Metadata of the chapter that will be visualized in SpringerLink

Book Title	Advances in Natural, Human-Made, and Coupled Human-Natural Systems Research	
Series Title		
Chapter Title	A System Model and Tools for Modernization of Federal and Regional Digital Services of Statistics and Data Analytics in Education	
Copyright Year	2021	
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Switzerland AG	
Corresponding Author	Family Name	**Kovalev**
	Particle	
	Given Name	**Evgeny E.**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Moscow Pedagogical State University
	Address	Moscow, Russia
	Email	ekovalev@yandex.ru
	ORCID	https://orcid.org/0000-0002-3015-5084

Abstract	The paper considers the main directions of digitalization of the education system in Russia, analyzing the infrastructural and technological aspects of the new strategic directions of teacher education—"Pedagogy based on data" and "Education management based on data." The author analyzes information systems and services in education, considering the possibility of using technologies for analytical data processing. Currently, the problem of using data analytics tools in education is the lack of uniform formats and tools for integrating information systems. Data in education accumulates in various solutions that do not share the results of the educational process. Thus, it is impossible to build a clear system of end-to-end interaction between systems on common platforms. The author aims to solve this problem by modeling an integrated solution for data analysis at all education levels. The paper proposes technical and organizational solutions for creating integrated analytics services that allow accumulating and transmitting data on educational results during the transition of a student between learning levels. The analysis tools are based on the use of big data technologies. Technologically, it is advised to integrate the proposed solution into the national data management platform.

Keywords (separated by '-')	Digitalization - Information systems and services - Integration - Education management - Data management - Data analysis - National data management system in the Russian Federation

A System Model and Tools for Modernization of Federal and Regional Digital Services of Statistics and Data Analytics in Education

Evgeny E. Kovalev

Abstract The paper considers the main directions of digitalization of the education system in Russia, analyzing the infrastructural and technological aspects of the new strategic directions of teacher education—"Pedagogy based on data" and "Education management based on data." The author analyzes information systems and services in education, considering the possibility of using technologies for analytical data processing. Currently, the problem of using data analytics tools in education is the lack of uniform formats and tools for integrating information systems. Data in education accumulates in various solutions that do not share the results of the educational process. Thus, it is impossible to build a clear system of end-to-end interaction between systems on common platforms. The author aims to solve this problem by modeling an integrated solution for data analysis at all education levels. The paper proposes technical and organizational solutions for creating integrated analytics services that allow accumulating and transmitting data on educational results during the transition of a student between learning levels. The analysis tools are based on the use of big data technologies. Technologically, it is advised to integrate the proposed solution into the national data management platform.

Keywords Digitalization · Information systems and services · Integration · Education management · Data management · Data analysis · National data management system in the Russian Federation

1 Introduction

Recent global changes taking place in the digital technological order in today's world led to a sharp avalanche of information. International Data Corporation (IDC) predicts that the digital universe will reach 40 ZB in 2020, up to 5 ZB from the previous forecast (Goepfert et al. 2020). In total, since the beginning of 2010, the volume of data grew 50 times, and around the world will be created and used 2.8 ZB

E. E. Kovalev (✉)
Moscow Pedagogical State University, Moscow, Russia

S. G. Maximova et al. (eds.), *Advances in Natural, Human-Made, and Coupled Human-Natural Systems Research*, Lecture Notes in Networks and Systems 234,
https://doi.org/10.1007/978-3-030-75483-9_121

of data. The current trend in the development of digital technologies for extracting the necessary information in a form convenient for processing and perception is now creating large data arrays and the use of analytical tools, especially when making management decisions. Solutions of this type allow us to optimize operational processes, apply forecasting and strategic planning tools, and work with arrays of heterogeneous data obtained from various sources.

Another trend is the growth in the total volume of information, which occurs due to automatically generated data—by 2020, their volume increased by about 15 times (Goepfert et al. 2020). In this case, large amounts of useful data are lost. Today, less than 3% out of 23% of potentially useful data is used that could be used with Big Data technologies (Deloitte Insights 2019). The general trend in the development of information and communications technology [ICT] in the process of digital transformation is the integration of various levels of systems to conduct the necessary analysis and forecasting of the behavior of system components and the entire environment.

