Research Article

Pavel Kaygorodov, Nikolay Skvortsov, Dana Kovaleva, and Oleg Malkov*

A new version of the binary star database BDB: Challenges and directions

https://doi.org/10.1515/astro-2022-0215
received December 15, 2022; accepted January 22, 2023

Abstract: In accordance with the principles of open science, the results of the research should remain available for further investigations and practical interaction with them in the research community. Such results include collected initial and obtained data, specifications and implementations of methods applicable to them including data integration rules, and other possible resources. At the same time, the results of research on data should be curated to maintain their relevance, compliance with domain development, and standards in order to make them as useful as possible in the long term. A research support system in the research domain of physics of stellar systems is being developed to provide relevant data and methods, classify them, make them findable, avoid multiple integration of heterogeneous data, and integrate them into ongoing research in the astrophysical community. New catalogues are being added on regular basis, and this process is accompanied by correct cross-identification of the entities. New observational types of binary stars with their features are added to binary star database (https://bdb.inasan.ru).

Keywords: data quality, principles of FAIR-data, formal specifications, data reusability

1 Introduction

Research in communities of specific disciplines is based on obtaining observational and secondary data and extracting new knowledge from available data using available and developed methods. Different research groups repeatedly use them in different combinations for different research directions. The bottleneck of most studies is a set of heterogeneous data sources that have to be integrated and reconciled with each other at the level of their representations, conditions, and observed entities. Thus, the efforts of open science aimed at the availability of research data to researchers face not as much their inaccessibility as non-triviality of using them in further research.

Resources developed on the principles of open science should provide not only research data but also support different sides of using data in further research, reproducing the same or similar results using available methods and data. In accordance with the principles of open science, the results of the research should remain available for further investigations and practical interaction with them in the research community.

Source data and results of research on data should be continuously curated to keep them relevant to contemporary scientific results and demands, compliant with domain development and standards in order to make them as useful as possible in the long term. For this purpose, interoperability of data with domain standards and supporting potential users of data with necessary tools for solving problems appearing in the domain using data and previous results are necessary parts of open research resources.

In the domain of stellar system astrophysics, the creation of resources collecting data on binary and multiple stellar systems from various sources became the results demanded by researchers. Pairs of stars and whole stellar systems were matched and identified among all data sources. These results were published as two open resources: the list of binary star identifications (ILB, Skvortsov *et al.* 2017) and the database of binary and multiple stars (binary star database [BDB], https://bdb.inasan.ru, Kovaleva *et al.* 2015a).

Today, the result of that work requires updating and development in accordance with the demands of the community formed by these resources. Over the past few years, new instruments and observations have become available

* **Corresponding author: Oleg Malkov,** Department of Physics of Stellar Systems, Institute of Astronomy, 48 Pyatnitskaya St., Moscow 119017, Russia, e-mail: malkov@inasan.ru
Pavel Kaygorodov: IT Department, Institute of Astronomy, 48 Pyatnitskaya St., Moscow 119017, Russia
Nikolay Skvortsov: Department of Methods and Tools for Data Acquisition and Processing, Federal Research Center "Computer Science and Contro" of the Russian Academy of Sciences, 44-2 Vavilov St., Moscow 111333, Russia
Dana Kovaleva: Department of Physics of Stellar Systems, Institute of Astronomy, 48 Pyatnitskaya St., Moscow 119017, Russia

in astrophysics of stellar systems, new known binary stars of different types have been identified, possible ways have appeared for refining the structure of some stellar systems, and the range of interests and problems being solved in this domain has expanded as a result of the formation of an interacting research community. Therefore, in order to maintain the relevance of the created resources, support and expand the community using the database of binary stars for research, it is necessary to significantly develop the BDB project itself.

This article is intended to analyze the challenges and directions of the development of the BDB project from the perspective of the open science principles and to propose ways to solve the target tasks as a result of this analysis. The development of BDB should allow continuous curation for its relevance to the domain research and support of the research community using BDB. The article is divided to sections as follows. The next section is dedicated to the state of art in open science in general and in astronomy/astrophysics particularly. Section 3 describes the database of binary and multiple stars BDB and the demands of the research community it usually meets. The requirements to BDB development are analyzed in Section 4. In Section 5, the new approaches to the interface development are presented. The applied problems are discussed in Section 6, and Section 7 contains conclusions.

