doi 10.26089/NumMet.v22r422

The use of ontologies for solving scientific problems (by example of geophysics)

B. M. Glinskiy
Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
ORCID: https://orcid.org/0000-0002-0119-6370, *e-mail:* gbm@sscc.ru

G. F. Zhernyak
Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
ORCID: https://orcid.org/0000-0002-0621-9551, *e-mail:* zhernyak@sscc.ru

G. B. Zagorulko
A. P. Ershov Institute of Informatics Systems, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
ORCID: https://orcid.org/0000-0003-2155-1357, *e-mail:* gal@iis.nsk.su

P. A. Titov
Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
ORCID: https://orcid.org/0000-0002-0378-7465, *e-mail:* tapawel@gmail.com

Abstract: The paper covers an intelligent support system that allows to describe and to construct solutions of various scientific problems. In this study, in particular, we consider geophysical problems. This system is being developed at the Institute of Computational Mathematics and Mathematical Geophysics of the Russian Academy of Sciences (ICMMG SB RAS) and the Institute of Informatics System of the Russian Academy of Sciences (IIS SB RAS). The system contains a knowledge base, the core of which is a set of several interconnected ontologies such as the ontology of supercomputer architectures, the ontology of algorithms and methods. Ontology can be viewed as a set of concepts and how these concepts are linked. As the result, the authors present an ontological description of two geophysical problems using the intelligent support system: 1) the seismic wavefield simulation and 2) the reconstruction of a seismic image through pre-stack time or depth migration. For a better visual understanding of the described system and the obtained results, the paper also contains several schematic diagrams and images.

Keywords: supercomputers, geophysics, intelligent support, ontology, modeling.

Acknowledgements: The work was funded by the budget project of the ICMMG SB RAS 0251–2021–0005 and the RFBR grants No. 19–07–00085, No. 20–01–00231. All numerical tests were conducted on the SSCC of ICMMG SB RAS.

For citation: B. M. Glinskiy, G. F. Zhernyak, G. B. Zagorulko and P. A. Titov, "Using ontologies for solving scientific problems (by example of geophysics)," Numerical Methods and Programming. **22** (4), 333–343 (2021). doi 10.26089/NumMet.v22r422.

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

1. Introduction. Every specialist in any branch of knowledge that implies the work with massive data and is linked to the use of supercomputers faces the issues: 1) what technique (a set of techniques) is appropriate to use for solving the problem? 2) which supercomputing systems are required, best suited or available? 3) what the most effective way to utilize supercomputing technologies for numerical tests is? It is necessary to derive the maximal benefit from the hardware means available for solving the concrete tasks by optimizing the process of finding the solution. There are several components to solving the problem effectively. Components that are not related to each other but should be considered together at the same time. This means that a specialist should be well oriented in fairly different areas, for instance, both in methods and algorithms used in solving the problems and in the supercomputing systems and thechnologies, in their effective joint utilization.

That is where the ontological approach comes in. Ontology can be viewed as a set of concepts and how these concepts are linked. For example, in geophysics by ontology we understand a systemized knowledge base that allows researching a medium structure: methods, algorithms, hardware, software and how they can be used together. This approach will be covered in paragraph 4, for two specific geophysical problems.

The presentation of information utilizing the ontological methodology has grown in its popularity over recent years [1, 2]. The necessity to create such systems of intelligent support was advanced in various scientific fields such as physics [3], astrophysics [4-6], biology [7], botany [8]. In regards to geoscience, the use of ontologies was discussed in [9–14]. These papers are mostly focused on geology, data processing, and geographic mapping. In this paper, we focus on the ontological representation for geophysical tasks solution, what components contribute to it and how these components are connected.

Therefore, we have a set of algorithms, methods, program codes that can be used for a task in question, and we have also a set of developed supercomputing hardware and software. Their combination enables specialists to create effective program components for solving and studying the original problem. And such combination is a seriuos problem. Nowadays, the ontological representation is attracting the attention of a growing number of scientists. However, navigating through ontologies usually requires certain level of expertise in several subject areas. Therefore, it is highly desirable to create a user-friendly interface and intuitive navigation system.

The authors think that creating a wide knowledge database on various specific subject areas is an urgent problem. If this database will be properly organized as a system of intelligent support, then it could be a powerful support for researchers. So, the main work of authors is developing, updating and improving such a system. It is considered a constantly evolving and growing object and one can always add new elements and links to it. For example, the user can choose a scientific field and find his task described in the system. The user can see what mathematical and numerical methods there are, what supercomputer architectures and parallel software can be used and which are best suited for different numerical methods. If there are program components that can be used, or someone has already got the experience in solving the problem — the user will also know this, since there will be corresponding links to pieces of code and literature.

This investigation's novelty consists in developing a free-access resource that provides the relationship between two different areas: geophysics and supercomputers. It also has intuitive navigation system and user-friendly interface.

As an example, this paper also describes how the system can be applied to the two different geophysical problems: 1) the seismic wavefield simulation and 2) the reconstruction of a seismic image through pre-stack time or depth migration.

2. Bases of intelligent support system. The published work [15] touches on the means of the ontology concept in any general subject area. The proposed concept of the general intelligent support system (ISS) is covered in detail in [16]. Fig. 1 represents the upper level of the system objects and interconnections between them.

