

Date: 22.03.2017

Revision D, based on ver. C 0.5

A contribution model for funding of the national e-infrastructure

Executive summary

The purpose of this paper is to introduce a funding model that includes contributions from the users within higher education and public research and payment from commercial¹ users. Users in this context refer to research projects.

A main principle in Norwegian R&D policy is that operational cost for the use of research infrastructures should, as far as possible, be financed by the R&D projects using the infrastructures. It is thus a requirement from the Norwegian Research Council that research infrastructures supported by the *National Financing Initiative for Research Infrastructure* should as the main principle, include user contribution as an element in the funding of the operational cost.

There is also a need to find new sources of income for the Norwegian e-infrastructure as the basic funding and investment funding does not keep up with increasing needs for e-infrastructure. A contribution from the projects will release basic funding from the universities that can be used for investment. Currently the four oldest universities provide approx. 44.4% of the funding, while using approx. 83% of compute resources² and approx. 75% of storage resources.

The infrastructures for High Performance Computing and Scientific Data Storage have previously been funded without a requirement for direct funding contribution from the research projects. This will now change. However, there is a need for a transition period in order for the research projects to adapt to this new reality. When applying for funding for research projects from the Norwegian Research Council, there is a requirement that the projects allocate funds for expenses related to using research infrastructures. It is now important to allocate funds for using HPC and storage infrastructure in these applications in order to be able to pay for the usage. During 2016 a survey was conducted among the research projects using the e-infrastructure. Because of this, some adjustments have been made to the proposal, amongst other excluding EU-projects and reducing the number of projects that will be included initially.

The model introduces four different categories:

- A. Large projects with funding of 15 MNOK or more, from the Norwegian Research Council, paying for operational expenses
- B. Non-commercial projects needing Dedicated Resources, paying for capital and operational expenses
- C. Commercial research and industry which will pay the full cost price
- O. Non-contributing projects which are smaller projects not in any of the above categories

The cost for the services is based on cost models consisting of actual cost elements from service suppliers for operational cost and investment costs.

¹ E.g. "Oppdragsforskning"

² Average of allocation periods 2016.1, 2016.2 and 2017.1

This model will be introduced during 2018 and the existing research projects will get a reasonable time to adapt to these new rules and make provisions for this in their future applications for funding. In agreement with the Research Council, projects within the SFF and INFRASTRUCTURE programs will start to contribute according to the agreed and approved budget. Other new projects in a category mentioned above, can expect to pay from the allocation period starting in October 2018, depending on their funding arrangements with the Research Council

The objectives of the model are to work towards Realistic expectations, Responsible use of resources, User driven demand, Fair resource distribution and Encourage use of the e-infrastructure.

In general, this model should not have any consequences for the work of the Resource Allocation Committee.

Contents

Executive summary	1
Background.....	4
Current funding	4
The Research Council requirement for financial user contribution	5
Objectives of model	5
Realistic expectations and Responsible use of resources.....	5
User driven demand.....	5
Fair resource distribution	5
Encourage use of strategic resources.....	5
Services covered by this model	6
Categories of users, contributors and target groups.....	6
0) Small projects with RCN funding below 15 MNOK	6
A) Large projects with RCN funding above 15 MNOK	6
B) Non-commercial Projects needing Dedicated Resources (Special or extraordinary needs) .	6
C) Commercial research and industry	7
The cost of resources and services	7
Investment (CAPEX) in hardware	7
Operation cost of services (OPEX)	7
Development of services	8
Administrative costs	8
Implications and consequences for Resource Allocation Committee	8
Examples of potential consequences for projects.....	9
Compute projects	9
Storage projects	10

Background

Traditionally the research infrastructures in Norway have been funded either directly by the Ministry of Education and Research or by grants from the Research Council of Norway (RCN), both for investment and for operation. Some universities have also contributed funding to research infrastructures. Before 2015, the infrastructure for High Performance Computing (HPC) were hosted, partly funded and operated by the four oldest universities in Norway. The Research Council funded part of the investment and operation directly and had the right to allocate HPC resources to research project through a Research Allocation Committee (RAC). The research projects using the HPC-service did not pay for any service or contributed any other funding. However, the universities contributed with in-kind funding (housing, man hours) covering part of the costs. In 2010 it was decided to set up a Norwegian Storage Infrastructure for digital scientific data. This was funded by a grant from the Research Council and resources were allocated to research project through the same Research Allocation Committee as for HPC.

