Elecrtonic version is available on
https://fmm.ru/Journal/NDM57_2024_eng
Fersman Mineralogical Museum RAS

New data on minerals, V. 58, iss. 2 (2024), 26–34

NDM

ArDI – the system of mineral vibrational spectroscopy data processing and analysis

Shendrik R.Y.[1], Plechov P.Y.[2], Smirnov S.Z.[3]

[1]*Vinogradov Institute of Geochemistry SB RAS, Irkutsk, r.shendrik@gmail.com*
[2]*Fersman Mineralogical Museum RAS, Moscow, pplechov@gmail.com*
[3]*V.S. Sobolev Institute of Geology and Mineralogy SB RAS, Novosibirsk*

ArDI (Advanced spectRa Deconvolution Instrument) is a web application for processing and analyzing vibrational spectra of minerals (https://ardi.fmm.ru/). The application is designed for the rapid and reliable identification of minerals in geological samples. ArDI allows users to process spectra, search for similar spectra in databases, and upload reference mineral spectra to the database for its expansion. ArDI has development prospects in several directions, including improving the ergonomics of the interface and algorithms for automatic processing of Raman scattering (RS) spectra, filling the database of reference spectra, and integrating the database of reference spectra with the information system of the Fersman Mineralogical Museum of the Russian Academy of Sciences and other information systems dealing with minerals. The ArDI toolkit can be used for rapid mineral diagnosis and interpretation of individual vibrational bands in spectra. It can be useful in mineralogy, raw material and cut gemstone expertise, as well as in medicine, pharmaceuticals, and forensics.

Keywords: ArDI, spectra processing system, Raman spectroscopy, Fersman Mineralogical Museum, reference collection, new mineral, mineralogy.

Introduction

One of the most important characteristics of a mineral is its crystal structure, which can be reflected in X-ray diffraction data and/or vibrational spectroscopy data. Among the methods of vibrational spectroscopy, the most popular are infrared spectroscopy (IR) and Raman scattering spectroscopy (RS, also known as Raman spectroscopy). Raman spectroscopy has been developing very actively in recent years due to the emergence of inexpensive but powerful LED lasers and sensitive CCD matrices. This has led to the development of compact instruments and has allowed Raman spectroscopy to become a routine method with several advantages over other common methods of studying matter (IR spectroscopy, X-ray diffractometry, etc.). One of the obvious advantages of Raman spectroscopy is that it can acquire spectra from microscopic inclusions, and many rare minerals have been found and described only in this form. Additionally, Raman spectroscopy does not require special sample preparation and can be applied even for the analysis of inclusions in minerals. The compactness of modern laser sources and detectors allows the creation of instruments that can be easily integrated into desktop microscopes and spectrometers, rapid diagnostic devices, unmanned remote sensing platforms, and so on. Despite several advantages of Raman spectroscopy over IR (localization, ease of measurement, high sensitivity to some extra-framework complexes, etc.), Raman and IR spectroscopy are complementary methods. Often, modes active in IR are inactive in Raman, and vice versa. Moreover, currently, instruments for local IR analysis using specialized microscopes are more accessible than Raman microscopes for similar tasks. New opportunities for the use of local IR spectroscopy are also opened up by the new generation of

synchrotrons. Therefore, it seems effective to develop universal software products for the processing of vibrational spectra without binding them to a specific method.

We anticipate that in the near future, the methods of vibrational spectroscopy will become the basis of the strategy for geochemical and technological research and development in a wide range of human activities, both on Earth and in the study of objects in the Solar System.

To describe the mineral diversity in the 19th and 20th centuries, numerous reference books and card indexes were created. To date, six thousand mineral species are known, and about a hundred new ones are added each year. The paper-based format is not very convenient for rapid updates, and it is also poorly suited for spectral information. The transition from paper to electronic media allows solving many problems, including contributing to the preservation of forests on our planet. Therefore, since the late 20th century, the mineralogical community has been engaged in creating databases that, with the development of information technology, are becoming more comprehensive. The goals of such databases and information systems are to accumulate knowledge about minerals and to assist in their rapid diagnosis.

