

Mathematical Model of Formation of a Unified Digital Platform of Scientific and Educational Resources

Medennikov V.I.* Flerov Y.A.

Dorodnitsyn Computer center of FRC "Informatics and Control" of RAS, Vavilova 44-2, 119333, Moscow, Russia
**Corresponding author. Email: dommed@mail.ru*

ABSTRACT

The article considers a mathematical model that proves the potential possibility of creating a single information Internet space of scientific and educational resources on the basis of a single digital platform, the need for which is due to the rapid increase in the volume of information arrays in educational and scientific activities, the advent of modern digital technologies, and the need for these resources of all layers users. The work analyzes the simulation results and experience in the formation of such a space. The proposed project is the implementation of the ideas of A. Kitov. and academician Glushkov V.M. on the development of the National Automated System (NAS) for the collection and processing of information for accounting, planning and managing the national economy in the USSR in the 60s of the twentieth century in the field of science and education. Such a space should also become an instrument for improving the quality of human capital, its assessment, and the impact on the socio-economic situation in the regions and the country.

Keywords: *scientific and educational resources, digital economy, mathematical model, Internet space*

1. INTRODUCTION

The need to create a unified information Internet space for scientific and educational resources (UISSER) in connection with the epidemic of coronavirus, which forced the offline business and public services to be transferred online, caused this problem to be transformed from a purely theoretical format at conferences and scientific journals to a practical plane. Of particular relevance to the problem was the need to transfer students to distance learning, and employees of organizations to work remotely. At the same time, the ever-increasing adoption of digital technologies in the economy and social life requires a significant number of highly qualified employees, retraining of many specialists, radical updating of both production and management, from low-level workers to senior managers, the transition to digital methods management. In this situation, science, together with education, must offer scientifically-based advanced digital technologies that are ahead, both in time and in efficiency, of spontaneous, chaotic, non-integrated technologies offered by the market. Otherwise, soon we will hear many critical comments about the unfulfilled hopes for the digital economy. As, for example, this happened in the agro-industrial complex, the era of the original, non-integrated design, development and implementation of information management systems. So, in [1], it is stated that "attempts to solve managerial problems at the expense of computers led to enormous costs of labor and funds, and all this went into oblivion, informatization of agriculture brought only harm and did

not bring any effect in the country's GDP". Based on this critical approach, the director of the Institute of Agrarian Problems and Informatics A.V. Petrikov even made a decision about the unnecessary topics of research on the digital economy of the agro-industrial complex and closed it at his own institute. Moreover, he went further and offers to close information departments in agricultural universities, justifying such a decision by the fact that special IT organizations should deal with the digitalization of the agro-industrial complex.

Digitalization of the economy inevitably makes this process shift towards integration of both information systems (IS) and information resources (IR), which is possible only on the basis of relevant standards, ontological modeling of the presentation of information resources, management functions, and also on the basis of an integrated approach to all stages of the design, development and implementation of digitalization systems. An integrated approach is also needed in the training of relevant specialists.

2. SUBSTANTIATION OF FORMING A UNIFIED INFORMATION INTERNET SPACE OF SCIENTIFIC AND EDUCATIONAL RESOURCES

In this regard, science proposed a scientifically-based project based on mathematical modeling of a single information Internet space of a country's digital interaction (SIISCDI), integrating a single digital platform (SDP) of the country and UISSER, as well as design automation

tools for applied information and computing systems [2, 3, 4, 5, 6, 7].

Obtained by mathematical modeling, such a single digital platform is a combination of all information arrays of technological, primary and statistical accounting of most sectors of the national economy into a single cloud database using a universal information system for collecting, storing and using it, a unified system of classifiers, registers of all material, intellectual and human resources. UISSER, on the other hand, is a combination of information scientific and educational resources (ISER) in a single cloud database, which have a three-fold role: supporting research, increasing the level of education (sometimes retraining) for all segments of the population, and an effective system for transferring scientific and educational knowledge to the economy through unlimited access to this knowledge not only to traditional users in the person of scientists, students and teachers, but also to future applicants and employers, government agencies, producers, business, management, other categories of the population. Such a space should also become an instrument for improving the quality of human capital (HC), its assessment, and the impact on the socio-economic situation in the regions and the country [8, 9].