In Russia, such a breakthrough direction is the project for creating and operating a national data management system (NDSM) for the National project "Digital economy." It establishes uniform requirements for data management as a set of mandatory or recommendatory rules for execution by public sector bodies and organizations' management of government data at each stage of their life cycle. It also considers such activities as the creation of (1) a digital analytical platform for the provision of statistical data, (2) a unified register of objects of statistical observation, and (3) a unified register of forms of statistical observation, indicators, and relevant information resources (Government of Russian Federation 2019).

In the digitalization of education, such transformations are necessary to form current competencies and identify optimal ways to develop the entire education system. According to the project "Digital School" of the National Project "Education," it is necessary to create by 2024 a modern and safe digital educational environment that ensures high quality and accessibility of education of all types and levels. The project also provides for the automation of workflow, reporting, and accounting, digitalization of the learning process with access to individual trajectories, and continuous online teacher training (Government of Russian Federation 2018).

Within the project "Digital School" ("Digital Educational Organization"), it is necessary to use technologies of "big data," "cloud" data storage, and artificial intelligence to ensure a complete electronic document flow of the educational organization, including (1) conducting administrative and economic, financial and economic activities ("Management," "Office work," "Cloud accounting," "Electronic reporting," etc.), and (2) ensuring the educational and disciplinary process ("Electronic diaries," "Electronic assessment journal," "Teacher's electronic office," "Student's electronic portfolio," "Online education," etc.) (Government of Russian Federation 2018).

At the same time, based on long-term forecasts of the development of society and technology, it is necessary to highlight the key direction of personnel training, which is necessary to ensure transformations of education digitalization. Such a

509811_1_En_121_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:12/6/2021 Pages: 11 Layout: T1-Standard

change presupposes the development of a new area of pedagogical knowledge—"Data-Driven Pedagogy" (Fiofanova et al. 2020).

Besides, it should be considered that digitalization should cover both the open state data system and data related to the non-formal and informal education of a person. Until now, digitalization covers only formal education, while the role of the non-formal and informal will only increase and store many data that must be considered in analytical processing. In sum, it will contribute to the breakthrough scientific, technological, and socio-economic development of the country, the creation of opportunities for self-realization, and the disclosure of each person's talents.

2 Materials and Methods

The current state of information systems and services in education is characterized by a large number of discrete software products. Each of them accumulates digital traces of the learning process and generates output data in its own, sometimes ad hoc, formats. This makes it impossible to perceive the picture of the state of the education system and makes it difficult to exchange data between systems. Some of these systems do not generate data in formats suitable for reuse and application by third-party systems. In this regard, the primary task is to analyze the available data in education, assess and classify them, and model the data integration process in the form of big data technologies suitable for use.

According to the Learning Activity Management System [LAMS] concept, the following basic data classification was used in the analysis of data in education (Government of Russian Federation 2019):

1. **By data types**. In the public sector, the following types of data with different norms are distinguished:

   - Regulatory and reference information. References and classifiers should be applied wherever a reference or classifier can be generated. If it is frequently used in the public sector, it should be uniform and used by all participants in information interaction. There are backbone reference books and classifiers—all-Russian classifiers. They should also be accessible using information systems;
   - Registry data. Maintaining registers of data (lists of objects that can be objects or subjects of legal relations or facts) should switch to the type of register keeping in the form of a "registry model": confirmation of a legal fact is an entry in the register (transition to paperless interaction). To do this, it is necessary to define reference data, improve their quality, and implement mandatory use. Only then will the transition to the "registry model" be the most "painless." The reference data will have increased requirements for quality and protection due to their high importance;

- Reporting. Reporting refers to data provided by individuals, legal entities, state authorities, and local authorities on a regular (or conditionally regular—on an event) basis. The collection of reports should be minimized. Minimization can be carried out in three directions: (1) reduction of reporting forms (through systematization), (2) transition to the collection of reports in other ways (for example, using Internet of Things technologies), and (3) obtaining consent from the reporting provider to reuse the provided data;
- Other types of data: unstructured (audio, video), “streaming”, “large”, etc. Due to the increasing typing of data, any type requires special treatment and may also have special requirements.