Databases are being compiled by individual researchers and research groups, as well as by companies that produce spectroscopic instruments. The former are usually accessible to members of these groups, while the latter are provided with the software supplied with commercial instrument models. The problem with these two categories of databases is their "locality" and the limited sources of spectral information (usually from a single laboratory). A much more advantageous form is a public database, access to which on certain conditions is available to a wide range of specialists and users interested in the functionality of this complex. Moreover, a large number of spectra in reference spectral libraries have been obtained on old equipment and require a serious revision. As a result, the identification of minerals using Raman spectroscopy is still difficult - it requires very high qualifications or often leads to errors.

One of the first mineralogical databases of this type was the database of the California Institute of Technology (http://minerals.gps.caltech.edu/), created and maintained through the efforts of Professor G. Rossman. To date, this database contains data on the optical, infrared, Raman, and Mössbauer spectroscopy of minerals, as well as various related references. Later, the Raman spectroscopy data were transformed into the RRUFF database (https://rruff.info/) [Lafuente et al., 2015] - a joint project of the California Institute of Technology and Arizona State University, where they were combined with X-ray diffraction data. The RRUFF database has gained popularity in the mineralogical community due to the large amount of data and the availability of a tool for searching for the necessary information by the name of minerals, their chemical composition, and bibliographic data. Another important feature that has stimulated the popularity of this resource was the presence of a special application, Crystal Sleuth, which allowed calculations on user X-ray diffraction data and comparison of these data and Raman spectroscopy data with what is uploaded to the project database. In addition, the application allowed the application of basic Raman spectrum processing operations: automatic baseline subtraction, removal of outliers in the spectra, cutting or extracting part of the spectrum, and searching by it. The RRUFF database is being supplemented with new data and currently contains spectral data on 2351 minerals. However, the update of Crystal Sleuth was discontinued after 2009.

In 2017, another public database was created, similar to the RRUFF project, based on the combination of the capabilities of Raman spectroscopy and X-ray diffraction for the identification of various materials, including mineral species [El'Mendili et al., 2019]. It is called the Raman Open Database (ROD) (https://solsa.crystallography.net/rod/). The database is supported by teams from several academic institutes and universities in France, Italy, and Lithuania, and is integrated with the COD crystallographic database (http://www.crystallography.net). One of the areas of application of ROD is the development of a system that combines ultrasonic drilling and automated analysis of mineral and chemical composition, including Raman spectroscopy of the core. Such systems are especially in demand when studying the planets of the Solar System using unmanned devices. Currently, this database contains 1133 records. The ROD also includes a search tool that allows you to search both the database and the related bibliographic references. However, there are no ways to automatically compare user data with the already uploaded data, which does not allow using ROD as a convenient tool for mineral diagnosis. Despite this, participation in the SOLSA project (https://solsa-dem-up.eu/en) implies that the collected database should be part of an automated system that should identify minerals.

Review-reference books on vibrational spectroscopy in the form of a series of books are periodically published by N.V. Chukanov and his colleagues. The books [Chukanov, 2013; Chukanov, Chervonnyi, 2016; Chukanov, Vigasina, 2020] contain the most complete library of infrared spectra of minerals. Many of these spectra were obtained on reference samples of studies of new minerals from the collection of N.V. Chukanov and are of great value. In the book [Chukanov, Vigasina, 2020], in addition to new data on IR spectra, the characteristics of Raman scattering spectra (position of the main spectral bands)

of more than 2000 minerals are given, significantly supplementing the RRUFF database. Unfortunately, there are currently almost no open, convenient, and complete electronic databases of infrared spectra of minerals and tools for working with them.

There are also other projects that contribute to the accumulation of huge arrays of empirical data that are of great importance not only for the identification of mineral species, but also for solving more complex problems related to the determination of crystal structures, features of the mineral composition, and oriented not only to geological problems, but also important for the field of materials science, including for the creation of modern high-tech devices. It is necessary to mention the WURM project (https://www.wurm.info/), within the framework of which vibrational spectra for 325 minerals have been calculated from first principles [Caracas, Bobocioiu, 2011].

In this article, we present a new project called ArDI (Advanced spectRa Deconvolution Instrument), undertaken by the Vinogradov Institute of Geochemistry SB RAS (Irkutsk), the A.E. Fersman Mineralogical Museum of the RAS (Moscow), and the Institute of Geology and Mineralogy SB RAS with the Central Siberian Geological Museum (Novosibirsk). The core of the project is an application that allows processing spectra, searching for analogues of the loaded spectrum in available databases, and relying on its own public database being created by the project participants.

