

УДК 550.8.028

Единая система хранения и доступа к геофизическим данным. Традиции и новые подходы

© 2021 г. П.Г. Бутырин, С.А. Красилов

ФИЦ ЕГС РАН, г. Обнинск, Россия

Поступила в редакцию 09.11.2021 г.

Аннотация. Представлены особенности развития систем сбора данных в пределах Информационно-обработывающего центра (ИОЦ) Федерального исследовательского центра «Единая геофизическая служба Российской академии наук» (ФИЦ ЕГС РАН). Приведены исторические сведения, включая технические детали, связанные с систематизацией архива и эволюцией форматов геофизических данных. Учтены исторические, территориальные особенности, а также опыт развёртывания подобных информационных систем в пределах ФИЦ ЕГС РАН и за рубежом. Предложена новая концепция построения информационной системы, которая учитывает требования к масштабируемости, воспроизводимости на различных объектах и использование стандартных программных средств. Выполнена большая работа по формированию однородного архива волновых форм и инвентаризация метаданных по сейсмическим станциям, результатом которой стала возможность включения ИОЦ ФИЦ ЕГС РАН в состав международных центров обработки геофизической информации на основе FDSN. Создание распределённой системы сбора и обработки с использованием облачного сервиса позволяет абстрагироваться от территориальных особенностей сбора и хранения геофизической информации, что повышает производительность сервиса доступа к данным и степень технической готовности ключевых узлов системы.

Ключевые слова: сейсмологический мониторинг, система сбора данных, информационная система, геофизические данные, универсальный доступ.

Для цитирования: Бутырин П.Г., Красилов С.А. Единая система хранения и доступа к геофизическим данным. Традиции и новые подходы // Российский сейсмологический журнал. – 2021. – Т. 3, № 4. – С. 77–87. DOI: <https://doi.org/10.35540/2686-7907.2021.4.05>

Введение

Федеральный исследовательский центр «Единая геофизическая служба Российской академии наук» (ФИЦ ЕГС РАН) проводит многопрофильные геофизические исследования на основе данных собственных сейсмических станций и станций мирового сообщества. За период более чем 25 лет накоплен уникальный архивный материал волновых форм, бюллетеней и каталогов сейсмических событий. Волновые формы представлены как в аналоговом виде (сейсмограммы на фотобумаге, микрофишах и микрофильмах), так и в виде объёмного цифрового архива. Материалы архива расположены не только в Центральном отделении (ЦО) в г. Обнинске, но и в региональных филиалах. Следует отметить значительную разнородность носителей, многообразии форматов и представления информации. Кроме того, возрастающая скорость обнов-

ления технических средств сбора и организации хранения информации увеличивает разнородность форм доступа и значительно усложняет структуризацию. ФИЦ ЕГС РАН эксплуатирует порядка 150 сейсмических станций, а также постоянно получает информацию из международных сетей, что увеличивает размер архива на 3 ТБ ежегодно. В то же время необходимым элементом проведения современных геофизических исследований является обмен и совместный анализ сейсмологических данных как между подразделениями ФИЦ ЕГС РАН, так и с международными сейсмологическими центрами.

Перечисленные выше обстоятельства сформировали потребность строгой стандартизации сбора, хранения и доступа к геофизической информации. Главной идеей этой концепции является использование уже готовых, проверенных мировым сообществом архитектурных решений и программных средств. Итогом развития

системы в этой концепции явится интеграция ФИЦ ЕГС РАН в состав сети мировых центров данных.

Цели и задачи

Основная цель проекта – создание универсальной отказоустойчивой платформы для обеспечения работы сервисов публикации волновых форм, каталогов и метаданных сейсмических сетей и станций, входящих в состав УНУ «Сейсмоинфраструктурный комплекс мониторинга арктической криолитозоны и комплекс непрерывного сейсмического мониторинга Российской Федерации, сопредельных территорий и мира». Для достижения поставленной цели сформулированы следующие задачи:

- обеспечить хронологически структурированное хранение метаданных;
- обеспечить публикацию волновых форм в форматах miniSEED и ASCII;
- обеспечить доступ к каталогам сейсмических событий и сейсмологическим бюллетеням в форматах QuakeML и ISF.

Историческая справка по системам сбора и хранения

Развитие цифровых систем сбора, передачи и накопления сейсмологических данных в ФИЦ ЕГС РАН (в тот момент Центральной опытно-методической экспедиции Института физики Земли АН СССР – ЦОМЭ ИФЗ АН СССР) началось с установки четырёх широкополосных цифровых станций IRIS («Обнинск», «Аркти», «Кисловодск» и «Гарм»), поставленных в 1987–1994 гг. в рамках участия в Международном эксперименте по обмену данными ТЭГНЭ-1/2 [Старовойт и др., 2017]. До этого момента цифровые данные были представлены в незначительном объёме, в основном, в экспериментальном режиме. На начальном этапе передача данных отсутствовала, архивы формировались методом копирования на ленточные носители на самих станциях с дальнейшей пересылкой их в Обнинск. Но уже с 1990 г. появилась возможность организовать передачу информации с части каналов по проводным линиям в Экспериментальный международный центр данных (ЭМЦД) в г. Обнинске, оснащённый на тот момент компьютерами PDP-11/84 и SUN-2/SUN-3, объединёнными в локальную сеть Ethernet 10Mb, и аналоговыми модемами для передачи данных со станций в центр по выделенным телефонным линиям. В дальнейшем количество стан-

