Big Data
National Research Programme

Call for proposals

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Swiss National Science Foundation
Wildhainweg 3
P.O. Box 8232
CH–3001 Berne
Tel +41 (0)31 308 22 22
E-Mail nrp75@snsf.ch
www.snsf.ch
www.nrp75.ch

Content

What are National Research Programmes (NRPs)?

Research carried out by National Research Programmes consists of research projects that contribute to the solution of contemporary problems of national importance. Under the provisions of Article 10, paragraph 2, of the Federal Act on Research and Innovation of 14 December 2012 (version of 1 January 2015) the Federal Council selects the topics and foci to be researched in NRPs and mandates full responsibility for implementing the programmes to the Swiss National Science Foundation.

The Federal Ordinance on the Federal Act on Research and Innovation of 29 November 2013 (version of 1 January 2015, art. 3 V-FIFG) describes the NRP funding scheme as follows:

"[1] The National Research Programmes (NRPs) of the Swiss National Science Foundation (SNSF) are a means of generating and conducting coordinated research projects that pursue a common goal.

[2] Topics of research are generally appropriate for National Research Programmes if:

a. Swiss research can make a significant contribution to the resolution of the problem;

b. solutions require research contributions from multiple disciplines;

c. research on the problem can be expected to produce research results that have practical applications within a five-year period.

[3] In exceptional cases, an NRP may also be used for the targeted creation of additional research potential in Switzerland.

[4] The following criteria are also taken into consideration in setting forth the topics of National Research Programmes:

a. the programmes can provide the scientific basis for decision-making by the government and administration;

b. the programmes can be conducted with international collaboration."

1. Summary

The data available in digital form is expected to increase at an explosive, exponential rate throughout this decade. It has been suggested that to be competitive, societies and businesses must be able to base their decisions and operation on data. Data has been termed the new oil, the idea being that data is an indispensable resource that will serve the role in the future that oil served in the past. At the same time, this development affects the privacy of individuals and renders entities vulnerable to unauthorized access to their data. Indeed, big data is slated to have a profound effect on society. We are in the midst of a revolution of the way we live, work, and interact.

The characteristics of big data, notably the sheer volumes of data, in many cases render state-of-the-art solutions in computing ineffective and call for novel solutions. Further, big data has the potential to enable new, as yet unexplored, applications. And the effects on society of big data need to be understood and shaped.

National Research Programme 75 “Big Data” (NRP 75) supports research that aims to provide foundations for the effective and appropriate use of big data across all aspects of society, including in business, public administration, and science. The anticipated research projects will provide scientific foundations for computing infrastructures capable of making effective use of big data; will enable the effective use of big data in specific applications; and will perform research with actionable outcomes on the effects of big data on society, covering economics, law, and education.

The total funding of NRP 75 is CHF 25 million for the projects with duration up to 48 months. On 24 June 2015, the Federal Council approved the new National Research Programme "Big Data" (NRP 75) and mandated the Swiss National Science Foundation (SNSF) to conduct the NRP.

2. Introduction

Background

Almost all areas of everyday life are accompanied, guided, and influenced by computing and communication devices that are embedded in large networks, most notably the Internet. These devices produce increasing amounts of data that are stored and processed. For example, consumer electronics such as smartphones can record audio, video, and user locations. Online services, including e-stores and apps, are recording user behavior, and also industrial and scientific settings are being equipped with technology that is capable of generating increasing volumes of data.

There is consensus that society and businesses can benefit substantially from the ability to base their functioning on these large amounts of data. Notions such as “data-driven business” and “data-driven society” have been advanced, suggesting that businesses that are able to base their decisions and operation on data are capable of being more competitive than those that are not. Indeed, data has been termed “the new oil,” the idea being that data is an indispensable resource that will serve the role that oil served in the 1900s. Data is viewed as an immensely valuable, but as yet relatively untapped, resource.

The vast and increasing volumes of data in many cases render state-of-the-art solutions in computing ineffective and call for novel solutions that are capable of contending with the specific challenges of big data. This results in highly relevant research questions in the area of Computing and Information Technology, where innovative foundational research is the key to achieving solutions that can process and manage big data efficiently and effectively.

Big data is slated to have a profound effect on society. We are in the midst of a revolution of the way we live, work, and interact within society. Questions on privacy, mapping of data with regard to property rights, and intellectual property rights are pressing, and it is challenging for regulatory measures, both locally and globally, to keep up with the changes in society that are brought on by technology. Likewise, citizens face challenges in understanding the implications of big data. As a result, big data calls for research in areas such as law, social sciences, and education.

