

Open Data Management at St. Petersburg Polytechnic University

The Polytechnic University was established in February 1899, more than 120 years ago. Now, it is one of the largest universities in the country, the leader in the field of higher engineering education in Russia, with more than 30,000 students, in the structure of 12 institutes and 34 higher schools.

As for the university library, it appeared immediately upon the foundation of a new type of engineering university. From the very beginning, the foundation was formed by leading scientists in the profile of the university. The drawings depict the reading room in 1900 and then 100 years later. The atmosphere of academia has been preserved: antique tables and chairs, bookcases.

About the fund, the volume of the printed fund is decreasing by about 500,000 over the past 10 years. Subscriptions to electronic databases are growing; a significant part of which is the national subscription of the Ministry of Education and Science of Russia. But the main priority in the university (and in the report) is the resources collected by the library. The repository of the university is called the SPbPU Electronic Library, the resource is registered in the national database registry.

In 2005, the Fundamental Library was transformed into an Information and Library Complex. This additionally absorbed the IT and digital resource structures. On the national context of working with open resources in scientific and educational activities, the general trend of the development of “openness” is reflected both in the national and university context. The desire to open everything is gradually supplemented by different aspects and options of openness. The types of access to metadata and resources themselves are separated, the types of access are standardized, methods of linking data in the global space are being developed, and knowledge graphs are being built focused on different types of tasks to be solved.

For Russian universities, the era of active work with electronic resources came 20 years ago. This was when the Ministry of Education and Science of Russia began to require mandatory access to the electronic library system (or ELS) for university licensing and program accreditation. This term became known as the repository (electronic library), which contains educational resources on disciplines taught at the university. On this wave, new structures began to appear in publishing houses that provide access to ELS with digital copies of the educational publications produced. Aggregators began to appear next. But some universities, especially large ones, began to create their own ELS. The Polytechnic University also created an Electronic Library in 200 by order of the rector.

About 15 years ago, there was a demand from the Ministry to create an electronic information and educational environment at the university, where a separate paragraph indicated the need for electronic educational resources to be present in this environment. Moreover, it was separately indicated that access to educational electronic resources for students and teachers should be provided from any device connected to the Internet.

In 2015, also by order of the Ministry of Education and Science, there was a requirement to keep all graduate works of students and postgraduates in the university's ELS.

Thus, there were quite a lot of incentives to work with electronic resources in the educational process. And these requirements were checked during regular accreditation inspections. Since 2020, a national project on the digital transformation of universities has been carried out in the country, and all the results achieved over the past 20 years in terms of the use of electronic resources in scientific and educational activities should be implemented in the digital platform of the university.

Considering the national context in terms of open data and scientific activity, open data was introduced by a separate legislative act - this is open (state) data on the activities of government agencies, grouped into 16 areas of activity. There is no science among them! In the global Open Data Barometer system, Russia's position is fairly well represented. On open science and its modern interpretation, we pay attention to only one point - phrases about the need to create infrastructures and ensure connectivity in various aspects of the applicability of this term are repeated.

Only the tip of the iceberg is really open now – the texts of articles and data present in articles in the form of appendices or materials included in the text. So much data is missing! It is not for nothing that publications have appeared about the crisis of reproducibility of experiments.

Also on the slide is a fragment of the OECD's recommendations on whether to transform "open access" into "enhanced open access to achieve a balance between cost constraints, privacy, priority protection, prevention of data misuse, and others.

We are impressed with the approach of saying that in addition to the openness of the resource content, equally important is the fact that the resource is involved in various global structures that allow you to find the resource through its properties represented in a generic way. This approach means creating a space based on persistent identifiers. The initiators of this approach are the largest agencies registering DOI - DataCite and CrossRef. Without dwelling on the details, we note that the most important requirement is the openness of metadata, the indication of permanent identifiers in them for authors, organizations, other involved objects, as well as the guaranteed availability of a page with metadata at the address on the Network with which the permanent identifier is associated. That is, the degree of responsibility of the parties entering the open space of a new level increases.

