Building data marts to analyze university faculty activities using power BI

Sergei Karabtsev*, *Roman* Kotov, *Ivan* Davzit, and *Evgeny* Gurov

Federal State Budgetary Educational Institution of Higher Education «Kemerovo State University», Digital Institute, Kemerovo, Russia

Abstract. Evaluating the performance of university faculty is a hard task because of the diversity of the work performed. The authors assume that the founder of the university evaluates the effectiveness of the university according to performing the teaching staff. The aim of the study is to improve the monitoring of key performance indicators of the university teaching staff based on data management. The authors present an information system to support decision-making related to the teaching staff at Kemerovo State University. The authors of the paper describe creating such a system using Business Intelligence technologies step by step. The authors identified data sources, designed the structure of the data mart and built ETL-processes for its filling, implemented various analytical dashboards. Implementing the information system in the daily activities of the university allows responding promptly to changes in the key indicators, forecasting their further change, deciding on activation of efforts in the chosen direction or types of work.

1 Introduction

Teaching staff in higher education institutions comprise many tasks and activities and traditionally divides them into academic (teaching) and extracurricular activities. Teaching activities mainly include conducting various types of classes, supervising practical training, graduate and research work of students, working as a member of state or attestation commissions. Extracurricular activities include preparation of scientific articles and theses of reports, preparation and submission of applications for participation in competitions and grants, scientific review of articles and monographs, performance of certain types of works for the department, professional development, development of a working programme of disciplines, preparation for classes, etc. We can refer such tasks to scientific, educational and methodological, organisational and pedagogical types of work. Educational activities exceed extracurricular activities in terms of hours per academic year. It is not possible to plan and record the hours of activities for teachers without the use of automation tools. University departments of higher education institutions use their own information systems or purchase specialised solutions. The best known commercial solution is 1C:University PROF (over 400 universities in Russia use this product) from 1C and the automated system "Study Load" (over 1000 educational organisations of higher and secondary education use various system

* Corresponding author: skarab@kemsu.ru

modules) from MMIS Lab. Such information systems are an integral part of the planning process for institutes, departments, and teaching staff. The systems do not provide up-to-date information on the fulfilment of planned indicators, as they either do not receive primary data on the results achieved, or they receive it with a great delay, for example, as a teacher's report at the end of the academic year. We cannot make managerial decisions, influence the situation with non-fulfilment or delayed fulfilment of the plan, e.g. for scientific or teaching and learning publications.

Overburdening of teaching and non-teaching staff (faculty) with teaching and non-teaching activities causes delays in the completion of planned tasks. We can measure learning and extracurricular activities in terms of the hours allocated in the rate or proportion of the rate at which a faculty member works in a higher education institution. The traditional information systems used in HEI planning cannot provide this kind of analytical information promptly. We need to use modern approaches based on data analysis. One such approach is the construction of data marts and data visualisation using Business Intelligence (BI) tools [1, 2, 3].

A well-known variety of BI tools is OLAP - On-Line Analytical Processing, which some use in descriptive analytics [4, 5]. Descriptive analytics is a field of mathematical statistics, whose methods focus on extracting, transforming, systematising and summarising data from various sources to discover interpretable dependencies in it. Descriptive analytics focuses on historical data, looking for answers to the questions of what happened, why it happened and what is happening at the current moment. Higher education institutions in their information systems process and store a huge amount of data on the learner and his/her academic performance, on the workload of the teaching staff, his/her research activities, on the financial costs and budget planning of the HEI. Most of this data is in an electronic form in dispersed operational sources (e.g. resources on the Internet) and databases, and sometimes on tangible media. The dispersion of data makes analysis and decision-making much more difficult, as stand-alone databases or data sources cannot provide merged information quickly and visually. Understanding the value of analytical processing of data, which is the foundation for informed decision-making, has changed attitudes towards data. Where once accumulated data was a by-product of business activities, today it has been transformed into a vital asset of the organisation, playing a key role in managing the organisation and its business processes [6, 7].

This study improves the monitoring of key performance indicators of university teaching staff based on data management. We present an information system for decision support (Decision Support System, DSS) related to the teaching staff at Kemerovo State University (KemSU).

The system is part of an emerging information situation centre built using BI technologies [7]. The objectives of the study are to develop a decision support information system for monitoring the performance of university teaching staff and implementation in the daily activities of university managers. The information system should comply with the digitalization strategy adopted in the university and provide answers in digital or graphical format to the following questions:

- the number of teaching staff rates and the number of teaching hours, the average teacher load in hours, the amount of hourly fund, the number of actual rates and hours in them;
- amount of funds spent on teaching staff remuneration, average age and percentage of young teachers, availability of academic degrees and titles;
- the structure of extracurricular workload, the number of planned and already published works indexed in the Russian Science Citation Index (RSCI), Scopus and Web of Science (WoS);
- the number of planned and already published study guides, methodological works.

The system should provide all information by various categories: institutes, departments, and faculty members. Users should be able to filter information by year, type of activity, conditions of teacher employment. The information system should be able to work at different levels of decision-makers: the rector and members of the rectorate, institute directors, heads of departments. Decision-makers should have access to data through a thin client - a web-browser.