The main research directions of proposals solicited by this call are threefold. While fundamental technological and methodological research and societal and regulatory aspects are addressed, a third focus is on applications of big data in areas such as personalized medicine, transportation, and digital humanities that require interdisciplinary research involving the computing and information technology domain on the one hand and the application domain on the other hand.

Practical significance

Big data holds substantial promise for businesses, public administration, science, and society as a whole.

Many enterprises have embraced big data projects in order to increasingly base their operation on data, an underlying idea being that decisions based on good data beat decisions based on opinions. In many businesses, big consumer data and data from social media may help to develop different aspects of marketing: products and services as well as sales promotion and advertising are brought more closely into line with actual needs. Direct client contacts can be improved based on a personalized and tailor-made offering. In addition, there is a greater emphasis on value-based development that may lead to more valuable and profitable products and services.

It is believed that big data will enable increasingly personalized medicine. Big data offers possibilities for more accurate diagnoses of individual health risks as the same medical data (such as e.g. high blood pressure) do not implicate the same risks for people with differing genetic predispositions. Likewise big data offers new possibilities for more targeted, individualized therapies which take both genome data and life characteristics into consideration. New technologies capable of extracting large amounts of data from samples or biopsies make it further possible to discover and use previously unknown factors involved in disease as drug targets or disease biomarkers.

In general terms, big data offers a new level of detail in our understanding of systems and processes. For example, in transportation, big data may enable a more detailed and up-to-date understanding of a transportation infrastructure, which may in turn be used for improving the utilization of the infrastructure and for improving the infrastructure itself. Similar holds for other infrastructure like water or power supply and communication. Big data can also be used for real-time management, e.g. in order to avoid traffic jams or global power failures.

The Internet of Things (IoT) will increasingly be the key source of big data in the future. A modern IoT infrastructure is expected to be in place in Switzerland before very long. The Swisscom telecommunications provider has recently decided to move forward with the installation of a Low Power Wide Area Network (LPWAN), an infrastructure for the Internet of Things. A pilot aims to equip the cities of Zürich and Geneva with this infrastructure by June 2015, and we can expect Switzerland to soon have a state-of-the-art infrastructure for the Internet of Things. This development presents new opportunities and challenges. The appropriate use of big data will substantially increase their positive impact.

In the future, big data will strongly facilitate disaster and emergency prevention and aid government in fulfilling its security tasks. Based primarily on IoT data, experts will be able to predict natural disasters and the course they may take with greater precision. In the event of an emergency, decision-makers will be able to rely on the integrated processing of facts, experiences and simulations. In addition, it will be possible to identify attacks on critical national infrastructures, organised crime structures, terrorist cells and vulnerable young people earlier and with greater precision. In this context, an optimal form of cooperation between humans and computers is crucial, because the spoon-feeding of big data to human decision-makers could also lead to bad decisions with catastrophic knock-on effects or even trigger disasters in the first place.

Big data also calls for new societal considerations, including regulatory and educational considerations. With the advent of big data, societal values such as privacy and solidarity will perhaps need to be reconsidered. As data is a valuable asset, regulations that enable appropriate protection of big data and big data rights are important for the effective exploitation of big data. Furthermore, the increasing digitization and more detailed capture of the lives of citizens, including their consumer behaviors, their financial status, and their medical records, affect their privacy. Legal provisions are needed to provide adequate protection against abusive profiling. Considerations of the balance between personal privacy rights and the exploitation of personal data represent an important theme. In the context of big data, national options for action are limited by international events, processes and debates; in addition, they need to comply with recognized technical and legal provisions. The international aspect is particularly relevant with regard to the area of data security vs privacy rights. Furthermore, big data enabled personalization – such as targeted campaigning and price discrimination – will have a big impact on economy and society, potentially increasing both social diversity and economic differences. Citizens need to understand how data is generated and used in order to behave purposefully and in order to be able to formulate their own informed opinions on privacy and the use of big data in general.

Overall, the intelligent use of data is slated to substantially impact decision making in business, public administration, science, and society; and value creation from big data is believed to become a major economic driver. As a result, it is important for any society to ensure that it is capable of reaping the benefits of big data along with its institutions and businesses. That said, it must also be able to anticipate the risks and to introduce effective counter-measures in good time.

Related initiatives

The importance of research in big data has been recognized internationally. New big data initiatives have been launched by funding bodies, and research institutions have established new activities

in relation to big data, e.g., by the creation of new professorships and the re-targeting of existing activities to embrace the opportunities and challenges offered by big data.