The core of the system is a knowledge base module. It contains ontologies of computational methods, parallel technologies and available hardware architectures. Rules allow one to perform output and obtain information that is not explicitly presented in the knowledge base.

The information and analytical Internet resource (IAIR) module will be described in the next section.

The library of software components contains ready-to-use code fragments that can be implemented during a problem-solving process.

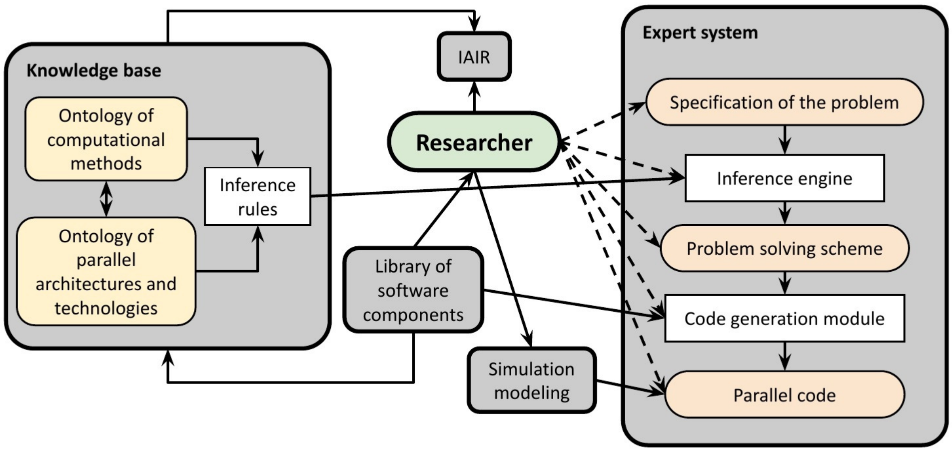

Figure 1. The upper level of the intelligent support system

Another vital part of ISS is a set of modules called the expert system. Throughout the process of solving a problem, it implies a constant interaction between a user, a knowledge base and a library of software components. The user should first specify his compute-intensive problem. Then the inference engine allows one to implement the inference rules to construct the best ways for solving the problem. The inference engine works as an advisor providing different methods within the available knowledge base. The code generation module is responsible for supporting the construction of a parallel code. It provides the necessary elements of the software library that will help during the programming process of the problem-solving scheme.

The module of simulation modeling provides the user with the ability to evaluate the scalability of his parallel code. One can get a time estimate of program execution on a different number of cores on a specific parallel architecture. Thus the user can choose the most effective time-wise and the capacity-wise way to solve his problem, taking into account its features. Indeed, computing time is costly. For example, if there is one big numerical test to be conducted, then the user should simply get as much computing power as he can afford. But more often a series of numerical tests is carried out. Therefore, there is a problem of optimizing time and energy (and costs) spent for each test compared to time and energy spent for all tests. That's where code performance simulation comes in. This approach has already been used in a few geophysical problems [16, 17].

We should note that ontologies are described with the aid of the OWL language [18]. This includes a description of classes and their properties, as well as specific objects that fill the knowledge base — methods, algorithms, software, parallel architectures, etc. The language SWRL (Semantic Web Rule Language) [19] is used for setting the rules inside the ISS.

3. The informational and analytical internet resource. The information and analytical Internet resource (IAIR), which is a part of the ISS under consideration, is a system with a web interface that contains structured information related to the knowledge area "Support for solving compute-intensive problems on supercomputers" and provides meaningful access to it. The systematization of information and the functioning of the IAIR is carried out on the basis of the ontology of this knowledge area. The ontology concepts hierarchy is shown in the left panel of each page of the IAIR. When one selects any ontology concept, a list of its objects will be displayed in the central part of the page (Fig. 2). The user can select an object of interest from the list and get its description.

Figures 3 and 4 present a description of the balance method (also, might be called the balance technique). The method allows meeting the law of energy conservation during the process of constructing a numerical solution to the problem. Figure 2 shows that the balance method is an element of the subclass "Numerical Method" of the class "Research Method". This class is highlighted in ontology. Figure 3 shows a table representation of the balance method. It has a brief description and links to other elements of the ISS, such as problems existing in the ISS that utilize this method, software that was created using this method, as well as the literature that describes the implementation of the method. Fig. 4 shows a graphical representation of the considered method, with all logical links, related classes (subclasses) and elements. Therefore, one can fully understand the method's place in the system.