Such a project, of course, is the most important component of the digital economy program, since scientific development is the main factor influencing the growth of the economy, one of the drivers for improving the quality of the most valuable strategic resources at present, human capital. And human capital is a determining factor in the level of social development of states. The ideas of SIISCDI are connected with the appearance and significantly increased capabilities of the Internet with all the technologies accompanying it. The implementation of SIISCDI will enable a large number of users to access various information systems with a significant expansion of the number of tasks for automation. Moreover, SIISCDI allows the integration of both IS and IR not only at the level of individual enterprises, but also at the level of entire industries, states and all of humanity.

The UISSER project is the realization of the ideas of A.I. Kitov, who is turning 100 years old this year, and Academician V.M. Glushkov on the development of the National Automated System (NAS) in the 60s of the twentieth century in terms of science and education [10, 11]. Significantly increased opportunities and the level of development of both the scientific component of the NAS and the software and hardware of the Internet currently make it possible to realize these ideas on a full scale.

The refusal of this project by the Government of the USSR played a negative role in the future at the stages of computerization, electronization, and computerization of the country and continues to play at the current stage of digitalization of the economy, education, science and society as a whole. The rejection of integration, ontological technologies in the design of databases

containing various types of informational scientific and educational resources (ISER) leads to a significant increase in the cost of their design and operation and makes it difficult to find the necessary information for an increasingly growing audience of ISER users.

Currently, as a result of the requirements of the Ministry of Education and Science, Rosobrnadzor to educational organizations about the content structure of their sites (in the form of a self-examination report), universities are forced to normatively improve their sites. It should be noted that the self-examination report does not contain data reflecting the ISER, while no similar requirements for scientific organizations by the Ministry of Education and Science and Rosobrnadzor are observed. Thus, Russian scientific and educational organizations only develop Internet technologies under the influence of the state. And since officials, as a rule, are not well versed in information technologies, universities and research institutes without strict requirements put forward for them to present the necessary content on websites in a certain structured form on the basis of integration, typification and standardization calmly ignore these opportunities.

Types of ISER are based on an analysis of the websites of research institutes, universities, information and consulting services in comparison with a survey of the necessary information resources for business. As a result of such a comparison, the intersecting set of the following generalized types (without ontological modeling) of IR was identified: development, publication, application software packages (ASP), DB, consulting activity, regulatory information, distance learning [2, 8].

Figure 1 shows the functional structure of UISSER with a list of various subprojects-services, the number of which will constantly increase with the accumulation of information. In [12], a more detailed description of the UISSER study with an analysis of the results of mathematical and ontological modeling of ISER integration is given, the effectiveness and the possibility of forming such a scientific and educational space are substantiated. The practical implementation of UISSER was carried out in 2007-2008. within the framework of the portal being developed by the Russian Academy of Agricultural Sciences (RAAS). At the same time, over 12 thousand scientific publications from subordinate research institutes, over 2.5 thousand developments over the past 10 years, over 0.4 thousand consulting services in accordance with the thematic section and the list of relevant consultants were integrated on a portal from a single methodological position. Note that at that time E-library contained a significantly smaller number of publications, while other types of scientific and educational resources were not contained at all, as there are none now [6]. The experiment was terminated due to the reform of the RAAS, as well as the lack of funds to continue improving the development.