Leading up to 2015, the Research Council and the four oldest universities agreed to set up a new organisation, UNINETT Sigma2 AS, which was given the overall responsibility for the e-infrastructure in Norway. The company's objectives are increased value-creation, by means of providing national e-infrastructure services in the fields of high-performance computing and storage of digital scientific data. The primary target groups are universities, university colleges, and research institutions in Norway. The company shall also serve as the means by which the same institutions execute in-house e-infrastructure projects, i.e. instead of setting up their own e-infrastructure projects, they formed a partnership to execute these. The Research Council and the four oldest universities provide the funding. It was a requirement from the Research Council to gradually introduce a model where also the research projects contribute funding for operation of the research infrastructures.

In addition, it shall be possible for commercial organisations to purchase services, regulated by the limitations in the COMMISSION REGULATION (EU) No 651/2014³. When the e-infrastructure is used in this context, the users should pay market prices⁴.

A first draft of this document was circulated in 2016 and all projects using the e-infrastructure were given the chance to provide feedback on the document. The main concerns in the feedback were: too quick introduction of the model, disadvantage when participation in projects with EU funding, a risk of large projects to be split into smaller projects to come under the threshold of payment or having smaller local systems. There were also supportive feedback including a broader range of the community contributing, aligning the e-infrastructure with other research infrastructures, better awareness by the user in planning the resources for a (large) project can certainly help the infrastructure services to meet those requirements in good time before the project start.

Current funding

The current funding is composed of the following elements:

- Research Council of Norway: 25 MNOK covering development and operation, international activities
- The four oldest universities: 50 MNOK covering investment, development and operation

³ <http://www.forskningradet.no/no/Statsstotterregelverket/1254004171884>

⁴ See chapter «Categories of users... A) Commercial research and industry» for details

- Research Council of Norway: Competitive infrastructure funding, with with an average of 35 MNOK/year for 2016-2017 and 57.5 MNOK/year 2018-2019.

The Research Council requirement for financial user contribution

The contract UNINETT Sigma2 AS has with the Research Council of Norway has the following requirements:

- Research infrastructures supported by the National Financing Initiative for Research Infrastructure should as the main principle include user contribution as an element in the funding of the operational cost.
- User contribution is an important principle for a sustainable funding of the services.
- User contribution should also contribute to the development and delivery of services which are requested by users.
- The introduction of the model for user contribution should not be a hindrance for projects to get free allocations based on scientific merit.

Objectives of model

Beyond the requirement from the Research Council to introduce a wider contribution model/user payments, it serves some other purposes.

Realistic expectations and Responsible use of resources

- When the research projects are not contributing to the funding of a service and consider it a free resource, there is a risk of unrealistic expectations for large resource allocations. This has already been evident in some of the discussions with research projects regarding expectations of very large storage areas. The consequences of large allocations are more evident for storage projects where data are stored over several years.
- Research projects with a need for large amounts of resources should not get disproportional or otherwise unreasonable amount of resources. This implies that these large projects should contribute in order to partly justify their large allocations.

User driven demand

- The willingness of large research projects⁵ to contribute to the funding of the e-infrastructure will reflect the importance of the services for the research.

Fair resource distribution

- The model must take into consideration the difference between CPU services (resource spent over a very limited period of time and can be reused) and project storage (resource spent over a longer period of time and cannot be reused until the project is completed), hence the need for an incentive to keep the usage under good stewardship and control. This objective implies some contribution for the /project storage.
- There should not be any profit for the company in this model.

Encourage use of strategic resources

- The model should not discourage researchers to deposit data in the archive, implying the /archive storage should be free.
- No general contribution for CPU hours from non-commercial research projects/users because we want the service be used for good research and make it easy to start using

⁵ Proposed definition of large projects: The top 20 projects in terms of resource allocations.

the services. To promote an efficient use of resources can also counter a growth of many small local systems.

