

ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА ПО КРИСТАЛЛОХИМИИ И МИНЕРАЛОГИИ WWW-MINCRYST: 18 ЛЕТ НАУЧНОМУ ИНТЕРНЕТ-ИНСТРУМЕНТАРИЮ

Д.А. Варламов¹, Т.Н. Докина¹, Н.А. Дрожжина¹, О.Л. Самохвалова¹

¹ *Институт экспериментальной минералогии РАН, г. Черноголовка, РФ*

Информационно-вычислительная система (ИВС) WWW-MINCRYST (<http://database.iem.ac.ru/mincryst>, <http://mincryst.iem.ac.ru>), предназначенная для работы с максимально возможным количеством пользователей и поэтому изначально базирующаяся на использование интернет-технологий, в качестве базы данных, размещенной в среде Интернет, возникла достаточно давно и стала одним из пионерских интерактивных научных Интернет-ресурсов как в России, так и в мире (в области наук о Земле). Первый полностью работоспособный вариант WWW интерфейса был создан в рамках гранта РФФИ № 96-07-89162 и представлен пользователям в декабре 1997 года [1], таким образом система успешно функционирует и активно востребована пользователями уже почти 18 лет.

Ниже описывается созданная в 1997 году и активно развиваемая вплоть до настоящего времени ориентированная на Интернет-пользователя информационно-вычислительная система (ИВС) WWW-MINCRYST со свободным доступом. Цель создания ИВС WWW-MINCRYST с ее обширным инструментарием – обеспечить максимально широкий круг пользователей во всех областях науки, оперирующих с кристаллическим веществом (геология, геохимия, кристаллография, физика твердого тела) достоверной и актуальной информацией в области кристаллохимии и кристаллографии минералов, их синтетических аналогов и элементов, а также дать дополнительный толчок в развитии методов обработки и визуализации кристаллоструктурных данных.

Первоначально (в конце 90-х) ресурс был призван обеспечить интерактивный доступ Интернет-пользователям к накапливавшимся с 1985 года в рамках локальной ИВС МИНКРИСТ данным по кристаллическим структурам (прежде всего литературным, а также авторским аналитическим — всего около 3500 объектов на момент старта проекта). В ходе проекта в состав ИВС вводились программные разработки по обработке кристаллоструктурных и кристаллохимических данных, их анализу и мультимедийной визуализации. Главным инициатором и основным идеологом работ был заведующий группой рентгеноспектрального анализа Института экспериментальной минералогии РАН выдающийся кристаллограф А.В. Чичагов (увы, безвременно ушедший от нас в 2010 году).

За 18 лет развития ИВС WWW-MINCRYST нарастил свой информационный фонд почти в 2,5 раза (почти на 6000 объектов) и по праву вошел в первые ряды рентгеноструктурных и кристаллохимических баз

данных, связанных с изучением минерального вещества, как по оценкам пользователей, так и по мнению составителей отраслевых каталогов информационных ссылок.

ИВС WWW-MINCRYST зарегистрирована в государственном регистре баз данных (НТЦ "Информрегистр") под номером 0229805169 (регистрационное свидетельство № 4873 от 11.02.99).

Основным компонентом ИВС является собственно база данных или Информационный фонд (на июнь 2015 года — около 9300 информационных объектов для почти 4000 уникальных фаз, включая 3700 природных минералов и около 300 синтетических фаз, не имеющих пока натуральных аналогов). Информационный объект состоит из 4 базовых таблиц с 60 фиксированными и до 130 дополнительными полями (число последних варьирует в зависимости от сложности структуры) и 4 бинарных файлов. Поддерживается иерархическая система индексов и ряд вспомогательных таблиц. Общий объем базы данных составляет около 500 Мбайт.

Информационный фонд в ИВС WWW-MINCRYST содержит информацию о большинстве минеральных видов (около 3700 из почти 5000 официально признанных на май 2015 года Международной Минералогической Ассоциацией – <http://www.ima-mineralogy.org/Minlist.htm>), кристаллические структуры которых расшифрованы к настоящему времени. Помимо природных объектов, в базе данных представлены синтетические минералы – их структурные аналоги, отличающиеся по составу (например, с заменой одного из катионов), и неорганические соединения (силикаты, фосфаты, бораты и др.), близкие по свойствам к природным веществам. Информационный фонд содержит данные структурных работ из более 130 иностранных и отечественных журналов за период от 1930-х годов по настоящее время. Ежегодный прирост новых или заново переопределенных/уточненных кристаллических структур минералов и их аналогов достаточно значителен, чтобы требовалась постоянная актуализация информационного фонда (в среднем, до 350-450 структур в год). В последние годы основной акцент был сделан на новые минералы, а также на кристаллические структуры, описанные в советских и российских журналах (которые в последние десятилетия были менее доступны web-пользователю, нежели зарубежные).