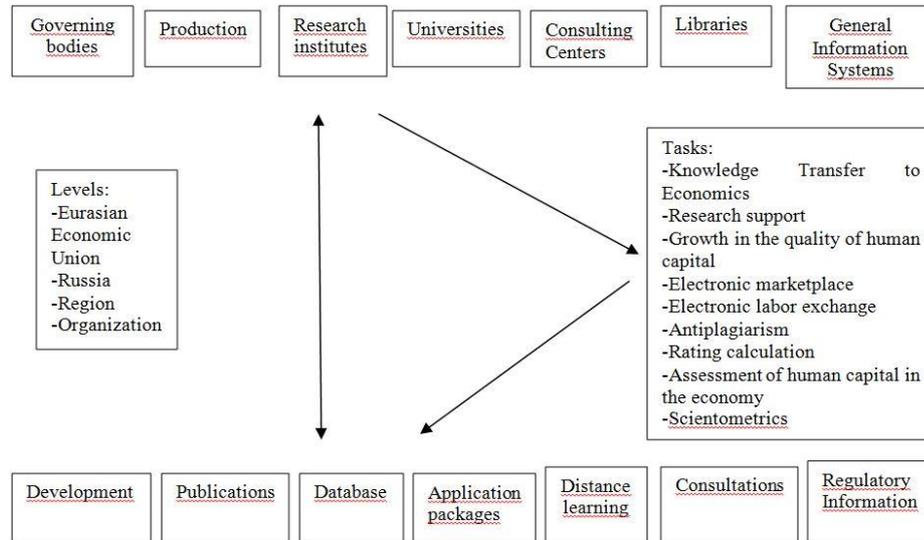


Figure 1 Functional structure of UIISSER

3. MATHEMATICAL DESCRIPTION OF THE UIISSER MODEL

3.1. The experience of forming UIISSER

It is not known how much the NAS project was mathematically calculated, but with respect to UIISSER it was necessary to confirm by the scientific examination the fundamental possibility of creating such a space by mathematical modeling of various options for integrating ISER for any potential volumes and number of users. Modeling was carried out on the example of organizations involved, to one degree or another, on agricultural topics.

In the model, and in practice, during the formation of the UIISSER, the three most feasible of the possible ISER integration options were used. In the first, ISER was placed in the form of catalogs in the UIISSER common database with a single provider. In this case, if the user is interested in any information found, similar to the catalog of a traditional library, he is redirected to the site of the keeper of the full-text information resource indicated in the catalog. In the second case, all ISER was hosted by a single provider in a common database. In the latter version, a mixed strategy was implemented, in which part of the ISER, for some reason, for example, the requirements of some privacy of the information, or the corporate interests of individual authors and organizations, was placed in the catalog form, the other in full, both in full text and in the form of directories.

It was taken into account that the quality of service by the user's provider is usually measured by several parameters: network reliability, time delays in transmitting information, statistical characteristics of delays, and throughput. Since we want to find out the global

characteristics of the network when transferring information by a large number of owners to one provider without having specific volumes of this information (only a small part of it is poorly structured on the sites), we will simulate processes on a sufficiently large time interval, taking month per unit of time. At this interval, only network bandwidth is relevant.

When modeling UIISSER relied on the characteristics of the currently most common site content management system (CMS) "1C-Bitrix". This was shown by an analysis of the sites of agricultural universities: 17 out of 54 sites (31.5%) were developed using this software product.

3.2. Constants and Parameters

i - an index defining the type of information transmitted (text, picture, video, etc.) $i \in I$;

j - an index defining membership in the category of organizations-owners of information (research institutes, universities, etc.) $j \in J$;

m - number of provider of CMS 1C-Bitrix services $m \in M$;

n - ISER type, $n \in N$;

k - number of the organization that owns the information $k \in K_j$;

l - code of the form of the stored information $l \in L$;

d_{im} - the available load m of the provider on i the transmitted information (in Mb);

V_{ijkl} - volume of the i transmitted information l of the stored form k - of the organization j - of the category of the organization that owns the information (in Mb);

D_{im} - throughput capability of the m - 1-st i 2-nd transmitted information (in Mb);

z_{ijmkl}^1 –specific expenses for transferring to the provider of CMS 1C-Bitrix services, units of i the transmitted information, the l stored form, the j category of the information owner organization of the m provider, k organization (in rubles / Mb) ;

z_m^2 - specific costs of maintaining the site of the m - provider (in rubles / unit);

z_{jk}^3 - specific cost of maintaining the site j - category of the organization that owns the information when storing the content with the provider k - organization (in rubles / unit);

P_i^1 - specific value (in time unit) of calls to the i transmitted information;

P_{il}^2 - the specific value (in units of time) of page views i - the transmitted information l - the stored form;

P_{ijkl}^2 - the specific size of page views j - the category of the organization that owns the information i - the transmitted information l - the stored form k - the organization

Then, $P_{il}^2 = s \cdot \sum_{j,k} P_{ijkl}^2$ where s determines the gain of views due to the integration of the transmitted information (in our case, the calculations showed that $s = 2,5$ in the case of the transition of organizations to a standard site);

P_m^3 - the average volume of the page of the site of the m provider (in Mb);

C^0 - financial resources allocated to transfer information to one of the providers m (in rubles / unit); it is suggested that all providers providing CMS 1C-Bitrix services have stored information in a single structured form on a single data model.