2. **By the nature of data use**. Data that a state body transfers to other state bodies and organizations to pursue the powers of these state bodies and organizations, and data that the state body creates and uses for official purposes. There will be different requirements for maintaining data required only for internal use and those transmitted in some form to any recipients since the quality of the transmitted data is more important and easier to control.
3. **By the type of information resource in which the data is maintained**: (1) federal information resources, (2) regional information resources, and (3) municipal information resources.
   Such data represent:

   - An ordered set of state data contained in information systems or posted on the official websites of state bodies on the Internet;
   - An ordered set of state data, formed as a result of a collection of reports by state bodies or other persons determined by Federal Law;
   - A set of state data formed as a result of the collection of statements by state bodies of a constituent entity of Russia;
   - A set of municipal data contained in municipal information systems or posted on the official websites of local governments in the information and telecommunications network Internet;
   - State registers, cadastres, databases, and data banks for the formation of which state bodies are authorized.

4. **By data access level**. The legislation defines types of data according to the level of restriction of their availability.

The research makes it possible to systematize the main information resources that collect and aggregate data on the educational process, identify possible means and opportunities for their integration, and consider existing technical means, classifications, and data use recommendations.

The author considers the formats of the provided data as one of the criteria for systematization, which can be used for exchange between information systems and their subsequent integration.

509811_1_En_121_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:12/6/2021 Pages: 11 Layout: T1-Standard

Data sets currently provided by information systems:

1. Contact information of educational institutions and basic information about institutions;
2. The number of students in organizations carrying out educational activities in educational programs of primary general, basic general, and secondary general education;
3. The share of students in general education programs, additional general education programs for children, and educational programs of secondary additional education, for whom a digital educational profile and an individual training plan are created using the federal information and service platform of the digital educational environment, in the total number of students in these programs;
4. Aggregated data on the forms of the Russian Classification of Units of Measurement: (1) on the number of educational institutions; (2) a contingent of students; (3) the number of graduate students; and (4) a number of teachers; (5) several students per teacher in the constituent entities of Russia (according to the SSRF handbook);
5. Information about the in-depth study of individual subjects;
6. Information about the material base and technical condition of buildings of institutions;
7. Data on specialized training in institutions implementing general education programs;
8. Information about the sources of funds received by institutions implementing general education programs;
9. Expenditures of state (municipal) and non-state institutions implementing general education programs;
10. Budgets of educational institutions;
11. Information about inspections, results of inspections;
12. Registers of educational programs, standards, and methodological developments;
13. Navigator of universities;
14. Exam Calculator;
15. List of Olympiads with the ability to save the results of queries and configure alerts;
16. Consolidated register of licenses of educational institutions;
17. Ratings and Independent Evaluation Results.

The author concludes that the main disadvantages of the processed analytical data are their depersonalization. There is practically no possibility to identify personal results and information about the results of educational activities of a particular student. Besides, there is an insufficient set of data for the formation of data sets to integrate information systems of different levels, for example, in the transition of students to the status of students.

Also, from a systematic approach and the possibilities of using data analytics point of view, the following are other disadvantages of existing solutions:

- Discreteness of platforms and piecewise data analysis;
- A large amount of unstructured data, own (other than classified) data presentation formats;
- Impossibility to complete reuse of data;
- Lack of integration and the ability to exchange data between platforms without preliminary processing and adaptation;
- Lack of logical links between assessment criteria at different levels of education;
- Poor data visualization;
- Poor ability to collaborate on existing platforms, implement project work, and crowdsourcing, as well as the possibility of replicating the obtained results.

As a result, the education system at various levels weakly interact with each other and the participants in the educational process. This leads to the impossibility of building in a single format, a general picture of the education system's state and the implementation of continuity between its levels.

In this regard, it is necessary to recommend executive authorities, software developers, and professional communities to consider the possibility of developing information services that can format and lead to uniform quantitative and qualitative indicators of various metrics and methods for assessing the state of the educational process results both at the student and education level. Such services should identify and eliminate data duplication, highlight the need to transfer data from one part of the system to another when students move between educational levels, integrate (instead of separate) ICT tools for analyzing statistical data. This common data analytics generation toolkit can be visualized, transferred to other systems or a new level of education, and used in the automation of making organizational and managerial decisions in education.