2 Related research

The movement of open science focuses on open publications, research data, software, research infrastructure, and other directions. The desire for open science involves organizing free access to collections of research data, methods and tools of data processing and analysis for research, as well as providing access to research results, including tools for reproducing them, and the ability to use them in other research. Research communities often use multiple sources of available observational data as well as results contributed during different investigations. They cannot pay too much attention to data heterogeneity by reducing time required directly to the research. Shared tools, standardized formats, interfaces, and domain-specific languages are developed to manage research data and to provide their usage in domain communities.

Communities of researchers such as FORCE11 and GO FAIR made significant contributions and promoted the establishment of an open science. FORCE11 focuses on digital publications including data, references, software, models, protocols, workflows, and other resources. It participates in discussing the guiding principles FAIR-data (Wilkinson *et al.* 2016) gathering summaries of data curation and preservation practices. They declare that data and other resources related to them should be well-identified, semantically defined with shared vocabularies and ontologies, accompanied by provenance information, comply with known protocols, standards, and data models used by domain communities, and have clear access rules. Open resources in data-intensive science should support researchers and machine agents with discovering and accessing relevant data, and then integrating and reusing them. So, the FAIR principles are becoming the basis of open science.

GO FAIR (https://www.go-fair.org/) community promotes following these principles in research projects and providing technologies to implement the principles. The community participates in research and development of the European Open Science Cloud (EOSC, https://eosc-portal.eu/sites/default/files/eosc_declaration.pdf), which is created to collect research data resources and provide services for making them finable, accessible, interoperable, and reusable. EOSC supports the conception of FAIR Digital Objects as identifiable containers for data, metadata, and any resources of different kinds related to research data. Research projects should show their conformance with FAIR, make data management plans for sharing data, and register their results in EOSC. EOSC can coordinate and support the federation of various data of data infrastructures in different disciplines. Multidisciplinary research infrastructures like EOSC typically provide basic services for long-term preservation of research data repositories, sharing, searching, and access for reusing, and they can have workspaces for data-intensive research, domain-specific services, and computing resources.

Independent or domain-specific research data repositories can be created and evaluated for conformance to FAIR. GO-FAIR collects a network of best practices of implementation of FAIR and describes projects in different research domains. The FAIRsFAIR community explains the provisions of FAIR, their purposes and consequences on the simple understandable level for wide data producers, and provides assessment metrics, which can be used for trivial evaluation of any repositories and collections of data.

In astronomy/astrophysics as a research domain, all collected historical data remain relevant and valuable for analysis with the latest obtained data. There are open sky surveys and catalogues of observational data collected over time with different instruments, and databases of specific purposes that are used in conjunction to obtain full history and enrich data of observation of astronomical objects. The concept of a virtual observatory implies that different astronomical data resources are integrated into a single environment together with data analysis tools and a computing service. Long before the dawn of the concept of research infrastructures, virtual observatories providing access to open data sources became a basis of research. The International Virtual Observatory Alliance (IVOA, www.ivoa.net) has developed formats, domain-specific languages, and standards for interoperability of astronomical data.

Unfortunately, global identification systems are not popular in astronomy. Data centers and repositories of astronomical catalogues use own identification systems for registered catalogues. Each catalogue can use own identification systems so that known astronomical objects can have a set of identifiers of different catalogues. Some catalogues contain several fields with references in particular identification systems. Even though most of the identifies are based on positional information about objects, researchers often have to cross-match objects in several catalogues not by identifiers but by positional and other observational parameters.

Data interoperability is supported by widely used standard formats (such as FITS, VOTable). These formats encapsulate data and rich astronomical metadata. There are semantic specification standards, thesauri, and conceptual models. Unified content descriptor (UCD) is the most commonly used controlled vocabulary for semantic annotation of fields in astronomical catalogues. It does not exclude description ambiguity and variance but allows expressing general astronomical concepts for semantic annotation metadata. Conceptual models standardize structures of the most common subdomains of astronomy such as coordinate systems, photometry, spectroscopy, registered astronomical events, and others. They include general elements for objects and characteristics used in these subdomains. However, they are not usually used for data representation, but their names are used for semantic annotations in some catalogues. ProvenanceDM is based on the W3C PROV-DM provenance model that defines relations between entities, actions, and agents. Observation Core Data Model (ObsCoreDM, Louys *et al.* 2017) can be perspective for FAIR representation of observational data in astronomy. It joins UCDs, features of several standard conceptual schemes defining observation concepts, and provenance metadata and allows querying them simultaneously.