For example, the German so-called "Schwerpunktprogramm 1736" on Algorithms for Big Data was started recently. Also, an initiative by the German "Bundesministeriums für Bildung und Forschung" sets rules in relation to the management and analysis of big data in the scope of the German programme "IKT 2020 - Forschung für Innovationen."

In the USA, the National Science Foundation has recently launched a program called *Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA).* With an estimated budget of some $26.5 million, this program solicits calls for research on foundations and innovative applications of big data.

The European Commission has also launched a range of big data activities in the context of Horizon 2020. For example, the ICT 2015 call with a €0.56 billion budget includes a "Big Data" topic. As another example, the European Commission and Europe's data industry announced in autumn 2014 a commitment to invest €2.5 billion in a public-private partnership that aims to put Europe at the forefront of the global data race. Envisioned outcomes include 100,000 new data-related jobs in Europe by 2020 as well as 10% lower energy consumption, better health-care outcomes, and more productive industrial machinery.

A recent report from the World Economic Forum argues that it is imperative to chart a course that makes it possible to utilize big data for sustainable development. The report covers aspects such as the needs for access to data, shared policy frameworks, legal protection, capacity building at all levels, and for the recognition of individuals as both producers and consumers of data.

Target audience / targeted roles

- **Scientists in computing, information technology, mathematics, and statistics** who are interested in foundations that enable value creation from big data.
- **Scientists in all disciplines of science** where the use of big data holds promise. Such disciplines span social sciences and the humanities, including law and economics, natural and technical sciences, and mathematical sciences. For example, this includes scientists in medicine who are interested in personalized medicine, lawyers who are interest in digital forensics, and computer scientists who are interested in cyber security.
- **Forecast experts, strategic planners, and decision makers in industry and government** who are interested in using big data to identify hidden trends and correlations and to perform forecasts. This includes entrepreneurs, investors, and C-level managers in industry as well as councilors and board members of agencies at federal and cantonal levels and in the big cities.
- **Public regulators** who want to understand which public infrastructure services and which regulatory measures are needed. This includes Members of the Federal Parliament and the political parties as well as strategic decision makers in the State Secretariat of Economic Affairs (SECO), the Federal Office of Communication (OFCOM), the Federal Office of Public Health (FOPH), the Federal Office of Justice (FOJ), the Federal Office for Civil Protection (FOCP), the Federal IT Steering Unit (FITSU), and cantonal government.

- **Operative managers** in industry and public administration who are interested in using big data for real-time control activities, including sales and marketing experts in industry and experts in traffic and conflict handling.
- **Program leaders and teachers in institutions in secondary, tertiary, and continuing education** as strong needs for educational programs on big data are anticipated.

3. Goals of NRP 75

The NRP has four goals:

1. Advances in computing and information technology

The first goal of the National Research Programme is to achieve new scientific results in computing and information technology.

The programme aims to foster advances in areas of computing and information technology of critical importance for the exploitation of big data. Such areas include, but are not limited to, the following: i) data analytics, including, e.g., interactive analysis and visualization, ii) data management techniques and systems, and iii) data security, access control, and personal privacy. Furthermore, the programme aims to foster advances that enable new applications of big data. Emphasis is on advances where the benefits can be demonstrated on actual big-data workloads.

Expected output:

- Publications in high-quality peer-reviewed outlets
- New technologies in the field of data infrastructure and algorithms
- patents

2. Addressing societal, economical, regulatory (both local and gobal), and educational challenges

First, the programme aims at enabling research that offers actionable insight into the interplay between big data, economics, and society, including regulatory and educational aspects. Second, the programme aims to engage and cultivate exchange between scientists and key stakeholders – e.g., administrations at the federal and cantonal levels and civil society organizations.

Research may address the economic and social benefits of big data and how these can be exploited or how their exploitation can be fostered. It may study the nature and dynamics of socio-technological data and computing ecosystems in and across organizations as well as the effects of big data on the economy in general, on social and cultural life, and on democracy. Research may further address ethical issues of big data usage, and how education can enable individuals to safeguard their privacy and other rights. The identification of efficient regulatory requirements and challenges needs to be informed by local and international approaches. Emphasis is on research that analyses the anticipated effects of big data and explores their policy implications both nationally and internationally.

Expected output:

- Publications in high-quality peer-reviewed outlets

- Legislative activities on the political level and in government administration
- Education programs to enable citizens to safeguard their rights
- Improved public debate and awareness about expectable economic, social, and political changes

3. Enabling of applications

The programme aims to enable new or improved applications of big data that promise substantial benefits to society or the economy.