2 Research methodology

The development of an information system to support decision-making requires preliminary work on the survey of the university under the goals and objectives of the study, coordinated with the digitalization strategy adopted at the university. The conducted survey determines the level of IT-infrastructure of the university, processes, types and sources of data for analytics, stakeholders. Figure 1 shows the classical approach to implementing such level of DSS systems and includes three stages, which provide searching the sources of necessary data, extraction and pre-processing before placing in a single data warehouse, construction of multidimensional data model and creation of analytical report mappings [2, 5].

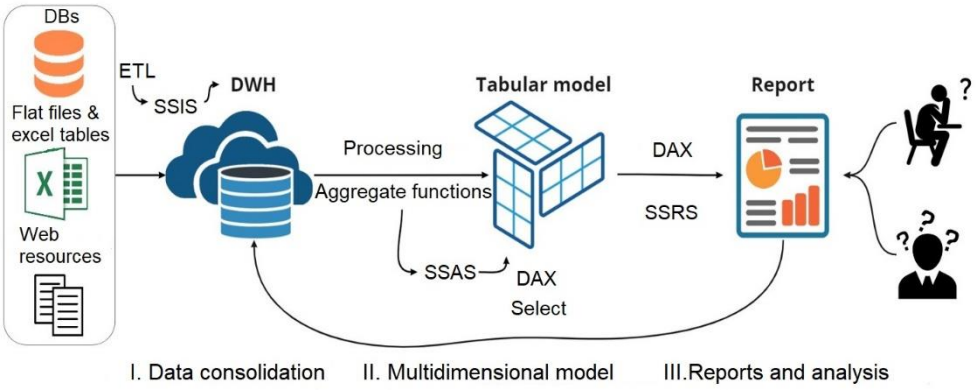


Fig. 1. Conceptual scheme of realization business-analysis in organization.

We chose as a development tool DSS BI-platform from Microsoft, one of the top three analytical and BI-platforms, according to Gartner. The platform comprises core components: SQL Server 2019, Analysis Services, Integration Services, Data Tools and Power BI Report Server.

2.1 Data consolidation

The consolidation stage begins with a search for data sources suitable for analysis. We identified about 25 data sources, some of which are not automated systems. For example, the scientific department of the university twice a year collects information about scientific publications Scopus and Web of Science, based on the reports of teachers and departments in .docx and .xlsx format. Employees manually search the websites of scientific citation systems elibrary.ru, scopus.com, publons.com in case there is a need for an urgent update of information about the works (articles) of a particular faculty member. Most information systems do not have a developed API for interaction with external systems, so the extraction of data from them is possible only at the level of direct reference to the providing their work DBMS.

The analysis of the data sources and the available technical capabilities for accessing them highlighted the following sources and approaches to data extraction (Table 1):

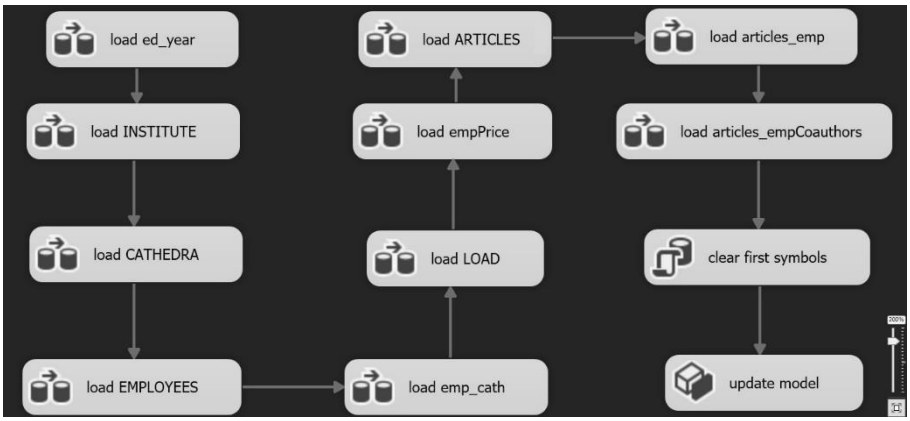
Table 1. Data sources and data extraction approaches.

№	Data source	Retrived data	Data extraction approaches
1	Information system for calculating teacher workload (official website of the developer https://www.mmis.ru/)	Surname, name, patronymic of the lecturer, institute, department, volume of educational and extracurricular workload, plan of publication activity	Using direct database access MS SQL Server
2	Accounting information system (official website of the developer https://v8.1c.ru/stateacc/)	Surname, name, patronymic of the teacher, position, academic degree and title, official salary and other payments	Using direct database access MS SQL Server
3	Website Scopus Preview https://www.scopus.com	Name of the journal, information about the publication (title, author and co-authors, volume, issue, year, pages)	Parsing json-file by lecturer's ORCID and authorid
4	Website Clarivate https://publons.com/wos-op/	Name of the journal, information about the publication (title, author and co-authors, volume, issue, year, pages)	Using API with authorization token and lecturer's ResearcherID, parsing
5	Website RSCI https://www.elibrary.ru	Name of the journal, category, information about the publication (title, author and co-authors, volume, issue, year, pages), type of publication (scientific article, textbook, monograph, etc.)	Parsing xml-file by lecturer's SPIN-code
6	Website Scimago Journal & Country Rank https://www.scimagojr.com/	Name of the journal, quartile, year	Parsing csv-file by journal's name
7	Electronic information educational environment KemSU	Surname, name, patronymic of the lecturer, lecturer's ORCID, SPIN-code, AuthodID and ResearcherID	Using direct database access Oracle

