TIP Working Party
CO-CREATION PROJECT
2019-2020

Case study from the Russian Federation

HSE Joint Laboratories

OECD

Disclaimer

This study is released under the responsibility of the authors of the document. It does not represent the official views of the OECD or of the governments of its member countries. The Organisation cannot be held responsible for possible violations of copyright resulting from any written material in this case study.

Please cite as:

Executive summary

Joint research laboratories set up within National Research University Higher School of Economics (HSE): Laboratory of Methods for Big Data Analysis (LAMBDA) founded in 2015 with Yandex – Russia's most valuable digital company and the biggest search engine; and Laboratory of Financial Data Analysis (LFDA) opened in 2019 in cooperation with Sberbank – Russia's largest state-owned bank. Both laboratories conduct research in the field of big data, develop new algorithms for big data analyzing and practical application using artificial intelligence tools. They mainly focus on basic research for applied purposes.

In the each initiative, there are two core partners: research team located at the university and a business partner, additional partners include a transnational organization (European Organization for Nuclear Research – CERN) and other Russian research institutions. The focus on solving applied problems, which are important both from the scientific and industrial point of view, as well as the combination of scientific, innovative and educational activities contribute to the dynamic and mutually beneficial development of cooperation. In both laboratories partner interaction builds on strong relations on the personal level and regular communication.

The joint labs are less formalized which at the one hand makes it difficult to identify common performance indicators and universal management practices and on the other hand allows them more freedom to act. Partner agreements are of formal nature and kept in general terms with have little or no direct impact on the activities of joint laboratories. A data restricted access strategy even among co-creation partners is widespread. IP related issues are not always fixed in the contract. KPIs are not formally set out in the agreements between partners, projects implementation depend mainly on the desire of the current management.

The systematization of different management approaches revealed the great role of strong relations on the personal level, mutual trust and regular (not necessarily structured) communication for the success of joint co-creation initiatives. Another success factor is the carrying out teaching activities and active involvement of students in R&D. It contributes to the development of human resources for both research teams and companies.

By Dirk Meissner, Valeriya Vlasova

HSE Joint Laboratories

The case originates from National Research University Higher School of Economics (HSE) – the leader in Russian education and one of the preeminent economics and social sciences universities in Eastern Europe and Eurasia.

General description

Joint Laboratories set up within NRU HSE are the result of partnership between the university and the largest companies in Russia in the specific field. *Laboratory of Methods for Big Data Analysis (LAMBDA)[1]* founded in 2015 with Yandex – Russia's most valuable digital company and the biggest search engine; *Laboratory of Financial Data Analysis (LFDA)[2]* opened in early 2019 in cooperation with Sberbank – Russia's largest state-owned bank and one of the leading financial institutions worldwide. Established as a part of the Faculty of Computer Science, both laboratories conduct research in the field of big data, develop new algorithms for big data analyzing and practical application using artificial intelligence (AI) tools.

Research focus and objective of the co-creation initiative

Laboratory of Methods for Big Data Analysis. The LAMBDA's aim is creating an excellent research center to improve expertise in big data processing and analysis. It focuses on using machine learning and data analysis methods to solve issues in fundamental sciences. The research team works with the leading scientists in physics and astrophysics in search of answers to the mysteries of the universe. The emphasis on problem solving in fundamental sciences is combined with a strong interest in commercializing the results obtained.

Initially, the team consisted of employees from a scientific division of Yandex. For many years, Yandex worked in close cooperation with the European Organization for Nuclear Research (CERN), participating in its experiments. Now cooperation is developing based on LAMBDA, where a high level of scientific expertise and computational capabilities (partly provided by Yandex) are concentrated.

In 2018, HSE has joined the LHCb collaboration at the Large Hadron Collider and became an official LHCb associate member at CERN. Interactions with CERN allows LAMBDA (HSE) becoming a part of international research network (as today CERN has 23 Member States) and researching the events of the Large Hadron Collider, thus working with foremost scientific tasks and raising the efficiency of data analysis.

Laboratory of Financial Data Analysis. The idea of establishing the Laboratory was a continuation of joint projects with Sberbank to provide consulting services and research (Fig. 2). The main area of LFDA research is the analysis of large volumes of financial and behavioural data. LFDA focuses on the application of machine learning methods to financial services, such as credit scoring, using techniques such as the interpretation of complex neural network models, learning reinforcement, natural language processing, and competitive networks in the removal of directed information from samples.