The goal thus is to strengthen the use of big data in applications where the management of big data is promising. One example is personalized medicine where also privacy and security are major issues. Other examples can be found in transportation, consumer behavior, and leisure and tourist management, for instance. Another example is digital humanities, a research area that integrates computing into the humanities and social sciences. For example, we would consider research that aims to facilitate innovative techniques for data mining, analysis, and visualization and integrates human feedback into the process of discovering knowledge from data available in the humanities and social sciences. Another area is civil protection, where natural disasters could be predicted earlier, allowing for efficient and coordinated measures to be taken, or attacks on key national infrastructures repelled at an early stage.

Expected output:

- Demonstrations of concrete applications

4. Strengthening of research

The programme aims to promote and strengthen research related to big data in Switzerland.

Switzerland has world-class researchers in computer science and information technology. The programme aims to generally strengthen the Swiss research capacity in relation to big data by enabling networking among researchers with interests that relate to big data. It is part of this goal that the programme becomes a visible counterpart to corresponding initiatives in Germany, the United States, Great Britain, and the European Community, thereby creating a momentum for increased big data research in Switzerland.

Indicators:

- Number of supported Ph.D. students
- Number of international exchange activities
- Number of organized national and international networking events

4. Main research topics

To achieve the four goals, the program contains three modules that correspond to the first three goals. The fourth goal, the strengthening of the Swiss research capacity in big data, will be achieved by organizing open national and international meetings that involve the participants in the research projects funded by the programme.

Module 1: Computing and information technology (CHF 9 mio)

Big data holds a tremendous potential for society and businesses alike. However, to unlock that potential, a range of challenges to computing and information technology must be addressed. The following are non-exclusive examples of areas where research is solicited.

Data analytics, data mining, and machine learning: Functionality capable of extracting value from data is essential in that it is what justifies the investments needed to collect and manage the data. The characteristics of big data present many challenges.

To be able to extract value from high-dimensional data in the terabyte range and beyond, it is often necessary to be able to further scale up existing algorithms, in particular in machine learning. For example, in order to contend with newly collected and continuously changing data, approaches and algorithms are needed that enable easy and fast scaling. One example challenge is how to exploit approximation without jeopardizing explanatory power so that algorithms with satisfactory computational complexity are achieved.

Different applications also offer new challenges. In some applications, sensor networks produce streams of real-time data. This data needs to be interpreted and integrated before automatic learning algorithms can take over and form optimal input-output mappings for prediction and decision making. Research is needed on how to scalably interpret and integrate such data based on domain and application expertise.

Data management services: Data management systems and infrastructures also face challenges due to the characteristics of big data. These systems use specific techniques for data storage, for securing the data against loss due to failures, for enabling concurrent and low-latency data access, and for supporting transactions.

The volumes of data and also the data arrival rates and latency requirements call for the effective and efficient use of distributed computing and of the capabilities offered by modern and novel computing hardware. Many techniques that were developed for systems with disk-resident data must be re-thought to work well in settings where much data is kept in increasingly large-capacity main memory.

Put together, big data offers new challenges to transaction support, indexing, and query processing; and techniques such as lossy and non-lossy approximation, sampling, and anytime computing can be expected to play increasingly important roles.

Security, access control, and trust: A big-data management environment should not only support the extraction of value from data. It must also offer a range of infrastructural services.

Perhaps most notably, it must be able to enable the security of the data and protect the privacy of the individuals that the data concern.

The reliance on cloud computing and outsourcing calls for efficient and scalable techniques that support the secure storage and management of data on external servers. Likewise, efficient and scalable techniques are needed that are capable of providing guarantees on the correctness of the computations when the underlying servers are untrusted. A big data infrastructure may also need support for real-time security monitoring to safeguard against malicious activity.

Because different kinds of data should be shared only with specific parties, access control mechanisms are needed that render it possible to guarantee that data is accessed by authorized parties for specific uses and purposes only. Access control is particularly challenging for highly dynamic and complex data that needs to be controlled at varying granularities. Big data also increases the risks of undesirable inferences that put the privacy of individuals at risk. To mitigate such risks, techniques are needed that address issues related to data linkage, knowledge of external information, and the exploitation of analysis results. Techniques are also needed that enable individuals to keep ownership of the data that concerns them. They should be able to access, modify, and delete their data, and they should be able to control the uses of their data.

Means are needed that make it possible to keep track of and reason about the provenance of data. Such means can offer insight into the veracity, trustworthiness, and quality of the data.

