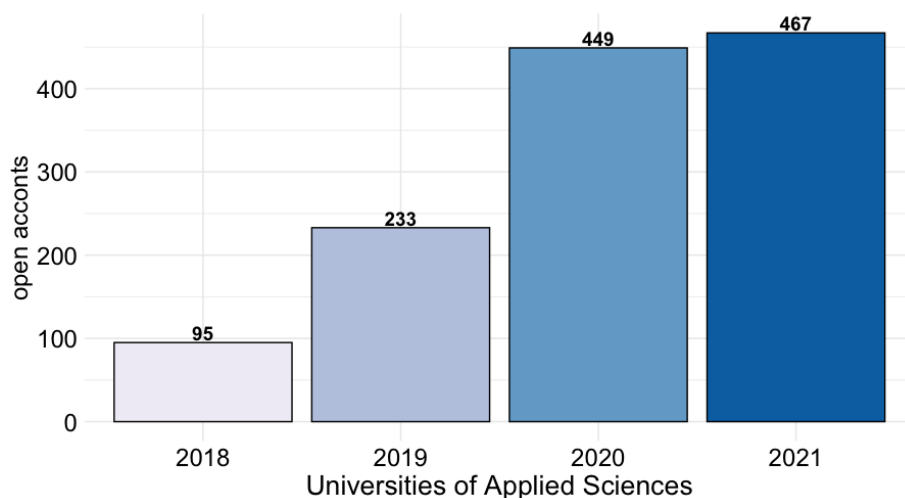
Figure 2.6. Number of users from universities of applied sciences in 2018–2021.

Data source: CSC user statistics.

Universities

Universities are the biggest customer segment using CSC's services. There has been growth in the number of users also for universities (Figure 2.7) during 2018–2021, although it has been less than for governmental research institutes and universities of applied sciences. About 15% of the universities' research personnel use CSC's services [1]. One of the goals for the programme was to respond to the new needs for services from the research community and new research fields, which is reflected in the growth. Examples of research themes are presented in section 2.4.
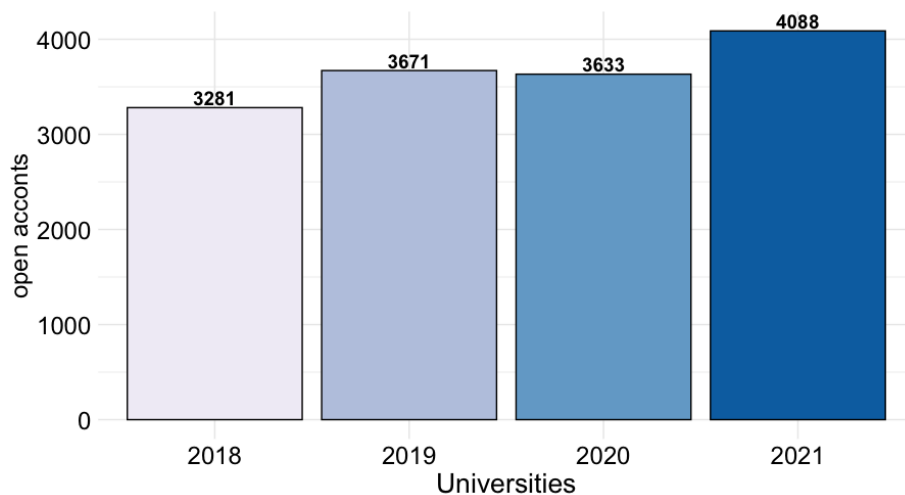

Figure 2.7. Number of users in universities in 2018–2021.

Data source: CSC user statistics.