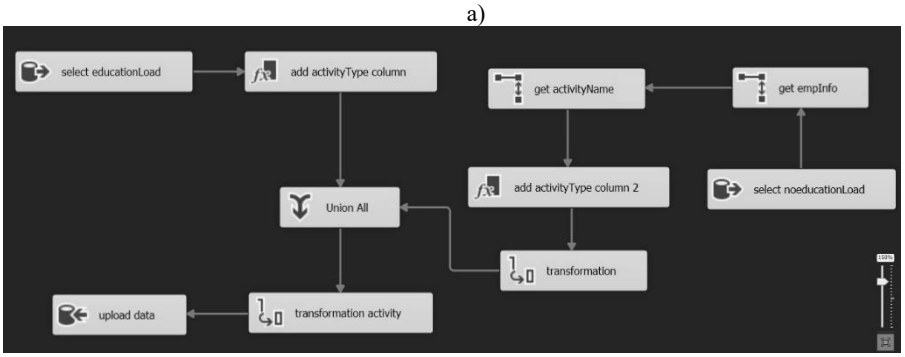
After determining the sources of data, it is necessary to design the structure of the data warehouse, able to ensure the implementation of the functions of the developed information system. We replaced the data warehouse with a special case - Data Mart, the simplicity in implementation and the volume of stored data. One of the most powerful information flows in the data mart is the input flow associated with transferring data from sources [8]. The data mart checks, cleans, sorts, groups and enriches the information by adding new attributes when transferring it. This process in the literature is called ETL - Extraction, Transformation and Loading.

We used SQL Server Integration Service (SSIS) to implement an ETL process, creating packages with control flows and data flows. The control flow can transfer data to a data mart table and execute in parallel or sequentially with other flows as needed. A data flow details a control flow with a specific sequence of actions: retrieve from a data source, perform a series of transformations, and load into a data receiver. The data receiver can be a target table of the data mart or some intermediate table or a temporary file that can later act as a source of transformed data. SSIS packages are convenient for us to use when transferring data from structured sources, such as relational databases (Table 1, sources #1, 2, 7), csv, json, or xml

files (Table 1, source #6). Figure 2, a) represents the container with the data control flows of the SSIS package.



a)



b)

Fig. 2. a) SSIS Control flows; b) SSIS data flow.

This container selects, converting and uploading information about the teacher's academic and extracurricular workload from the sources listed in rows 1 and 2 of Table 1. Container comprises 10 tasks "Data Flow Task", 1 task "Execute SQL Task" and 1 task "Analysis Services". Figure 2, b) shows the data flow for the "load LOAD" control flow. The data flow specifies the control flow action in specific steps.

Besides SSIS packages, the authors of the paper created their own ETL mechanisms implemented in Python3 using Pandas, BeautifulSoup, Pyodbc and other libraries to extract data from web resources (Table 1, sources #3, 4, 5). We set up a regular xml upload of data from the Elibrary resource (https://www.elibrary.ru) to retrieve all the university's faculty publications from the Russian System of Scientific Citation (RSCI). To do this, it is necessary to have an account with extended rights of action on behalf of the organization that has an agreement with the scientific electronic library. The "To Organizations" section of the website contains a subsection "List of Publications of an Organization", where you can set parameters, specifying the university subdivision, subject, journals, types of publications, dates of publication, etc., and generate the xml-file. Figure 2 shows an example of the structure of the resulting file. Next, we parse the file using Python, tag the data, and save it to a temporary table in the database, which we then import into a mart using the SSIS package.