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

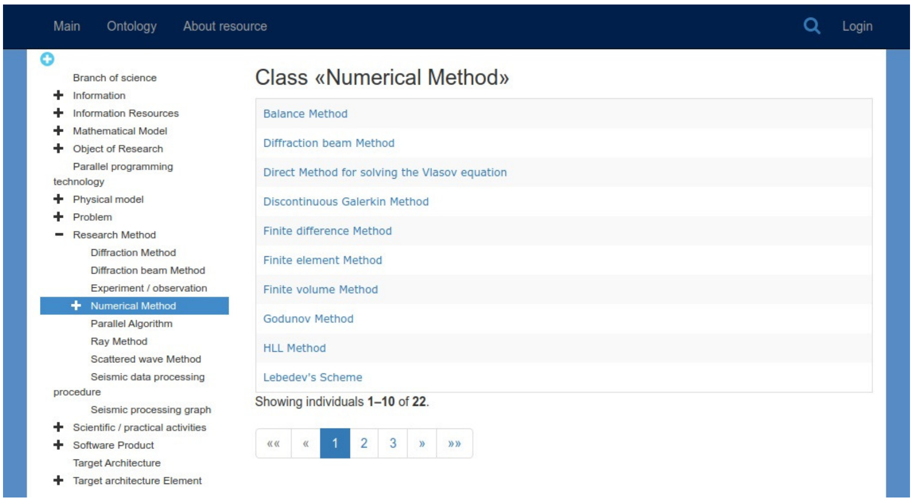

Figure 2. Subclass “Numerical Method” and its elements

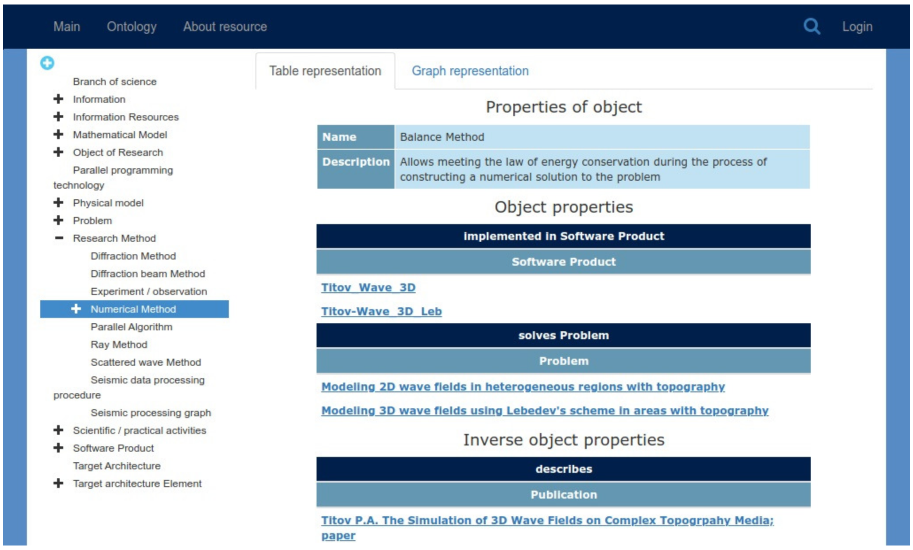

Figure 3. The table representation of the balance method and its links in the system

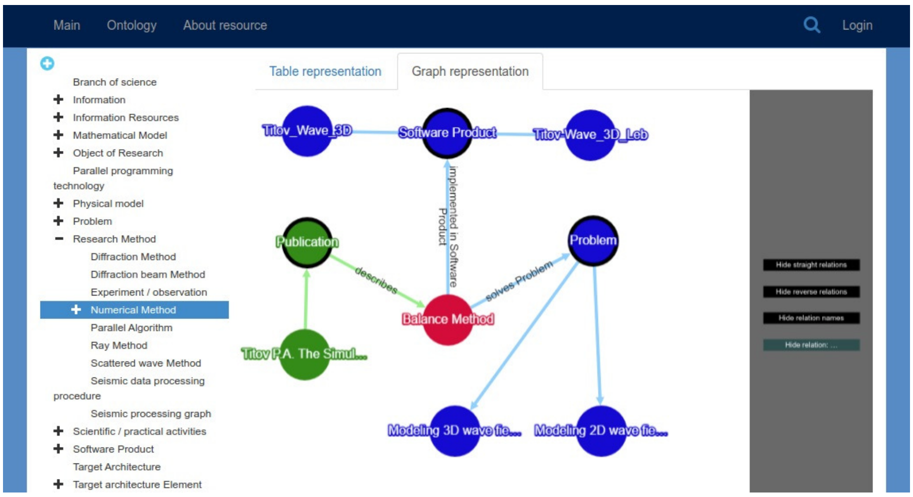

Figure 4. The graph representation of the balance method and its links in the system

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

This way of presenting the information seems to be highly intuitive and should not cause any confusion for inexperienced users. The website design has yet to be discussed and improved, but both logical and navigational parts of the IAIR have already been finished (https://uniserv.iis.nsk.su/spsc/).

4. The ISS for solving geophysical problems. In this section, the authors present their view on the ontological representation of the information on geophysical problems. The ontology is multileveled and includes other ontologies as well. Note that ontologies are widely used in oil and gas surveys [20], geophysical studies [21] and others. But the accessibility of these systems is restricted. Geophysics and geology are highly industrial scientific areas and almost any useful software and data are proprietary, including databases.

Figure 5 represents the upper level of ontology for solving compute-intensive problems on supercomputers. Further, we overview the main components that generally contribute to the original problem-solving process and in subsections 4.1 and 4.2 we touch on a more detailed view of two specific problems.