Services covered by this model

Services initially covered by this model is compute services and storage services. New services that are under development e.g. Visualisation, Data Analytics, Portals, should be assessed with regards to this contribution model, before they are put into production.

Categories of users, contributors and target groups

So far, three different categories of contributors have been identified and different policies will be applied to each of these. The following principles set the foundation for the contribution model

0) Small projects with RCN funding below 15 MNOK

- For compute services (CPU hours) there is a as a general rule no payment (research project contribution) for smaller projects.
- For storage services the general principle should be that research projects must contribute when they use the services for active data (i.e. /project area), but there will be a free quota of 10 TB for smaller projects. Storage for long term archiving is free of charge

A) Large projects with RCN funding above 15 MNOK

These projects have funding from RCN and have dedicated funding for use of e-infrastructure. This will include projects in the SFF and INFRASTRUKTUR programs and projects with a RCN funding above 15 MNOK in other RCN programs.

The following policy is applicable to this category:

- Cost coverage including operational cost and an administrative fee to Sigma2.
- Their allocation is decided by the Resource Allocation Committee
- The policy will only apply to new projects or projects getting renewed funding after 2018, with the exception of SFF and INFRASTRUKTUR projects awarded in 2017.
- Payment will be handled by invoicing the project according to an agreement with the Research Council. For SFF and INFRASTRUCTURE projects this means:
 - The project will pay the budgeted amount for e-infrastructure expenses in the approved budget with the Research Council. This is a contribution to the operational cost.
 - If the project overspends or underspends resources, this will not change the amount the projects pay. This gives predictability for expenses for the project and for income for Sigma2.
 - Sigma2 will provide resources according to grants by the Resource Allocation Committee.

B) Non-commercial Projects needing Dedicated Resources (Special or extraordinary needs)

There are some projects with special needs in terms of e.g. a certain predictability for allocations within a specific timeframe due to for instance international obligations or dependencies for other services. Some examples are CMIP6, WLCG, Portals or on-demand-compute. The timeframe should be aligned with the allocation periods.

The policy for this category will be:

- Cost coverage including capital expenditure and operational cost for the required CPU hours and an administrative fee to Sigma2 (Special requirements might incur extra cost).
- Access to a fixed number CPU cores or storage space
- They can also apply for an allocation from the Resource Allocation Committee for additional non-dedicated resources.

C) Commercial research and industry

This includes assignments taken on by universities or research institutes under commercial terms^{6, 7}. For these projects, the following policy is applicable:

- Full cost coverage including capital expenditure (share of investment cost), operational cost
- Request for capacity made before 1st March or Sept 1st in order to have guaranteed quotas in the following periods. Requests outside these dates can be considered
- Pay a yearly or bi-yearly fee for a fixed agreed quota
- The quota is evenly distributed over the period for HPC usage⁸, outside of the Resource Allocation Committee quotas
- The total quota available for this category must within the COMMISSION REGULATION (EU) No 651/2014 and is decided by the Sigma2 Board

Another general principle is the need for research projects contributions when Sigma2 or one of our partners is managing and operating a service on behalf of a user community, e.g. portals or other specific services.

Special development of services should be kept outside of this model.

The cost of resources and services⁹

The following cost elements are usually considered relevant for e-infrastructures:

Investment (CAPEX) in hardware

This is the hardware needed to produce services. The cost for this is normally depreciated over 4 years, which gives a yearly cost for the hardware. The cost also include the procurement costs.

Operation cost of services (OPEX)

When it comes to the operation of the services the following cost elements are included:

- Housing (floor space, physical infrastructure)
- Electricity
- Computer network connections
- Maintenance
- Some basic license costs
- Staff for operation of the e-infrastructure and basic support

⁶ Considered as economic activity as opposed to non-economic activity

⁷ Including contracts or assignments from municipalities or county government (kommune eller fylke). Assignments on behalf of any ministry of the Norwegian government will need to cover operational expenses.