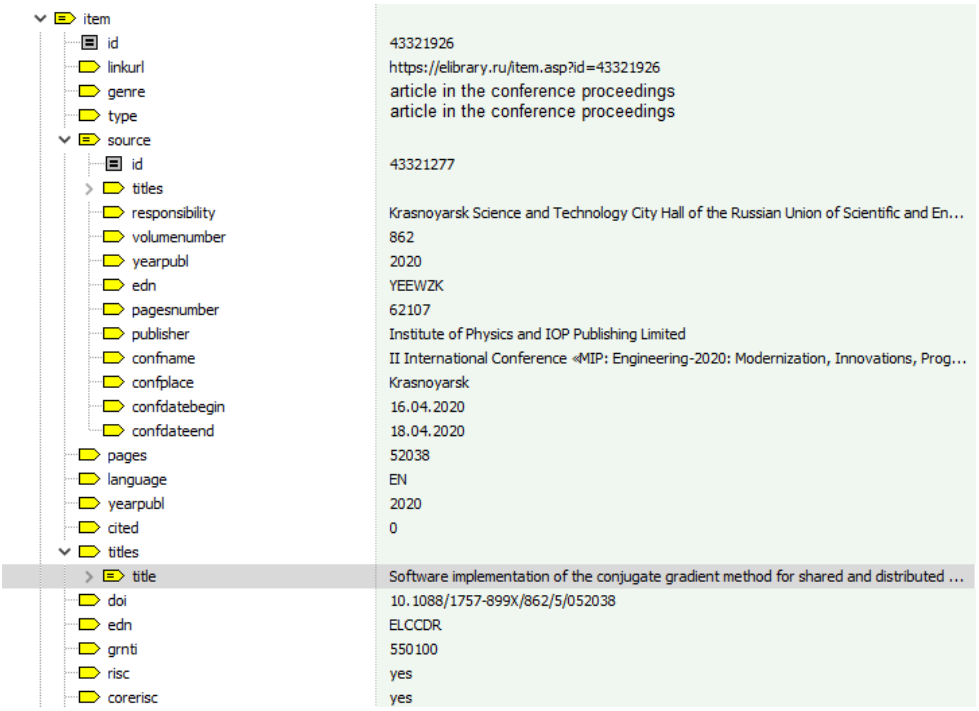


Fig. 3. Data in XML format.

We have also written functions in Python that use a ready-made API at https://publons.com/wos-op/api/v2/academic/ to retrieve data from the Web of Science citation system. To use the API, a university must have a paid subscription to a digital library and must get a special Authorization Token from the API developers. API and ResearcherID of all authors are retained in a short-term table for transporting to the data mart.

You can receive the list of published papers in the Scopus citation system the same way as from the RSCI system: formation of a csv-file, parsing by Python, transfer of data into temporary tables. It is necessary to purchase a paid subscription to form a csv-file. After unloading the data from the sources, KemSU's data mart contained about 500,000 rows. The ETL process is one of the most time-consuming processes for creating a data warehouse and can take up to 60% of development time.

2.2 Multivariate model

Data analysis requires the construction of dependencies between different parameters, otherwise known as measurements. Constructing dependencies between different parameters represents the data as a multidimensional model, which visually can be correlated with an n-dimensional cube. Each dimension can have a logical relationship to other dimensions. At the intersection of measurement axes in n-dimensional cube, there are data - "measures", quantitatively characterizing analyzed facts. Thus, the model for data analysis contains two dominant entities - measures and measurements. We found in a systematic literature comparison that the most frequent methodology for designing a data warehouse (storefront) in a relational DBMS and storing multidimensional data in it is the "star" or "snowflake" scheme [10]. Often works contain a schema with controlled denormalization of relationships between entities. Figure 4 shows the ER model of a data mart.

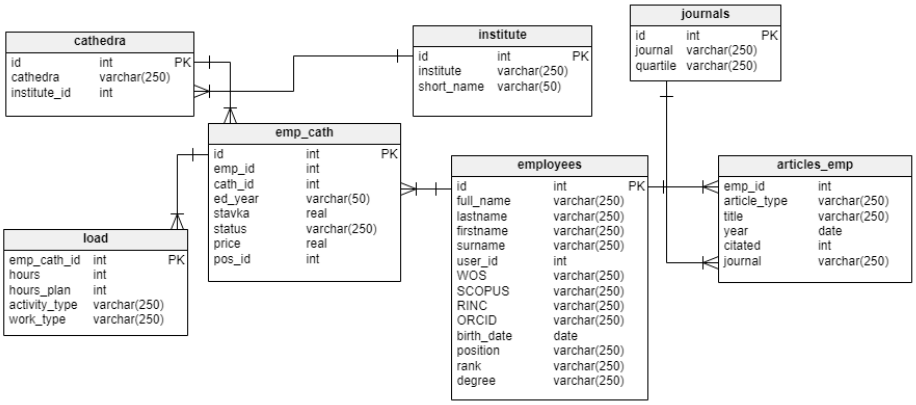


Fig. 4. Relation database schema.

The tables emp_cath and articles_emp are fact tables, and the remaining tables describe measurement tables. The database schema, together with the data, is the basis of the multidimensional data model. We used its simplified version for simplicity and speed of the development - tabular model of data. Several dozens of computable columns and measures complete such a model. Table 2 gives an example of some computable columns and measures.

Table 2. List of measures and calculated column.

Measure / calculated column name	**DAX-expression**
Measures	
Number of VAK articles of the 1st quartile	COUNTAX(FILTER(articles_emp, AND ([article type] = "VAK", [quartile] = "1")), [title])
Total employees	COUNT(EMPLOYEES[full_name])
Average age	SUM([Age]) / SUM([Flag Age])
Hourly	SUMX(FILTER(LOAD, AND ([Status] = "h - hourly", LOAD[activity_type] = "Learning activities")),[hours])
Hours per job	SUMX(FILTER(LOAD, AND ([Status] <> "h - hourly", LOAD[activity_type] = "Learning activities")), [hours])
Calculated columns	
Young employee	if(AND(INT([Age]) < INT(40), NOT ISBLANK([Age])), 1, 0)
Quartile	IF(OR(ISBLANK(articles_emp [Quart]), articles_emp [Quart] = 0), "Without qurtile", CONVERT(articles_emp [Quart], STRING))
Age	DATEDIFF([birth_date], NOW(), YEAR)
Article Type	IF(articles_emp[VAK] = 1, "VAK", IF(articles_emp[RSCI] = 1, "RSCI", IF(ISBLANK(articles_emp[article_type]), "No data", IF(articles_emp [CORE RINC] = 1,"CORE RINC", IF(articles_emp[RINC] = 1,"RINC", articles_emp[article_type])))))

Besides measures and calculated columns, dimension hierarchies are built in a multidimensional model. For example, university employees work in departments, and departments are part of the structure of institutes, thus forming a hierarchy "Institute -> Department -> Employee".