Moving on to the experience of the Polytechnic in the existing context, let's focus only on one object of the university's information infrastructure – the repository, called the SPbPU electronic library. There are three most important requirements for a successful repository implementation. First, modern technology must be used, including a next-generation search interface, a single search window, and seamless access from anywhere in the world. Secondly, actions to fill the repository and use its resources should be integrated into the university processes. At the same time, the recipients of the resource will not necessarily be people, in other systems, and the further away, the greater the volume. Third, we are obligatorily following the new rules of behavior in the digital university, helping to get the most out of the services provided.

As for the SPbPU electronic library, at the insistence of students and learners, all electronic resources involved in the educational process are available through a single digital library window. According to the sources of content, it can be seen that the Electronic Library performs different functions: (1) the ELS – it houses electronic educational resources for university programs, (2) the institutional repository – it contains conference materials, journal articles, scientific reports and arrays of data (datasets) begin to appear, (3) an archive preserving scientific and the cultural heritage of the university in digital form, these are the personal collections of outstanding polytechnic scientists, photographs, memoirs, and the complete collection of the issue of the Polytechnic newspaper for over 120 years.

Some streams where automated resource transfer systems that have been developed in the EL. Draw your attention to the fact that for each stream there is a specificity in the order of transmission, in the set of metadata. The copyright holder transferring the resource can choose the type of access – from placement in a closed repository to open to everyone on the Internet. If necessary, license agreements are concluded, sometimes with the use of technical means, which is permitted by Russian legislation in terms of copyright.

The resources of the Electronic Library are grouped into collections. Pie charts show some characteristics of a common array: a significant part of the resources are created at the Polytechnic University; most of the resources are accessible by the login-password of the unified identification system of the Polytechnic University, less than half of the resources are publicly available. Of the resources coming from the Polytechnic, the overwhelming share is created digitally.

One of the streams of transfer of graduate works of students and postgraduates is presented. The process involves (1) students who upload the work to the website, create a description of it and sign a license agreement, (2) curators from departments who check the correctness and completeness of the data entered by the student and (3) bibliographers who control the correctness of the conversion of the file into a protected format and the presentation of metadata in the bibliographic record. It should be noted that this collection is not available in the public domain since 2020 by the decision of the university administration.

Despite the fact that many resources are closed, the Electronic Library is in the top hundred of the world ranking of institutional repositories from Google Scholar. Let's pay attention to two inflections on the graph of the statistics of calls to resources. The first one was in 2012, when a new search interface appeared in the Electronic Library that meets modern requirements. The second is 2017. When resources began to be assigned DOI identifiers. The decline in statistics was associated with the execution of the decision of the university management to close individual collections to external users. But we should note that the Electronic Library management technologies have made it possible to fulfill the requirement within a few days.

In 2017, the National Center for the Identification of Scientific Data was established at the Polytechnic University, and in 2021 – the national consortium DataCite. The developed pilot version of the Scientific Data Repository allows not only to place resources for long-term storage, but also to create their descriptions with the involvement of permanent identifiers and

reference books, using the universal metadata schema DataCite.mSingle sign-on, i.e. the use of the Polytech identification system account when working with Russian and foreign restricted access resources, when working with its own electronic library, is based on the use of federal authentication technology and the FEDURUS national federation, which is part of the eduGAIN global interfederation,

There is much involvement of electronic library resources in the processes of universities by the example of the formation of a digital profile of university teachers and researchers. Many resources fall into the profile, which means they are taken into account when determining the KPI and the level of remuneration of a scientist is accepted only if they are available in the Electronic Library of the university. It is impossible to introduce a new discipline into the programs if the required educational publications are not available in the library fund. Preferably in electronic form.

In conclusion, the Polytechnic's approach to ensuring improved/enhanced openness is in compliance with FAIR principles. When many results are scattered among different sources or simply lost, to a new model where more and more "dark data" will be stored and made available to the empowered. Along with words of gratitude to the organizers and participants – a photo of the rear facade of the same main building of the university, but in the autumn twilight, when not everything is clearly visible. And the call for libraries to disclose dark data so that the research is truthful, the results are verified and available for further use.