Strasbourg Astronomical Data Center (CDS, http://cdsportal.u-strasbg.fr/) has built a research data infrastructure with unified access to astronomical catalogues VizieR (http://vizier.u-strasbg.fr/viz-bin/VizieR), interactive atlas of the sky Aladin (http://aladin.cds.unistra.fr/aladin.gml), and the database of astronomical objects SIMBAD (http://simbad.u-strasbg.fr/simbad/). It supports access points for specialized query language (ADQL), the querying protocol (TAP), and other tools for data analysis. Although the interfaces are unified, catalogue structures remain heterogeneous. Representation of some fields such as coordinates and magnitudes is standard and easily integrated using multi-source queries. However, various catalogues have heterogeneous formats of fields and values intractable for automatic processing, and some catalogues have internal value heterogeneity.

Astronomical community has been intensively using these open resources and ideas of virtual observatories for many years despite the difficulties with object identification, heterogeneity problems, and the lack of semantic approaches. But the existing implementations of a virtual observatory are successful only in certain domain characterized by uniformity of observations and data and are stalled in domains with complex and ambiguous relationships between objects and observations. Astrophysics of stellar systems provides an example of such a difficult domain to implement with existing tools due to the heterogeneity of data on different observational types of binary objects, the complexity of cross-matching of objects, the lack of a unified system of designations, *etc.*

3 The BDB

Data on binary and multiple stellar systems can be obtained from a number of astronomical catalogues and databases. However, the well-known database of astronomical objects SIMBAD does not contain binary star parameters among the search criteria, and not all binaries (especially visual) are included in it. Another well-known service of the Strasbourg Astronomical Data Center, the VizieR catalogue database does not contain all the important catalogues of binary stars. The authors of numerous catalogues and databases of binary stars of certain observational types provide information only about the particular type of binary, usually not caring about cross-identification with other sources. The multiple systems catalogue (Tokovinin 2018) contains only systems with multiplicity three and higher and only objects with known stable orbits. A large number of catalogues and databases containing information about the parameters of binary stars of various types, including visual ones (WDS, http://www.astro.gsu.edu/wds/), orbital (Chulkov and Malkov 2022), interferometric (INT4, http://www.astro.gsu.edu/wds/int4.html), eclipsing (Avvakumova and Malkov 2014), spectroscopic (SB9, http://sb9.astro.ulb.ac.be/), as well as some others.

The BDB (Kovaleva *et al.* 2015a) was developed as a data collection that includes information about all types of binaries, and unified presentation, and complete information about multiple systems. It is currently the world's main source of data on binary and multiple systems of all observational types (visual, spectroscopic, eclipsing, *etc.*) and it contains astrophysical, photometric, and positional parameters of hundreds of thousand components of stellar systems of multiplicity from 2 to 20 and more. Information about these types of binary stars is included in the BDB from heterogeneous data sources: astronomical catalogues and surveys. By solving the complex problem of cross-identification of the components of double and multiple stars, the BDB provides previously unachieved connectivity in the domain of double stars. The use of BDB in solving various astrophysical and stellar astronomy problems makes it possible to synthesize data from heterogeneous sources, opening new opportunities for researchers. The BDB web-page is shown in Figure 1.

For all its originality and versatility, BDB is created using the astronomical data resources already available in the world. To build BDB, the experience of the well-known and widely used SIMBAD, VizieR, and their approaches, standards, and formats were taken into account, but catalogues that are not included in VizieR were also integrated. Links from BDB to full descriptions in binary star catalogues, surveys, and general-purpose databases were provided. Existing identification systems, identifiers, and cross-identification algorithms were insufficient for resolving stellar systems and encountered a large number of problems of cross-identification of components found in SIMBAD and original catalogues.

To match and identify multiple stellar systems and all their components and pairs, a special algorithm taking into account positional and observational data from various sources was developed and the identification list of binaries (ILB, Kovaleva *et al.* 2018) was created. Special identification system (BSDB, Kovaleva *et al.* 2015b) was registered, and every stellar system and its components were assigned with identifiers in single system. To keep provenance of data on formed stellar systems, ILB links BSDB identifiers to all identifiers of all participating catalogues. Most representative catalogues of binary systems and several single star surveys were connected to ILB.