3 Research results

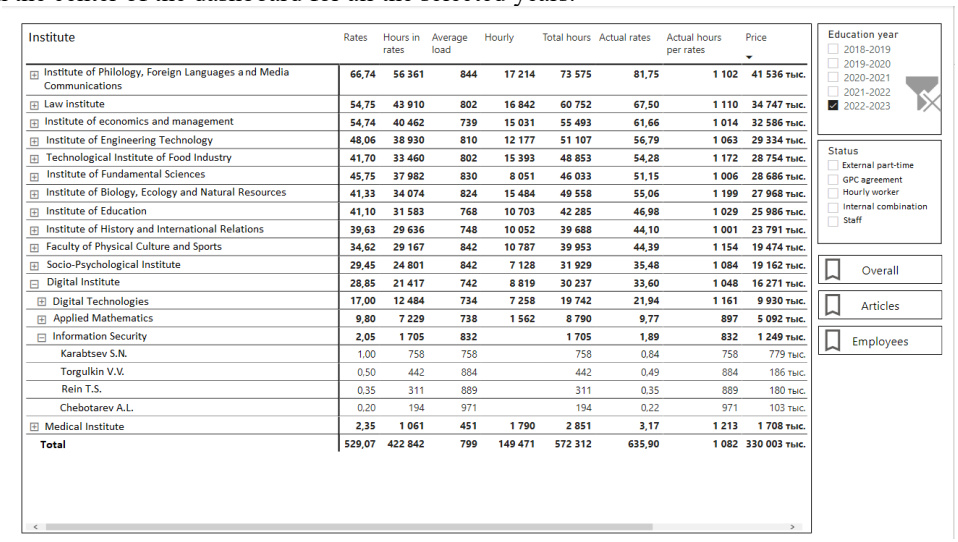
Descriptive analytics in Business Intelligence is the first stage, followed by visualization of the results. Visualization in this context is a set of methods for presenting raw information and data analysis results in a graphical form that is easy to perceive and interpret. You can use visualization to monitor the construction and operation of various analytical models, hypothesis testing and other analysis purposes. The Power BI Desktop component, with the Power BI Report Server, allows you to create and host interactive dashboards (dashboards). Dashboards visualize and analyze data, show the change of key indicators of the organization for the period. We have created, based on a multivariate data model, a series of dashboards. To avoid disclosing confidential data of the university and its employees in the article, we distorted the numerical values and replaced the names of teaching staff with their unique identifiers in the information system.

3.1 Dashboard «Load and teaching rates»

Dashboard "Load and teaching rates" to visualize and analyze the data on the amount of teaching load of the teaching staff (Figure 5).

On the left side of the dashboard is the hierarchy "Institute -> Department -> Employee". We built the measures in such a way that when you move through the elements of the hierarchy, drill-up and drill-down operations are performed to generalize and, conversely, to detail the data. Information about the number of rates, teaching hours in the rate, the calculated average value of the teaching load, the amount of hourly fund and the total number of teaching hours, the calculation of actual hours for a particular teacher (including hourly work), the amount of pay for the actual amount of hours, as well as the key indicator of monitoring the university - the percentage of young employees under 40 years of age and the average age of employees - is in the central part of the dashboard in tabular form.

The right side of the dashboard contains various filters that allow you to display only the necessary information. The first filter is the academic year - you can specify either a particular year or several periods. You can hold down the ctrl key on your keyboard to show the data in the center of the dashboard for all the selected years.