Module 2: Societal, regulatory (both local and global), and educational challenges (CHF 5 mio)

Big data is slated to impact society profoundly. On the one hand, big data holds the potential to enable society and businesses to be increasingly data-driven rather than opinion-driven, which is expected to increase efficiency and profitability. Likewise, big data is expected to enable new advances, e.g., in the area of personalized medicine and the development of personalized services and products more generally. In this perspective, the more data, the better; and the more the available data is shared, the better. On the other hand, data is often a valuable asset that needs to be protected from competitors in a global economy. Further, with the increased digitization of all aspects of our lives and the sharing of this data, including financial, medical, and political data, big data also impacts the privacy of citizens.

Big data applications often operate in a societal context that needs to be taken into account for them to achieve their full potential. For example, applications must comply with current legal regulations and must also take into account issues of social acceptability. It will be important to investigate the international context because of the cross-border data flows as well as local and global regulatory approaches.

Research in this module aims to offer an understanding of the effects of big data on society and businesses, to identify regulatory challenges and suggest solutions, and to educate citizens on the nature of big data, thus affording them a foundation for behaving rationally in relation to big data. Research topics in Module 2 include, but are not limited to, the following.

Societal acceptability of big data: It is important to understand which factors affect the acceptability of different applications of big data by public and private sector actors. For example, web-based services often collect data on their users. When interlinked, this data can result in detailed profiles of individuals. This capability holds a potential for profitable innovation, but can also lead to fear of losing control of one's data and individuality, which may lead to political regulation or market withdrawal. It is important to achieve a detailed understanding of legitimate concerns regarding acceptability, and it is relevant to identify educational and monitoring strategies that enable a balance between the benefits and risks offered by big data.

Regulatory challenges: Through de-anonymisation and integration with other data, non-personal data can be converted into personal data that is, in some cases, highly sensitive and confidential. This can result in restrictions on the use of data during their life cycle if privacy laws are respected. However, as non-personal data generated in Switzerland or elsewhere can also be transmitted to countries that do not have adequate data protection laws in place, de-anonymised data regarded as highly confidential under Swiss law could be available abroad and used by third parties without any restrictions. This shows that the protection is limited if data protection legislation takes effect late in the life cycle, as it is the case in Switzerland and the countries of the EU. Above and beyond this problem, there is little doubt that profiles of inhabitants of Switzerland are already circulating internationally. Inhabitants of Switzerland, while benefiting from a protected private space within the country, are at the very least “transparent customers” at an international level. What is more, the risk of intrusions into private space is also on the rise in Switzerland because data managers often do not have the specific know-how needed to distinguish between what is legally permissible and what is unpermissible when it comes to big data.

The goal must be to transform data protection so that its impact no longer depends on the contingencies of the data life cycle or on expertise that may or not may be available. This would guarantee a certain degree of protection for those affected, as long as no intentional breaches of the law occur. Research is necessary to determine the features this level of protection would ideally encompass and the measures needed to guarantee its provision as far as possible. In view of the future legal regulation of such privacy protection, the results of this research must take account of additional aspects, such as the balancing of legal interests in the private and public sphere as well as the balancing of privacy laws with the advantages of big data applications. Moreover, the legal situation should also be reassessed as regards other aspects of data handling, e.g. with regard to data usage rights that are enforceable in practice. To this end, we must first understand the extent to which big data represents a qualitative change that calls for new or changed regulations. One challenging aspect is that, as outlined above, much data will cross national boundaries, meaning that the international legal context and imminent changes due to international agreements, e.g. TTIP, also need to be considered.

As another highly significant aspect from a regulatory point of view, big data will enable new levels of personalization. In this setting, is price discrimination through “predictive pricing,” where different customers are charged differently for the same product or service, a legitimate commercial tactic? For instance, the envisaged price discrimination in the insurance sector results in a situation where clients can only benefit from low premiums if they allow insurance firms to monitor their lifestyles, while the solidarity principle in health insurance is abandoned because the premiums are calculated on the basis of individual risks. Also questionable are services for employers enabling them to estimate the probability of maternity, illness or family problems on the

basis of the information provided by candidates in their application documents. This not only reduces the career prospects of persons with undesirable prognoses, via statistical discrimination it may also lead towards a shrinking economy. Similar risk assessment services are already a reality in many areas or planned as future innovations (granting of loans, dating agencies, etc.) It is therefore necessary to study the side-effects of personalization-based innovations, to identify the resultant need for regulation and to prospectively analyse the effects of potential regulatory measures.