The knowledge base of the ISS originates from the field of science, in our case from geophysics. The main objects of the research are a physical phenomenon (wavefield) and a physical object (medium). Both objects can be studied through real experiments and observations or numerical experiments or their combinations. Real experiments involve sampling the medium and recording wavefield data on a set of receivers, and numerical experiments involve mathematical modeling using modern supercomputing technologies. A medium is described by the so-called "digital twin", which is the closest discrete representation of a real object. To describe the wavefield propagation, an approximate mathematical model is used (linear equation system of elasticity [22]). The solution presentation might be analytical, semianalytical or numerical which is the general case. The numerical method for a specific discrete domain decomposition is used to develop a parallel algorithm. A program code is created for a parallel algorithm to perform numerical tests on a given parallel architecture. For the most optimal performance, the code should consider the computation domain decomposition, what calculator (computing device) is used and what parallel programming technologies are necessary.

The upper level of ontologies of parallel architectures, computational algorithms and methods are covered in [5, 6] in detail.

4.1. Modeling of wave propagation in complex-built media with topography. In this subsection, we offer an ontological representation of the problem-solving process of numerical wavefield modeling and how the expert system logic works in this case. All utilized components and logical links are highlighted with a bold frame in fig. 6.

The problem lies in the numerical modeling of 3D waves in media with surface topography [22]. The subject of the study is wavefield propagation via numerical experiments. All parameters and structure of the medium are known. The medium is isotropic and solid, the area under study has a curved surface (hills or mountains, for example). When seismic waves move through the medium they create a deviation of particles from their balanced state. The common way to study the wave field is to measure the displacement of the particles over time or the velocity of this displacement. Both ways are best suited for different medium structures. For