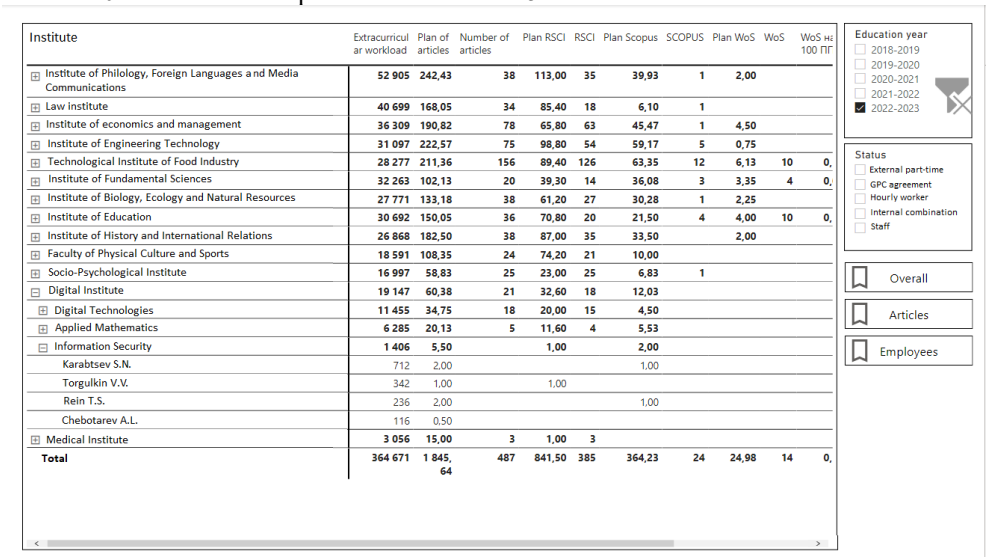
Institute	Rates	Hours in rates	Average load	Hourly	Total hours	Actual rates	Actual hours per rates	Price
⊞ Institute of Philology, Foreign Languages a nd Media Communications	66,74	56 361	844	17 214	73 575	81,75	1 102	41 536 тыс.
⊞ Law institute	54,75	43 910	802	16 842	60 752	67,50	1 110	34 747 тыс.
⊞ Institute of economics and management	54,74	40 462	739	15 031	55 493	61,66	1 014	32 586 тыс.
⊞ Institute of Engineering Technology	48,06	38 930	810	12 177	51 107	56,79	1 063	29 334 тыс.
⊞ Technological Institute of Food Industry	41,70	33 460	802	15 393	48 853	54,28	1 172	28 754 тыс.
⊞ Institute of Fundamental Sciences	45,75	37 982	830	8 051	46 033	51,15	1 006	28 686 тыс.
⊞ Institute of Biology, Ecology and Natural Resources	41,33	34 074	824	15 484	49 558	55,06	1 199	27 968 тыс.
⊞ Institute of Education	41,10	31 583	768	10 703	42 285	46,98	1 029	25 986 тыс.
⊞ Institute of History and International Relations	39,63	29 636	748	10 052	39 688	44,10	1 001	23 791 тыс.
⊞ Faculty of Physical Culture and Sports	34,62	29 167	842	10 787	39 953	44,39	1 154	19 474 тыс.
⊞ Socio-Psychological Institute	29,45	24 801	842	7 128	31 929	35,48	1 084	19 162 тыс.
⊟ Digital Institute	28,85	21 417	742	8 819	30 237	33,60	1 048	16 271 тыс.
⊞ Digital Technologies	17,00	12 484	734	7 258	19 742	21,94	1 161	9 930 тыс.
⊞ Applied Mathematics	9,80	7 229	738	1 562	8 790	9,77	897	5 092 тыс.
⊟ Information Security	2,05	1 705	832		1 705	1,89	832	1 249 тыс.
Karabtsev S.N.	1,00	758	758		758	0,84	758	779 тыс.
Torgulkin V.V.	0,50	442	884		442	0,49	884	186 тыс.
Rein T.S.	0,35	311	889		311	0,35	889	180 тыс.
Chebotarev A.L.	0,20	194	971		194	0,22	971	103 тыс.
⊞ Medical Institute	2,35	1 061	451	1 790	2 851	3,17	1 213	1 708 тыс.
Total	529,07	422 842	799	149 471	572 312	635,90	1 082	330 003 тыс.

Fig. 5. Dashboard «Load and teaching rates».

The "Status" filter specifies the type of employment for a teaching position at the university. The terms of employment may be external part-time, civil law agreement, hourly, internal part-time, and full-time faculty. The dashboard elements named "Overall", "Articles" and "Employees" are hyperlinks to other dashboard pages.

3.2 Dashboard «Articles»

The "Articles" dashboard visualizes and analyzes data on the number of hours of extracurricular work of the teaching staff to publish scientific articles and methodological papers (Figure 6). The structure of the dashboard is like that described above. Besides the number of hours of extracurricular work, the central part of the report displays the total number of planned articles and implementing this plan at the current time. For example, the Institute of Education has produced 36 scientific papers out of 150.05 planned. The number of works can be a fractional number, since each teacher plans to spend a certain number of hours from extracurricular activities and in co-authorship with other teachers of the department or institute. The value of 150.05 comprises 70.8 works indexed in RSCI, 21.5 works indexed in Scopus and 4 works from WoS. As seen, at the current moment (April 25, 2023g.) plan for publications in Scopus partially fulfilled (4 papers), but the number of works in WoS 10 and exceeds the planned number of 2.5 times.