2.3 Growing research areas and needs

Usage in different research fields

Figure 2.8 shows the number of active projects in 2018–2021 for all main research fields. The fields that have been growing the most in terms of active projects [2] are agricultural sciences and engineering and technology. The number of their active projects in 2021 is more than three times of what it was in 2018. Natural sciences has been the biggest research field using CSC's services in terms of users, projects and used resources.
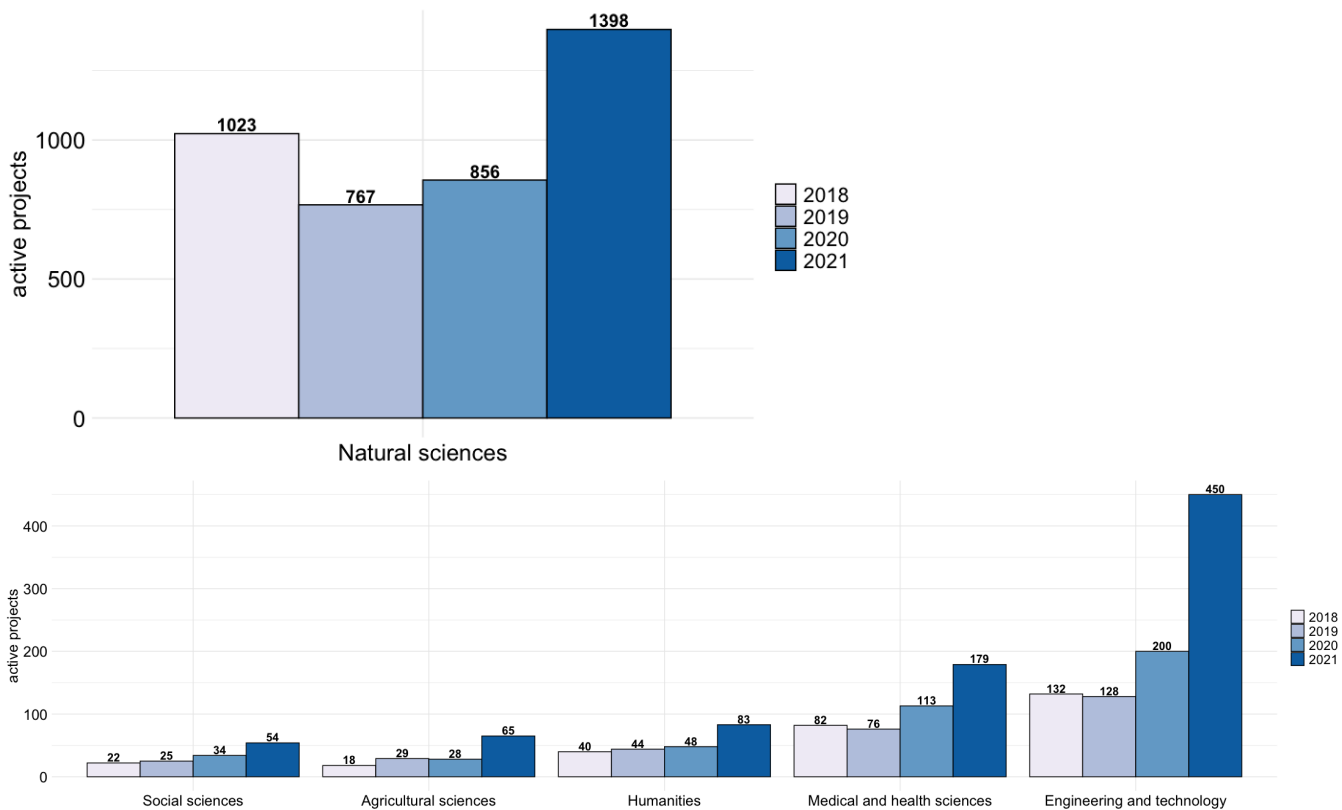
Figure 2.8. Number of active projects in the main research fields in 2018–2021. Active project is a project that has used data management or computing resources in a particular year. The fields 'other' and 'unknown' are not included in the figure. The number has varied between 80-700.

Data source: CSC user statistics.