ций IRIS было увеличено до 12, соответственно вырос поток входных данных. Для организации непрерывной передачи данных с гарантированной доставкой на тот момент не существовало разработанных протоколов и систем передачи, поэтому первоначально был использован FTP-протокол с копированием часовых фрагментов данных по расписанию. Настройка и разработка передачи данных в Обнинском ЭМЦД в первые три года проходила при активном содействии специалистов Калифорнийского университета Сан-Диего. За это время была разработана и внедрена система передачи непрерывных данных в режиме, близком к реальному времени (Near Real-Time System – NRTS) [Chavez, Berger, 1997], (версии 1.x–2.x). Если версия 1 представляла собой просто набор скриптов, использующих FTP-протокол в качестве транспортного, то версия 2 уже составляла набор программ на языке C/C++ и использовала в качестве транспорта протокол XFER, специально разработанный Д. Чавезом для гарантированной доставки данных со станций в центр. Передача данных велась в исходном формате станций IRIS без дополнительных преобразований. Формат записи самих цифровых данных за этот промежуток времени также претерпел значительные изменения – от МК.4 до МК.8, что потребовало для унификации обработки иметь целый набор конвертеров. Накопление данных осуществлялось в так называемые «дисковые кольцевые петли», типовой размер которых составлял семь суток.

Параллельно в эти же годы шло развитие и внедрение отечественных цифровых систем сбора сейсмологических данных. Так, в 1989–1992 гг. в районе Кавказских Минеральных Вод была развёрнута аналоговая телеметрическая система сбора с цифровой регистрацией РТСС [Антонов и др., 1992] на восемь пунктов, разработанная в ЦОМЭ ИФЗ АН СССР (ныне ФИЦ ЕГС РАН) под руководством В.Н. Мишаткина. Весь набор программ, используемых для сбора данных, был разработан коллективом сотрудников ЦОМЭ с использованием доступных на тот момент Intel-совместимых компьютеров для операционных систем DOS и Windows, поэтому использовать уже имеющуюся транспортную систему NRTS не представлялось возможным. Формат сбора и хранения данных также был не унифицирован. Для организации оперативной передачи данных этой системы в центр в Обнинске была разработана собственная система RMM, которая функционально повторяла NRTS v1.x (набор скриптов на файловом SMB-протоколе).

Набор архивных данных в ЦОМЭ за этот период времени представлял собой совокупность разнородных носителей и форматов, не имеющую общей описательной части, что не позволяло организовать быстрый поиск и доступ к архивным данным.

Начиная со второй половины 1990-х гг. в сейсмологическом сообществе стал распространяться протокол передачи данных LISS (Live Internet Seismic Server) [Hutt, Bolton, 2021] на основе ставшего де-факто стандартом сбора данных протокола SEED (Standard for Exchange of Earthquake Data) [SEED, 2021] и его подмножества miniSEED (без описательного заголовка), который обеспечивал передачу данных по протоколу TCP/IP. Протокол LISS не имел возможности повтора передачи «сбойных» пакетов и не гарантировал возобновления передачи с места разрыва (за исключением небольшого буферирования), но при встроенной в формат SEED системе упаковки STEIM-1/2 позволял при небольших размерах пакета передавать достаточные объёмы данных без сбоев. А адаптация Д. Чавезом приёмной части пакета NRTS для приёма LISS-пакетов позволила использовать эту систему для гарантированной доставки данных в центр сбора. Это позволило подклю-

чить к общей системе приёма и накопления данных в Обнинске станции мировой сети, передающие данные по протоколу LISS, значительно увеличив набор передаваемых данных в реальном времени (рис. 1).

В мае 1994 г. произошло преобразование ЦОМЭ ИФЗ АН СССР в Геофизическую службу РАН [Старовойт и др., 2017], одной из целей которого было проведение планомерной замены во всех подразделениях аналоговой аппаратуры на цифровую как в системах сбора, так и организации архивов. Одним из результатов этих работ стало внедрение цифровых систем сбора совместной разработки коллективом Геофизической службы и ООО «Геотех» – цифровых станций SDAS [Старовойт и др., 2004], в которых в качестве формата сбора был использован формат, совместимый с системой PTCC, а в качестве транспортного протокола – LISS. Это позволило органично провести подключение этих станций к системе NRTS для передачи данных в Обнинск. Перенос кода системы NRTS на платформу Windows одним из разработчиков SDAS А.П. Акимовым значительно упростило эксплуатацию системы сотрудниками сейсмических станций, не обладающими навыками работы с Unix-подобными системами.