Big data ecosystems: Data, computing infrastructure, and all involved stakeholders constitute a socio-technological ecosystem that may be studied at different levels of scale, ranging from the view of the individual over the organizational view to the view of national and international economy. Some data is public, some is shared in groups, some can be purchased, and some is kept private. It is important to understand the dynamics of such ecosystems and how they can be developed and sustained. Monopolistic behavior of actors in ecosystems, social self-organization of ecosystems, and innovative sharing economies are aspects of particular interest. A specific topic in the context of big data ecosystems is open data for the public good. The general availability of data creates potential benefits as it maximizes the valorization of data. Research is solicited on best practices for making data available for the public good.

Economic and societal impact of big data: Big data offers opportunities for social and business innovation. From an entrepreneurial perspective, it is important to understand how innovation can be achieved through big data. From the perspective of national economy, it is important to understand the impact of anticipated innovations on market. Further, it is important to understand the impact of changes in markets on society from a sociological perspective. In addition, big data creates new opportunities for orientation, lifestyle improvement towards personal preferences, and more opportunistic relationship management for individuals and groups. The resulting social change is important in a psychological and anthropological perspective. All these perspectives are of high relevance to political policy making.

Education: The Web is playing an increasingly important role in many aspects of the lives of individuals, including their interaction with the authorities, their information needs, their needs for education and training, their social lives, and their lives as consumers. At the same time, the Web is becoming an increasingly important platform for the collection and sale of data about individuals by entities such as data brokers. It is important that individuals understand these developments, including how data about them is collected, integrated, and used. Further, it is important that this collection and use of data is regulated appropriately and that technical support for such regulation is available. The scope of the NRP includes projects that establish foundations that address the needs for a transparent Web and the rights of its users.

Methodological implications: It is important to understand the implications for research methodology of having massively more data available in relation to a given problem. Some observe that the scientific method was only developed because data was scarce, inexact, and expensive to collect, and they proceed to claim that the “data deluge” renders the scientific method obsolete. This then leads to the “end of theory.” Others argue that without theory against which hypotheses can be tested, no insights can be gained with any degree of certainty.
(Projects addressing this type of methodological implications are expected to involve scientists from the area of data analytics, for example via collaborations with projects in module 1.)

Module 3: Applications (CHF 9 mio)

Many domains embody applications that may benefit from the use of big data. In transportation, for example, the collection of data presents a huge potential for different applications that relate to the management of a transportation infrastructure. In the medical domain, detailed data on patients may enable more precise diagnoses and may enable understanding the effects of targeted therapies, thus enabling increasingly personalized medicine. In such application domains, research in the area of computing and information technology must work hand-in-hand with expertise in the application domain to achieve effective solutions.

In the humanities, available big data is often noisy. Techniques for making data machine readable such as optical character recognition (OCR) and speech recognition are error prone. Old-fashioned fonts and faded or stained prints render OCR difficult, and hand-written documents typically cannot be OCRed. Similarly, old tapes render speech recognition problematic. Moreover, historical languages cause challenges. These data characteristics call for robust techniques, e.g., for data mining, that work when the data is erroneous and diverse.

Natural events, e.g. the course taken by a storm, the amount of rain and water in a thunderstorm, but also the course of an illness, a large gathering of people at a sporting event, can today be predicted and analysed in view of specific aspects (size and density, quality of the event) based on a large number of heterogenous data sources that are interconnected. Such early warning systems, used efficiently, can be of great use in civil protection. Apps available for mobile phones can transmit early disaster warnings. Developing such apps and encouraging the public to use them, i.e. making them more socially acceptable and fostering the necessary technical and psychological competencies – these are further issues that could inspire interesting interdisciplinary and transdisciplinary research projects in Module 3.

The above are just a few examples of domains with applications of big data where current methods, algorithms and infrastructures are not capable of handling the challenges posed by big data.

This module solicits proposals for innovative applications that require close collaboration between researchers from the area of computing and information technology and domain experts to solve the specific challenges. This also includes synergies of research located both in Module 1 and Module 2.

5. Research Guidelines

Challenges posed by big data

“Big data” is often defined as any data that cannot be handled by today’s solutions, techniques, and technologies. The effective use of big data offers challenges that are caused by different characteristics of the data, notably the volume, variety, velocity, and veracity of the data, although other challenging aspects may also be envisioned. Here, “variety” refers to the fact that it is often important or desirable to be able to integrate data from multiple sources. Doing so can be challenging for a range of reasons. Next, “velocity” often refers to the arrival rate of the data and may also refer to the speeds with which processing must occur. Finally, “veracity” refers to quality

of the data. These characteristics yield challenges in a range of areas, including, but not limited to, the following:

- data analytics for vast amounts of dynamic data and with stringent time requirements
- cost-effective data management and query processing, e.g., of vast amounts of dynamic data
- access control and privacy protection for person-identifiable data
- data security, e.g., in cloud computing where data management is outsourced to third parties
- certification of correctness and regulatory compliance of computations by third parties

Big data also offers challenges that are specific to different application areas.