Figure 2.9 shows the positions of top ten (in 2021) research fields in terms of active projects and used resources. Physical sciences, biological sciences and computer and information sciences have been among the top three research fields during the past couple of years. In figure 2.8 it was seen that engineering and technology has been growing during 2018–2021. One of the fields growing has been electronics that was eleventh biggest field in terms of active projects in 2018 and is fourth biggest in 2021. Among the natural sciences, other natural sciences has been growing and has risen from 25th to eight biggest field in terms of active projects. In terms of used resources, while there has been little changes during the years, pharmacy has increased its use of resources and has risen from 25th position to seventh. Materials engineering changed its position from 12th in 2018 to 21st in 2019 but has raised to eight in 2021. The position of both basic medicine and geography and environmental sciences lowered in terms of active projects but has risen in terms of resource use. There are also smaller fields that don't yet have many projects or a lot of resource use but that have been growing during 2018–2021 (Figure 2.10).
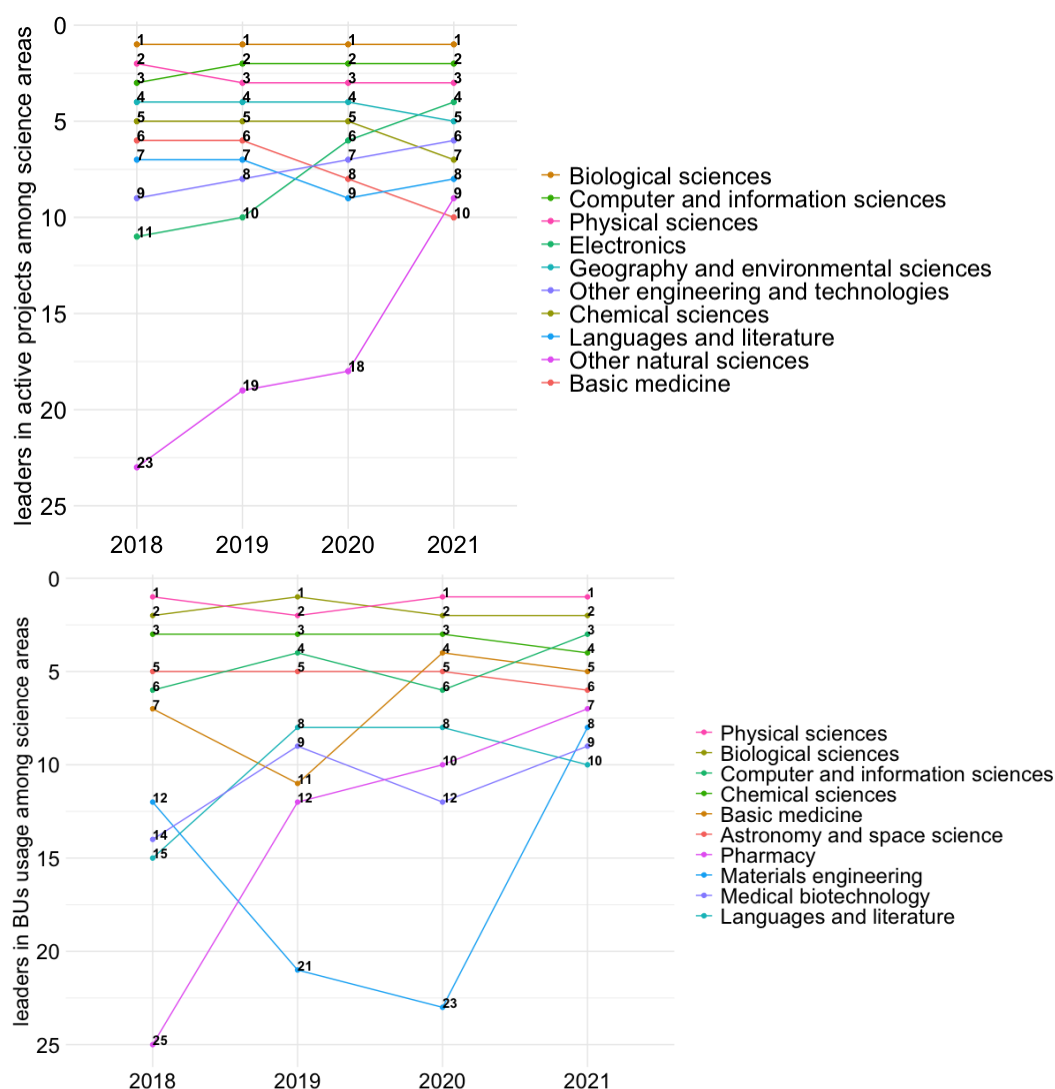
Figure 2.9 Positions of top ten (in 2021) research fields in 2018-2021.