One of the goals is to make the database available on various types of devices, including portable ones, which are controlled by different operating systems. These include personal computers running several generations of Microsoft Windows, Linux, various generations of BSD systems, macOS, etc., as well as mobile devices whose screen sizes and resolutions allow working with scientific graphs. Thus, a modern application for searching and processing spectra requires significant universality, that is, the ability to use it on devices with different operating systems. This is difficult to achieve in the classical desktop application paradigm, even within the Microsoft Windows system family. One possible solution is to use web applications that run directly in a web browser.

Web applications have several significant advantages over traditional desktop applications in solving scientific tasks. Firstly, web applications allow working with data and running computations directly in the browser, which significantly simplifies accessibility and access to the programs. Instead of needing to install and update software on each computer, users can simply open a link to the web application and start working. Secondly, web applications support multiple platforms, meaning they can be run on various operating systems such as Windows, macOS, BSD or Linux. This provides flexibility and convenience in usage. Thirdly, web applications can easily exchange data and results over the Internet, allowing scientists to utilize databases hosted on different servers; work remotely with colleagues and share research findings. Additionally, web applications typically have a more simple and intuitive user interface, which makes them easy to use even for users without specialized programming or system administration skills. Overall, web applications provide a convenient and flexible solution for scientific tasks, making them accessible and user-friendly for all users. Indeed, web applications are already successfully operating and being used for organizing scientific research [Shevchenko, 2022].

As mentioned earlier, a factor contributing to the widespread use of mineralogical databases is the availability of an application that allows for quick searching of the database, providing results in a convenient form, as well as performing necessary manipulations with the spectral information: baseline construction and subtraction, peak position determination, spectrum deconvolution, etc. It is important that this application can work, operating with all available databases, and at the same time does not lead to significant resource consumption on the user's computer. The Crystal Sleuth application of the RRUFF project possesses some of these capabilities. However, the set of operations with the spectrum is greatly limited there, and, as mentioned earlier, the application has not been supported since 2009. Another feature of this application is that it operates with a copy of the database downloaded to the user's computer, rather than accessing the dynamically updated database hosted on the project's server.

This article describes the current state of the ArDI project and discusses the prospects for its development.

Methods

Python 3.11 (https://www.python.org/downloads/release/python-3110/) was chosen as the development environment, Flask (https://flask.palletsprojects.com/en/3.0.x/) for its basic internal part (backend), and Dash (https://dash.plotly.com/) for the user presentation part (frontend). Functions built into the numpy (https://numpy.org/) and scipy (https://scipy.org/) libraries are used for curve processing. Spectrum decomposition into individual peaks (deconvolution) is performed using the lmfit package (https://lmfit.github.io/lmfit-py/).

Currently, the Dash frontend is gaining increasing popularity due to the convenience in data representation and tools for working with graphs. The Python language contains a sufficiently large number of libraries for working with data, is cross-platform, and has a relatively low entry threshold, which will further allow relatively easy increase in the number of developers. Critical time-consuming functions have

been rewritten in C++ and integrated as a static library into the application. In particular, the Asymmetric least square smoothing (ALS) method, implemented according to the algorithm proposed in [Eilers, 2005] with minor modifications, is used for baseline calculation. HDF5 format databases (https://www.hdfgroup.org/solutions/hdf5/) are used for storing spectra.

The ArDi application is packaged in a Docker container (https://www.docker.com/), which allows easy scaling of the project and ensuring its connectivity. Access to the container is provided through the NGINX reverse proxy (https://www.nginx.com/). Interaction with the reverse proxy is provided through the WSGI (Web Server Gateway Interface) web application server. Several options are available: UWSGI (https://uwsgi-docs.readthedocs.io/) or Gunicorn (https://gunicorn.org/) packages.

Description of the web application interface

The ArDI project's web application (https://ardi.fmm.ru) solves several tasks.

1) Processing of spectra to improve search results: smoothing, cropping, baseline subtraction, removal of outliers and unwanted peaks, etc.

2) Searching for similar spectra in existing databases.

3) Uploading reference mineral spectra to the database to expand it.

The structure of the ArDi web application is presented in Figure 1. The computational part of ArDI consists of three modules. The first module, labeled as Deconvolution in Figure 1, is responsible for the initial processing of the spectrum and its deconvolution - decomposition of the spectrum into constituent peaks and calculation of the baseline.