BDB has become a popular tool in the research community in the domain of stellar system physics for further research of binary and multiple stars. Now curation of BDB encounters some typical challenges and demands from researchers who work in the domain and solve different research problems for timely analyzing and updating stellar systems with recent data in the domain, studying various characteristics and behavior of stellar systems.

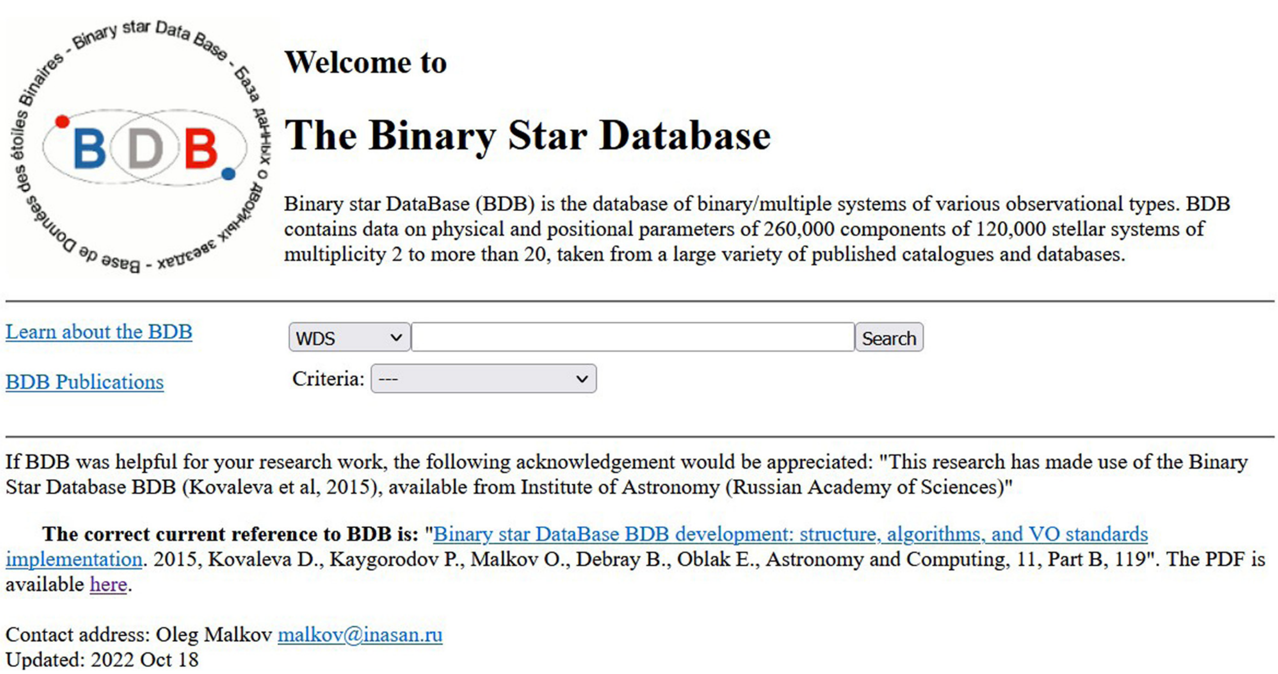

Figure 1: The BDB webpage.

First of all, binary systems inserted in the catalogues were up to date, but the possibility should always be open for the new binary catalogues and surveys and new approaches discovering stellar systems. Binary and multiple stars make up, according to various estimates, from 50 to 90% of the stellar population of the Galaxy, and provide researchers with heterogeneous and extensive data for investigation of the physics and evolution of stars in general. Currently, thanks to the development of observational capabilities and the publication of large surveys, and new level of computing technologies that have made possible the efficient processing of large amounts of data, it has become possible to bring the study of the population of binary stars to a new level.

New sky surveys make it possible to detect new binary and multiple star systems. In three published releases of the Gaia sky survey, a significant number of binary and multiple stars with a common proper motion have been found. Some of these stars were known earlier, and some systems were partially known. Each of the surveys has its own identification type. The Gaia catalogue of binary and multiple stars is published in 2022. Unlike catalogues of stars with a common proper motion, it will contain mainly information about close pairs. In addition, there is a flow of information about binary stars (mostly eclipsing) from surveys related to photometric variability monitoring. A certain amount of new data for inclusion in the project can also be expected from spectroscopic surveys.

Publication of new catalogues and updates of the versions of registered catalogues require new source integration, and even domain development including enhancement of canonical data schema.