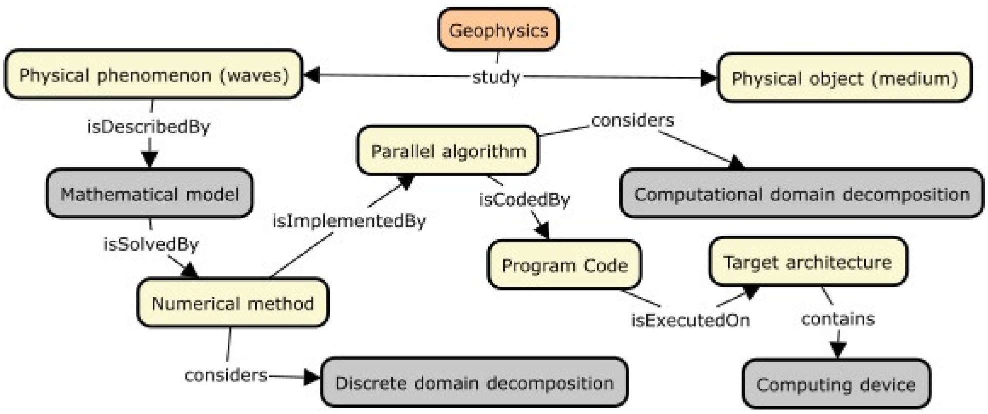

Figure 5. The upper level of the geophysical ontology

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

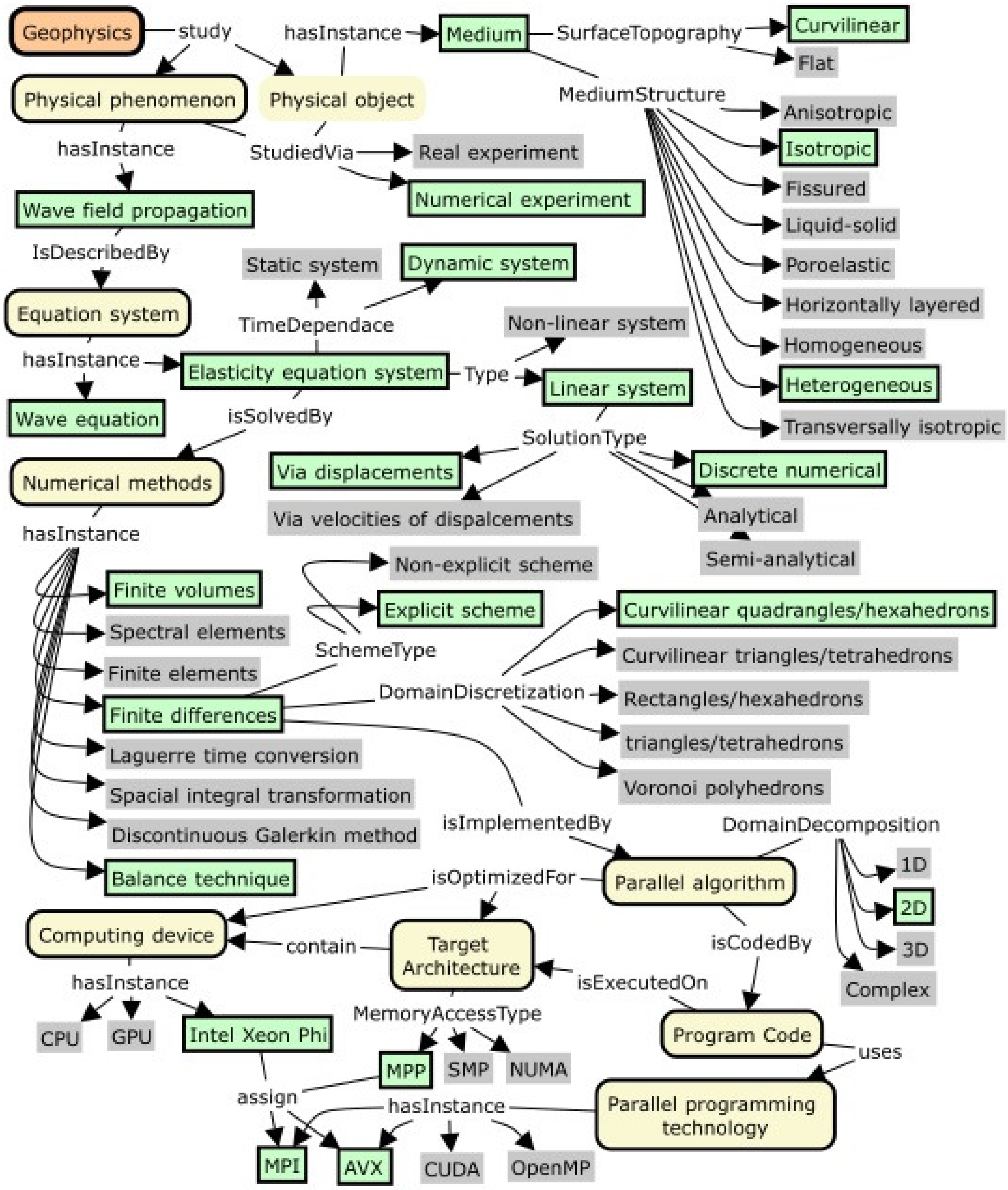

Figure 6. Highlighted elements with thick edges are used for solving the problem of numerical seismic field simulation

the mathematical model, a linear elastic system is chosen. It contains of 3 equations describing the particle displacement. The structure of the medium is solid and the use of the system with 9 equations (in the case of displacement velocities) is excessive. This approach allows us to save on RAM during numerical tests on a parallel architecture. Discrete decomposition of the computational domain is represented by curved hexahedrons [23]. This allows a good consistency between the boundaries of the numerical domain and the physical one. Therefore it provides better accuracy of the free surface boundary conditions when we develop a finite-difference scheme. More details can be found in [22].

As for a numerical method, there are used the balance technique, finite volume method and finite differences method to create an explicit finite difference scheme. An explicit scheme makes a good base for a highly scalable parallel algorithm, which is optimized for the 3D domain decomposition. Due to the massive data of real-scale

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

digital models, the 3D decomposition is the most suited approach to handle the workload for used computational cores, as we increase the model size. Based on the algorithm, a parallel code is created that is executed on a supercomputer architecture. In this case, Intel Xeon Phi 7290 KNL accelerators are used, which are installed at the NKS-1P cluster of the SSCC SB RAS (http://www.sscc.icmmg.nsc.ru/). The software technologies that we used for code optimization are the MPI library that organizes data exchange between computers and AVX technology for code vectorization. The module of simulation modeling provided the tools needed to investigate the code for scalability and performance [17]. Results show over 90 percent scalability up to 32,000 cores but beyond it the performance starts to decline which means that the code is memory bound and needs further optimization.

4.2. The ontology for the problem of processing experimental seismic acquisition. Another important problem is interpreting the data obtained in the fieldwork from the observation system. The data recorded on receivers are scattered with much noise in the signal. So, processing and presenting the data in a usable form is an urgent task.

Commonly, the field data is packed in the SEG-Y format files that do not include seismic observation system parameters. On the contrary, three types of SPS files have their full description. While processing, samples are required from the CSP file (a common source point), the CRP (a common receiver point) and the CMP (a common midpoint). Index files are used to select the required gathers. They consist of trace numbers for each gathering. Usually, the input data are scattered over multiple files. The index files are, basically, a multi-key database of a dataset. This way of organizing the data is a novelty from one of the authors. The same technique and logic are used almost in all widely known commercial packages of geophysical processing, but there are no publicly available papers on this subject [24].

Figure 7 shows the ontological representation of one schematic procedure (the DMO processing stage) used in obtaining the so-called seismogeological section, which is an image of the medium reconstructed after processing the data from the field observation system. It is done with the aid of a ray-diffraction wave field model which is based on restoring the averaged velocity model, commonly via longitudinal waves (p-waves). Next, these velocity models are used to construct stacked seismic profiles. The seismic processing graph is a set of procedures required to reconstruct seismogeologial sections. It includes many components with feedback. All the components of the graph are used coherently one by one. A parallel algorithm and a code are created, optimized for the CPU. For this problem, the RAM demand is relatively small (<100GB) even for real-scale tasks, therefore it is sufficient to use one computing node of the cluster NKS-1P of SSCC SB RAS (a node with two eight-core Intel Xeon E5-2697A v4 processors and 128 GB of RAM). For a single node, the OpenMP package is sufficient to organize parallel execution of the code. The AVX technology is used for vectorization as well.

The graph of seismic processing was applied to real field data of one of the West Siberian regions. The reconstructed seismogeologial sections are presented in fig. 8 and fig. 9. The obtained snapshots are of practical value for professional researchers in geophysics and geology. It does not represent a "real" slice of the medium, since the horizontal axis is in milliseconds and the vertical axis is spatial. But one can see the multi-layered heterogeneous structure of the studied region. Similar images are used in industrial research when a company develops new oil deposits.

