

Integration and data portability in the programming of digital analytical services of the education system

Fiofanova Olga Aleksandrovna

Russian Academy of National Economy and Public Administration under the President of the Russian Federation, Moscow, Russian Federation, fiofanova-oa@ranepa.ru

Kovalev Evgeniy Evgenievitch

Moscow State Pedagogical University, Moscow, Russian Federation, ekovalev@yandex.ru

ABSTRACT

It analyzes the development trends of information and communication technologies for the analysis of big data in the process of digital transformation of education - the integration of various levels of the educational data analytics system to conduct the necessary analysis and forecast the behavior of the components of the education system as a whole. The problem of the current state of information systems and services in education is analyzed - the presence of a large number of discrete software products, each of which accumulates digital traces of the learning process and generates output in unsystematized. The formats methods of data analysis in integration electronic services and systems are analyzed: big data mining, data fusion and integration, genetic algorithm, predictive analytics, spatial analysis, split testing. Integration digital applications based on system interoperability and data portability, using class software middleware and open systems methodology. Modern systems of the middleware class are able to process data based on universal formats and provide multi-channel data transfer between all application components. The methodology of open systems for analyzing big data in education is analyzed. The essence of the methodology of open systems is characterized, consisting in the fact that during their construction, docking should be provided using standard interfaces between all components of the systems. The paper analyzes the management of creating a specification for a protocol for the interaction of multi-format systems based on the use of XML technology. The prospects of using an open system methodology, building integration using standard interfaces using Application Program Interface (API) standards and External Environment Interface (EEI) standards are described to solve the problems of system interoperability, data portability. The study uses methods of genetic data analysis, modeling of digital integration of data portability applications.

KEYWORDS

information systems and services in education, integration of digital resources and services, analysis of big data in education, interoperability of digital systems and data portability.

1 Introduction

Currently, the fourth industrial revolution is unfolding, which entails not only leading scientific and technological developments, but also a qualitative change in the culture of work. [1]. Studies show that educational reforms carried out in recent decades have been insufficiently effective. Today, the number of jobs where performers are required to have a high level of general literacy and the ability to solve problems using a computer has increased significantly compared to the mid-1990s. At the same time, the number of workers capable of performing similar work at a high level has not increased. [2].

The urgent task of the education system is to respond to the challenges of modern society. To solve the problems posed by the fourth industrial revolution, all spheres of education must go through a digital transformation, which is already happening in the economy and social lifestyles. This transformation should be aimed at creating a personalized, result-oriented model of the organization of the educational process, based on an analysis of the needs of the economy and society in modern competencies.

In the process of digital transformation of education, it is necessary to form and introduce new models of educational organizations, which are based on the development of digital tools, sources of information and services aimed at improving organizational and infrastructural conditions for implementing the necessary transformations and making managerial organizational and pedagogical decisions. A flexible response of the education system to challenges is also necessary for the transformation of the system based on the analysis of data on the state of society.

The general trend in the development of ICT in the process of digital transformation is the integration of systems of various levels to provide the necessary analysis and forecasting of the behavior of system components and the environment as a whole. The relevance of the task is emphasized by the fact that in the field of digitalization of education, such a change is necessary to determine the competencies being formed and to determine the optimal ways for the development of the entire education system. Thus, according to the Digital School project of the National Education project, it is necessary to create a modern and safe digital educational environment by 2024, ensuring high quality and accessibility of education of all types and levels. In addition, the project provides for automation of document management, reporting and accounting, digitalization of the learning process with access to individual trajectories, and continuous online teacher training.

Based on long-term forecasts of the development of society and technology, we can highlight the current areas in the field of education, which require transformations in the field of digitalization (National Project "Education") [3]:

- approval of the Digital School Standard, approval of the Standard for the creation and operation, content of sites and information systems of educational organizations;
- Creation and functioning of a unified information system “Digital School”;
- identification of promising and outgoing professions and skills;
- development of innovative methods and tools for creating an educational environment using digital tools;
- development of the integration capabilities of electronic services and information systems in education to solve the problems of building analytical reporting;
- the possibility of flexible transformation of the educational process at all levels in connection with changes in the external environment (economic factors, isolation, viruses, etc.);
- development of a new field of pedagogical knowledge - “Pedagogy based on data” [4,5];
- the formation of knowledge in breakthrough areas of technology in the medium and long term development prospects.