Solving research problems using BDB needs flexible data processing, stating data processing tasks, making complex queries with entire set of types and attributes in them, batch analysis of stellar systems, and support for domain standards, mostly IVOA ones.

It is necessary to develop utilities that serve more effective interaction of specialists in the community, and interaction of the BDB with the world archives of astronomical data and the tools of the IVOA.

New BDB should become a research support system, equipped with the necessary data and tools. The creation of the system to support scientific research in physics of stellar systems will not only make the data more accessible, the results of previous work reproducible and interoperable, but also allow continuing intensive research and develop research applications for the study of astrophysical characteristics of stellar systems. It can significantly increase the efficiency and productivity of research.

4 Proposed approaches

The central part of the system is the conceptual schema of the research domain that should be extended for representation of information on objects of subject domains, research problem statement, and research application development using the research support system. Semantic integration of multiple relevant data sources is based on their mapping into the conceptual schema. Data manipulation is performed with data from repositories and collections transferred into the conceptual schema. During transformation of source data into the conceptual representation their structure is changed, and entities from different sources are matched with each other and with previously identified entities. Reuse of scientific results obtained earlier is based on their publishing using the conceptual model of the subject domain. In particular, data are mapped into the conceptual model, and methods are bound with such data. Access to data of subject domain is also provided using the conceptual model of the domain. Research problems and applications are developed over the conceptual model of the subject domain independently on the data sources. Results of research including data, methods, and programs and results of data integration and transformation between data representation standards should be published using their descriptions in the conceptual model of the subject domain required for their search and reuse.

We are going to reuse and enhance the results obtained during the development of the database of binary and multiple stars (BDB) that includes the conceptual model of binary stars, algorithms of multiple system cross-matching, the database of known components and pairs, the identification list of binary stars, mappings of the integrated catalogues into the conceptual model, and the user interface structure.

Extending the domain of the astrophysical research of stellar systems leads to the definition of new concepts and development of the conceptual schema of the domain. It should include all essential observational and astrophysical parameters of systems, pairs of different kinds, and components, knowledge about their dependencies, identification criteria and possible behavior.

The support of research should include tools for integrating data sources based on their semantic mapping into the conceptual model, management of data on stellar systems using the conceptual model, required interfaces for the development of the research applications, support of embedding the system into the virtual observatory, and others.

Data structures in astronomy should be annotated in terms of the UCD controlled vocabulary. The UCD describes elements of data source schemas to be integrated in the research support system. The methods for the integration of the heterogeneous data sources include unification of the source data models in the canonical data model (to overcome data model heterogeneity), and the source schema mapping into the conceptual schema of the subject domain to overcome schema heterogeneity. Though astronomical catalogues in general utilize the relational data model, the data model mapping can be required for the integration of virtual observatory standards and for the mapping into the object data model. Schema mapping in its turn may be complicated due to the features of data semantics and specific usage of similar attributes in different catalogues. Generally speaking, the application of methods for the heterogeneous data sources integration allows us to transform data from heterogeneous sources into the uniform representation in terms of subject domain specifications for their analysis during research investigations.

Enrichment of data on stellar systems in BDB with new observational data requires development of the algorithms of the identification of compound entities and their components providing correct modification of previously described systems with new components, relating new data to known structures, discovering new systems, and binding correct identifiers to systems, pairs, and components.

The methods for the entity resolution including the algorithm for the resolution of multicomponent entities characterized by different sets of parameters are applied for the cross-match of stellar systems, pairs, and their components. It is assumed that the same entity can be considered to be of different types in different catalogues. Systems, pairs, and components can be absent in some catalogues or be added when new data are published. Catalogues may contain errors of different nature. During the update of data from new sources, it is required to estimate the relevance and completeness of data.

The methods for the identification of entities of different types and for the establishment of their identifiers with each other are developed having in mind requirements for matching identifiers of the same systems, pairs, and components in different catalogues. Astronomical catalogues are datasets having global identifiers for their entities.

Every entity (a system, a pair, a component) can have different identifiers in different catalogues. Catalogues refer to identifiers of the same entities in other catalogues. Note that such identity is often established incorrectly due to the different interpretations of entities. Establishing correct relationships between identifiers is based on cross-matching of entities. It is also required to keep persistent identifiers when new data on stellar systems are published.