Интерфейс ИВС WWW-МИНКРИСТ представляет собой многоуровневую информационную двуязычную (русский и английский, включая содержание и языки интерфейса) систему. В нее входят следующие компоненты: (а) комплексные поисковые интерфейсы, использующие как критерии поиска (в различных комбинациях) названия минералов, химический состав, кристаллоструктурные параметры, литературные ссылки и вспомогательную информацию, причем возможны комбинации поисковых признаков; (б) модули мультимедийного представления информации (интерактивные полиэдрические и шаровые структуры, линейчатые и непрерывные спектры — через Java-апплеты WWW-CrystPic и WWW-MixiPol;

(в) классификационные схемы; (г) модули организации взаимосвязей с внешними информационными ресурсами через систему динамически формируемых ссылок; (д) WWW-ориентированный инструментарий разработчика (включая систему импорта, проверки и редакции данных, а также архивации и резервного копирования).

Технологически ИВС WWW-MINCRYST как web-сервис реализована на "классической" связке Linux-Apache-Mysql-PHP (LAMP) с использованием JavaScript и (для интерактивных апплетов) Java (на базе Sun/Oracle JDK). Размещена ИВС на сервере баз данных Институт экспериментальной минералогии РАН (<http://database.iem.ac.ru>), доступ нелимитирован. На старте проекта ИВС WWW-MINCRYST стала в 1999 году (в реализации на ОС Digital Unix) одной из первых интернет-ориентированных баз данных в России, использовавшим данную технологию, причем за 15 лет она нисколько не потеряла свою технологичность на стороне сервера. В 2015 проведена миграция базы данных с проприетарной СУБД MySQL на свободно распространяемую MariaDB (<https://mariadb.org/>) для сохранения свободного статуса ИВС. В настоящее время ИВС размещена на сервере со следующими параметрами – CPU Intel Core i5-2400@3.10GHz x 4 ядра, RAM 24 Gb, disk array SSD 128 Gb + 2x2 Tb WD RE 2000, GPU Nvidia Kepler GTX 770, программно - Apache HTTPD server 2.4.6, PHP script language 5.4.16, MariaDB database 5.5.40.

В теоретической основе ИВС WWW-MINCRYST лежит ряд принципиальных "идеологических" моментов, реализация которых и обеспечила ему имеющееся признание кристаллографической и минералогической научной общественности. Синтез информации о кристаллической фазе, рассматриваемой как "монокристалл" и/или как "поликристалл" с заменой экспериментальных поликристалл-стандартов расчетными – базовая идея WWW-MINCRYST, которая и делает его универсальной, комплексной и оригинальной разработкой, поскольку позволяет существенно расширить информационную базу сравнительно с чисто аналитической. На базе введенной информации программный пакет эксперта-разработчика позволяет автоматически сформировать вторую, производную от первой (аналитической), базу расчетных поликристалл-стандартов, проводя синтез двух типов информации о кристаллической фазе. Связка "Кристаллическая структура фазы и ее расчетная поликристалл-рентгенограмма" является информационной основой ИВС WWW-MINCRYST и служит важнейшим инструментарием в руках пользователя (в особенности рентгеновского кристаллографа). Стержневая идея WWW-MINCRYST о синтезе двух типов кристаллоструктурной информации о кристаллической фазе реализует принципиально новый подход к формированию всей кристаллоструктурной информации о веществе и организации доступа к ней.

Основным источником информации для ИВС служат оригинальные журнальные статьи, опубликованные в открытой печати. Извлеченная кристаллоструктурная информация помещается (в соответствии со

специальным форматом записи) в ASCII-файл с последующей программной экспертизой (проводимой на локальных ПК) по результатам расчета межатомных расстояний и других кристаллоструктурных характеристик и, в случае положительного решения, импортируется в ИВС специальными средствами группового или одиночного импорта. Возможен импорт кристаллоструктурной информации через общепринятый в кристаллографии специализированный файл (Crystallographic Information File, CIF), содержащий кристаллоструктурные характеристики расшифрованной кристаллической фазы (формат – <http://www.sdsc.edu/pb/cif/cif.html>).