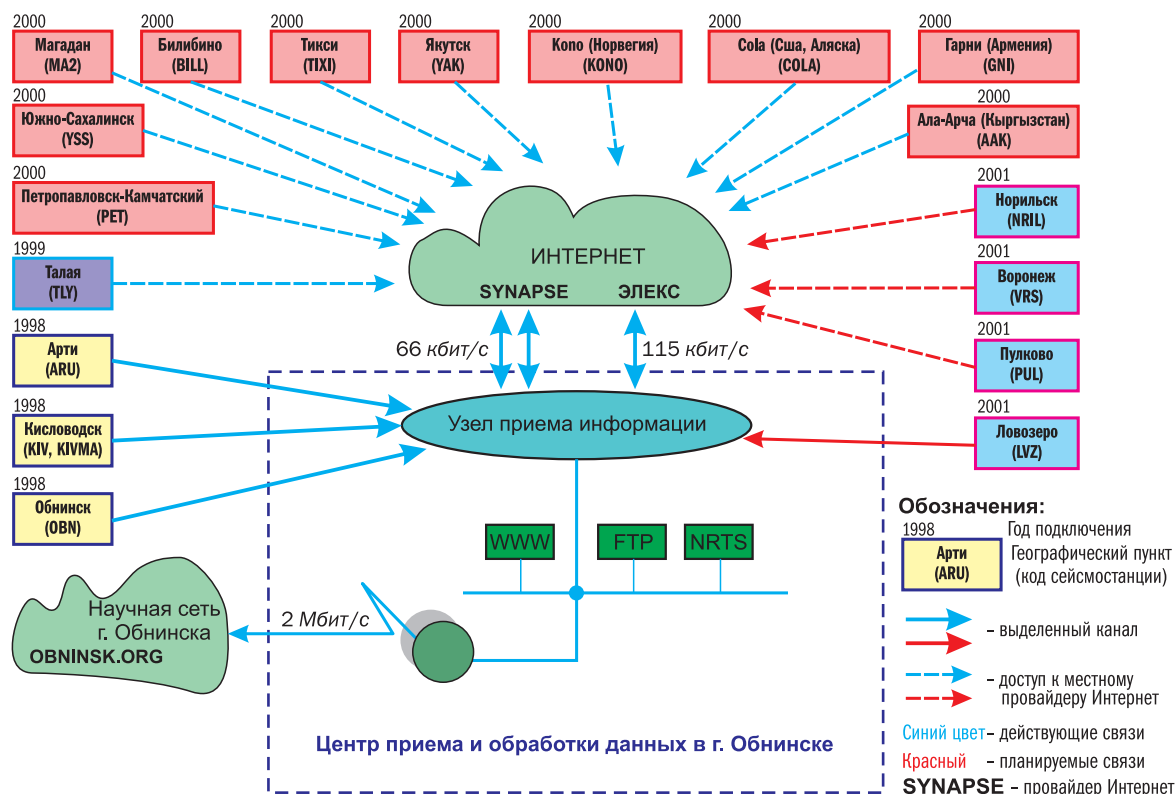


Рис. 1. Структура системы сбора информации с сейсмических станций в Информационно-обработывающем центре в 2001 г.

В начале 2000-х гг. была произведена модернизация системы NRTS на версию 3.x, в которой был введён новый протокол ISI, позволявший инкапсулировать данные других форматов, не меняя их структуры. После внедрения новой версии NRTS для цифровых систем сбора и передачи сейсмологической информации эта система стала основной. В это же время остро встал вопрос реорганизации архива сейсмологических данных, так как разнообразие форматов и структур хранения данных, отсутствие централизованной системы получения сведений о самих сейсмических станциях и их аппаратуре не позволяли организовать оперативные поиск и получение требуемых фрагментов волновых форм из имеющегося архива (рис. 2).

Параллельно с системой NRTS в период с конца 1990-х гг. в сейсмологическом сообществе активно развивались и другие системы накопления и передачи данных, которые имели и дополнительный функционал, позволяющий не только принимать, накапливать и передавать данные, но и получать информацию о самих сейсмических станциях (координаты, описание места установки и аппаратуры, амплитудно-частотные характеристики (АЧХ) каналов в исторической последовательности и др. – так называемое Inventory) и, при наличии, сведения об организации хранения данных в архиве и доступа к ним. Это такие системы как Antelope, EarthWorm, SeisComP, CD1.x и другие [Documentation ..., 2021; CD-1.1 Formats ..., 2021].

Большая часть из них начинала разрабатываться как открытые системы (Antelope, EarthWorm, SeisComP), а CD1.x была разработана для гарантированной и защищённой передачи данных в системе контроля за проведением ядерных испытаний и была доступна только для организаций, авторизованных в Международном центре данных СТВТО без возможности распространения. Разработчики Antelope, EarthWorm (на период с 2006 по 2016 г., сориентировавшись на использование коммерческой базы данных Oracle) после успешного применения на значительном количестве станций и центров отказались от свободного распространения и перешли на коммерческое использование. Система NRTS не обладала расширенным функционалом, что ограничивало область её применения и распространения, кроме того, в 2014 г. основной разработчик Д. Чавез прекратил её развитие и поддержку. В результате основные пользователи NRTS – институты сети IRIS – приняли решение отказаться от её применения, а в качестве альтернативы выбрать систему SeisComP.

С учётом всех аспектов (необходимость реорганизации архива, отсутствие перспектив для используемой системы сбора NRTS, необходимость интеграции с мировым сейсмологическим сообществом) для ФИЦ ЕГС РАН стала актуальной задача выбора единой системы приёма, накопления, передачи и хранения сейсмологической информации. Опираясь на то, что из всех используемых в настоящее время систем только



Рис. 2. Структура архива сейсмологических данных в 2006 г.

SeisComP обладает наибольшим функционалом при наличии некоммерческой версии лицензии, а также имеется немалый опыт её практического применения во всём мире, налажены хорошие контакты с разработчиками, к тому же доступны исходные коды, выбор был сделан в пользу этой системы.

Конструктивные элементы и решения

При разработке новой концепции построения информационной системы были рассмотрены уже существующие и успешно развивающиеся системы, использующие как разработки в пределах ФИЦ ЕГС РАН [Чеброва и др., 2020; Костылев, 2021], так и международную платформу SeisComP [Еманов и др., 2019].

Для решения поставленных задач нами выбрана платформа SeisComP на базе операционной системы Linux Centos7.

В качестве однородного формата хранения данных устанавливается формат miniSEED. Волновые формы расположены по правилам иерархии SeisComP Data Structure (SDS) [Waveform archives, 2021].