The previously implemented algorithm does not completely resolve the structural conflicts of the high-multiplicity system and requires an analysis of the system of universal matching criteria associated with components and pairs. But the main reason for the need to enhance the algorithm for identifying entities and principles of work with identifiers is related to obtaining new knowledge about objects already included in the database. For example, when new data apper, an object that had been previously considered a component may turn out to be a pair, which entails correcting the type, attributes, linking previously known data with the pair as a whole or with its components. Also, the appearance of new data requires the introduction of new types of identifiers into the identification system. For example, Gaia identifiers are different in each data release (three have been published so far), and in the course of identifying objects of known systems with Gaia, it is necessary to solve the problem of matching the Gaia source to known systems, pairs or components, as well as to supplement the database with information about previously unknown components, pairs, and systems. These new entities will need to be assigned new BSDB identifiers and organize a search for alternative identifiers in the main identification systems. In June 2022, the Gaia catalogue of binary and multiple stars is published, which contain much information about close pairs. It will also require a change in the principles of identification of close pairs, which until now have been quite trivial.

Development of the methodology of publishing of data, developed methods, and results should make them accessible for further reuse after research applications having finished their work. New source data should be registered and related to existing data, the results should be reproducible, and the methods should be reusable for the new observational data and for solving the derivative problems.

Wide support of subject domain standards allows us to register research results, provides for their accessibility and interoperability. In astronomy and astrophysics IVOA standards on conceptual schemas, data formats, protocols, and interfaces should be applied for the integration of source data and publication of results. The following standards recommended by the IVOA are considered to be applied within the project:
- ConeSearch protocol for data search describing sky position and an angular distance;
- Table access protocol (TAP) for the access to relational data;

- UCDs controlled vocabulary for the description of data semantics in astronomy;
- IVOA provenance data model;
- VOTable format for the representation of the output data as tables and metadata headers;
- Universal worker service pattern for the persistent batch queries;
- VO registry for registration of developed catalogues and databases.

The new version of BDB is still in development and currently queries with VO protocols are not implemented enough to show them in the article. Note that the well-known atomic and molecular database VAMDC (Ryabchikova *et al.* 2011, Albert *et al.* 2020) also complies with IVOA protocols.

The construction of a system of scientific research based on the principles of open science makes it possible to provide the community of researchers in physics of stellar systems with sufficiently complete data on stellar systems for today, the results of work on the formation of stellar systems based on information from disparate data sources on pairs of different observational types, as well as to provide the tools to solve problems in the domain and the possibility of publishing the results. The creation of such a system and, in general, the principles of building such systems are very relevant. On the one hand, the system uses a popular data resource in the domain of stellar systems. On the other hand, today in many parts of the world it is required to provide research on the principles of open science.

5 An approach to the interface development

Currently, the BDB can be accessed with web browser and (work in progress) via REST API over HTTP protocol. We are using a traditional architecture with backend server, written in Python and a frontend, written in (mostly autogenerated) HTML and JavaScript. The backend is using a PostgreSQL database to store all data.

The BDB project is interdisciplinary, relating not only to the astronomy but also to the computer science, so, we are trying to push forward new methods and approaches in interface development. Looking backward in years, we can see what early WWW (Web 1.0) was just a structured collection of linked documents. One can read them and navigate between them with hyperlinks. The Web 2.0 changed the paradigm, adding database capabilities to the WWW, so, now, the typical website (including the current version of the BDB) is just an interface to a database and shows some formatted data records, fetched from it, allowing users to perform operations over the database, like selecting data by some criteria, inserting new records (typically by making new posts and comments), and modify and (sometimes) delete the stored data. The typical architecture of Web 2.0 site is based on Model-View-Controller (MVC) architectural pattern.

But now we see the outlines of a new paradigm (may be, it will be called Web 3.0 after a while) where a site will be more like an application – it is running in a browser pushes a significant amount of logics to the client side, using servers just as a source of data to process and display. The typical examples are Facebook, Discord, Slack, Skype, and other web servers, where almost the entire interface is dynamic and controlled by local JavaScript (JS) code. Even their desktop and mobile applications are just JavaScript applications, sharing most of the code with the web versions. This approach has lot of advantages – the interface interactions become more smooth and the very complicated interfaces can be realized with less efforts compared with MVC approach.

Looking forward, we are planning to investigate the advantages of the site-as-application model and starting to develop a new version of the BDB in this approach. We want to completely eliminate the use of HTML in the web-interface and construct it directly by adding and manipulating with DOM elements directly by JavaScript functions.