Информационная запись для индивидуального кристаллического вещества содержит информацию о названии (в соответствии с классификацией International Mineralogy Association или рекомендациями по наименованию неорганических веществ IUPAC), химическом составе, симметрии, параметрах элементарной ячейки, координатах атомных позиций с изотропными температурными факторами и заселенностями, информацию о межплоскостных расстояниях, HKL-индексах и интенсивностях сильнейших рефлексов рентгенодифракционной картины поликристалл-фазы, а также ссылки на соответствующие публикации по расшифровке или уточнению кристаллической структуры. Запись может быть специфицирована по полезным свойствам, особенностям химического состава и структуры, P-T условиям синтеза, принадлежности к условным минеральным группам. Каждая запись содержит "монокристалльные" и "поликристалльные" характеристики кристаллической фазы. Минералы классифицированы в соответствии с таксонами структурно-химической систематики минералов А.А. Годовикова, кристаллохимической классификации М. Чириотти, в настоящее время вводится классификация по структурным типам минералов по Г.Б. Бокию. Для 2400 фаз сделаны экспресс-оценки потенциальной энергии кристаллической решетки. В настоящее время создается система расчета содержаний элементов в соответствии с идеальной и реальной формулами минерала-образца.

В ИВС WWW-MINCRYST реализован импорт одобренной экспертами информации через web-интерфейс как в виде единичных записей, так и пакетов записей (до нескольких сот) с входным контролем данных, что позволяет проводить постоянную актуализацию информационного фонда. Также реализована возможность редактирования записей в on-line режиме, их удаления, замены служебных файлов записей, архивации и восстановления базы данных через web-интерфейс.

Кратко об основных возможностях ИВС WWW-MINCRYST.

В ИВС реализована достаточно эффективная система поиска фаз по отдельным наборам или комплексу параметров: название минерала (полное или частичное, по спецификации, в составе общеупотребительных групп типа цеолитов, алфавитные списки), химический или элементный состав в различных комбинациях (присутствие/отсутствие элементов и их комбинации), кристаллоструктурные характеристики (симметрия, пространственные группы,

параметры элементарной ячейки, межплоскостные расстояния $d(hkl)$), что в сочетании с химическим (элементным) составом дает возможность прямого интерактивного качественного рентгенофазового анализа. В систему поиска добавлен поиск по классификационным параметрам нескольких кристаллохимических и структурных классификаций (Годовиков, Чириотти, Бокий). Система поиска по этим параметрам обеспечивает поиск (и группировку) минералов по указанным низшим и промежуточным таксонам классификаций. По ряду параметров поиска ИВС WWW-MINCRYST до сих пор не имеет аналогов среди минералого-кристаллографических баз данных.

По результатам поиска пользователь получает ссылку на список фаз, отвечающих поисковым критериям с указанием ID записи, названия фазы, формулы и пространственной группы с возможностью последующего перехода по гиперссылке. В случае единичного успешного ответа пользователь сразу автоматически переводится на найденный объект.

В «карте» найденного объекта представлены основные данные (как хранимые в БД, так и динамически рассчитываемые при формировании «карты») – название, формула, параметры ячейки, различные кристаллографические данные (кол-во рефлексов, рентгеновские плотности, коэффициенты поглощения); CPDS карта – 20 максимальных рефлексов, их hkl позиции и интенсивности; базовые атомные позиции и их заселенности; полная информационно-расчетная карта (рентгеновские плотности, все рефлексы и т.п.); рассчитанные энергии решетки; принадлежность к различным классификационным схемам; источники данных (ссылки на публикации, поправки и т.п.); автоматически формируемые ссылки на внешние ресурсы.

Для всех записей через апплет WWW-Crystpic (Java-3D с использованием библиотек OpenGL) доступны динамически создаваемые интерактивные изображения моделей кристаллических структур в шарах-сферах и в полиэдрических проекциях (до 138 позиций и до 1500(!) атомов на структуру) в соответствии с основными канонами минералогии и кристаллографии. Программа позволяет делать всевозможные манипуляции с моделью структуры, включая масштабирование, непрерывное и/или автоматическое дискретное вращение вокруг "экранных" осей X,Y,Z; ориентацию по кристаллографическим осям, hkl -фрагментацию структуры (на hkl -ориентированные фрагменты толщиной $d(hkl)$), наращивание элементарных ячеек вдоль любых выбранных направлений для формирования "сверхструктур" и мотивов, а также прямой "ручной" и автоматизированный для малых полиэдров (тетраэдров и октаэдров) расчет любых межатомных расстояний и углов (плоских и телесных) в структуре. Программа изображает любые полиэдры, включая "дефектные" с необычно малыми ("плохими") межатомными расстояниями. Более детальное описание всех возможностей апплета приведено здесь – <http://mincryst.iem.ac.ru/rus/crystpic.php>. В настоящее время существуют проблемы с работой с внешними библиотеками OpenGL (на которых реализован апплет) и параноидальными настройками безопасности