The following options are available for the initial processing of spectra: 1) smoothing and noise reduction; 2) removal of "cosmic" peaks (cosmic rays); 3) selection of the required region on the spectrum and cropping of the spectrum; 4) shifting the spectrum along the x-axis; 5) scaling the spectrum along the y-axis; 6) conversion of the spectrum to an absorption spectrum when processing transmission spectra, which can be further used in the processing of IR spectra. The results of all operations with the spectra are reversible. The processed spectrum can be saved locally in .csv format and reloaded.

An important function of this module is the decomposition of the loaded spectrum into components and the calculation of the baseline (Fig. 3). The baseline calculation is performed using the ALS method [Eilers, 2005] with minor modifications. In this method, the baseline parameters are set by only two values: p - asymmetry and λ - smoothness, which is an absolute advantage in automatic parameter selection. A small change was made to the ALS algorithm, which allows setting the baseline parameters for two parts of the spectrum. For example, if the region of OH valence vibrations requires low asymmetry values and high smoothness values, then for the low-frequency region, conversely, a decrease in smoothness and an increase in the asymmetry coefficient are required. The deconvolution parameters are adjusted using the least squares method. It is possible to choose different distributions to describe the observed peaks in the spectrum. The initial deconvolution parameters are entered in the form of an editable and exportable table.

The second module (Figures 4, 5) is used for working with databases. ArDI uses HDF5 format databases. HDF5 is a vector database. Each record represents a multidimensional array or vector that stores the sample spectrum, with the sample name, acquisition parameters, geographic location, number, chemical formula, etc. as the record parameters. Access to the record is by a key that includes the sample name, its unique number, and the record number in the database. The HDF5 format is convenient for storing large amounts of vector data and has fairly fast searching and access to records by their key [Gosink, 2006].

When working with databases, it is possible to load all records contained in them, sort records by name, number, chemical elements in the sample, its classification according to Nickel–Strunz [Strunz, 2001], which is currently the basic one used by the Commission on New Minerals, Nomenclature and Classification (CNMNC) of the IMA and is supported by the Mindat.org (https://www.mindat.org/strunz.php) and Webmineral.com (https://webmineral.com/help/StrunzClass.shtml) projects. The found spectra can be plotted on a comparison graph (Fig. 5), with the possibility of filtering and selecting spectra.

An important function of the data management module is the ability to search the databases for spectra similar to the one uploaded by the user. The search can be performed on the entire spectrum or on one of its parts. The Euclidean metric is used as a measure of spectrum similarity. For this, the spectra are reduced to the same dimension, and then the cosine of the angle between the multidimensional vectors representing the ordinate values of the spectra is calculated. The similarity of the spectra is determined as the cosine of the angle between their ordinate values:

$$Similarity \equiv cos\alpha = \frac{\sum_{i=1}^{N} y_i^{exp} y_i^{base}}{\sqrt{\sum_{i=1}^{N} (y_i^{exp})^2} \sqrt{\sum_{i=1}^{N} (y_i^{base})^2}} \quad (1),$$

where y_{exp} is the spectrum being searched for, ybase is the spectrum from the database, and N is the number of points in the spectrum. The spectrum

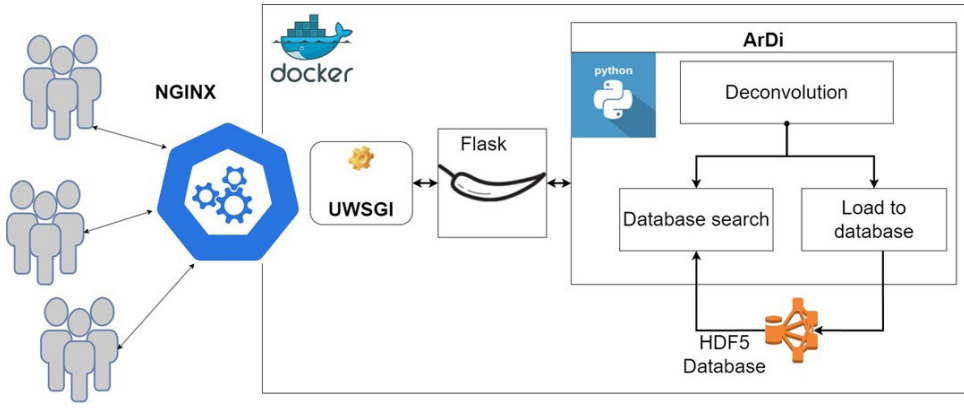

Fig. 1. The structure of the ArDI web application.