To make the conception more general, the browser just establishing a WebSocket connection with the server (backend) and executes all the JavaScript code, transferred from the server, returning results back. This will allow us to neglect any APIs or protocols during the communications and make it very flexible. Using this way, we can define new functions on the browser side to call them later, and they can call each other, realizing any interface logic on the browser side. To be even more general, we can use a specific programming language, which can be executed on the backend and partially can be translated into the JavaScript, to be executed in the browser.

The Common Lisp (CL) is very flexible and mature language, having a lot of libraries and supporting a number of architectures. There exists a JSCL library (https://github.com/jscl-project/jscl) for the Common Lisp, which is a self-hosted Lisp-to-JavaScript compiler, translating (with some limitations) plain CL code into JS. Based on JSCL, we developed the OMGlib library (https://github.com/hemml/OMGlib), which is doing all low-level job, like maintaining connections with browsers, compiling a code into JS, *etc*. The Lisp macros are used to make a browser-side code execution seamless – we are just defining a new function with defun-f instead of standard defun and it will be executed in the browser when called. All the parameters are implicitly serialized before the call using a standard Lisp writer and unserialized on the other end with Lisp reader; the result of the execution is serialized/unserialized too. If the browser-side function calls another browser-side function, it will be called directly within the browser, without any interaction with the backend. If the browser does not have a code for the called function, it will query it from the backend for the first time and use this code later. If we are redefining a browser-side function on the backend, it instructs all of the connected browsers to forget the old code and the function will be requested again when (if) called. This significantly simplifies the development, because we can define new functions and redefine old ones "on the fly," without page refresh and with no backend restarts, seeing the result immediately.

OMGlib allows us to work with code via Read-Evaluate-Print-Loop (REPL), but all of the changes are applied immediately, so, we need two copies of our code – local development version and production one, which is updated less frequently. Each production code update causes a server restart, so all connected clients will lose their connections and will have to reconnect. After the reconnection, browser still has functions from previous version, which may lead to problems.

OMGlib offers a better way to maintain version updating. The special daemon omgdaemon is spawned and works as a reverse-proxy, which accepts HTTP connections and, based on special cookie, connects clients to specific versions. Here is one special version with name "devel," where developer(s) can do any development works, using remote connection via traditional lisp swank library. When the version becomes good enough to be pushed into production, the (commit-production) function is called to save a lisp image to disk with a unique version name. After that, all new connections will be redirected to this version and new "devel" version will be spawned again. If there were previous versions with clients connected, they will receive notification (the (commit-notify new version function will be called on each version running)) so they can offer version update to the clients, but clients may not be enforced to immediate update.

6 Applied problems

Applied problems in the domain of astrophysics of stellar systems can be solved using developed tools for research support. For instance, the problems to be solved include determining astrophysical parameters of stars, constructing fundamental relations between stellar parameters, and determining the maximum multiplicity of hierarchical stellar systems.

For instance, accurate direct determination of the masses and other physical parameters of as many stars with different characteristics as possible (different mass ranges, different types of binary – wide and close, *etc.*) is a task that has not lost relevance for decades. The use of BDB promises a qualitative leap based on the most effective use of the data already obtained, which has not been analyzed comprehensively before. Such a task becomes especially interesting and relevant in the perspective of integrating data from the Gaia space mission.

Binary stars are very common; perhaps 90% of stellar systems are binaries; however, there are still many puzzles to solve about their formation, structure, and evolution. Binaries provide accurate data that are very difficult to determine for single stars, such as masses and radii: the direct determination of the mass of any star requires measurable gravitational interaction between at least two stars. Other properties of binary stars (and advantages of their studying) are the following. The components of binaries have the same age (useful for the tests of theoretical models of evolution), were born at the same place and time (useful for the tests of chemical abundance), are at the same distance (useful for the tests of distance calibrations), and suffer the same interstellar extinction (useful for the tests of interstellar reddening).

The relation between the mass of a star and its luminosity on the main sequence (mass-luminosity relation) is a fundamental law that is used in various fields of astrophysics. For single objects, the mass–luminosity relation allows astronomers to convert a relatively easily observed quantity, luminosity, to a more fundamental characteristic, mass, which yields a better understanding of the object's nature. It is especially important for the construction of the initial mass function from the luminosity function of stars. Independent stellar mass and luminosity determination is possible only for components of some types of binary systems. Visual binaries with known trigonometric parallax provide masses of (mainly low-mass) components. Another main data source is detached main-sequence eclipsing binaries, with the spectrum lines of the two (mainly intermediate- and high-mass) components.