Метаданные станций хранятся в формате SC3INV v.11 (SeisComP3 inventory XML – текстовый структурированный формат представления метаданных сейсмических станций для использования непосредственно в системах сбора данных SeisComP), а также в формате Dataless (SEED vol.).

Данные принимаются как в автоматическом, так и в ручном режиме. Автоматический режим предполагает приём данных с использованием плагинов (автоматических потоковых конвертеров) в составе пакета SeisComP: SeedLink, NRTS, CD1.1, REFTEK.

В ручном режиме предусмотрено использование набора программ для автоматизации конвертирования информации из различных форматов в формат miniSEED. Часть программ свободно предоставляется компанией IRIS для операций с miniSEED-файлами [Software ..., 2021]:

- Msmod – программа структуризации данных и модификации информации внутри miniSEED-файлов;

- Dataslect – основная программа для поиска вырезки фрагментов и структуризации данных;

- Rdseed – программа конвертации из miniSEED с требуемые форматы;

- Msi – программа инспектирования данных внутри miniSEED-файлов;

- Ringserver – программа публикации данных со сложной схемой выборки.

Некоторые программы были разработаны в ФИЦ ЕГС РАН:

- All2mseed – конвертация данных из множества различных форматов в miniSEED с последующей модификацией и структуризацией хранения;

- UgraAC3_1w – конвертация данных из формата UGRA в miniSEED.

Все дополнительные приложения и утилиты для автоматического и ручного конвертирования данных, обслуживания HTTP-сервисов созданы на языке программирования Python. Для быстрого доступа к данным и манипуляций с фрагментами волновых форм используется библиотека ObsPy [ObsPy ..., 2021]. Для мониторинга работы сейсмических каналов в пакете SeisComP предусмотрен механизм анализа спектров мощности микросейсмического шума с использованием пакета ObsPy.

Особенности опытной эксплуатации

В октябре 2020 г. Единая система хранения и доступа к геофизическим данным (далее Единая система) в ЦО ФИЦ ЕГС РАН вступила в фазу опытной эксплуатации. На этот момент система включала:

- центральный узел сбора данных, который интегрировал сбор информации со всех необходимых сейсмических сетей;

- архивы волновых форм и фрагментов сейсмических записей в различных форматах, характерных для соответствующих систем сбора;

- систему хранения данных (СХД) или архив волновых форм на основе miniSEED-файлов и специализированной структуры расположения файлов SDS;

- базу данных, ориентированную на формализованный поиск фрагментов волновых форм.

При этом система решала следующие задачи:

- обеспечение резервирования всех видов сейсмических данных (данные при создании СХД копировались и конвертировались в miniSEED с сохранением исходных записей);

- обеспечение двух подходов к хранению данных: реальное время на центральном узле сбора в течение двух недель и архив в формате miniSEED без ограничения времени хранения;

- автоматизированный интерактивный поиск фрагментов сейсмограмм по различным видам запросов (ранее такой сервис отсутствовал);

- контроль и журналирование доступа к сейсмическим данным.

Организация доступа

Для организации управляемого интерактивного доступа к геофизическим данным была выбрана концепция FDSNWS (International Federation of Digital Seismograph Networks Web Service). В рамках этого подхода была зарегистрирована сейсмическая сеть RU [RU: Seismic network ..., 2021]. Было принято решение развивать Единую систему для включения её в состав FDSN (Federation of Digital Seismograph Networks), то есть создавать FDSN Data Center. Такое направление интеграции соответствует мировым тенденциям в сфере сбора и управления сейсмологической информацией [Ottemöller et al., 2021], что подтверждается развитием FDSN-сервисов в составе программного обеспечения для сбора данных SeisComP, выбранного в качестве основного.

На начальном этапе (с 2020 г.) и по настоящее время для доступа к геофизическим данным используется фрагмент сервиса FDSN dataset, обеспечивающий автоматизированный и интерактивный доступ только к волновым формам, тогда как полноценный FDSN-сервис позволяет также получить сведения о каталогах, бюллетенях и метаданных сейсмических станций. Однако даже в ограниченном виде сервис позволяет выполнять выборки данных больших объёмов с высокой производительностью. Ввод в строй центрального узла сбора на основе SeisComP4 позволил использовать встроенный сервис FDSNWS для предоставления метаданных сейсмических станций, участвующих в сборе.

Данные из системы могут быть получены в автоматическом и ручном режимах через HTTP-запрос (HTTP-форму). В автоматическом режиме данные предоставляются в формате miniSEED за любой интервал времени в пределах имеющихся волновых форм. Метаданные могут быть получены через HTTP-запрос (HTTP-форму) в виде SC3INV_xml или FDSN_xml.

Начата реализация «Пилотного проекта системы сбора для совмещённого тестирования новой платформы и системы сбора и обработки, работающей в службе срочных донесений, с целью верификации разрабатываемых организационно-технических решений». В рамках данного проекта был подготовлен специальный сервер на платформе SeisComP4, который обеспечивает:

- интеграцию непрерывных записей глобальных сейсмических сетей II, IU, IM;
- автоматизированное детектирование вступлений;

- автоматическую локацию сейсмических событий;

- автоматический расчёт динамических характеристик событий;

- интерактивную обработку и верификацию сейсмических событий.

Развитие пилотного проекта и работа в приложениях пакета SeisComP4 показали бесперспективность интеграции старой системы обработки WSG [Акимов, Красилов, 2020] в новый пакет программ SeisComP4 [Documentation ..., 2021], т.к. все необходимые инструменты для автоматизированной интерактивной обработки и интерпретации сейсмологической информации присутствуют в комплексе SeisComP4, необходимо только произвести настройку, провести опытную эксплуатацию комплекса и обучить персонал.