The main obstacle standing in the way of creating unified information and analytical space is the lack of the ability to transmit in electronic form and in established formats information about learning outcomes to a specific student and provide a reliable assessment of the transmitted learning outcomes. This does not allow creating a single technological platform for storing and exchanging information and introducing a completely electronic document flow between participants in educational relations. The situation is aggravated by mergers and acquisitions, leading to the inheritance of new information systems and applications, an extremely heterogeneous IT landscape containing applications and software components from different manufacturers that are implemented on different platforms and often duplicate separate functions.

To achieve single information and analytical space, it is necessary to integrate and interoperate data, ensure its availability, while information systems must interact with each other in the same language. A prerequisite for this is uniform rules for data interpretation and a single data ontology (information exchange model), considering the specificity of the education industry, which will unify data management technologies. According to uniform established rules, all information services and systems must connect to the data management, analysis infrastructure, and exchange data. 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 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 to create the National Data System [NDS] and its technical features.

The basic algorithm for data integration and processing assumes the following:

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 or 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 text, which can contain dates, numbers, and facts. This leads to difficulties in analysis, especially when using traditional programs designed to work with structured data), extraction and removal of "noise," a 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, and transform such data and its constituent parts (elements) without human intervention, and 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. Introduction of the Application Programming Interface (API);
8. Application of criteria for evaluating analytical data;
9. Formation of analytical data in formats suitable for (1) consumer and decision making, (2) reuse, and (3) 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.

Pre-design work involves considering the following factors:

- Study of the potential demand for relevant datasets by potential consumers;
- An assessment of the degree of readiness, characterized by the availability of the necessary data in electronic form, as well as the readiness of the organizational, technical, technological, and other means necessary for the publication of data sets;

- Publication costs (financial, time, and labor) are required to publish datasets and keep them updated.

As a proposed solution, it is necessary to develop a single portal with entry points for system participants at various education levels to download and exchange data and obtain statistical information. After the accumulation and cleaning 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 interaction mechanisms between participants in educational relations, it will be possible to build a quality management system based on Deming principles, adapted to assessing the education system and continuous quality standards.

In the light of the indicated solutions for the modernization of existing systems, it is necessary to use the technology of accumulation, processing, and big data analysis. They allow (1) to process large volumes of data in comparison with the "standard" scenarios, often in different formats, and (2) to work with rapidly arriving data with a quick update period in large volumes. Moreover, such data is continuously growing, allowing us to work with structured and weakly structured data in parallel and different aspects.

The relevance of big data use is confirmed by the fact that education policy began to be built on educational analytics, new analytical and managerial methods (Fiofanova 2019, 2020a).

The integration of information and services based on collecting relevant information from numerous heterogeneous applications and databases is relevant. This solution will also allow supporting end-to-end processes at different system levels for various categories of information consumers. The ability to use the functionality of already created and legacy systems for their support and adaptation is also important.

3 Results

System model of a single information resource based on the integration of information services of various levels of education and the proposed algorithm for multi-level integration and data processing:

1. Data collection at the level of integration of functional subsystems of information systems, information services, and streaming data;
2. Cleaning, structuring, and formalization of data;
3. Application of analytical subsystems and decision support tools.
4. Application of analytical subsystems and decision support tools;
5. Application of quality assessment and performance indicators;
6. Data visualization, generation of reports and data sets in the form required for end-users, government information systems, and services;

7. Formation of tools for joint work on the received data sets and organization of document management;
8. Formation of databases, registers, and data repositories suitable for reuse;
9. Setting up the personalization of consumers of analytics results.

When developing and implementing the model, the next stage should solve the problem of converting educational results (when the results from the database at the school level will be converted into the data of applicants during the transition to the next stage of education (college, university)). So far, such solutions are represented by objects of piecewise informatization. They are implemented only based on individual local information systems of universities and exist only at the level of sufficiently closed data of individual universities (i.e., Higher School of Economics, Omsk State Technical University). At the same time, the data collected by the local information systems of universities cannot be integrated into a single system. The main collected data sets include (1) information on the results of the Olympiads and creative events, (2) information on the preparation of applicants for the Unified State Exam, (3) and the development of additional education.