Types of projects

The program welcomes applications for research projects on big data as indicated in Modules 1-3. Research in Module 1 should address fundamental problems in the field of big data that advance the state-of-the art methods, algorithms, and infrastructures. Typically, but not exclusively, researchers may come from computer science, mathematics, statistics, or engineering. Preference is given to projects that involve the invention of innovative new methods and techniques that also deal with the empirical evaluation of prototype implementations on big-data workloads.

Research in Module 2 is typically conducted by scientists from sociology, law, political science, educational science, ethics, psychology, economics, or other relevant fields. It is anticipated that projects in Module 2 aim for actionable outcomes, including specific policy recommendations.

For applications that fall within Module 3, preference is given to those that are driven by real-world big data workloads and that involve prototype implementations and empirical evaluations of the contributions. Proposals should:

1. Identify and characterize the value that may be derived by big -data research for their application domain
2. Identify and characterize the kinds of big data workloads to be supported
3. Identify and describe the data characteristics that current techniques and systems are incapable of contending with and that the research will enable support for
4. Describe the synergetic effects between the involved application and computing domains

Projects in Module 3 are interdisciplinary, bringing together experts from computing and information technology and the relevant application domains including societal and regulatory issues.

Cooperation

Collaboration with other projects within and outside the NRP, including international cooperation, is encouraged.

Promotion of junior researchers and setting of research structures

NRP 75 aims to strengthen scientific competencies by supporting junior academics and scientists. Projects that emphasize this aspect by, e.g., enabling dissertation projects or strengthening basic research structures and platforms, are encouraged.

6. Submission procedure and project selection

General conditions

One call for proposals is foreseen. In the event of significant thematic gaps, a second call for proposals may be launched.

Research conducted of NRP 75 is limited to a maximum of 48 months. The average budget of a project is expected to range between 300‘000 to 600‘000 CHF. This range is meant as a reference point. Smaller and larger budgets are possible.

To allow for optimal coordination, approved projects must start no later than six months after the date of the approval.

A two-stage submission procedure is used: Pre-proposals are submitted first, a selection of proposals is then invited so submit a full proposal. The Steering Committee expects pre- and full proposals to be submitted in English, unless it can be shown that either German or French is intrinsically better suited to the research topic. Before handing in the proposal in German or French, the programme manager of the NRP 75 must be contacted for approval. For evaluation purposes, the pre- and the full proposals must be written in the same language.

Cross-border research projects are supported if the competence of researchers from abroad is essential for realising the project. As a rule, the share of financing requested for researchers abroad may not exceed 30% of the overall budget, and the person responsible for the project abroad may not be assigned the role of corresponding with the SNSF. For applicants from abroad, the norms and salary rates of the relevant country will be applied mutatis mutandis, with the SNSF maximum rates generally serving as the upper limit.
Before submitting a proposal for a cross-border research project, please contact the programme manager of NRP 75.

All forms, rules of procedure, and instructions for the submission of proposals can be found on the mySNF portal under ‘information/documents’ after selecting the corresponding NRP and creating a new application.

Online submission on *my*SNF

Pre- and full proposals have to be submitted on the mySNF portal (www.mySNF.ch). For this, user-registration is needed. User accounts obtained in the past are valid and provide access to all the funding instruments of the SNSF. It is recommended to request new user accounts as early as

possible; however, they need to be requested no later than five working days before the submission deadline viathe mySNF portal.

Pre-proposals

The deadline for the submission of pre-proposals is **13th January 2016**.

In addition to the data that has to be entered directly on mySNF, the following documents need to be uploaded:

- Project description (as a PDF file)
  Applicants must use the document template provided on the mySNF portal under 'information/documents'. The project description must not exceed six pages.
- Short CVs and publication lists of all applicants (as PDF files)
  The CVs must not exceed a maximum of two pages each. Each publication list must contain the five most relevant publications only. Links to full publication lists may be included.

Project descriptions and CVs exceeding the indicated length will not be considered.

Full proposals

The deadline for the submission of full proposals is **24th June 2016**.
In addition to the data that has to be entered directly in mySNF, the following documents need to be uploaded:

- Research plan (as PDF file)
  Applicants must use the document template provided on the mySNF portal. The research plan must not exceed 20 pages.
- Short CVs and publication lists of all applicants (as PDF files)
  The CVs must not exceed two pages each. Links to publication lists may be included.