новых версий Java, которые решаются вмешательством пользователя. Однако, из-за этого сейчас необходим перенос апплета на новые программные платформы типа встроенных средств HTML5. Примеры моделей кристаллических структур цеолитов (до 600 атомов) приведены на рис.1

Модуль в виде апплета WWW-Mixipol предназначен для графического представления полных расчетных спектральных профилей поликристалл-рентгенограмм с возможностями манипулирования спектрами для разных источников излучения и разных типов спектральных шкал. Также модуль способен формировать рентгенограммы смесей фаз (до 6 фаз одновременно) при возможности варьирования относительными содержаниями компонентов смеси. Пример расчетных спектров смеси шести фаз приведены на рис.2

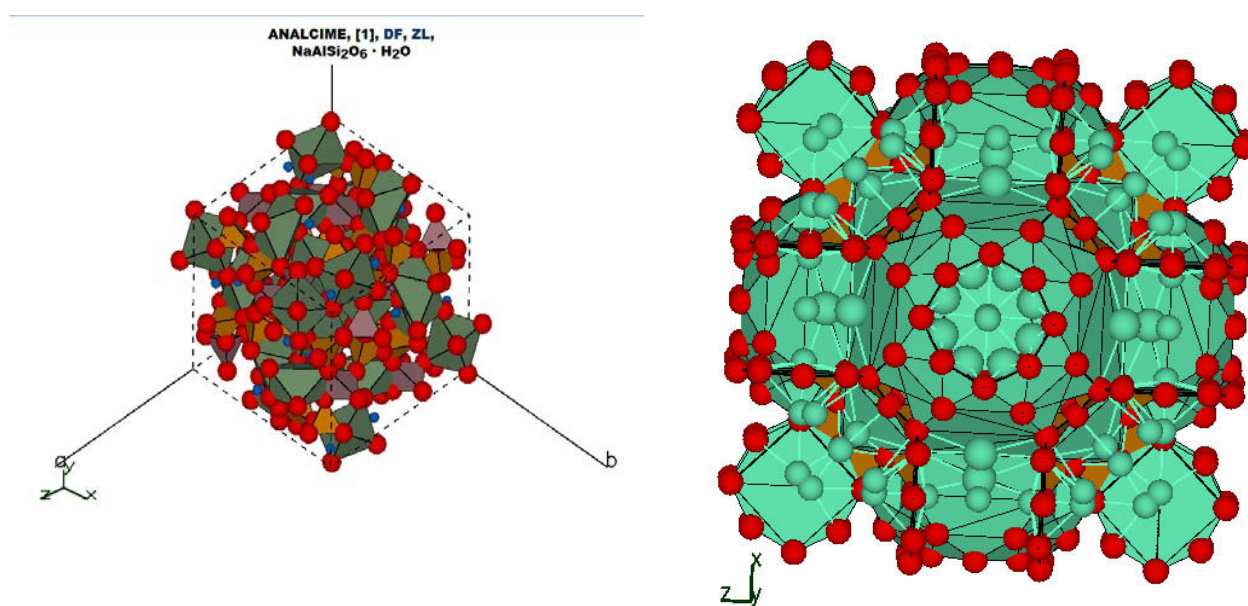


Рис. 1. Примеры интерактивных моделей кристаллической структуры (цеолиты разных типов, смешанное представление в шарах и полиэдрах, размеры атомов уменьшены вдвое)

Как для структур, так и для спектров минералов предусмотрены и упрощенные варианты представления в виде так же интерактивных традиционных шаровых структур и линейчатых спектров.