Описание сейсмических станций

В 2021 г. проводилась активная инвентаризация метаданных сейсмических станций, используемых при обработке в ФИЦ ЕГС РАН. В рамках этой задачи был организован репозиторий в системе управления данными сейсмических станций SMP (Station Management Portal) [Station ..., 2021].

Репозиторий является базой данных с системой управления компонентами и версиями. Он ориентирован на формат SC3INV, который без дополнительных преобразований включается в систему сбора данных на основе пакета программ SeisComP. В течение года метаданные по станциям сетей II, IU, GE, IM, AN и других, используемых в производственном процессе, были введены в Единую систему в требуемом формате. Предусмотрены механизмы их быстрой и точной актуализации. Особое внимание было уделено станциям сетей Российской Федерации, которые подготавливались с учётом всех исторических изменений. Была разработана методика создания и верификации комплекса метаинформации по сейсмической станции. При этом были учтены как исторические сведения, так и особенности представления данных в современных системах предоставления доступа.

Также была разработана, внедрена и отлажена система потокового конвертирования данных сейсмических станций сети RU, оснащённых сейсмическими регистраторами производства ФИЦ ЕГС РАН UGRA и SDAS. Система является каскадной:

- разнородные данные конвертируются в формат miniSEED в файлы длиной один час

и раскладываются по каталогам с кольцевой ротацией данных;

– сервис `ringserver` один раз в 5 с сканирует каталоги и создаёт буферы данных, предназначенных для публикации на сервисе `SeedLink`;

– центральный узел сбора ФИЦ ЕГС РАН выполняет подключение к сервису `SeedLink` от `ringserver` и консолидирует множество данных, обеспечивая централизованную раздачу и запись в архив.

Представленная на рис. 3 структурная схема Единой системы демонстрирует все основные элементы сбора, хранения и предоставления геофизической информации. Видно, что определяющими являются центральный узел, файловый архив и облачный сервис.

Развёртывание и поддержка этих компонентов являются приоритетными задачами, в них закладывается максимальное количество материальных и интеллектуальных ресурсов.

Узел сбора данных работает в виртуальной среде на основе гипервизора ядра `ESXi 5.0`. Для поддержания бесперебойной работы предусмотрен резервный узел сбора с полной копи-

ей конфигурации без использования виртуализации. Таким образом, сформирован «холодный» резерв. Запуск дополнительного узла сбора осуществляется в течение 20 минут. Миграция и восстановление виртуальной машины основного узла составляют 40 минут. Возобновление сбора на основном узле занимает ещё 10 минут, таким образом «холодное» резервирование даёт два перерыва в сборе данных в сумме 30 минут. Этот показатель больше, чем время обработки события в службе срочных донесений. В связи с этим нами предполагается развернуть резервный узел в виртуальной среде, кластеризовать виртуальные машины для сбора данных и выполнять «горячее» резервирование с перерывом в процессе сбора и передачи данных не более 5 минут.

Файловый архив выполнен на основе отказоустойчивой системы сбора данных с полным территориально распределённым зеркалированием. Дисковый массив оснащён системами мониторинга технологических параметров. Всё это обеспечивает своевременную реакцию на технические инциденты и полное восстановление информации в случае аварии.