Institute	Extracurricular workload	Plan of articles	Number of articles	Plan RSCI	RSCI	Plan Scopus	SCOPUS	Plan WoS	WoS	WoS на 100 ПГ
⊞ Institute of Philology, Foreign Languages and Media Communications	52 905	242,43	38	113,00	35	39,93	1	2,00		
⊞ Law institute	40 699	168,05	34	85,40	18	6,10	1			
⊞ Institute of economics and management	36 309	190,82	78	65,80	63	45,47	1	4,50		
⊞ Institute of Engineering Technology	31 097	222,57	75	98,80	54	59,17	5	0,75		
⊞ Technological Institute of Food Industry	28 277	211,36	156	89,40	126	63,35	12	6,13	10	0,
⊞ Institute of Fundamental Sciences	32 263	102,13	20	39,30	14	36,08	3	3,35	4	0,
⊞ Institute of Biology, Ecology and Natural Resources	27 771	133,18	38	61,20	27	30,28	1	2,25		
⊞ Institute of Education	30 692	150,05	36	70,80	20	21,50	4	4,00	10	0,
⊞ Institute of History and International Relations	26 868	182,50	38	87,00	35	33,50		2,00		
⊞ Faculty of Physical Culture and Sports	18 591	108,35	24	74,20	21	10,00				
⊞ Socio-Psychological Institute	16 997	58,83	25	23,00	25	6,83	1			
⊟ Digital Institute	19 147	60,38	21	32,60	18	12,03				
⊞ Digital Technologies	11 455	34,75	18	20,00	15	4,50				
⊞ Applied Mathematics	6 285	20,13	5	11,60	4	5,53				
⊟ Information Security	1 406	5,50		1,00		2,00				
Karabtsev S.N.	712	2,00				1,00				
Torgulkin V.V.	342	1,00		1,00						
Rein T.S.	236	2,00				1,00				
Chebotarev A.L.	116	0,50								
⊞ Medical Institute	3 056	15,00	3	1,00	3					
Total	364 671	1 845, 64	487	841,50	385	364,23	24	24,98	14	0,

Fig. 6. Dashboard «Articles».

The report also contains indicators for monitoring the university - the number of scientific papers in RSCI, Scopus and WoS per 100 employees. Figure 7 shows the dashboard element with interactive pie charts.

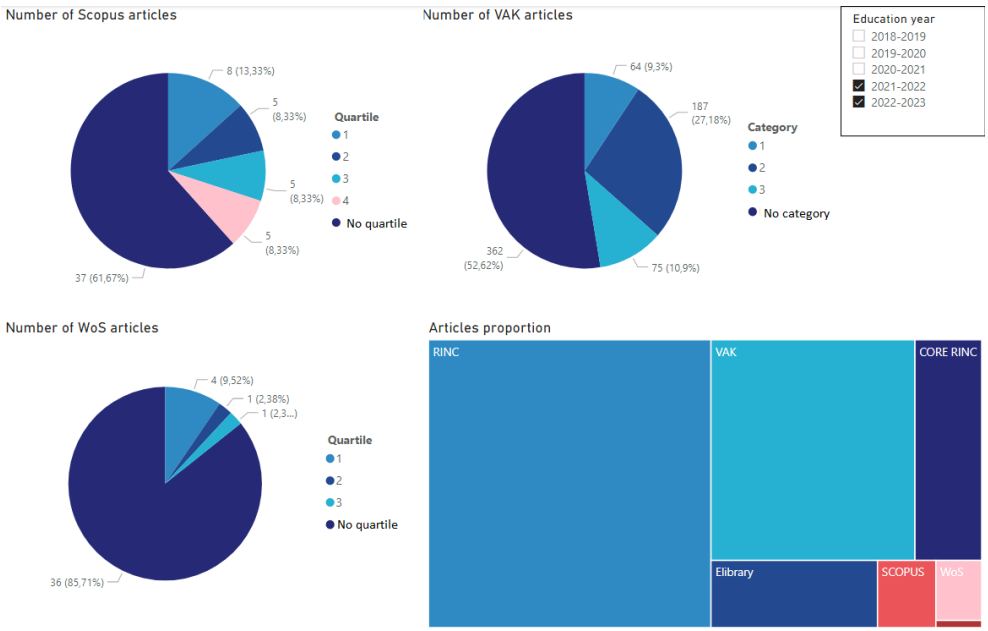


Fig. 7. Interactive element of dashboard «Articles».

Pie charts in absolute value and percentage show the number of articles reviewed in RSCI, Scopus and WoS by quartiles (categories) with filtering by year. The "Articles" dashboard also includes an interactive element displaying the contribution of each institute to the total number of articles by indexing system, year, and quartile.

4 Discussion of results

Various university departments and decision-makers used the analytical dashboards that they built. The Science Department gets information from the Articles dashboard to compile reports on the university's publication activity. The university needs this information to take part in research grant competitions. This information is the basis for summing up the annual results for the title of the best department at KemSU. The university's Best Department statute contains 12 criteria, 5 of which you can get from the dashboard.

Heads of departments use dashboards to evaluate the performance of an extracurricular plan by faculty members of the department. When a publication appears in the scientific citation system (RSCI/Scopus/WoS), it will automatically appear in the analytical dashboard (within only a month), and the department chair will evaluate the reported and actual publication activity of each faculty member. Department heads can also assess a faculty member's workload using the "Load and teaching rates" dashboard before issuing recent assignments for the department.

The built data mart allowed to construct a new analytical dashboard "Dissertation Council" in minimum time to evaluate and regularly monitor the compliance of all members of dissertation councils with the requirements for the availability of peer-reviewed research papers during the last five years (Figure 8).