In this case, we introduce a number of constants:

b_{il} - the average volume of the transmitted file i - transmitted information l - stored form (in MB);

g_{il} - the average number of calls for i - the transmitted information l - the stored form with all providers providing CMS 1C-Bitrix services;

r_{ijknl} - the number of transferred files for the i - information n - the type of ISER for the j - category of the organization that owns the information l - the stored form k - organization at its provider;

a_{iknl} - the average value of calls n of the ISER type for the i - the transmitted information l - the stored form at provider;

V_{inl} - a feature indicating the presence of the i - transmitted information l - the stored form of the n - type of ISER;

$V_{inl} = 1$, if there is available i - the transmitted information l - the stored form n - the type of ISER; 0 - otherwise

Then

$$g_{il} = s \cdot \sum_{n,k} v_{inl} \cdot a_{iknl} ;$$

$V_{ijkl} = b_{il} \cdot \sum_n r_{ijknl}$ - volumetric characteristics;

$$P_i^1 = \sum_l g_{il} .$$

3.3. Variables

x_{ijmkl} - an increase in the load on the m provider due to the placement of the i transmitted information in its DB, l stored form, j category of the organization that owns the information, k organization (Mb);

$y_{ijmkl} = 1$, if the k organization of the j category of the organization owning the information places the i transmitted information of the l stored form of the m provider, otherwise 0.

3.4. Model equations

$d_{im} + \sum_{l,j,k} x_{ijmkl} \leq D_{im}$ - restrictions imposed on the throughput capabilities of the m - provider for the i - transmitted information;

$x_{ijmkl} = (P_i^1 \cdot P_{il}^2 \cdot P_m^3 + g_{il} \cdot b_{il}) \cdot y_{ijmkl}$ - equation determining the increased load m of the provider due to the placement i of the transmitted information in its DB, l stored form, j category of the organization that owns the information, k organization (Mb);

$\sum_m y_{ijmkl} \leq 1$ - an expression stating that the entire array of information should be stored only with one of the m providers.

Financial constraints on efforts to move information to a provider providing CMS 1C-Bitrix services

$$C^1 = \sum_{i,j,m,l,k} z_{ijmkl}^1 \cdot V_{ijkl} \cdot y_{ijmkl}$$

Are the following: $C^1 \leq C^0$.

3.5. Optimization criteria

There can be two such criteria in a model:

The first criterion $w = \sum_{i,j,m,l,k} V_{ijkl} \cdot y_{ijmkl} \rightarrow \max$ determines the maximization of the volume of movement of information to the provider that provides CMS 1C-Bitrix services.

The second criterion $-C^2 \rightarrow \min$ – determines the minimization of the cost of servicing ISER for 1C-Bitrix providers, where

$$C^2 = T \cdot \left(\frac{1}{I \cdot J \cdot L} \sum_{i,j,m,l,k} z_m^2 \cdot y_{ijmkl} + \frac{1}{I \cdot M \cdot L} \sum_{i,j,m,l,k} z_{jk}^3 \cdot (1 - y_{ijmkl}) \right)$$

T - the contractual service period of UIISSER).

To find a solution to the formulated problem, one could apply standard graph methods [13]. However, taking into account the fact that at present there are very few providers providing CMS 1C-Bitrix services, it is quite possible to use a two-stage solution search procedure. First, the optimization problem is solved sequentially for each provider, fixing a specific value of m . At the next stage, the most acceptable provider is selected by brute force, based on the previously selected criteria.

3.6. Initial data of the model

For model calculations, three basic options were generated from the information available on the websites: annual volume, 5-year volume (information for the last 5 years), full volume (information for all years), and two basic options for the forecast number of website visitors - the current and forecast numbers. As an example, the Table 1 shows the forecast number of visitors to UIISSER for organizations related to agricultural topics.

Table 1 Prognosed number and type of UIISSER visitors

Visitors	Monthly visits
Farms	210,000
Farms workers	1,100,000
Students	31,100,000
Management	240,000
Scientific workers	1,340,000
Others	31,400,000
TOTAL	65,390,000

4. RESULTS OF CALCULATION

Numerous model experiments of forming a single information Internet space of scientific and educational resources [12] gave great confidence in the capabilities of providers that provide technical and software services for

developing and maintaining sites on the most famous 1C-Bitrix software, to accumulate all ISER in a single UIISSER database, produced by agricultural universities and research institutes over the past five years, to one of them with a sufficient degree of response efficiency to simulated numerous requests various users. Economic efficiency, which is of great importance in conditions of underfunding of science and education, looks like this - only on the development and maintenance of sites the annual savings will amount to more than 1 billion rubles, which will be ensured by the integration and typing of IR and sites within the framework of UIISSER.