Fig. 2. Automatic peak identification on the Raman spectrum of almandine (sample FMM_1_57724).

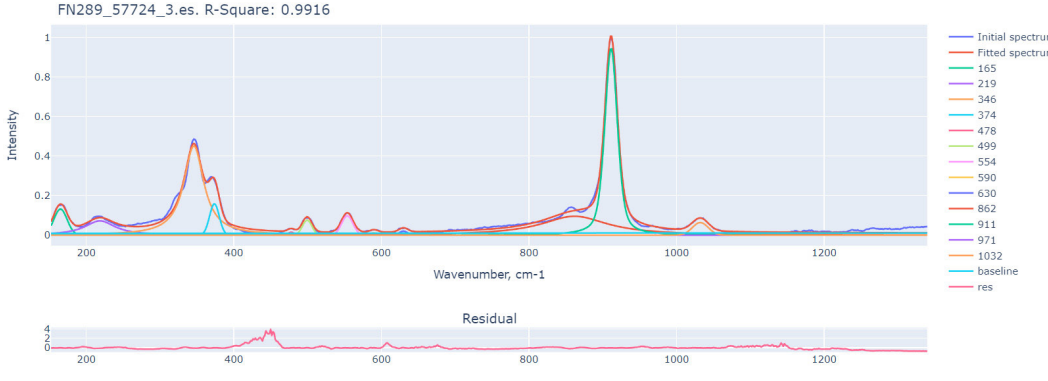

Fig. 3. Deconvolution of the Raman spectrum of almandine (sample FMM_1_57724), performed in automatic mode.

similarity metric is output in the search results, and it lies in the range from 0 to 1, where 1 means fully identical spectra, and 0 means completely different from each other. In general, this metric is similar to the Pearson correlation coefficient, but the use of the Euclidean metric is more mathematically correct [Tan, 2005].

When working with Raman spectra, the appearance of artifacts poses a significant challenge. Methods for interpreting and processing such spectra are quite extensively covered in the literature [Chukanov, Chervonnyi, 2016]. The most common one is sample luminescence, the intensity of which is several orders of magnitude higher than the intensity of Raman light scattering. The capabilities of ArDI allow reducing the contribution of luminescence bands and extracting the Raman spectrum directly. In particular, transition metal ions mainly have a relatively wide band in the 2000–5000 cm-1 region when excited in the green part of the spectrum. Such bands are effectively removed using the baseline subtraction procedure. Trivalent lanthanide ions have relatively narrow bands over a wide spectral range. However, their position is known and weakly dependent on the matrix. Thus, these bands can be removed after deconvolution, using the tool for removing "extra" peaks.

Another important problem is the study of anisotropic samples, when the Raman or IR reflection spectra of an anisotropic sample can differ significantly from each other when the orientation of the sample changes. In this case, Pearson or Euclidean correlation metrics of the spectra may give a low degree of similarity for the same sample when its spatial orientation changes. A possible solution to this problem is the use of spectrum clustering using machine learning methods. We plan to integrate such capabilities into ArDI.

The third module (Fig. 6) is responsible for uploading spectra to the database. In addition to the original sample spectrum and the spectrum with the baseline subtracted, the result of decomposing the spectrum into peaks is recorded in a separate database. Information about the geographic location of the sample, the owner, the upload date, the acquisition parameters, and an arbitrary comment is also saved in a JSON file. The sample image and its chemical

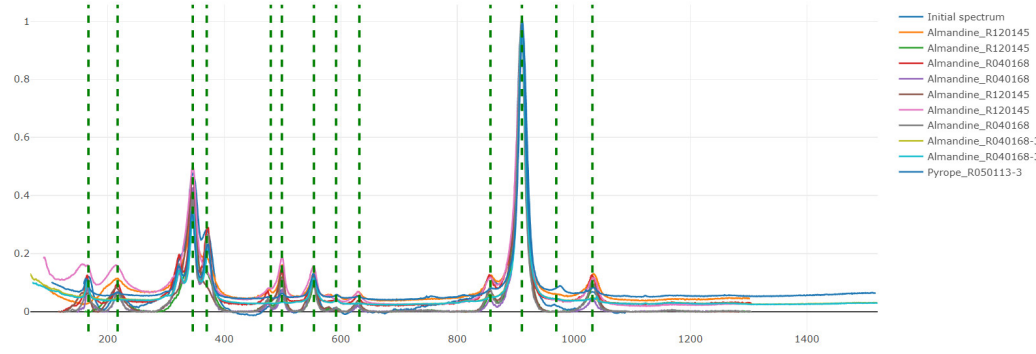

Fig. 4. The result of the search for similar Raman spectra for almandine (sample FMM_1_57724) among the spectra in the RRUFF database.