The implementation of these directions is possible only when analyzing the accumulated data in education and adapting changes in educational policy based on these data.

2 Methods

The methodological foundations in the field of educational data mining are the use of methodological approaches that allow you to collect digital traces and analyze them:

- a) research methodology for the development of e-learning systems and electronic portfolios, digital footprints in education, including the study of brick and click models (a mix of traditional and e-learning), the study Modular Object Oriented Dynamic Learning Environment (modular object-oriented dynamic training platforms) and Digital twin (digital twins in education),
- b) the methodology of designing and researching artificial intelligence technologies - the analysis of educational data, including the design and study of public repositories of educational data,
- c) research on the development of methods for analyzing a new type of educational data (for example, a new method of social network analysis and predictive models of student performance using big data),
- d) a research methodology for the structure of competencies, concepts and practices of the competence development of data analysis professionals in education, including research and evaluation of the effectiveness of programs (for example, “Big data in education”, “Practical learning analytics”, “Data, analytics and learning”).

Scientific novelty of the research is the systematization of the current state of information systems and services in education. The author analyzes the presence of a large number of discrete software products, each of which accumulates its own digital traces of the learning process and generates output data in unsystematized formats. To enrich knowledge, it is proposed to use methods for analyzing accumulated data in integration electronic services and systems: intelligent analysis of big data, data fusion and integration, genetic algorithm, predictive analytics, spatial analysis, split testing, integration of digital applications based on the interaction of systems and data portability with using middleware class software and open systems methodology. The proposed solution is modern middleware class systems. They are able to process data based on universal formats and provide multichannel data transfer between all components of the application. Compliance with the open systems methodology must be a prerequisite for implementation. It lies in the fact that when building solutions, integration and docking of models should be ensured using standard interfaces between all components of the systems. The article proposes management of the creation of the specification of the protocol for interaction of multi-format systems based on the use of XML technology. The prospects for using the open systems methodology, building integration using standard interfaces using the Application Programming Interface (API) and external environment standards (EEI Standards) to solve the problems of systems interaction and data portability for their reuse are described.

The relevance of the study is substantiated by the need to modernize educational data base management systems (DBMS) for analyzing educational data and making organizational, pedagogical and managerial decisions on the education and development of children at all levels of management.

3 Integration model of educational digital services

The current state of information systems and services in education is characterized by the presence of a large number of discrete software products, each of which accumulates digital traces of the learning process and generates output data in its own, sometimes unregulated, formats. This makes it impossible to perceive the state of the education system as a whole and makes it difficult to exchange data between systems. Also, some of these systems do not generate data in formats suitable for reuse and use by third-party systems.

The landscape of information systems (including data that is freely available in various social networks and educational services) that generate and process data in education is currently presented in the following form (table 1).

Table 1. Education Information Systems Landscape

Information Services and Resources	The level of education	Data set	The possibility of re-uses (Reusability)	Use case
General education statistics (ministry sites, departments of education of the RF subjects and regional ministries, supervisory authorities in education) https://fedstat.ru/ https://www.gks.ru/ http://stat.edu.ru/ http://eis.mon.gov.ru	All levels	Aggregated statistics by region and level of education, socio-economic indicators	Medium, text data formats	Analysis of the level of informatization, analysis of the number of students, analysis of the structure of employment in education
Basic information about educational institutions https://bus.gov.ru/ http://www.obrnadzor.gov.ru/ru/ https://abitur.cbias.ru/	All levels	Information about the address, contacts, constituent documents, basic numerical indicators of activity	Medium, text data formats	Creating visual maps and social portraits of territories, using spatial analysis to optimize indicators
Information on educational standards (federal and own) https://fgos.ru/ https://fgosreestr.ru/	All levels	Mandatory requirements for educational programs, text data sets, including criteria for the structure, conditions and results	Medium, text data formats, the need for data formalization	Evaluation criteria, building models for the analysis of educational programs
Personalized information on students, including personal data posted on the websites of educational institutions, profiles and activity on social networks, information from educational information systems https://resh.edu.ru/ https://uchi.ru/ https://foxford.ru/ https://vk.com/	All levels	Personal data, achievements of students, social environment	Low, very diverse output formats	Creation of individual recommendations, individual educational trajectories, personalized distribution of educational events
Performance indicators of educational institutions https://www.ucheba.ru/ https://abitur.cbias.ru/ http://indicators.miccedu.ru/	All levels	Inspection results, educational results, socio-economic information	Medium, text data formats, the need for data formalization	Creation of visual portraits of educational institutions in terms of assessing the performance of indicators, data exchange for the formation of ratings and comparative parameters
Information about inspections of educational institutions https://inspect.obrnadzor.gov.ru/ https://obrnadzor.gov.ru/	Professional education	General information, details of documents	No	Only as a basis for regulations
Public spending on education http://budget.gov.ru/ https://budget.edu.ru/ https://www.roskazna.ru/	All levels	Aggregated information at the level of subjects and departments	Low, the need for detailed items of expenditure and decomposition of budgets	Comparative analysis, identification of areas of the most effective financing
Budgets of educational organizations http://budget.gov.ru/ https://budget.edu.ru/ https://www.roskazna.ru/	All levels	Aggregated information at the level of educational organizations, information about wages, attracted finances, grants and economic activities	Low, the need for detailed items of expenditure and decomposition of budgets	Comparative analysis, identifying areas of budget development
Expenses of educational organizations https://zakupki.gov.ru/ https://budget.edu.ru/ https://www.roskazna.ru/	All levels	Aggregated information at the level of educational organizations, information on	Low, the need for detailed items of expenditure and decomposition of budgets	Comparative analysis, identification of areas of development and redistribution of costs