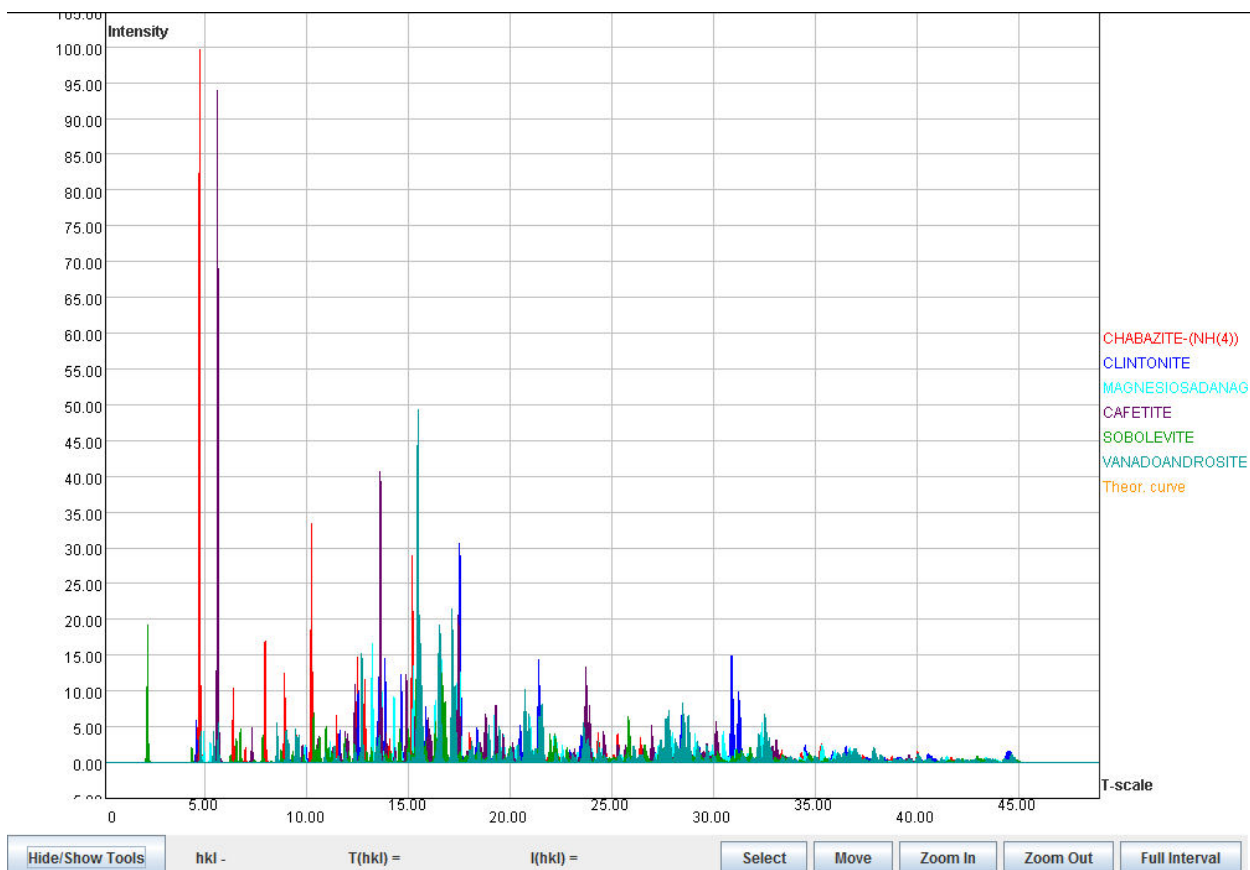


Рис.2. Пример комбинации расчетных спектральных профилей поликристалл-рентгенограмм (смесь 6 фаз)

Разработанный и интегрированный (пока частично) в WWW-MINCRYST расчетный модуль WWW-Хraypol в связке с модулями WWW-Crystpic и WWW-Mixipol позволяет квалифицированному пользователю извлекать и модифицировать "на лету" структуры из информационного фонда (не затрагивая его непосредственно), а затем рассчитывать новые варианты и получать графические изображения моделей кристаллических структур и квазиреальных полных профилей расчетных поликристалл-рентгенограмм, тем самым решая задачи по модификации реальных структур.

Одними из первых среди научных интернет-ориентированных баз данных для ИВС WWW-MINCRYST была разработана система динамически формируемых перекрестных веб-ссылок для связи объектов с записями для конкретных минералов в ведущих минералогических базах данных в Интернете. Система генерации динамических гиперссылок на внешние информационные ресурсы (на минералогические базы данных и поисковые системы) позволяет "прозрачно" для пользователя подключать большие внешние массивы данных, используя метод "генеральных" запросов. При этом пользователь сразу получает доступ к информации по интересующему его объекту, минуя стадии поиска или просмотра всей внешней базы. Кроме того, данный механизм реализует обратную связь, позволяя таким же образом

ссылаться этим базам уже на наши информационные объекты, что резко повышает востребованность ИВС внешними пользователями.