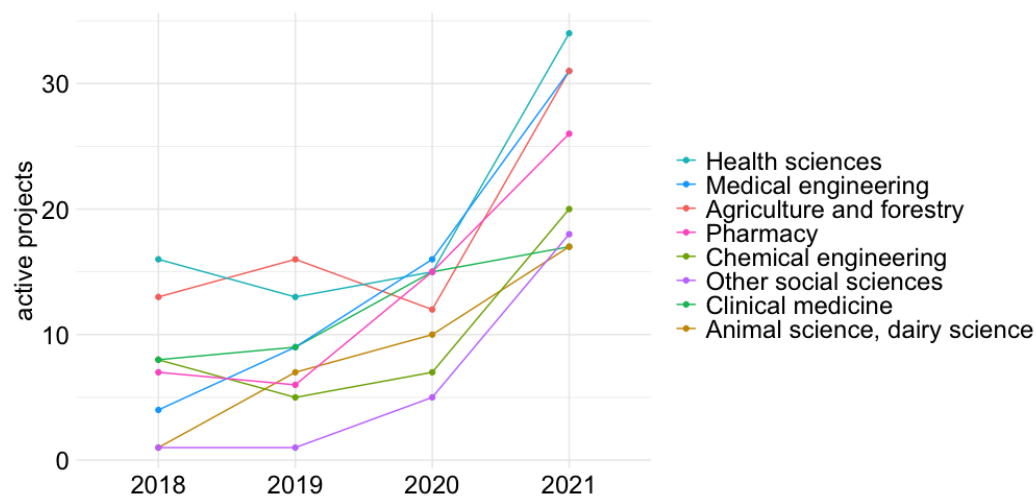Data source: CSC user statistics.

Figure 2.10. Small (15-35 active projects in 2021 research fields that have at least doubled their number of projects during 2018-2021. Active project is a project that has used data management or computing resources in a particular year.

Data source: CSC user statistics.

Services for new type of needs

Puhti-GPU was introduced in the fall 2019 and the number of projects utlising it has been increased since (Figure 2.11.). Puhti-GPU has been in use for about one quarter of the projects that are using Puhti in 2021. Mahti-GPU was introduced in 2021, and there are 65 projects using it (about one quarter of all Mahti projects). Based on preliminary analysis on the descriptions of all projects as of November 2021, there are about 200 projects that relate to artificial intelligence methodologies (machine learning, deep learning, computer vision, natural language processing) [3]. In addition, artificial intelligence can be used in many applications areas as well. Examples are provided in section 2.4.
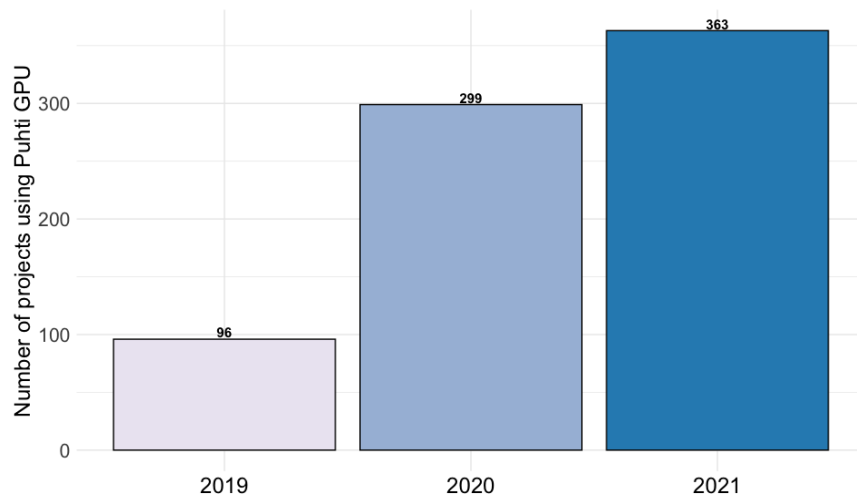

Figure 2.11. Projects using Puhti-GPU partition.

Data source: CSC user statistics.