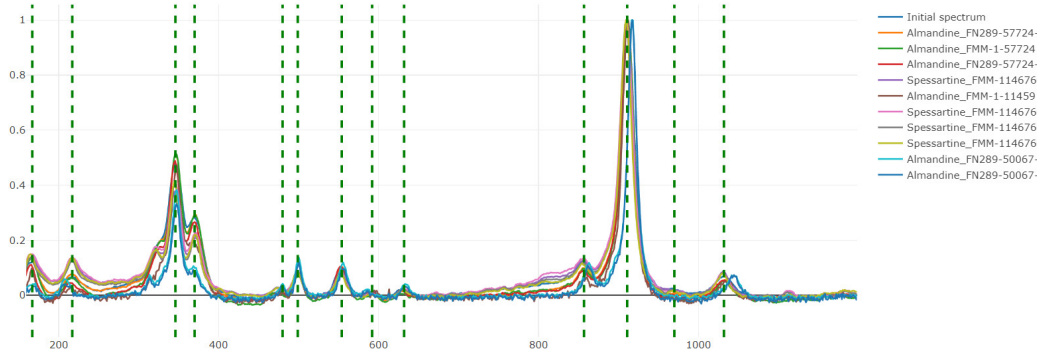

Fig. 5. The result of the search for similar Raman spectra for almandine (sample FMM_1_57724) among the spectra in the ArDI database being created.

composition can be uploaded as separate files. In the future, this data will be located on a static page with a description of the sample, similar to the existing ones on the RRUFF website. The mineral name is selected from the list of minerals approved by the CNMNC of the IMA (the list of approved mineral species is regularly updated), or it can be entered manually. For each sample, before uploading, it is necessary to specify its unique number. It is also possible to indicate specific shooting conditions, etc. In the future, a card with the composition and shooting conditions will be created for each sample. Currently, such information is stored in the database but not displayed on the screen.

Discussion and development prospects

Currently, the ArDI application is a convenient tool for rapid processing, searching for similar spectra, and uploading spectra to the database. It can be used both for routine mineral diagnostics and for the interpretation of individual vibrations in the spectra of known and new minerals. The development of the project is planned in several directions:

1. Improving the ergonomics of the interface and the algorithms for automatic processing of Raman spectra. In its current form, ArDI allows loading a spectrum, highlighting the vibrations on it, and subtracting the baseline based on the performed deconvolution in a matter of seconds. Sometimes there are problems with the deconvolution of "noisy" spectra, as well as with spectra that have bands in both the low-frequency and high-frequency regions. Improving the algorithms for automatic spectrum processing has already solved some of the problems, but work on further improving them continues.

2. Filling the database of reference spectra. We plan to review the spectra accumulated by the RRUFF and ROD projects. This revision will involve saving high-quality spectra and removing poor-quality spectra. Thus, we expect to obtain spectra of approximately two thousand minerals. The collection of the A.E. Fersman Mineralogical Museum of the Russian Academy of Sciences contains samples of about four thousand mineral species. First of all, we will take spectra of those minerals that are not present in the RRUFF and ROD databases, as well as spectra of minerals whose quality in the existing databases is criticized. Thus, we hope to create the most complete database of high-quality Raman spectra for minerals in the next year. The database will also be filled based on the collection (about 1300 mineral species) of the Central Siberian Geological Museum (Novosibirsk) and the working collections of employees of the IGM SB RAS and other scientific organizations.

3. We expect a number of methodological problems related to obtaining spectra of minerals that change under the influence of a laser beam, strongly luminescent minerals, and minerals with a high reflectivity index. We hope that a systematic approach will help solve some of these problems. For example, we have already managed to select shooting parameters for some sulfosalts, for which it was previously not possible to obtain Raman spectra. There are also minerals that do not have active vibrations in the routinely used frequency range. Unfortunately, such minerals will remain outside the current version of the system.