BDB contains data on all observational types of binaries and, consequently, will allow us to properly match the two mass–luminosity relations and also to compare theoretical mass–luminosity relations with empirical data.

These considerations also apply to mass–radius and other relations.

BDB contains data on very high multiplicity (up to 20 components and more), though not necessarily hierarchical, systems, including moving groups and (mini) clusters. To explain inconsistency in maximum stellar multiplicity between rather high grades predicted as possible by theoretical considerations and observational lack of systems with multiplicity higher than six, we will carry out a detailed study of the multiple systems contained in BDB. The information contained in the database will allow us, in particular, to eliminate (known or assumed from the probability filter) optical pairs, and add photometrically unresolved sub-components.

In addition to the tasks described above, BDB will be used to study the evolution of binary and multiple stars (in particular, close pairs, undergoing matter exchange and, as a consequence, emitting a strong X-ray radiation), to study merger processes of massive compact binary components in the final stages of their evolution (sources of recently discovered gravitational waves), to plan observations of binary systems, *etc*.

7 Conclusion

The scope of tasks to be solved within the enhancement of the Database of binary and multiple stars is extensive. They include knowledge formalization in the conceptual model, methodological developments, software architecture development and implementation, integration with the existing software tools, working with multiple data sources, and fundamental problem solving in the domain. The principles of solving these tasks were described in the article. We also described the approach of site-as-application the development of the new version of the BDB (https://bdb.inasan.ru/).

Acknowledgement: The authors are grateful to our anonymous reviewer whose constructive comments greatly helped us to improve the article.

Funding information: The authors state no funding involved.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and have approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The data underlying this article will be shared on reasonable request to the corresponding author.

References

Albert D, Antony BK, Ba YA, Babikov YL, Bollard P, Boudon V, et al. 2020. A decade with VAMDC: Results and ambitions. Atoms. 8(4):76.

Avvakumova EA, Malkov OY. 2014. Assessment of evolutionary status of eclipsing binaries using light-curve parameters and spectral classification. Mon Not R Astron Soc. 444(2):1982–1992.

Chulkov D, Malkov O. 2022. Visual binary stars with known orbits in Gaia EDR3. Mon Not R Astron Soc. 517(2):2925–2941.

Kovaleva D, Kaygorodov P, Malkov O, Debray B, Oblak E. 2015a. Binary star database BDB development: Structure, algorithms, and VO standards implementation. Astron Comput. 11:119–125.

Kovaleva DA, Karchevsky AV, Skvortsov NA, Malkov OY. 2018. Catalogue of identifications of binary stars, ILB. In: Shustov BM, Wiebe DS, editors. Stars and Satellites, Proceedings of the Memorial Conference Devoted to A.G. Masevich 100th Anniversary. 2018 Oct 15–16; Moscow, Russia. INASAN Science Proceedings. p. 397–401.

Kovaleva DA, Malkov OY, Kaygorodov PV, Karchevsky AV, Samus NN. 2015b. BSDB: a new consistent designation scheme for identifying objects in binary and multiple stars. Baltic Astron. 24:185–193.

Louys M, Tody D, Dowler P, Durand D, Michel L, Bonnarel F, et al. 2017. Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1. IVOA Recommendation 09 May 2017.

Ryabchikova TA, Pakhomov YV, Piskunov NE. 2011. New release of Vienna Atomic Line Database (VALD) and its integration in Virtual Atomic and Molecular Data Centre (VAMDC). Kazan Izdatel Kazanskogo Universiteta. 153:61–66.

Skvortsov N, Kalinichenko L, Karchevsky A, Kovaleva D, Malkov O. 2017. Matching and verification of multiple stellar systems in the identification list of binaries. In: Kalinichenko L, Manolopoulos Y, Malkov O, Skvortsov N, Stupnikov S, Sukhomlin V, editors. Data analytics and management in data intensive domains. DAMDID/RCDL 2017. Cham: Springer; p. 102–112, doi: 10.1007/978-3-319-96553-6_8.

Tokovinin A. 2018. The updated multiple star catalog. Astrophys J Suppl Ser. 235:6.

Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, et al. 2016. The fair guiding principles for scientific data management and stewardship. Sci Data. 3(1):160018. http://www.nature.com/articles/sdata201618.