5. Results and Discussion. The general system that is being developed with the framework of this project has several distinctive features:

- Allows access to well-structured information on solving different problems of computational physics.
- Open source: allows not only to use the presented information but also update it.
- Provides the access to problems, methods and hardware architectures not presented in existing program packages.
- Enables to use the system for scientific and educational purposes.
- Free access and no specific licensing compared to commercial products.

The novelty of the research covered in this paper consists in developing a free-access resource that allows connecting two different areas: geophysics and supercomputers. The authors present a concept of the intelligent support system for solving geophysical problems on supercomputers. The ontological information representation in this subject area is described. The system incorporates algorithms and methods, useful software and available supercomputer hardware. The main purpose of the system is to save time and provide organized knowledge. The user can significantly reduce the time required to familiarize himself with a cer-

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

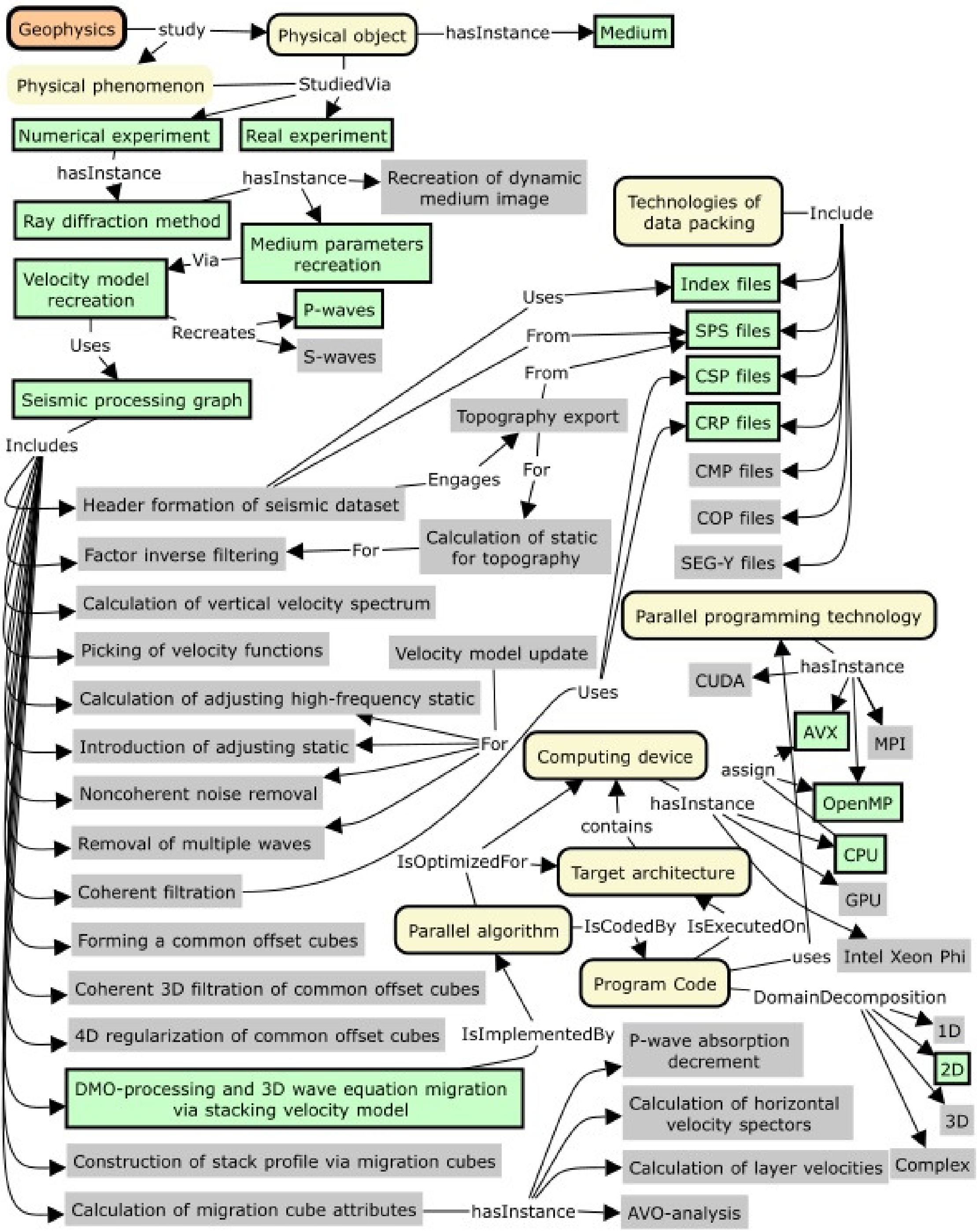

Figure 7. The ontology of the signal processing problem with the DMO-processing stage

2021, **22** (4), 333–343. doi 10.26089/NumMet.v22r422

Figure 8. West Siberian region. The stacked section after the prestack time migration