4. Coupling the database of reference spectra with the information system of the A.E. Fersman Mineralogical Museum of the Russian Academy of Sciences and other information systems on minerals. The Museum's information system contains data on 6,006 mineral species and 157,210 samples for 4,202 minerals. To date, this is the largest information system on mineralogy in the Russian-language

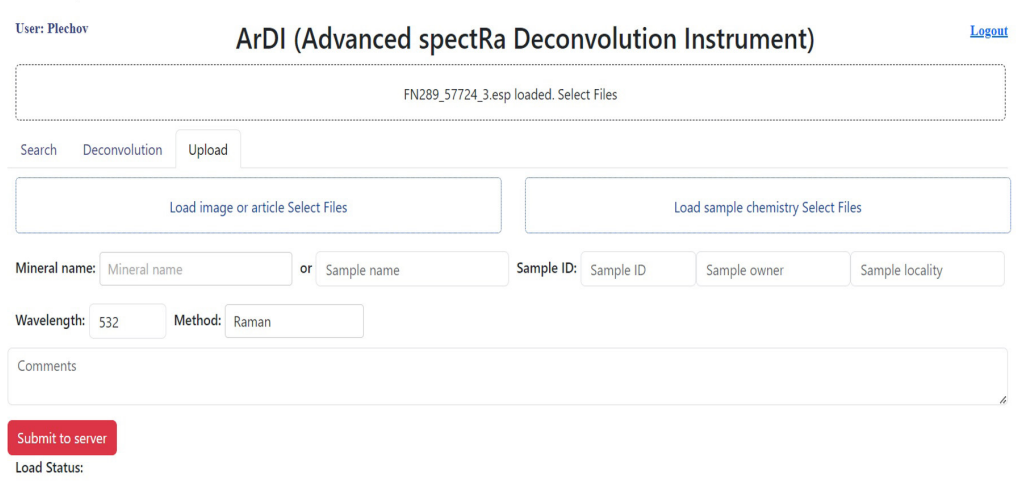

Fig. 6. The ArDI system interface for uploading a reference spectrum to the database.

segment of the Internet. This system is built taking into account the accumulation of diverse information about minerals and museum samples [Plechov et al., 2019], which allows such a coupling to be done quite easily. In the future, cooperation with other world-class mineralogical projects (RRUFF, Mindat.org) is possible. Cooperation with the main manufacturers of equipment for the implementation of developments in the standard software used in the acquisition of Raman spectra is also possible.

5) Currently, the ArDI project is voluntarily supported by three main participants. These are the A.E. Fersman Mineralogical Museum of the Russian Academy of Sciences, which provided computing power, software support, and a collection of minerals to replenish the reference spectrum database; IGC SB RAS (represented by R.Yu. Shendrik), who is doing the main work on the development and improvement of the interface; and IGM SB RAS, which is also providing the collection of the Central Siberian Geological Museum to fill the reference spectrum database and is involved in debugging the ergonomics and functionality of the interfaces. Thus, a consortium has already been formed, where each participant makes a tangible contribution and benefits from the results of joint efforts. We would like to expand this consortium by inviting all leading institutes and laboratories of the Russian Federation that could actively participate in one way or another. Active discussions on the terms of joining the consortium are currently underway. Some options are planned to be opened for a wide range of users, but some options will be available only to consortium participants or by subscription. The possible commercial use of ArDI will be regulated by the consortium participants. Our ideas about the future architecture of the consortium involve some quantitative assessment of each participant's contribution. Modern technologies allow automating such an assessment, for example, based on smart contracts [Zheng et al., 2020]. However, the use of blockchain technology for this purpose seems redundant for this task. We are consulting with IT specialists to find a solution that will meet the objectives of the consortium and correspond to the current level of information technology development.

6) The developed processing algorithms can be applied not only to Raman spectra, but also to many types of spectral information. We are conducting preliminary work on the processing of IR spectra measured under various conditions, diffractograms, optical absorption spectra, etc. With the expansion of the consortium, work on the creation of reference databases of various types of spectra is possible.