		economic activity		
Data on student performance from various services for the organization of the educational process, LMS, LCMS https://resh.edu.com/ https://uchi.com/ https://foxford.ru/	All levels	A large number of samples and statistical parameters	Medium, very diverse output formats	Analytics of the content of disciplines and competencies, maintaining internal ratings of an educational nature, the possibility of exchanging data with information services of employers and / or information resources of other education levels to take into account educational results
Exam statistics USE / OGE / GIA https://4ege.ru/ http://ege.edu.ru/ https://www.uche.ru/	the general	A large number of samples and statistical parameters	High enough text and table formats	Disciplinary analytics, the possibility of creating regional
Information about the olympiads https://www.turgor.ru/ https://olimpiada.ru/	the general	Samples from various information systems	Low, very diverse output formats	Analytics and the identification of gifted students, the formation of knowledge bases, formation of personalized information
Information about the achievements of students GIS Education https://resh.edu.ru/ https://uchi.ru/ https://foxford.ru/	All levels	Samples from various commercial and non-commercial information systems	Low, very diverse output formats	Analytics and the identification of gifted students, the formation of knowledge bases, the formation of personalized information, the exchange of data between educational institutions of different levels and the field of employment
Ratings of educational institutions http://edu.ru/ https://raex-a.ru/ https://ratingskol.rf https://russiaedu.ru/rating/	All levels	Aggregated assessment indicators by established methods	Medium, text data formats, the need for formalization and data granularity	Possibility of comparative analysis and identification of indicators requiring a causal analysis, the ability to add expert and independent assessments, create visual portraits and forecast indicators
Educational resources https://fpu.edu.ru/fpu/ https://fgosreestr.ru	All levels	Samples from various commercial and non-commercial information systems	Low, very diverse output formats	The possibility of applying expert methods for assessing and identifying best practices, the exchange of documents in uniform formats
Registers of educational programs / methodological developments https://fgosreestr.ru/ https://fpu.edu.ru/fpu/	All levels	Text and presentation materials, poor compliance with standards and the ability to automatically check for compliance	Low, very diverse formats of output data	The possibility of applying expert methods for assessing and identifying best practices, the exchange of documents in uniform formats
Graduate employment data http://graduate.edu.ru/	Professional education	Aggregated information with minimal decomposition	Low, the need for detailed parameters and data decomposition	Identification of areas of employment, assessment of compliance with directions, the possibility of applying a comparative analysis of the acquired competencies
Scientific activity data https://rosrid.ru/ https://cyberleninka.ru/	Professional education	Scientometric indicators, research areas, funds raised	Medium, there is data in formats for reuse, ready for exchange between systems	Creation of a unified database of scientometric indicators, data exchange, creation of project teams

The study showed serious shortcomings of existing solutions in terms of a systematic approach and the possibility of using data analytics:

- fragmented platforms;
- proprietary data presentation formats;
- impossibility of full reuse of data;
- lack of integration and the ability to exchange data between platforms without pre-processing and adaptation;
- lack of logical connections between assessment criteria at different levels of education;
- poor data visualization;
- poor opportunity for teamwork, crowdsourcing and replication of the results.