---

[1] https://cs.hse.ru/lambda/
[2] https://cs.hse.ru/big-data/bayeslab/daft/

Major banks are now introducing machine learning, both for basic tasks and for new areas such as chatbots. Sberbank, as the largest bank in Russia, needs specialised research, as methods – such as credit risk assessment, automation of support services, personalised marketing – often require further development when applied in the banking sector. LFDA provides Sberbank with solutions for all these issues.

Cooperation with Sberbank allows LFDA to exchange experience with corporate researchers and establish links with other research groups engaged in artificial intelligence (e.g., the Moscow Institute of Physics and Technology (MIPT), the Skolkovo Institute of Science and Technology (Skoltech), and ITMO University (ITMO: Information Technologies, Mechanics and Optics). Another advantage is the interaction with Sberbank itself, since the bank has the Laboratory for Artificial Intelligence, where professional researchers work.

Partners and their functional roles

Joint Laboratories with HSE involvement are physically located at the university campus under the administration of HSE. They feature shared funding involving international research teams. The laboratories focus on basic research for applied purposes and implement a combination of scientific, innovation and educational activities.

LAMBDA' research team is located at HSE and performs both research and educational activities in cooperation with Yandex and the European Organization for Nuclear Research (CERN) (see Fig. 1). HSE is responsible for setting the research priorities, considering both CERN's and own scientific interests and the needs of business. R&D is conducted by both HSE and CERN. In this CERN provides LAMBDA with the access to real data (researching of the events of the Large Hadron Collider) and international integration (working with foremost scientific tasks), while LAMBDA assists CERN in solving application problems. Yandex provides computational capabilities and engages mostly at the experimentation and development and commercialization stages. Funding is mainly provided by the state through grants (about 50%), the rest – through Yandex (35%) and HSE Academic Fund Programme (15%).

1   Close cooperation for many years
2   Creation of the Joint Laboratory based on Higher School of Economics
     – Laboratory of Methods for Big Data Analysis (LAMBDA)
3   HSE joined LHCb collaboration and became an official LHCb associate
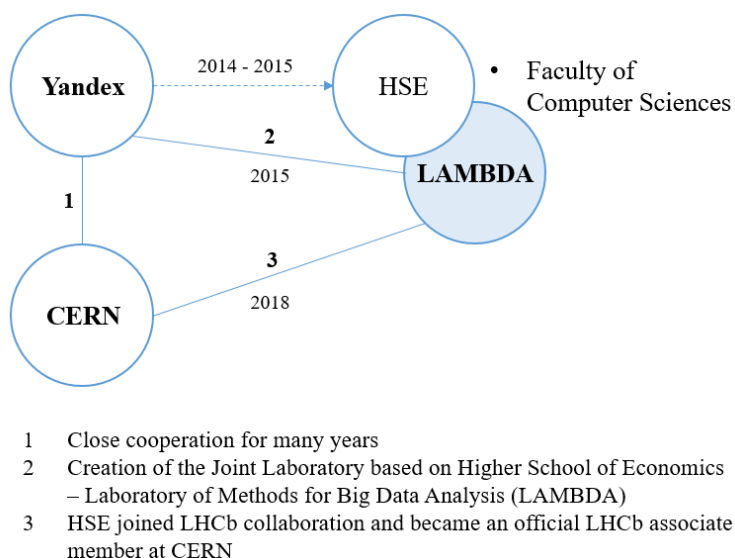   member at CERN

Figure 1. Framework of co-creation partners – LAMBDA

Source: Elaborated by the authors

LFDA's research team is also located at HSE (see Fig. 2). In addition to conducting R&D, it solves complex problems related to the application of machine learning and neural networks in practice and presents the results on international conferences. Sberbank is responsible for defining the vision and research objectives, provides funding and computational capabilities. Despite the focus on applying machine learning methods to financial services Sberbank does not provide access to real data, so the obtained results are mainly theoretical. The project involves frequent interaction with other Russian research institutes to verify and get a professional evaluation of R&D results.