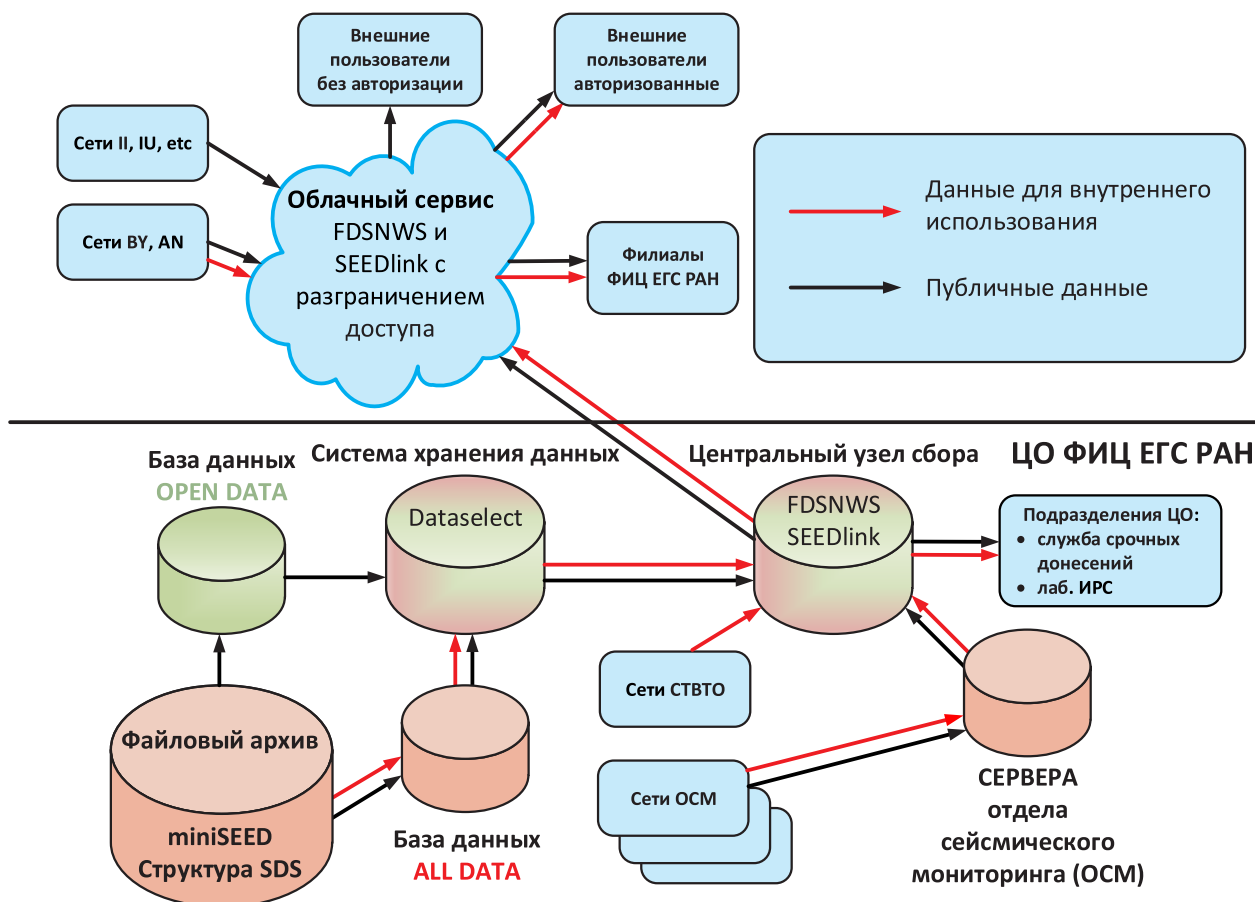


Рис. 3. Структурная схема организации данных в Единой системе хранения и доступа к геофизическим данным

Облачный сервис

Облачный сервис, который на первом этапе был вторичным по отношению к центральному узлу сбора, на текущий момент получил статус полноценного резерва на случай сетевых или аппаратных сбоев центральной системы. В дальнейшем облачный сервис станет первичным, благодаря всестороннему широкополосному отказоустойчивому доступу, а также высокой степени технической готовности, обеспечиваемой современным центром обработки данных компании МТС. Таким образом, решается задача минимизации накладных расходов на многократное копирование данных при сборе и публикации и значительно повышается отказоустойчивость системы. Практика показала, что наибольшее количество ретроспективных запросов к волновым формам укладывается в два последних года, поэтому в дальнейшем планируется расширение дискового пространства облачного сервиса для хранения информационного кольца с периодом накопления данных в расчёте на два года. Уже сейчас по отношению к внешним сетям облачный сервис является первичной системой сбора, и только часть информации, которая передаётся по сегментам внутренней сети, продолжает собираться на центральном узле.

На рис. 3 видно, что потоки данных условно разделены на публичные и для внутреннего использования, для которых предусмотрены открытый и авторизованный методы доступа соответственно. Данные предоставляются только двумя сервисами: SeedLink и FDSNWS. Для SeedLink не предусматривается авторизация, поэтому доступ предоставляется по «белому» списку IP-адресов. Сервис FDSNWS имеет штатные средства ограничения доступа, связанные со списками сетей, станций, каналов. Для сложных алгоритмов выборки реализована и активно используется схема формирования отдельных индексных баз данных, на основании которых выполняется выборка информации из единого центрального файлового архива.

Заключение

По результатам опытной эксплуатации Единой системы можно сделать следующие заключения.

- Создан автоматизированный высокопроизводительный сервис доступа к геофизическим данным на основе только стандартных программных средств с использованием программного обеспечения с открытым кодом.

- Сейчас в режиме, близком к реальному времени, в архив Единой системы поступают данные со 191 станции (из них 82 станции в составе сети RU), расположенных преимущественно на европейской части РФ. Данные некоторых станций добавляются вручную. Список станций может быть расширен за счёт привлечения американских, европейских станций и станций сетей восточной части России (Камчатки, Байкала и Алтае-Саянского региона).

- На текущий момент SDS-архив содержит 70% имеющихся исторических данных в формате miniSEED с возможностью автоматизированной раздачи. Постоянно происходит его дальнейшее пополнение.

- Облачное хранилище может предоставлять полный и ограниченный доступ ко всем станциям, собираемым в ЦО ФИЦ ЕГС РАН в режиме реального времени и ретроспективно.

- Использование облачного сервиса повышает безопасность и устойчивость доступа к информационным ресурсам ФИЦ ЕГС РАН и эффективно оптимизирует трафик при большом числе запросов или больших объёмах запрашиваемых данных.

- Уже сейчас ФИЦ ЕГС РАН может входить в число центров обработки данных на основе FDSN-сервисов.

Сервисы Единой системы имеют большой потенциал для дальнейшего развития. Среди ближайших наиболее перспективных задач можно отметить:

- разработку системы публикации сканов аналоговых сейсмограмм;

- интеграцию геодинамических данных в основной архив с возможностью вывода в режиме реального времени и публикации на основе FDSN;

- развёртывание системы автоматизированного контроля работоспособности каналов сейсмических станций на основе встроенных средств SeisComP.

Литература

Акимов А.П., Красилов С.А. Программный комплекс WSG «Система обработки сейсмических данных» / Свидетельство о государственной регистрации программы для ЭВМ № 2020664678 от 16.11.2020 г.

Антонов В.В. и др. Отчёт о сейсмических наблюдениях на Кавминводском полигоне Северного Кавказа Российской Федерации в 1991 г. – Обнинск: Фонды ГС РАН, 1992.