There is still no comprehensive software solution at the state level; therefore, the most critical vector for the further development of an integrated system should be an integration with the systems of universities for social support of potential applicants and making decisions about future students.

4 Discussion

The research shows the options for further use, namely:

1. Identification of potential applicants;
2. Formation and analysis of the competencies of schoolchildren for the development and modernization of their own educational standards and programs;
3. Integration with ICT of universities, subsystems Entrant, and Admissions Committee;
4. Formation of targeted advertising;
5. Analysis of potential consumers of services in territorial and social aspects;
6. Development of a socially oriented "environment of opportunities" for quality and affordable additional education.

Thus, summarizing the results obtained, we can conclude that the further development of educational data analysis is possible with the comprehensive implementation of education projects and their integration into the currently implemented National projects. In particular, within the National project "Digital Economy," it is necessary to:

- Improve the regulatory framework for the mechanism of accumulation, processing, and analysis of educational data, analytics of the education system, and data exchange between information systems and resources at different levels in education;
- Develop methodology and technologies for the analysis of educational data; mechanisms for integrating the services of analytical, educational data and educational statistics;
- Develop technological platform solutions and technological infrastructure of education for the accumulation and exchange of educational data between various existing systems in education;
- Integrate educational data at the level of general education institutions with additional education systems, the formation of a unified landscape of data formats for integration, data exchange, and extraction in formats suitable for making decision-making;
- Develop indicators to assess and consolidate data, a system for assessing the effectiveness of educational data.

5 Conclusion

The current research results can become the basis for the formation of documents on the management of analytical systems and data in education. The documents should describe the general target data management and determine the procedures for data analytics formation. The requirements to manage specific datasets and data formats should be a set of techniques used for each specific task. The methods should be refined and refined constantly, in the implementation of the flexible approach and continuous quality improvement. The development of methods should correspond to the current needs of the information and analytical system and be accompanied by personnel training with the information system at all levels.

Within the framework of the national projects “Human Resources for the Digital Economy” and “Digital School,” it is necessary to develop competencies and professional standards of technology to analyze educational data and its integration into professional development programs for education personnel; implementation of professional development programs for teaching and management personnel in the logic “Data-Driven Pedagogy,” “Data-Driven Education Management” (Fiofanova 2020b).

Acknowledgements The work was supported by the Russian Foundation for Basic Research (O. A. Fiofanova, grant No. 19-29-14016 “The methodology of Big Data mining in education and its integration into the training programs for teachers and heads of educational organizations, such as “Data-based Pedagogy”, “Data-based education management.”

References

Deloitte Insights (2019) High technology, communication, and entertainment. Forecasts for the development of industries in 2020. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/ru/Documents/technology-media-telecommunications/russian/tmt_predictions_RU_2020.pdf

Fiofanova OA (2019) Organization of educational programs. Training education management specialists based on data (Big data in education). Vocat Educ. Capital 6:27–33

Fiofanova OA (2020a) Analysis of the current state of research in the field of education management based on data. Values Meanings 1(65):71–83

Fiofanova OA (2020b) Methods of analysis of educational data and methods of their application in pedagogical and managerial practice in the field of education. School Technol 1:117–128

Fiofanova OA, Bokova TN, Morozova VI (2020) International comparative analysis of national state electronic educational platforms for schoolchildren. Revista Inclusiones 7(S2–1):51–61

Goepfert J, Minonne A, Vojtisek P, Shikita K, Elshewy E, Membrila R et al (2020) IDC's worldwide big data and analytics spending guide taxonomy, 1H19. Retrieved from https://www.idc.com/getdoc.jsp?containerId=US46089419

Government of Russian Federation (2018) Resolution "The national education project 2019–2024" (December 24, 2018 No. 16). Moscow, Russia

Government of Russian Federation (2019) Resolution "Concept of creation and functioning of a national data management system" (June 3, 2019 No. 1189). Moscow, Russia

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	··· under matter to remain	New matter followed by
Insert in text the matter indicated in the margin		or (x2)
Delete	through single character, rule or underline or through all characters to be deleted	or (x2)
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	