Services for secure managing of sensitive data was one of the needs set for the new infrastructure and driving use cases in the development. The programme facilitated gathering the needs from the users and piloting the services with them, and the services have been developed as part of the DL2021 environment. Two of the services, SD Desktop and SD Connect were launched for beta use during 2021, two others will follow. Offering the sensitive data services directly to the end-users has been a remarkable change as previously the service for sensitive data management (ePouta) was offered for organisations who could then offer it forward within the organisation. This change will also enable collaboration within sensitive data easier than before. The usage of the services is still small but steadily growing and used by researchers from life sciences but also social sciences, humanities and engineering.

The programme responded to a need for new type of services needed in research but also in education. A service that is well-suited for teaching and research within data and programming is Notebooks. It allows teachers to create a teaching environment or produce own teaching materials without install to University IT maintained machines or students' own laptops. The service is in use by universities and universities of applied sciences but also by the governmental research institutes. The number of users varies during the year, partly due to the nature of the service, but the trend is increasing and there was up to nearly 3 500 monthly starts of the Notebooks computing environment in 2020 (2 500 in 2019 and 1 500 in 2018).

2.4 Benefitting research, education and society

The goal for the programme was to support research and education, which in the longer term can lead to broader impacts on society as well. Majority of the respondents in the user survey use the services for pure fundamental or basic research, but almost half also for applied academic research. The services also support teaching and education and research, development and innovation work, and about 10% of the respondents also use them for studies.
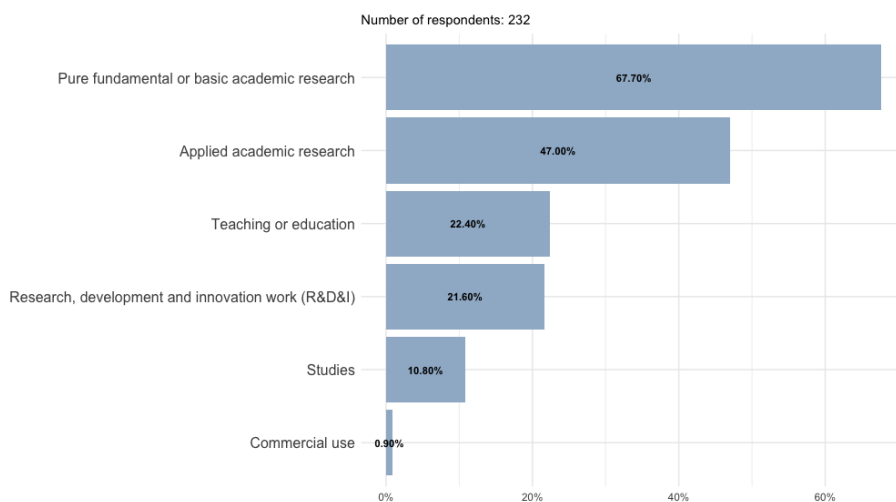
Number of respondents: 232

Figure 2.12. Type of work to which the services are used.

We asked: For what type of work do you use the services offered by CSC? N=232.

Data source: Survey for users of CSC's data management and computing services 2021.

CSC's services have benefited over 300 theses during 2018–2021 (October). The inquiry regarding information on theses was sent to 1 909 principal investigators of open academic end-user projects in October 2021. Altogether, 131 responses were obtained and thus 300 can be considered as the lower bound for the number of theses. About 40% of the theses were doctoral degrees and 40% for master degree.