As a result, the education system at various levels weakly interacts both among themselves and with participants in the educational process. This leads to the impossibility of building in a single format a general picture of the state of the education system and the implementation of continuity between its levels.

The main disadvantages that lead to a distortion of the decision to assess the state of the education system:

- various metrics and methods for assessing the condition;
- duplication of data;
- the need to transfer data from one part of the system to another when students move between levels of education;
- the need to use disparate ICT tools to analyze statistical data and poor decision-making automation.

The result is a highly heterogeneous IT landscape, containing applications and software components from different manufacturers that are implemented on different platforms and often duplicate individual functions. The situation is aggravated by the processes of mergers and acquisitions, leading to the inheritance of new information systems and applications.

As a proposed solution, it is necessary to develop a single portal with entry points for system participants at various levels of education to download and exchange data and obtain statistical information. After the accumulation and purification of data, it is possible to identify inter-component groups of indicators and criteria for assessing education, which can be transferred and adapted between the levels of the education system. With the further development and refinement of the mechanisms of interaction between participants in educational relations and relations in the field of education, it will become possible to build a quality management system based on Deming principles adapted to assess the education system and continuous quality standards (TQM). [6]. A prerequisite is also the justification of the possible integration of the developed models to improve data connectivity into the National Data Management System.

In the light of the indicated solutions for the modernization of existing systems, it seems necessary to use the technology of accumulation, processing and analysis of big data. Moreover, such data is experiencing constant growth, being able to work with structured and loosely structured data in parallel and in different aspects. [7].

In addition, the relevance of the use of big data is confirmed by the fact that educational policy begins to be built on educational analytics, on new analytical and managerial methods. [8]

The main methods of big data analysis and their possible use cases in education management are given in table 2. [9].

Table 2. Big data analysis methods and their possible use cases in education management

Method, its description	Possibility of application in the education management system
Data mining	The analysis of educational outcomes and socio-economic information in the field of education for clustering and classification, identification of rules and parameters dependencies, regression analysis and the detection and analysis of variance for the adoption of organizational and educational decisions
Crowdsourcing	The classification and enrichment of data by the forces of a wide, indefinite circle of people, the involvement of experts to assess parameters and results, joint generation of ideas and development of solutions.
Mixing and integration of data (data fusion and integration)	Integration of heterogeneous data from a variety of sources in order to conduct in-depth analysis (for example, data streams from social networks, educational networks, digital campuses, digital natural language processing, including tone analysis, etc.)
Machine learning, including learning with and without a teacher	Statistical analysis or machine learning to obtain comprehensive forecasts of basic criteria for the development of education
Artificial neural networks, network analysis, optimization, including genetic algorithms (genetic algorithm heuristic search algorithms)	Addressing optimization problems and simulation by random selection, combination and variation of the desired parameters to optimize the formation of budgets, evaluation educational trajectories and load
Pattern recognition	Distance learning, identification and memorization of students, data exchange when moving to another level of education
Predictive analytics	Calculation of the student's life value, an indicator that helps to understand the value will be brought by the student throughout the life cycle (including future admissions), in particular in the management of educational services. Development of optimal offers / recommendations based on the data obtained on user behavior on the site. Formation of an analytical forecast of tasks, which the student with a certain degree of probability will perform in the future. Forecasting the outflow of trainees. Building a trainee profile based on a sample of data
Simulation modeling	The construction of logical and mathematical models of the educational process and educational paths, the calculation of the complexity of models, adaptation of models to changing input data of students
Spatial analysis (a class of methods that use topological, geometric, and geographic information extracted from data)	Analysis of the geography of students and the placement of points of interest for organizing the learning process - information kiosks, libraries, access points to the network, analysis of territories for opening educational institutions based on population, infrastructure, etc.
Statistical analysis, time series analysis, A / B testing, split testing	Analysis and testing of educational programs to detect changes, which improve target performance, possible to use for the formation transitions from one level to another formation.
Visualization of analytical data, presentation of information in the form of figures, diagrams, using interactive features and animations both for obtaining results, and for use as initial data for further analysis.	Creating dashboards and visual indicators that prompt in real time the state of indicators and their development trends.