Supplementary documents (support letters, confirmation of co-operation or co-financing, forms regarding international cooperations, etc.) can be uploaded on mySNF.

Project selection

The Steering Committee evaluates the submitted pre-proposals and makes a final decision based on the selection criteria outlined below. In making its decisions, it may refer to international assessments. Authors not invited to submit a full proposal will be informed accordingly by means of a ruling.

In the second stage of the submission procedure, the Steering Committee will invite the authors of the selected pre-proposals to submit a full proposal. In the invitation, the Steering Committee might give recommendations or set conditions for the full proposal. Based on external reviews as well as on their own evaluation, the Steering Committee will propose full proposals to be approved or rejected by the National Research Council (Programmes Division and Presiding Board).

Selection criteria

The Secretariat of the Division Programmes checks whether the personal and formal requirements are met, before forwarding the proposal for scientific review (cf. Funding Regulations of the SNSF). Pre- and full proposals that do not meet the personal and formal requirements will not be processed further.
Pre- and full proposals will be reviewed on the basis of the following criteria, with Compliance and Scientific quality being the most important criteria:

- **Compliance with the goals of NRP 75**: proposals must reflect the programme's objectives as outlined in the call and comply with its overall framework.
- **Scientific quality**: proposals must fulfill international state-of-the-art criteria with respect to scientific quality as well as methodology. Proposals must contain an innovative component and be relevant as compared to completed or running research projects in the same field.
- **Inter- and transdisciplinarity**: projects with research questions addressed by different disciplines or that demand approaches that transcend the boundaries between science and practice must secure adequate cooperations between the actors, project management and the methodology.
- **Application and implementation**: the potential for practical application and implementation of results is a key element of National Research Programmes. Projects of high practical relevance are therefore given priority.
- **Personnel and infrastructure**: adequate personnel resources and an adequate infrastructure must be secured for the project.
- **Response to given comments**: the Steering Committee may give comments or suggestions or recommendations to the research teams when inviting them to submit a full proposal. (This criteria is applicable to the full proposals only.)

Budget

Total funds of CHF 25 million are available for this NRP. The provisional allocation of this funding between the different research modules and administrative activities is as follows:

Module 1 Computing and Information Technology	9 Mio CHF
Module 2: Societal, Regulatory, and Educational Challenges	5 Mio CHF
Module 3: Applications	9 Mio CHF
Knowledge Transfer and Administration	2 Mio CHF

Schedule

The following schedule is envisaged for NRP 75:

Call for pre-proposals	18 September 15
Submission of pre-proposals	13 January 16

Invitation to submit full proposals	end March 16
Submission of full proposals	24 June 16
Final decision on full proposals	October 16
Start of research	1 January 17

7. Contact

For questions regarding the submission of pre-proposals and full proposals, please contact the Programme Manager: Christian Mottas, nfp75@snf.ch or 031 308 22 22.
For questions concerning salaries and eligible costs, please contact the Head of Finances Roman Sollberger, roman.sollberger@snf.ch or 031 308 22 22.

Technical help with *my*SNF and electronic submissions

Hotline:
Tel. + 41 31 308 22 99 (Français)
Tel. + 41 31 308 22 00 (Deutsch)
Tel. + 41 31 308 22 88 (English)

E-mail: mysnf.support@snf.ch
*my*SNF Homepage: www.mysnf.ch

8. Actors

Steering Committee NRP 75

President
Professor Christian S. Jensen, Department of Computer Science, Aalborg University, Denmark

Members
Professor Sabrina de Capitani di Vimercati, Computer Science Department, Università degli Studi di Milano
Professor Erkki Oja, Computer Science and Engineering, Aalto University, Finland
Professor Reinhard Riedl, University of Applied Sciences, Berne
Professor Caroline Sporleder, Institut für Informatik und Göttingen Center for Digital Humanities, Uni Göttingen
Professor Rolf H. Weber, Chair for International Business Law, Faculty of Law, University of Zurich

Delegate of the division Programmes of the National Research Council

Professor Friedrich Eisenbrand, Discrete optimization Group Disopt, EPF-Lausanne

Programme Manager

Christian Mottas, Swiss National Science Foundation

Head of Knowledge Transfer

NN

Representative of the Swiss Federal Administration

Willy Müller, Federal IT Steering Unit FITSU, Federal Department of Finance

For the State Secretariat for Education, Research and Innovation (SERI), Berne

Claudine Dolt, SERI, Berne