CSC's services have also had impact on research, if considering the quality of publications benefited of CSC's services. The publications reported to CSC in connection with the resource application process during September 2020–June 2021 was analysed with respect to JUFO classifications and compared to all publications from Finnish universities 2019–2020 (Figure 2.13). Among the publications reported to CSC, there was about 2 100 publications reported, of which 1 600 had been published during 2018–2021. It was possible to identify the JUFO classification for about 1 050 of these. Based on the comparison it seems that higher proportion of the publications reported to CSC belong to JUFO classification levels 2–3 than of all publications by Finnish universities. It is good to note, however, that the publications reported to CSC may not include all publications benefitted from the services but a selection by the principal investigator of the projects.
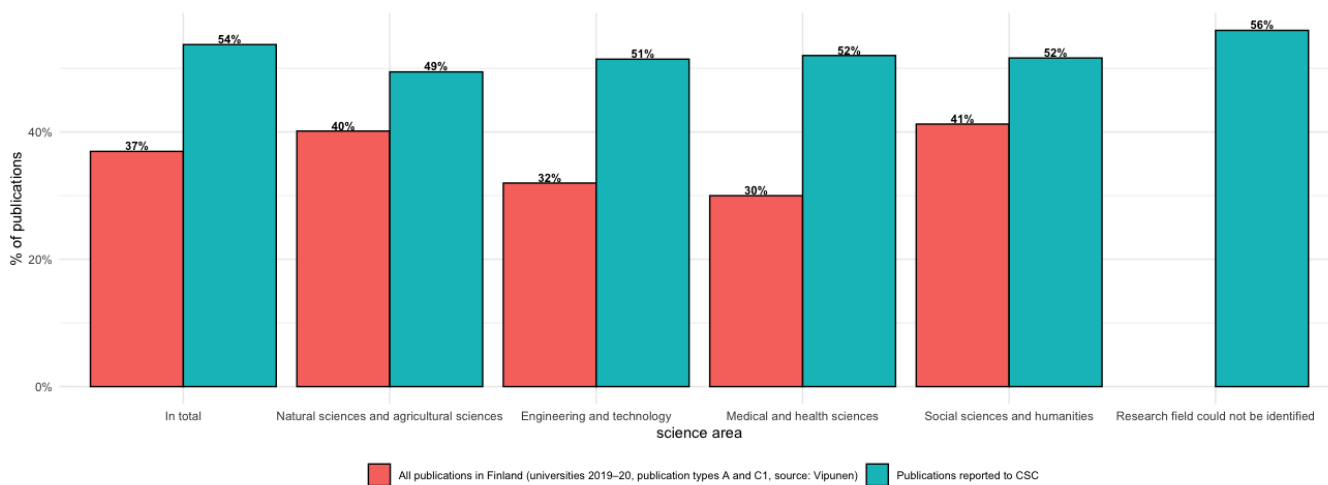

Figure 2.13. The share of JUFO classification levels 2-3 of the publications reported to CSC during September 2020–June 2021 compared to all publications in Finland published in 2019–2020.

Data source: Publications reported by the principal investigators of CSC's end-user projects and Vipunen (for all publications).

Examples of customer cases

The new services have been in place only for one to two years but they have already enabled research results that may have great impact on society as well. Examples of them are provided below.

COVID19 fast track

CSC wanted to do its part in the fight against COVID-19 and opened a fast track for projects researching this disease. The processing of applications was expedited and the computing queue to supercomputer Puhti was prioritized. The COVID-19 fast track was opened in March 2020 and a third of the capacity of the Puhti was initially allocated for the use of the fast track. The resource earmarked for COVID-19 research was significant. In the early stages of the fast track, Puhti was the only supercomputer that CSC was using, and use of the computational resources of the fast track was highest before Mahti was taken into use.

A total of 15 research projects were selected for the COVID-19 fast track. These projects studied the airborne spread of the virus, sought ways to prevent reproduction of the virus, explored potential drug therapies through virtual screening, studied virus mutations and identified virus variants through sequencing. The studies produced several peer-reviewed articles as well as plenty of new information on the airborne spread of the virus and droplet infections. Researchers have played a major role in managing the pandemic by helping health authorities to identify more sensitive virus variants through sequencing and considering recommendations for masking and indoor ventilation.

For further information, see:

Supercomputer Puhti vs. the Coronavirus: Review of CSC COVID-19 Fast Track Outcome

CSC's services benefit research on viruses and COVID-19 is more widely than the selected 15 projects for the Puhti fast track. Preliminary analysis on the descriptions of the projects indicates that there are about 50 projects that relates to viruses or COVID-19.