Еманов А.Ф., Бах А.А., Еманов А.А. Инженерно-сейсмологический мониторинг зданий и сооружений:

физико-математические основы метода, возможности, результаты // Научный вестник Арктики. – 2019. – № 7. – С. 34–43.

Костылев Д.В. Формирование единой системы сбора сейсмологической информации в Сахалинском филиале ФИЦ ЕГС РАН // Российский сейсмологический журнал. – 2021. – Т. 3, № 1. – С. 41–53. doi: 10.35540/2686-7907.2021.1.03

Старовойт О.Е. Сейсмологический центр в Обнинске в 1963–2003 гг. / Отв. ред. А.Я. Сидорин. – М.: ИФЗ РАН, 2017. – С. 35–44.

Старовойт О.Е., Габсатарова И.П., Мехрюшев Д.Ю., Коротин А.В., Красилов С.А., Галушко В.В., Коломиец Ю.Н., Пойгина С.Г., Каменская О.П. Исследование, разработка и создание в Российской Федерации системы сейсмических и геодинамических наблюдений для непрерывного национального и глобального сейсмического мониторинга. Отчёт по договору № 01.700.12.0094 от 01.10.2004. – Обнинск: Фонды ГС РАН, 2004. – С. 77.

Чеброва А.Ю., Чемарёв А.С., Матвеев Е.А., Чебров Д.В. Единая информационная система сейсмологических данных в Камчатском филиале ФИЦ ЕГС РАН: принципы организации, основные элементы, ключевые функции // Геофизические исследования. – 2020. – Т. 21, № 3. – С. 66–91. doi: 10.21455/gr2020.3-5

Chavez D., Berger J. The IDA Near Real Time System. – June 1997. [Электронный ресурс]. – URL: <http://www.gsras.ru/new/formats/NRTS/NRTS.pdf> (дата обращения 21.10.2021).

Documentation for the SeisComP // SeisComP 4.7.2 [сайт]. – URL: <https://www.seiscomp.de/doc/index.html> (дата обращения 21.10.2021).

CD-1.1. Formats and Protocols for Continuous Data. Draft 3. August, 1999 [Электронный ресурс]. –

URL: <http://www.gsras.ru/new/formats/IMS/CD%201.1%20Formats%20and%20Protocols%20for%20Continuous%20Data.pdf> (дата обращения 21.10.2021).

Elger K., Pampel H., Biskaborn B., Lantuit H. Open research data, Data portals and Data publication – an introduction to the data curation landscape // Polarforschung. – 2016. – V. 85, N 2. – P. 119–133.

Hutt, C.R., Bolton, H. Live Seismograms from the Net [Электронный ресурс]. – URL: <http://ftp.iris.edu/news/IRISnewsletter/volume1999no1/page02-04.htm> (дата обращения 21.10.2021).

Ottmöller L., Michalek J., Halpaap F. et al. UiB-NORSAR EIDA Node: Integration of Seismological Data in Norway // Seismological Research Letters. – 2021. – V. 92, N 3. – P. 1491–1500. doi: 10.1785/0220200369

RU: Seismic network of Russian Federation. FDSN Network Information // FDSN [сайт]. – URL: <https://www.fdsn.org/networks/detail/RU/> (дата обращения 21.10.2021).

SEED // IRIS [сайт]. – URL: <http://ds.iris.edu/ds/nodes/dmc/data/formats/seed/> (дата обращения 21.10.2021).

Station Management Portal [сайт]. – URL: <https://smp.gempa.de/> (дата обращения 21.10.2021).

Software and projects by the IRIS Data Management Center [Электронный ресурс]. – URL: <https://github.com/orgs/iris-edu/repositories> (дата обращения 21.10.2021).

ObsPy – a Python framework for Seismology [Электронный ресурс]. – URL: <https://github.com/obspy/obspy/wiki/> (дата обращения 21.10.2021).

Waveform archives [Электронный ресурс]. – URL: <https://www.seiscomp.de/doc/base/concepts/waveformarchives.html?highlight=sds> (дата обращения 21.10.2021).

Сведения об авторах

Бутырин Павел Генрихович, канд. техн. наук, ст. науч. сотр. Федерального государственного учреждения науки Федерального исследовательского центра «Единая геофизическая служба Российской академии наук» (ФИЦ ЕГС РАН), г. Обнинск, Россия. E-mail: pbg2000@mail.ru

Красилов Сергей Александрович, зав. сектора АПО ФИЦ ЕГС РАН, г. Обнинск, Россия. E-mail: krasilovs@gsras.ru

The unified system for storing and accessing geophysical data. Traditions and new approaches

© 2021 P.G. Butyrin, S.A. Krasilov

GS RAS, Obninsk, Russia

Received November 9, 2021

Abstract The features of the development of data collection systems within the Information Processing Center (IPC) of the Geophysical Survey of the Russian Academy of Sciences (GS RAS) are presenting. Historical information is given, including technical details related to the systematization of the archive and the evolution of geophysical data formats. The historical, territorial features, as well as the experience of deploying such information systems within the Federal Research Center of the Unified State Social Service of the Russian Academy of Sciences and abroad are taken into account. A new concept of building an information system is proposed, which takes into account the requirements for scalability, reproducibility at various objects and the use of standard software. A lot of work was done to form a homogeneous archive of waveforms and an inventory of metadata for seismic stations, which resulted in the possibility of including the GS RAS in the international centers for processing geophysical information based on FDSN. Creation of a distributed collection and processing system using a cloud service allows abstracting from the territorial features of collecting and storing geophysical information, which increases the performance of the data access service and the degree of technical readiness of key system nodes.