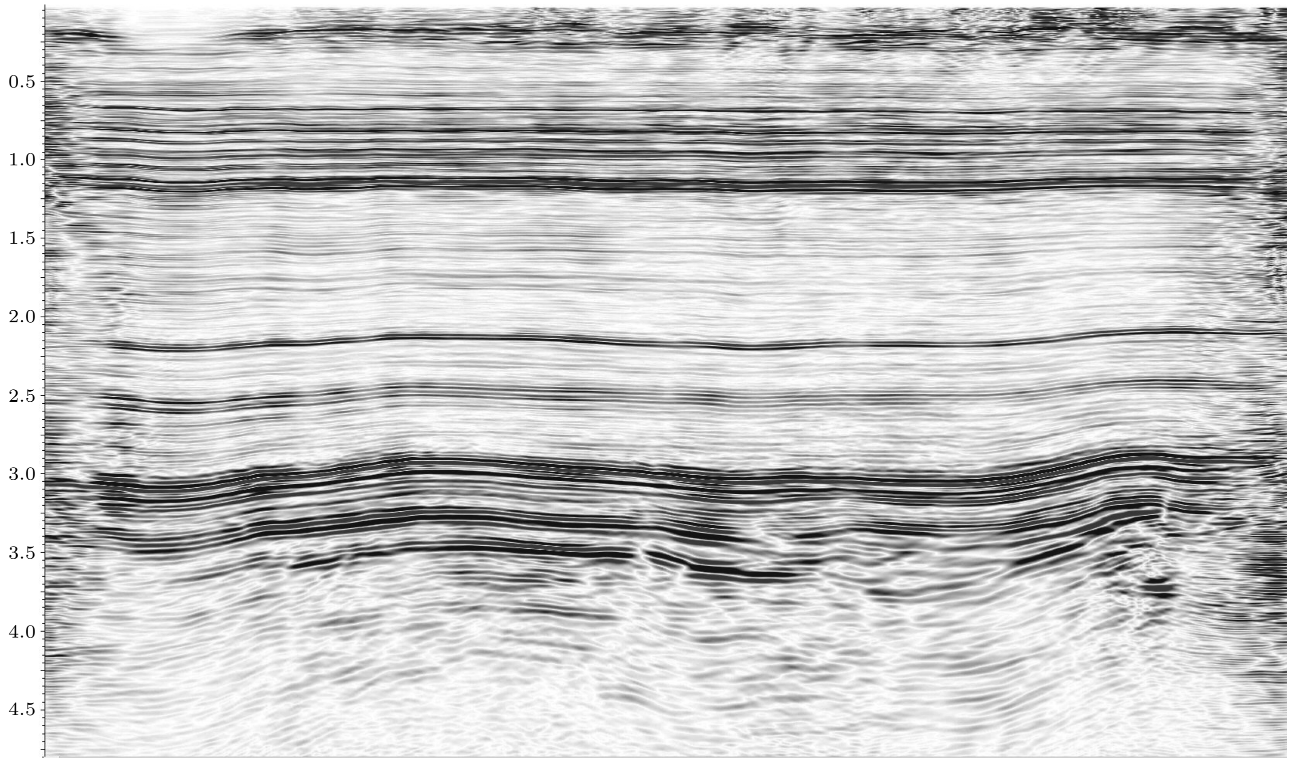

Figure 9. West Siberian region. The stacked section after the prestack depth migration

tain subject area and improve his knowledge base as a specialist and find the most suitable way to solve his scientific problem, depending on the features of the task. The authors also provide the ontological representation for two problems: 1) 3D modeling of the wavefield in areas with curvilinear surface and 2) reconstructing a 3D seismogeological section from data of one of the regions in West Siberia.

The internet resource https://uniserv.iis.nsk.su/spsc/ was created in the framework of this project. It provides an interactive interface for users and contains ontologies of various subject areas. It is routinely being updated with new information since the system is a "living" object.