1 One-, short-time projects over several years
2 Creation of the Joint Laboratory based on Higher School of Economics
   – Laboratory of Financial Data Analysis (LFDA)
3 Sberbank acts a platform for various research groups in Russia to
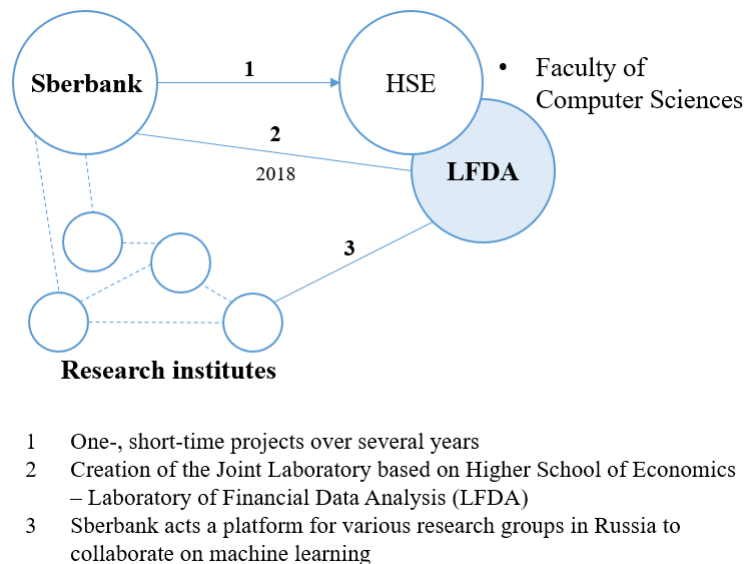   collaborate on machine learning

Figure 2. Framework of co-creation partners – LFDA
Source: Elaborated by the authors

Through the combination of scientific, innovative and educational activities contribute to the dynamic and mutually beneficial development of co-creation. This is achieved through carrying out teaching activities and active involvement of students in R&D for solving applied problems which turn out one of the *key success factors* of joint laboratories. The educational activities include:

– Organizing international academic seminars, hackathons and summer schools on big data analysis;
– Developing online courses on machine learning and interdisciplinary research projects with other HSE departments;
– Providing various internship opportunities and postgraduate students supervision.

Laboratory of Methods for Big Data Analysis. The keystone to success in building a strong research team is the involvement of students in projects and providing them with mentoring and feedback. The number of students involved in this practical training is increasing every year. In 2019, about 50 people were involved while in 2018, there were only 20 students. Together with Yandex, it is also planned to develop an online platform for supporting joint research projects and machine learning training for people with different levels of education.

Laboratory of Financial Data Analysis. LFDA plans to recruit students from the Master's Programme "Financial Technologies and Data Analysis" launched jointly with Sberbank in 2017, to work as interns in the Laboratory. Interns will be able to participate in developing solutions to

challenges in machine learning with the guidance of experienced researchers, as well as get to know the inner workings of the bank and interact with its developers and data scientists.

Management strategy

The joint labs are less formalized which at the one hand makes it difficult to identify common performance indicators and universal management practices and on the other hand allows them more freedom to act. Partner agreements are of formal nature and kept in general terms with have little or no direct impact on the activities of joint laboratories. A Memorandum of Understanding was signed between CERN and Yandex, later joined by HSE (LAMBDA). A Donation Agreement has been concluded between CERN and Yandex, which defines the general objectives of cooperation and financing flows. The contract was signed for five years and recently extended again. The cooperation between HSE (LFDA) and Sberbank is governed by annual contracts, but there is no reason to expect that they will not be renewed. In general, the responsibility for co-creation management lies always the research team (HSE).

Laboratory of Methods for Big Data Analysis

LAMBDA, additionally, assigns for each project an internal manager (who is often also a researcher) and an external expert from industry or science. These managers form the project team, solve administrative and communication issues, and conduct monitoring and performance evaluation of the project.

While key performance indicators (KPIs) are not formally set out in agreements between partners, internal project managers informally monitor the quality of work done and can make changes to the work plan if expected results are not achieved. There are various indicators reflecting the work done: number and quality of scientific publications, number of patent applications, intensity of participation in workshops, conferences and other public events, publication in the media of discoveries made in the HSE-Yandex cooperation. The results of the Laboratory's work are published in the leading international journals (e.g., the Journal of Instrumentation (JINST), Physical Review Letters), and are presented at such conferences as PyDataNIPS, KDD, ICML, CHEP, and PyData. Progress reports on project implementation are also regularly published. In cooperation with CERN, reports are submitted on a regular basis, as each project is important to the overall development of the experiment.