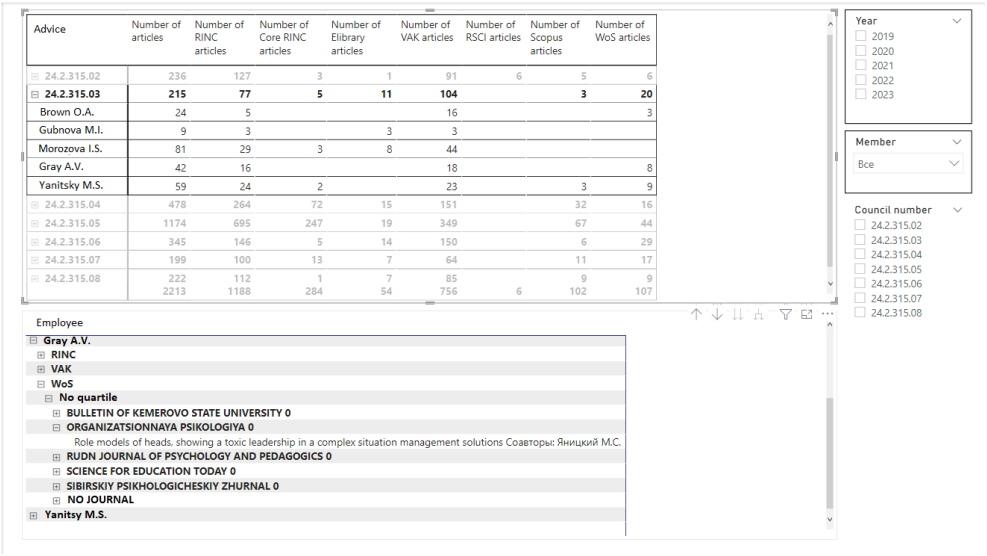


Fig. 8. Dashboard «Dissertation Council».

The central upper part of the "Dissertation Council" dashboard presents the codes and composition of the university's dissertation councils. Each member of the Dissertation Council keeps statistics on publications and their categories. At the beginning of a new calendar year, a check is automatically performed, and if a council member has an insufficient number of publications, the corresponding row or column in the report will be highlighted in red. This will signal the Vice Chancellor for Academic Affairs and the Chair of the Dissertation Council to take immediate action.

The authors of this paper systematically reviewed various scientific sources of literature and did not find a description of such an approach to automating functions related to decision-making among university employees. This study does not refer to the classical sphere of human capital management, as described in papers [11, 12]. These articles speak to the need to encourage organizations to move from reporting to true analytics. Our study is the next step above the scoring and rating systems of faculty activities with manual completion of data in Excel or similar formats [13].

5 Conclusion

This article put forward a method to investigate the tasks of university faculty, which applies advanced Business Intelligence tools, to automate and diminish the labor costs of data search and analysis. The proposed approach cannot carry out a comprehensive assessment of the activities of the teaching staff, as it is complex and diverse. The authors of the article focused their major efforts on the indicators that are most important in evaluating the performance of the teaching staff and the university. An important factor is the existence and readiness of data sources for automated processing. So, the desire to analyze faculty workload led to a purge and enrichment of faculty data in the workload planning system. The authors have also implemented a small webpage in the personal office of the teacher to collect identifiers in citation systems.

The authors plan to enrich the information system by adding dashboards responsible for research and development work, perform accounting of implemented grants and applications submitted for competition.

The research was conducted on the equipment of the Research Equipment Sharing Center of Kemerovo State University, agreement No. 075-15-2021-694 dated August 5, 2021, between the Ministry of Science and Higher Education of the Russian Federation (Minobrnauka) and Kemerovo State University (KemSU) (contract identifier RF----2296.61321X0032).

References

1. O. Moscoso-Zea, J. Castro, J. Paredes-Gualtor, S. Luján-Mora, IEEE Access, **7**, 38778–38788 (2019)

2. A. K. Hamoud, M. K. Hussein, Z. Alhilfi, R. H. Sabr, IJECE **11(6)**, 5301-5314, (2021)

3. M. B. Piedade, M. Y. Santos, *International Conference on Knowledge Discovery and Information Retrieval - Volume 1: KDIR* (IC3K, 2009)

4. P. N. Charikov, S. Yu. Fokin, J. Phys. **1515**, 022073, (2020)

5. A. S. A. Zina, T. A. S. Obaid, IJCSN **5(5)**, 824–827, (2016)

6. A. G. Savina, L. I. Malyavkina, OrelSIET bulletin **1(55)**, 85-92, (2021)

7. S. Karabtsev, R. Kotov, I. Davzit, E. Gurov, J. Appl. Inf., **17 (5)**, 125-142, (2022)

8. W. H. Inmon, *Building the Data Warehouse*. *4th edition* (John Wiley & Sons, Indianapolis, 2005)

9. S. Sharma, R. Jain, *Fourth International Conference on Advanced Computing & Communication Technologies*, 271-276, (2014)

10. S. Suman, P. Khajuria, S. Urolagin, MoSICom **659**, 30–39, (2020)

11. A. Margherita, HRMR **32 (2)**, 100795, (2022)

12. D. B. Minbaeva, HRM **57(3)**, 701–713, (2018)

13. V. G. Tronin, S. V. Skvortsov, UlSTU bulletin **1(77)**, 55-60, (2017)