Ice sheet modelling, input for IPCC 6th report

CSC and its supercomputers Puhti and Mahti participated to a large international ISMIP6 project (Ice Sheet Model Intercomparison Project) of 80 researchers and 38 research groups. They simulated the future evolution of Greenland and Antarctic ice sheets and their contribution to sea level rise. The research provides input for the Intergovernmental Panel on Climate Change's Sixth Assessment report which first part was published in August this year. Key findings: 1) the research confirms that Greenland and Antarctica's ice sheets could together contribute about 40 centimeters of global sea level rise if greenhouse gas emissions continue apace. Meltwater from ice sheets contribute about a third of the total global sea level rise. 2) Keeping the Paris Agreement target of limiting global warming to 1.5°C instead of continuing with the current contributions to mitigating climate change, sea level rise this century from the melting of ice could be halved.

For further information, see:

Halve the land ice contribution to sea level rise by following the Paris Agreement target of limiting warming to 1.5°C

Carbon emissions could add 40 cm to 2100 sea level rise

AI detects and grades prostate cancer nearly without error

Researchers from the University of Tampere and Karolinska Institute in Stockholm trained artificial intelligence to diagnose and grade prostate cancer. The artificial intelligence system could correctly identify biopsies containing cancer nearly without error. The supercomputer Puhti-AI, which was launched on a pilot basis in the early autumn of 2019, contributed to calculating the final phase of the work.

For further information, see:

Artificial intelligence detects and grades prostate cancer nearly without error

Antidepressants bind directly to a brain-derived neurotrophic factor receptor without help from serotonin

The effects of selective serotonin reuptake inhibitors (SSRIs) and other conventional antidepressants are believed to be based on their increasing the levels of serotonin and noradrenalin in synapses, while ketamine, a new rapid-acting antidepressant, is thought to function by inhibiting receptors for the neurotransmitter glutamate.

Neurotrophic factors regulate the development and plasticity of the nervous system. While all antidepressants increase the quantity and signalling of brain-derived neurotrophic factor (BDNF) in the brain, the drugs have so far been thought to act on BDNF indirectly, through serotonin or glutamate receptors.

A new study published in Cell demonstrates, however, that antidepressants bind directly to a BDNF receptor known as TrkB. This finding challenges the primary role of serotonin or glutamate receptors in the effects of antidepressants.

For further information, see:

Antidepressants bind directly to a brain-derived neurotrophic factor receptor without help from serotonin

Puhti-AI pilots and language technology

Two Puhti pilots were utilizing Puhti-AI's GPU nodes to develop language technology. To truly compete with human translation, the machine needs to acquire natural language understanding beyond the capacity of fixed symbolic rules and simple statistics and this is also the goal of MultiMT project. MultiMT uses parallel corpora (texts translated by humans) to discover meaning representations that are not tied to any single language by interpreting the semantics of over a thousand natural languages.

The DeepFin project is implementing BERT method into Finnish language. BERT (Bidirectional Encoder Representations from Transformers) is a natural language processing method originally developed by Google that enables a variety of language understanding tools. BERT is a bidirectional method, which means that when processing text, it looks both forward and backward in the sentence on each analysis layer. This allows deeper understanding and better prediction results. A major challenge in implementing the method for a variety of languages has been the sheer volume of text input and computational power required to train the model properly. Billions of words of text and supercomputer capabilities have made it possible to create a Finnish language model that can compete with models based on other approaches as well as the BERT implementations of other languages.

For further information, see:

Mahti pilot projects

Already in its pilot phase, Mahti benefited researchers in many different research fields. One of the pilot projects were related to the magnetic fields of the Sun and to show how a small-scale dynamo takes place in the sun and how it affects space weather. The Sun is a giant ball of gas, but it is also a huge magnet, and magnetic activities of different kinds cause eruptions of particles from the Sun. These eruptions can be seen on Earth as the Aurora Borealis, but they also affect space weather and disrupt the operations of satellites, electricity grids, communications, and aviation, among other things. For this reason, it is important to understand how Solar activity drives the space weather.

For further information about the Mahti pilots, see:

[1] Based on CSC's user statistics 2021 and statistics on research staff in Vipunen for 2019.

[2] Active project is considered a project that has used data management and computing resources, "billing units" in a particular year.

[3] The analysis was done for projects that were open in November 2021 by identifying themes using the keywords, description, methods and results of the projects. Each project were mapped to one theme that it matches best.