Laboratory of Financial Data Analysis. Key performance indicators (KPIs) are informal and include conference presentations that can be considered as publications, since research results are not published in scientific journals. Another KPI is the development of competencies in research areas important for Sberbank. For example, the HSE research team studies the information about key research groups, trends and prospects in the field of artificial general intelligence. In other words, the Laboratory reviews international experience and best practices. In general, KPIs are qualitative, not quantitative. One of the quantitative indicators is participation in international conferences from the list approved by Sberbank. The list includes the International Conference on Machine Learning (ICML), the Conference and Workshop on Neural Information Processing Systems (NeurIPS), the International Conference on Learning Representations (ICLR), the KDD Conference (KDD – Knowledge Discovery and Data mining). The project and its success are evaluated at the end of the study, but sometimes Sberbank may ask for adjustments while the project is still under way.

As there are no formal key indicators to determine the importance of the project, the beginning of its implementation depends mainly on the desire of the current management. Moreover, it is sometimes very controversial to assess the results. There is a risk that there will be

a discussion about the relevance and significance of the results obtained and their fulfilment of the initial expectations. It would be useful if the co-creation agreement could set out what the researchers have in common. If a purpose is defined, researchers will not be required to solve problems that do not fit the purpose.

IP regulation

IP protection strategies vary among laboratories, but in general, a data restricted access strategy even among co-creation partners is widespread.

In the LAMBDA case, IP related issues *are not fixed* in the memorandum and contracts. Yandex gets the rights to the achieved R&D results, but they are freely available to the research team (HSE). If the project is commercial, only those directly involved have access to the data. CERN periodically publishes the results on its website. This data is open and can be studied by everyone, while accepting the terms of the user agreement. Although all rights to research results are transferred to the business partner, LAMBDA may also use these results for its own purposes. This is not regulated by an additional contract, but is mainly due to the established trust between the co-creation partners. Informal communication promotes trust and openness.

In the LFDA case, IP related issues *are defined* in the contract. All intellectual property rights are transferred to the business partner (Sberbank). No specific IP enforcement mechanisms are used. Owing to specificity of the results in the field machine learning and artificial intelligence, the basic method of their protection is complexity. In general, the access to project results are open. The result is the development of a method that the research team (LFDA) presents at international conferences followed by the publication of a scientific paper. When publishing a theoretical model, i.e. without testing is on the real data, there is no need to restrict access to project results.

Communication between co-creation partners

In both laboratories partner interaction builds on strong relations on the personal level, mutual trust and regular (not necessarily structured) communication.

LAMBDA partners use both formal and informal channels of communication. Communication between the research team and CERN is structured and more formal. Meetings are held every four months and are organized in the framework of LHCb. The discussed issues are more structured, as they synchronize the work of hundreds of universities. Research team and Yandex communicate mainly on an informal basis, preferring frequent free exchange of news on current research findings, planned activities and publications. Process of project evaluation is also not formally regulated. Trust, openness and regular communication are key success factors in this joint initiative, as they promote respect for each party's obligations and effective teamwork to achieve common goals. Moreover, trust is indispensable to long-term cooperation.

LFDA partners prefer regular communication on a formal basis. They organise joint workshops and research seminars to discuss and verify the achieved R&D results. External participants, mainly from the scientific community, are also invited. At the end of each stage of the project, the research teams has to prepare a detailed report.

Challenges during the co-creation process

Inflexible regulatory system in the country may impose the challenge for further development. For example, there are inconsistencies between some norms of Russian and European legislation, which makes formal interaction difficult. This affects the preparation of accounting reports at HSE and the solution of administrative issues that require working with a

large volume of documentation. Moreover, joint cooperative projects are required to comply with national programmes and priorities. However, key performance indicators of such projects are not formalised, therefore project proposals may be underestimated. As a result, there is no strong support of co-creation initiatives from public authorities. Bureaucratic and unclear procedural requirements also hinder effective interactions. The need to compile many acts and reports in accordance with the state standards (GOST) increases the time for recording project results.

There appears a strong incentive for translating AI research into practical applications but the latter faces difficulties at the commercialization stage. The scientific breakthroughs in machine learning will be useful to the industry only in a few years. In addition, the conducted research does not always lead to the valuable results, sometimes it is just incremental improvements, which are technically much more difficult than what is currently in use. Therefore, the customer must have a very well developed technological infrastructure to adopt the results, but usually that is not worth the effort.

Furthermore it becomes more difficult creating a team of researchers capable of working on knowledge-intensive, complex projects on a long-term basis. Additionally, there is an issue of attracting young specialists.