In this regard, the integration of information databases and services, based on the collection of numerous heterogeneous applications and databases of relevant information, is relevant. Such a solution will also allow supporting end-to-end processes at different levels of the system for various categories of information consumers. Also important is the ability to use the functionality of already created and inherited systems to support and adapt them.

Let us characterize the system model of a single information resource based on the integration of information services at different levels of education and the proposed data integration algorithm.

A single information resource model based on the integration of education information services data includes several levels and elements:

- 1) the level of analytics and the formation of statistical reporting: a) BI & Data mining; b) decision support tools; c) repositories and (or) data marts; d) report generator;
- 2) functional subsystems: a) ERP; b) management of educational outcomes and the educational process; c) equipment management; d) electronic document management system;
- 3) level of applied applications: a) editor of processes; b) analytics task editor; c) document route editor.

Technically, the application integration approach is based on the use of middleware class software and open systems methodology. Modern systems of the middleware class are able to process messages based on universal formats and provide multi-channel message transfer between all application components. The essence of the methodology of open systems is that when they are built, docking should

be provided using standard interfaces between all components of the systems. The preferred solution for managing and creating specifications for the protocol of interaction of multi-format systems is the use of XML technology.

To achieve a single information and analytical space, it is necessary to integrate and interoperate data, ensure the availability of data, while information systems must interact with each other in the same language. A prerequisite for this are uniform rules for data interpretation and a single data ontology (information exchange model), taking into account the industry specificity of education, which will unify data management technologies. All information services and systems must connect to the data management and analysis infrastructure and exchange data according to uniform established rules. The data management infrastructure may not be designed to store the data itself; in this case, it performs technical and technological functions, storing only information about the data: their description (passports), data registers, data use accounting, data transfer rules and quality control. In part, these functions should be performed by modernized information systems of the e-government infrastructure or departmental management information systems. It is also necessary to use the recommendations for the creation of the National Data Management System and its technical features.

The basic algorithm for data integration and processing assumes:

1. Extraction of structured data (conforms to a data model, has a well-defined structure, follows a sequential order and can be easily accessed and used by a human or a computer program). Extraction of information based on ontologies, a terminological dictionary of synonyms / relationships
2. Cleansing of unstructured data (does not have a predefined data structure, or is not organized in an established order. Unstructured data is usually presented in the form of text, which can contain data such as dates, numbers and facts. This leads to difficulties in analysis, especially in the case of using traditional programs designed to work with structured data), extraction and removal of "noise", transformation of the maximum possible types of unstructured data, selection of data suitable for analytics (Text files and documents. Photos, drawings and other graphic information. Biometric data).
3. Obtaining data in a machine-readable format that allows information systems to identify, process, transform such data and their constituent parts (elements) without human intervention, as well as provide ranked access to them for system users, including public access.
4. Data validation. Formation of a reliable assessment of the transmitted learning outcomes (trust data register). Formation of metadata that makes it easier to extract the necessary data for analysis.
5. A selection of related data, which can store semantic queries and show data that affects the selection.
6. Obtaining analytical data.
7. Application of the Application Programming Interface (API).
8. Application of criteria for evaluating analytical data.
9. Formation of analytical data in formats suitable for the consumer and decision making, suitable for reuse, for accumulation in databases.
10. Uploading data in formats for exchange between systems, visualized data, and generation of reports in established forms to support the electronic document management system.

The new generalized integration algorithm looks like this:

- 1) standards of application programming interfaces (Application Program Interface (API) Standards) - determine the interaction of application software with a computer system (application platform) and are designed to ensure portability of applications between system levels;
- 2) standards of the external environment (External Environment Interface (EEI) Standards), which defines the interaction of an information system with its external environment and is designed to solve the problems of system interoperability, software reuse and data portability.

The generalized implementation algorithm of the proposed approach is as follows:

- a) development and implementation of an exchange document based on the XML language,
- b) development of components and software for the exchange between various information systems and / or subsystems;
- c) development of specifications for various layers of metadata that describe data in each of the subsystems integrated into information exchange processes;
- d) development of scenarios of information exchange based on XML-schemes, providing the ability to work with files in a single universal format with standard XML-tools.
- e) highlighting input formats, such as:
 - non-electronic document (paper document, scan copy, graphic image, speech and image recognition, information search results on web resources).
 - electronic unstructured document (rubrication and classification, extraction of certain types of data from an array of information - dates, geolocations, addresses, numbers, etc.)
 - electronic structured document (data conversion from the existing format to XML).
 - selection and development of output data formats. The result should be formats intended for reuse - HTML, RTF, ODT, TEX, PDF and others.