7) As the verified information in the spectrum database accumulates, it is planned to use machine learning methods for searching and processing. In particular, this will open up new approaches to the search for mineral species and the interpretation of vibrational spectroscopy data. It is expected that this approach will allow successful work with spectra containing peaks from several mineral species and move towards phase analysis of samples based on spectroscopic data, similar to X-ray phase analysis. We also hope that machine learning methods will be able to help in working with spectra of anisotropic minerals and in comparing spectra obtained on different equipment and under different shooting conditions.

The main application of the resulting product is the rapid and reliable identification of minerals in geological samples. This is extremely in demand in mineralogy itself (verification of found mineral species and search for new, still unknown to science minerals). In addition, there is a wide application in the examination of raw materials and cut precious stones, as well as products from them. These objects have a special material value and require non-destructive analysis methods (Raman spectroscopy in most cases does not destroy the sample). The use of ArDI will increase the efficiency and effectiveness of

An important prospect, in our opinion, is the possibility of integrating ArDI into the software of modern devices used in scientific research and the development of new technologies. Based on the method, it is possible to design compact devices useful for mineralogists in the field and for mineralogical mapping at mining sites. In addition to mineralogy and gemology, the developed approach can be widely used in medicine, pharmaceuticals and forensics, since Raman spectroscopy is already widely used in materials science and for the identification of organic substances.

Acknowledgements

The authors thank N.V. Chukanov and E.A. Pankrushina for carefully reading the manuscript at the review stage and providing valuable constructive comments that allowed to improve the article.

References:

Caracas R., Bobocioiu E. The WURM project – a freely available web-based repository of computed physical data for minerals // American Mineralogist. 2011. V. 96. P. 437–443.

Chukanov N.V. Infrared spectra of mineral species: extended library. Springer Science & Business Media, 2013. 1726 p.

Chukanov N.V., Chervonnyi A.D. Infrared spectroscopy of minerals and related compounds. Springer, 2016. 1109 p.

Chukanov N.V., Vigasina M.F. Vibrational (Infrared and Raman) Spectra of Minerals and Related Compounds. 2020. Springer, 1376 p.

Gosink L., Shalf J., Stockinger K., Wu K., Bethel W. HDF5-FastQuery: Accelerating Complex Queries on HDF Datasets using Fast Bitmap Indices, 18th International Conference on Scientific and Statistical Database Management (SSDBM'06), 2006, Vienna, Austria. P. 149-158. DOI: 10.1109/SSDBM.2006.27.

Eilers P.H.C., Boelens H.F.M. Baseline correction with asymmetric least squares smoothing // Leiden University Medical Centre report. 2005. 1(1). P. 1–24.

El Mendili Y., Vaitkus A., Merkys A., Gražulis S., Chateigner D., Mathevet F., Gascoin S., Petit S., Bardeau J.-F., Zanatta M., Secchi M., Mariotto G., Kumar A., Cassetta M., Lutterotti L., Borovin E., Orberger B., Simon P., Hehlen B., & Le Guen M. Raman Open Database: first interconnected Raman–X-ray diffraction open-access resource for material identification // Journal of Applied Crystallography. 2019. V. 52(3). P. 618-625. DOI: 10.1107/s1600576719004229.

Lafuente B., Downs R.T., Yang H., Stone N. The power of databases: the RRUFF project. In: Highlights in Mineralogical Crystallography, T Armbruster and R M Danisi, eds. Berlin, Germany, W. De Gruyter, 2015. P. 1–30.

Plechov P.Y., Trousov S.V., Bychkov K.A., Konovalova K.A. Multilayered mineralogical information in spectroscopy of minerals. In XIX INTERNATIONAL MEETING ON CRYSTAL CHEMISTRY, X-RAY DIFFRACTION AND SPECTROSCOPY OF MINERALS. 2019. P. 43.

Shevchenko Y. Open Lab: A web application for running and sharing online experiments // Behavior Research Methods. 2022. V. 54. P. 3118–3125. DOI: 10.3758/s13428-021-01776-2

Strunz H. and Nickel E.H. Strunz Mineralogical Tables, 9th Edition. Berlin and Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung, 2001. 870 p.

Tan P.-N., Steinbach M., Kumar V. Introduction to Data Mining, Addison-Wesley, ISBN 0-321-32136-7. 2005. Chapter 8. p. 500.

Zheng Z., Xie S., Dai H.N., Chen W., Chen X., Weng J., Imran M. An overview on smart contracts: Challenges, advances and platforms // Future Generation Computer Systems. 2020. 105. P. 475–491.