References

1. M. Compton, P. Barnaghi, L. Bermudez, et al., "The SSN Ontology of the W3C Semantic Sensor Network Incubator Group," J. Web Semant. **17**, 25–32 (2012). doi 10.1016/j.websem.2012.05.003.
2. C. M. Keet, A. Ławrynowicz, C. d'Amato, et al., "The Data Mining OPtimization Ontology," J. Web Semant. **32**, 43–53 (2015). doi 10.1016/j.websem.2015.01.001.
3. V. Cvjetkovic, "Web Physics Ontology: Online Interactive Symbolic Computation in Physics," in *4th Experiment and International Conf. (exp.at'17), Faro, Portugal, June 6–8, 2017* (IEEE Press, New York, 2017), pp. 52–57. doi 10.1109/EXPAT.2017.7984405.
4. L. M. Sarro and R. Martínez, "First Steps towards an Ontology for Astrophysics," in *Lecture Notes in Computer Science* (Springer, Heidelberg, 2003), Vol. 2774, pp. 1389–1395. doi 10.1007/978-3-540-45226-3_188.
5. B. Glinskiy, Yu. Zagorulko, G. Zagorulko, et al., "The Creation of Intelligent Support Methods for Solving Mathematical Physics Problems on Supercomputers," in *Communications in Computer and Information Science* (Springer, Cham, 2019), Vol. 1129, pp. 427–438. doi 10.1007/978-3-030-36592-9_35.
6. A. Sapetina, B. Glinskiy, and G. Zagorulko, "Content of Ontology for Solving Compute-Intensive Problems of the Cosmic Plasma Hydrodynamics," J. Phys.: Conf. Ser. **1640** (1) (2020). doi 10.1088/1742-6596/1640/1/012019.
7. D. L. Cook, M. L. Neal, F. L. Bookstein, and J. H. Gennari, "Ontology of Physics for Biology: Representing Physical Dependencies as a Basis for Biological Processes," J. Biomed. Semant. **4** (2013). doi 10.1186/2041-1480-4-41.
8. D. D. Sokoloff and M. V. Remizowa, "The Use of Plant Ontologies in Comparative and Evolutionary Studies Should be Flexible," Am. J. Bot. **108** (6), 909–911 (2021). doi 10.1002/ajb2.1692.
9. X. Ma, *Ontology Spectrum for Geological Data Interoperability*, PhD Thesis (University of Twente, Enschede, 2011).
10. A. Mantovani, F. Piana, and V. Lombardo, "Ontology-Driven Representation of Knowledge for Geological Maps," Comput. Geosci. **139** (2020). doi 10.1016/j.cageo.2020.104446.
11. C. Wang, X. Ma, and J. Chen, "Ontology-Driven Data Integration and Visualization for Exploring Regional Geologic Time and Paleontological Information," Comput. Geosci. **115**, 12–19 (2018). doi 10.1016/j.cageo.2018.03.004.
12. L. F. Garcia, M. Abel, M. Perrin, and R. dos Santos Alvarenga, "The GeoCore Ontology: A Core Ontology for General Use in Geology," Comput. Geosci. **135** (2020). doi 10.1016/j.cageo.2019.104387.
13. U. Visser, H. Stuckenschmidt, G. Schuster, T. Vögele, "Ontologies for Geographic Information Processing," Comput. Geosci. **28** (1), 103–117 (2002). doi 10.1016/S0098-3004(01)00019-X.
14. F. Reitsma, J. Laxton, S. Ballard, et al., "Semantics, Ontologies and EScience for the Geosciences," Comput. Geosci. **35** (4), 706–709 (2009). doi 10.1016/j.cageo.2008.03.014.
15. Yu. A. Zagorulko, G. B. Zagorulko, and O. I. Borovikova, "Technology for Building Subject-Based Intelligent Scientific Internet Resources Based on Ontology," Soft. Eng. **7** (2), 51–60 (2016). doi 10.17587/prin.7.51-60.
16. B. Glinskiy, A. Sapetina, V. Martynov, et al., "The Hybrid-Cluster Multilevel Approach to Solving the Elastic Wave Propagation Problem," in *Communications in Computer and Information Science* (Springer, Cham, 2017), Vol. 753, pp. 261–274. doi 10.1007/978-3-319-67035-5_19.
17. P. Titov, D. Weins, and I. Chernykh, "Application of an Integral Approach to the Parallel Algorithm of 3D Wave Fields Simulation in Generalized Coordinates," in *Proc. Int. Conf. on Russian Supercomputing Days, Moscow, Russia, September 23–24, 2019* (MAKS Press, Moscow, 2019), pp. 14–25.
18. G. Antoniou and F. van Harmelen, "Web Ontology Language: OWL," in *Handbook on Ontologies* (Springer, Berlin, 2004), pp. 67–92. doi 10.1007/978-3-540-24750-0_4.
19. I. Horrocks, P. F. Patel-Schneider, H. Boley, et al., "SWRL: A Semantic Web Rule Language Combining OWL and RuleML," https://www.w3.org/Submission/SWRL/. Cited December 14, 2021.
20. N. Belaid, Y. Ait-Ameur, and J.-F. Rainaud, 2009. "SWAPT: Semantic Workflow Architecture for Petroleum Technology," https://www.scitepress.org/Papers/2009/18401/18401.pdf. Cited December 14, 2021.

21. V. Parekh, J. Gwo, and T. Finin, "Ontology Based Semantic Metadata for GeoScience Data," in *Proc. Int. Conf. on Information and Knowledge Engineering, IKE'04, Las Vegas, USA, June 21–24, 2004* (CSREA Press, Las Vegas, 2004), pp. 485–490.
22. P. Titov, "The Simulation of 3D Wave Fields in Complex Topography Media," in *Communications in Computer and Information Science* (Springer, Cham, 2019), Vol. 1129, pp. 451–462. doi 10.1007/978-3-030-36592-9_37.
23. V. Liseykin, *Difference Meshes. Theory and Applications* (Ross. Akad. Nauk, Novosibirsk, 2014) [in Russian].
24. R. E. Sheriff and L. P. Geldart, *Exploration Seismology* (Cambridge Univ. Press, Cambridge, 1995). doi 10.1017/CBO9781139168359.

Received
November 8, 2021

Accepted for publication
November 28, 2021

Information about the authors

Boris M. Glinskiy — Dr. Sci., Chief Researcher, Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences, Lavrentieva prospekt, 6, 630090, Novosibirsk, Russia.

Gennadiy F. Zhernyak — Ph. D., Leading Engineer, Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences, Lavrentieva prospekt, 6, 630090, Novosibirsk, Russia.

Galina B. Zagorulko — Ph. D., Researcher, A. P. Ershov Institute of Informatics Systems, Siberian Branch of Russian Academy of Sciences, Lavrentieva prospekt, 6, 630090, Novosibirsk, Russia.

Pavel A. Titov — Junior Researcher, Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences, Lavrentieva prospekt, 6, 630090, Novosibirsk, Russia.