Consider the levels of integration of software solutions when implementing the proposed approach based on data integration:

- 1) user level: a) client browser;
- 2) level of web services: a) access portal, b) entry point;

3) the level of digital applications: a) the document management subsystem; b) the process control subsystem; c) analytics and reporting subsystem; d) application functional systems;

4) the level of the data analytics server: a) unstructured databases; b) a showcase of documents and reports; c) a system of indicators and indicators; d) analytical tools; e) repository of processed data; f) structured databases.

Consider the ecosystem for the integration of information services in education based on big data analysis, which will include the following elements:

1) discrete databases of educational information systems and resources (information of various analytical resources);

2) services for the aggregation and integration of educational data and education data;

3) a platform for the integration of information services and data (a system of indicators, tools for analyzing big data on education, visualization of data selection, data format conversion);

4) settings management (status monitoring, optimization, personalization).

We highlight the main advantages obtained from the creation of the proposed solution:

1) can play the role of an aggregator of information of educational infrastructure facilities for repeated collective use;

2) provide a single entry point for obtaining information on the results of educational activities, aggregated and statistical indicators of its effectiveness;

3) contain relevant registers of information resources and data repositories necessary for the provision of educational services and indicators of their demand and capitalization;

4) provide step-by-step research planning and record volumetric, temporary and other characteristics of the educational infrastructure workload;

5) automate the preparation of reporting materials on the results of the use of collective infrastructure in a unified format suitable for data exchange with other state-level information systems;

6) can be supplemented by services of support, planning, creation and certification of educational systems and the results of educational activities;

7) implement the functions of monitoring the state of infrastructure, including control algorithms for the need to update equipment and service;

8) implement a set of interfaces that allow interacting with external, including state, information systems (ESIA - Unified Identification and Authentication System, EIS - Unified Procurement Information System, UNSI - Unified Regulatory Information System, etc.);

9) ensure integration with electronic trading floors in order to simplify and reduce the cost of procurement procedures;

10) support multi-language interface for overseas users.

The following cases can be examples of using the obtained analytical results:

Analysis of student performance data. The system, using regression analysis of data, allows to identify students who are in the so-called "risk group" (missed classes or showed low and unsatisfactory results), to predict visits and future successes, trends in learning outcomes. The system aggregates all the grades and achievements of students, finds problems, including potential ones. It will also be possible to analyze the participation of each student in a variety of activities conducted by the educational institution and evaluate its electronic portfolio. This parameter can be considered key in terms of academic success. The school, therefore, monitors the frequency of attending various events using virtual identification cards: if the student's involvement decreases, the school staff identifies the reason and can offer ways to solve the problem and increase the level of student involvement, and for those who show good results - recommend as potential applicants and recommend to participate in Olympiads, contests. The entry point can be a student's virtual card, which collects data on the location, time spent studying, participating in other educational and scientific events, is a single ID for entering the library and social services of an educational nature.

Personalization of training. One popular learning personalization strategy is to offer an additional online course to a lagging student. As the student answers questions, the analytic system will be able to predict his readiness for new topics and analyze the time spent on topics studied and necessary for new topics.

Converting educational results. The system, based on the analysis of aggregated data from various information services and systems, forms the educational results shown at one stage of education and transfers them to another when the student transfers to it. Thus, a digital trace of basic knowledge is formed and an individual learning path is predicted.

Predictive modeling. Educational institutions can form a model of potential applicants with respect to the selected parameters (such as, for example, academic performance in a number of subjects, final works, grades, availability of certificates and documents on participation in various events). With such a selection of the target group, it is possible to build targeted interaction and point offers for applicants and / or graduates, for example, when interacting with a bank of job vacancies.

Choosing a future profession. The system carries out predictive analysis and helps to decide on a career choice: the service studies the traits of the student's character, his success in learning, the experience of participating in various educational events, Olympiads and competitions. The system based on data analysis suggests the applicant to submit applications to the most suitable universities for him.