Наличие в ИВС WWW-MINCRYST данных по 9300 кристаллическим структурам, встроенного расчетного комплекса и развитых средств визуализации позволили помимо возможностей, ориентированных на поиск и предоставление информации, использовать ИВС в разработке нетрадиционных научных подходов к интерпретации и представлению некоторых кристаллических структур. Например, WWW Xraypol позволил выявить в традиционных структурах возможность гибкого использования полиэдров и почти автоматически формировать различные варианты структурных моделей минералов. Как выяснилось благодаря гибкому использованию полиэдров, позволяющему формировать различные варианты структурных моделей минералов, для части минералов можно не строго привязываться к традиционному катионно-анионному изображению, а формировать структуры на основе любых атомов, входящих в ее состав. Используется принцип: в ряде случаев кристаллическое пространство можно организовать в смешанном шаровом и полиэдрическом изображении на основе любых (а не общепринятых) атомов в структуре. Метод особенно эффективен для сложных "неправильных" бескислородных структур (фосфиды, сложные сульфиды и др.).

Востребованность WWW-MINCRYST хорошо подтверждается статистикой обращений (за 2014 год – 7,3 млн. успешных единичных запросов, более 77 Гб скачанной информации, около 48000 уникальных сайтов-клиентов), а также большим количеством отзывов, описаний и внешних ссылок на WWW-MINCRYST (см. раздел "Ссылки" на сайте).

О перспективах развития WWW-MINCRYST следует сказать отдельно. В связи с очень быстрым развитием web-технологий и (соответственно) браузеров, возникла необходимость кардинальной переработки нынешнего клиентского интерфейса, уже не отвечающего как новым технологиям (и зачастую не полностью совместимого с новым ПО), так и требованиям пользователей. Она будет включать изменение способа представления записи путем сведения разобщенных сейчас полей (основные данные, CPDS карта, атомные позиции, структура, спектры, ссылки и др.) в единое информационное пространство, формируемое на базе динамического HTML (с использованием технологий JSON, DHTML, HTML5, оверлейных структур). Предусматривается как переделка и рекомпиляция Java апплетов (в связи с устареванием и неполной совместимостью кода с последними реализациями Java VM), так и их замена на модули, использующие мультимедийные технологии стандартов HTML5 или Adobe Flash (с использованием оригинального авторского расчетного кода). Также будут изменены в сторону повышения дружелюбности и простоты) дизайн и элементы управления предоставляемой пользователю информации. Предусмотрено пополнение базы данных дополнительной сугубо минералогической информацией (фото минералов, рисунки

кристаллографических форм, дополнительные минералогические данные и др.). Несмотря на то, что по ряду параметров поиска WWW-MINCRYST до сих пор не имеет аналогов, будет расширен круг потенциальных поисковых запросов (прежде всего в области поисков по составу и кристаллографическим данным).

Разработанная и развиваемая на протяжении 18 лет ИВС WWW-MINCRYST является общедоступным, дружелюбным пользователю интерфейсом к большому объему кристаллоструктурной и кристаллохимической информации с развитыми средствами поиска, представления и обработки и может служить мощным инструментарием для всех исследователей в минералогии, кристаллографии, физике твердого тела, материаловедении и прочих смежных областях науки.

Проведение работ по ИВС WWW-MINCRYST в течение 1997-2014 годов было поддержано 6-ю грантами РФФИ (в том числе в настоящее время – грантом РФФИ № 15-07-08399-а, рук. Варламов Д.А.).

ЛИТЕРАТУРА

1. А.В. Чичагов, Д.А. Варламов, Е.В. Ершов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова Кристаллографическая и кристаллохимическая база данных для минералов и их структурных аналогов (WWW-MINCRYST) // Записки РМО, 2007, т.136, № 3, с.135-141
2. Д.А. Варламов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова WWW-MINCRYST: Интернет-ориентированная информационно-вычислительная система по кристаллографии и кристаллохимии минералов // Вестник ЮУрГУ, Сер. «Вычислительная математика и информатика», 2013, т.2, вып.1, с.26-32
3. Д.А. Варламов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова WWW-MINCRYST: информационно-вычислительная система по кристаллохимии и минералогии — техническая реализация, возможности, перспективы // Труды XVI Международной суперкомпьютерной конференции «Научный сервис в сети Интернет: многообразие суперкомпьютерных миров», 22-27 сентября 2014 года, Новороссийск – М.: Изд-во МГУ, 2014. (497 с.) с.402-405