⁸ Some arrangements might be made to accommodate special needs given sufficient notification period

⁹ http://www.uhr.no/documents/A_Norwegian_Research_Infrastructure_Resource_Model_270214.pdf

Development of services

There is no cost element for general services in the model at present. Development of special or specific services must contribute with separate funding. This can be done in:

- Discipline specific advanced user support within the current guidelines
- Separate projects if the service is for general use, e.g. service for data analytics.

Administrative costs

There are administrative costs for Sigma2 related to operating the services. In general, the ratio between compute services and storage services is 70% vs. 30%. However, since the purpose of this model is not full cost recovery, it is suggested to simplify this. The suggestion is to add a percentage to the cost of CPU or storage covering the administrative cost and use the same percentage for all services.

Adjustment of costs

Costs will be adjusted yearly for new projects only, according to the general price index, new investments or change of other related services. Projects that have already signed a contract with Sigma2 will keep the same price for the agreed period.

Implications and consequences for Resource Allocation Committee

In general, this model should not have any consequences for the work of the Resource Allocation Committee as covered under each category (A-C). The Resource Allocation Committee will still control the majority of the e-infrastructure resources and allocate these according to their mandate and guidelines.

Examples of potential consequences for projects

Here are some specific examples¹⁰ of how this model will work with the current estimates of the costs for compute and storage services.

Compute projects

This table shows examples of different cost pr. CPU hour for various allocations:

	Cost pr. CPU hour in NOK		
	Category A, Large projects with RCN funding above 15 MNOK (OPEX)	Category B, Dedicated Resources (OPEX + CAPEX)	Category C, Commercial research (Full cost)
Million CPU Hours Price	0.07	0.14	0.29
75	5,400,000	10,800,000	21,600,000
40	2,880,000	5,760,000	11,520,000
30	2,160,000	4,320,000	8,640,000
20	1,440,000	2,880,000	5,760,000
5	360,000	720,000	1,440,000
1	72,000	144,000	288,000

This cost includes basic user support, installation of software and storage up to 10 Terabyte.

Alternative table with examples of dedicated CPU cores:

	Yearly cost for given number of CPUs, in NOK			
	Annual CPU hours	Category A, RCN/EU funding (OPEX)	Category B, Dedicated Resources (OPEX + CAPEX)	Category C, Commercial research (Full cost)
#of CPU cores Price		0.07	0.14	0.29
6000	52,560,000		7,568,640	15,137,280
4096	35,880,960		5,166,858	10,333,716
1024	8,970,240		1,291,715	2,583,429
512	4,485,120		645,857	1,291,715
256	2,242,560		322,929	645,857

¹⁰ The first version of this document included an example of cost using cloud services, but it has become increasingly difficult to find comparable platforms.

Storage projects

This table shows examples of different cost pr. Terabyte for various allocations:

	Yearly Cost pr. TeraByte in NOK		
	Category B, RCN/EU funding (OPEX)	Category C, Dedicated Resources (OPEX + CAPEX)	Category A, Commercial research (Full cost)
Tera Bytes (TiB)/ Price	1090	2550	3645
1,000	1,090,000	2,550,000	3,645,000
700	763,000	1,785,000	2,551,500
360	392,400	918,000	1,312,200
80	87,200	204,000	291,600
70	76,300	178,500	255,150
30	32,700	76,500	109,350
25	27,250	63,750	91,125
10	10,900	25,500	36,450

Storage projects will in addition apply the following principles:

- no charge is demanded for projects that consume less than 10 TB.
- usage above 10 TB are charged for all space used
- costs are charged retrospectively and on a yearly basis, based on the actual resource usage

In addition to the resource service for active data (project area), the national e-infrastructure can provide specific services to research groups or communities. This type of service intends to provide users from the community/science discipline better access to the data stored on the e-infrastructure (providing more effective use of the resources). A service level agreement (SLA) and a yearly fee is negotiated for such specific services, as each such service, e.g. a portal or community data browser, consume a varying degree of hardware and human resources, and projects will need to contribute separately for initial setup and extra software installation.