Keywords Seismological monitoring, data collection system, information system, geophysical data, universal access.

For citation Butyrin, P.G., & Krasilov, S.A. (2021). [The unified system for storing and accessing geophysical data. Traditions and new approaches]. *Rossiiskii seismologicheskii zhurnal* [Russian Journal of Seismology], 3(4), 77-87. (In Russ.). DOI: <https://doi.org/10.35540/2686-7907.2021.4.05>

References

- Akimov, A.P., & Krasilov, S.A. (2020). [WSG software package “Seismic data processing system”]. Certificate of state registration of a computer program No. 2020664678. (In Russ.).
- Antonov, V.V. et al. (1992). *Otchet o seismicheskikh nabludeniakh na Kavminvodskom poligone Severnogo Kavkaza Rossiiskoi Federatsii v 1991 g.* [Report on seismic observations at the Kavminvodsky test site of the North Caucasus of the Russian Federation in 1991]. Obninsk, Russia: Funds GS RAS. (In Russ.).
- CD-1.1. *Formats and Protocols for Continuous Data. Draft 3. August, 1999* (2021). Available at: <http://www.gsras.ru/new/formats/IMS/CD%201.1%20Formats%20and%20Protocols%20for%20Continuous%20Data.pdf>
- Chavez, D. & Berger, J. (1997). The IDA Near Real Time System. Available at: <http://www.gsras.ru/new/formats/NRTS/NRTS.pdf>
- Chebrova, A.Yu., Chemarev, A.S., Matveenkov, E.A., & Chebrov, D.V. (2020). [Seismological data information system in Kamchatka Branch of GS RAS: organization principles, main elements and key functions]. *Geofizicheskie issledovaniia* [Geophysical Research], 21(3), 66-91. (In Russ.). doi: 10.21455/gr2020.3-5
- Elger, K., Pampel, H., Biskaborn, B., & Lantuit, H. (2016). Open research data, Data portals and Data publication - an introduction to the data curation landscape. *Polarforschung*, 85(2), 119-133.
- Emanov, A.F., Bach, A.A., & Emanov, A.A. (2019). [Engineering and seismological monitoring of buildings and structures: physical and mathematical bases methods, opportunities, results]. *Nauchnyi vestnik Arktiki* [Scientific Bulletin of the Arctic], 7, 34-43. (In Russ.).
- FDSN. (2021). RU: Seismic network of Russian Federation. FDSN Network Information. Retrieved from <https://www.fdsn.org/networks/detail/RU/>
- Hutt, C.R., & Bolton, H. (2021). Live Seismograms from the Net. Available at: <http://ftp.iris.edu/news/IRISnewsletter/volume1999no1/page02-04.htm>
- IRIS. (2021). SEED. Retrieved from <http://ds.iris.edu/ds/nodes/dmc/data/formats/seed/>
- Kostylev, D.V. (2021). [Formation of a unified system for collecting seismological information in the Sakhalin branch of the GS RAS]. *Rossiiskii seismologicheskii zhurnal* [Russian Journal of Seismology], 3(1), 41-53. (In Russ.). doi: 10.35540/2686-7907.2021.1.03
- ObsPy - a Python framework for Seismology (2021). Retrieved from <https://github.com/obspy/obspy/wiki/>

- Ottewill, L., Michalek, J., Halpaap, F. et al. (2021). UiB-NORSAR EIDA Node: Integration of Seismological Data in Norway. *Seismological Research Letters*, 92(3), 1491-1500. doi: 10.1785/0220200369
- SeisComP 4.7.2*. Documentation for the SeisComP. (2021). Retrieved from <https://www.seiscomp.de/doc/index.html>
- Software and projects by the IRIS Data Management Center* (2021). Retrieved from <https://github.com/orgs/iris-edu/repositories>
- Starovoit, O.E., Gabsatarova, I.P., Mekhryushev, D. Yu., Korotin, A.V., Krasilov, S.A., Galushko, V.V., Kolomiets, Yu.N., Poigina, S.G., & Kamenskaya, O.P. (2004). *Issledovanie, razrabotka i sozдание v Rossiiskoi Federatsii sistemy seismicheskikh i geodinamicheskikh nabludeni dlia nepreryvnogo natsional'nogo i global'nogo seismicheskogo monitoringa. Otchet po dogovoru № 01.700.12.0094 ot 01.10.2004* [Research, development and creation in the Russian Federation of a system of seismic and geodynamic observations for continuous national and global seismic monitoring. Report under contract No. 01.700.12.0094 dated 01.10.2004] (p. 77). Obninsk, Russia: Funds GS RAS. (In Russ.).
- Starovoit, O.E. (2017). *Seismologicheskii tsentr v Obninske v 1963–2003 gg. Otv. red. A.Ia. Sidorin* [Seismological center in Obninsk in 1963–2003. Resp. ed. A.Ya. Sidorin] (pp. 35–44). Moscow, Russia: IPE RAS. (In Russ.).
- Station Management Portal*. (2021). Retrieved from <https://smp.gempa.de/>
- Waveform archives* (2021). Retrieved from <https://www.seiscomp.de/doc/base/concepts/waveformarchives.html?highlight=sds>

Information about authors

Butyrin Pavel Genrikhovich, PhD, Senior Researcher of the Geophysical Survey of the Russian Academy of Sciences (GS RAS), Obninsk, Russia. E-mail: pbg2000@mail.ru

Krasilov Sergei Alexandrovich, Head of APO Department of the GS RAS, Obninsk, Russia. E-mail: krasilovs@gsras.ru