With the accumulation of IR and the development of digital technologies within the framework of the UIISSER, it will be possible to develop assessment tools in the form of artificial intelligence, big data, mathematical modeling, both human capital at various levels of government and the impact of the HC on social welfare and the development of Russian society. In addition, such a project will provide an opportunity to create an automated and independent tool for evaluating all the activities of scientific and educational organizations, common to them, in particular, in the interests of the state, regions, and business. By integrating some of the information available in the pension fund into the tax service, such a tool will make it possible to track the career of graduates of specific universities, their contribution to the HC, to the economy at all levels.

5. CONCLUSIONS

The introduction of the UIISSER platform in all scientific and educational organizations is the most urgent task of the country's digital transformation with a significant reduction in resources for such a transformation of science and education. The considered platform will become a powerful tool for bringing the most effective innovative solutions to the economy.

ACKNOWLEDGMENT

This work was supported by RFBR (grant № 20-07-00836)

REFERENCES

- [1] Система управления – основа реализации модели инновационного развития агропромышленного комплекса России: материалы Всероссийской научно-практической конференции, 2013. Москва/ отв. ред. I.G. Ushachev. – GNU VNIIESKH, 2013. – 180 s.
- [2] Ereshko F.I., Medennikov V.I., Muratova L.G. Modeling of a digital platform in agriculture. IEEE Xplore Digital Library. Eleventh International Conference Management of Large-Scale System

Development (MLSD), Moscow, Russia, 2018. DOI: 10.1109/MLSD.2018.8551894

[3] F. I. Ereshko, V. I. Medennikov and S. G. Salnikov. Modeling of unified information internet space of the country. IEEE Xplore Digital Library. Tenth International Conference Management of Large-Scale System Development (MLSD), Moscow, Russia, 2017.

[4] Ereshko F.I., Kulba V.V., Medennikov V.I. Digital platforms clustering model // 12th International Conference Management of Large-Scale System Development (MLSD), Moscow, Russia, 2019. DOI: 10.1109/MLSD.2019.8911012.

[5] Vyshinskij L.L., Grinyov I.L., Flyorov YU.A., SHirokov A.N., SHirokov N.I. Avtomatizaciya proektirovaniya prikladnyh informacionnyh sistem. Soobshchenie po programmnomu obespecheniyu. M.: VC RAN, 2003 g. 52 s.

[6] Flerov YU.A., Vyshinskij L.L. Informacionnye tekhnologii i vychislitel'nye sistemy. 2018. № 3. S. 29-41.

[7] Zatsarinnyy A.A. Kondrashev V.A., Sorokin A.A. Approaches to the organization of the computing process of a hybrid high-performance computing cluster in the digital platform environment.// CEUR Workshop Proceedings. Volume 2426, 2019, Pages 12-16.

[8] Medennikov V.I., Muratova L.G., S.G. Sal'nikov S.G. Modeli i metody formirovaniya edinogo informacionnogo internet-prostranstva agrarnykh znaniy.– Moskva : Izdatel'stvo GUZ, 2014. – 426 s.

[9] Ereshko F.I., Medennikov V.I., Sal'nikov S.G. Proektirovanie edinogo informacionnogo Internet-prostranstva strany // Biznes v zakone. Ekonomiko-yuridicheskij zhurnal. – 2016. – №6. – S. 184-187.

[10] Benjamin Peters. How Not to Network a Nation: The Uneasy History of the Soviet Internet / Benjamin Peters – SSHA : MIT Press, 2016. – 360 c.

[11] Glushkov, V. M. Makroekonomicheskie modeli i principy postroeniya OGAS / V.M. Glushkov – Moskva : Statistika, 1975. – 160 c.

[12] Medennikov V.I., Muratova L.G., S.G. Sal'nikov S.G. Metodika ocenki effektivnosti ispol'zovaniya informacionnyh nauchno-obrazovatel'nyh resursov. – Moskva: Analitik, 2017. – 250 c.

[13] Kulba V.V., Mikrin E.A., Pavlov B.V., Platonov V.N. Theoretical foundations of designing spacecraft information and control systems. –Moscow: Science. 2006. (in Russian).