Spatial analysis when planning the construction of new educational institutions. System of indicators: population density, time spent on the road, infrastructure, availability of connection points for communications, estimated workload of the educational institution, social portrait of the environment.

4 Conclusion

Thus, summarizing all the results obtained in the course of the study, we can conclude that further development of the analysis of educational data is possible with the integrated implementation of the system of projects in the field of education and their integration into the currently implemented National Projects.

In particular, in the framework of the national project "Digital Economy", it is necessary:

- improving the regulatory framework of the mechanism of accumulation and analysis of educational data, analytics of the education system as a whole and the exchange of data between information systems and resources in education;
- development of methodology and technologies for the analysis of educational data; development of mechanisms for integrating educational data services and educational statistics;
- development of technological platform solutions and development of the technological infrastructure of education for the accumulation and exchange of educational data;
- development of indicators for evaluating and consolidating data, a system for assessing the effectiveness of educational data.

Within the framework of the national projects "Personnel for the Digital Economy", "Digital School", it is necessary to develop a framework of competencies and professional standards in the field of educational data analysis technologies and integrate it into professional development programs for personnel in the field of education and implement professional development programs for teachers and management personnel in logic «Data-driven Pedagogy», Data-driven Education Management". [10, 11].

Thus, the problem of the current state of information systems and services in the analysis of education data is analyzed. The program interfaces, interoperability of systems and data portability are described to create the integration of digital services of big data analysis systems in education.

ACKNOWLEDGMENTS

The author thanks the Russian Foundation for Basic Research for the financial support of the grant project №19-29-14016 "Methodology for the analysis of bulk data in education and its integration into training programs for teachers and heads of educational institutions in the logic "Pedagogy based on data", "Management of education based on data".

This article was prepared as part of the grant "Methodology for the analysis of big data in education and its integration into training programs for teachers and heads of educational institutions in the logic "Pedagogy based on data", "Management of education based on data" N 19-29-14016 of the Russian Foundation for Basic Research in the competition for the best projects of interdisciplinary fundamental research "Fundamental scientific support of digitalization of general education".

REFERENCES

- [1] Osburg, T. Industry 4.0 Needs Education 4.0. [Electronic resource] URL: <https://www.linkedin.com/pulse/industry-40-needs-education-thomas-osburg> (date of the application: 10.03.2021)]
- [2] Elliott, S.W. Computers and the Future of Skill Demand. OECD Publishing. [Electronic resource] URL: <https://www.oecd.org/education/computers-and-the-future-of-skill-demand-9789264284395-en.htm> (date of the application: 20.03.2021)]
- [3] National project "Education". [Electronic resource] URL: <https://edu.gov.ru/national-project/> (date of the application: 01.04.2021)]
- [4] Fiofanova, O.A., Bokova, T.N., Morozova, V.I. International comparative analysis of national state electronic educational platforms for schoolchildren. *Revista Inclusiones*, 7; 2020. Pages 51-61.
- [5] Fiofanova, O.A. Organization of educational programs. Training education management specialists based on data. (Big data in education). *Vocational Education. Capital*, 6; 2019. Pages 24-30.
- [6] Kovalev, E.E., Kovaleva, N.A. Implementation of Models for Assessing Professional Competencies Using ICT Tools *Edukacja – Technika- Informatyka. Kwartalnik Naukowy* Nr. 4(26); 2018. Pages 276-282. DOI: 10.15584/eti.2018.4.37
- [7] Patil, D.J. Building Data Science Teams. Sebastopol: O'Reilly Publishing, Inc. [Electronic resource] URL: <http://radar.oreilly.com/2011/09/building-data-science-teams.html> (date of the application: 23.03.2021)]
- [8] Fiofanova, O.A. Analysis of the current state of research in the field of education management based on data. *Values and meanings*, 1(65); 2020. Pages 71-83.
- [9] Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. EMC Education Services, 2015. ISBN: 978-1-118-87613-8
- [10] Fiofanova, O.A. Methods of analysis of educational data and methods of their application in pedagogical and managerial practice in the field of education. *School technology*, 1; 2020. Pages 117-128.
- [11] Fiofanova, O.A. Big Data in Russian education: data analysis methods in education and human development, digital data services. Materials of the international conference "Digital Society as a Cultural and Historical Context of Human Development: From Digital Culture to Cyberculture". February 12-14, 2020, Kolomna. Kolomna: State Social and Humanitarian University.