

P10-2018-23

А. С. Кирилов, И. А. Морковников

О КОНЦЕПЦИИ ФАЙЛОВОГО ХРАНИЛИЩА
ДЛЯ СПЕКТРОМЕТРОВ **ИЯУ ИБР-2**

Кирилов А. С., Морковников И. А.

P10-2018-23

О концепции файлового хранилища для спектрометров ИЯУ ИБР-2

Работа посвящена концепции центрального хранилища данных измерений со спектрометров ИЯУ ИБР-2 ЛНФ ОИЯИ. В концепции учтены разнообразие структур и форматов файлов данных на конкретных спектрометрах, пожелания пользователей, проведена оценка объема требуемого дискового пространства. Рассмотрены вопросы организации сервиса для работы с данными в хранилище с учетом имеющейся системы автоматической регистрации результатов измерений Journal.

Работа выполнена в Лаборатории нейтронной физики им. И. М. Франка ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2018

Kirilov A. S., Morkovnikov I. A.

P10-2018-23

About the Concept of the File Repository
for Instruments at the IBR-2 Pulsed Reactor

The work is devoted to the concept of a central repository of measurement data from the IBR-2 instruments at the FLNP, JINR. The concept takes into account the variety of structures and data file formats on specific instruments, the users' wishes, and estimated amount of disk space required. The questions of the service organization for the work with the data into the storage were considered, taking into account the existing system of automatic measurements registration Journal.

The investigation has been performed at the Frank Laboratory of Neutron Physics, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2018

АНАЛИЗ ТЕКУЩЕЙ СИТУАЦИИ

Разработка и внедрение централизованного архива для хранения результатов измерений являются одним из основных направлений развития программного обеспечения спектрометров ИЯУ ИБР-2 [1]. Работы в этом направлении ведутся во многих нейтронных центрах [2, 3]. В настоящее время наиболее популярной является система ICAT [4]. Проект ICAT предоставляет каталог и соответствующие компоненты для поддержки архива экспериментальных данных в крупных научных центрах, связывающих все аспекты исследовательской цепочки — от предложений экспериментов до публикации результатов.

Прямое использование системы ICAT в ЛНФ невозможно по ряду причин. Прежде всего, данная система рассчитана на единый формат данных, содержащий всю необходимую для их интерпретации информацию. В других нейтронных центрах одним из самых популярных форматов в настоящее время является NeXus [5]. В лаборатории до настоящего времени единый формат не принят и, вероятно, в ближайшее время принят не будет.

ФОРМАТЫ ДАННЫХ СПЕКТРОМЕТРОВ

На спектрометрах ИЯУ ИБР-2 установлена унифицированная система управления SONIX+ [6, 7]. Первоначально экспериментальные данные сохраняются во внутреннем формате SONIX+, после чего конвертируются в представление, необходимое пользователям. Табл. 1 иллюстрирует текущее положение дел. Форматы различаются по:

- способу представления данных (binary, ASCII, SONIX+);
- составу файлов (гистограммы, списки событий, inf-файлы и т. д.);
- способам именования и структурам каталогов.

Это определено уникальностью программ обработки экспериментальных данных на спектрометрах.

ЦЕНТРАЛЬНОЕ ХРАНИЛИЩЕ ДАННЫХ ЛНФ

В лаборатории организовано хранилище данных на центральном сервере [8], которое обеспечивает сохранение, удаленный доступ к измеренной спектрометрической информации и поиск данных по параметрам эксперимента с помощью системы регистрации Journal [9]. В настоящее время этот сервис доступен не для всех спектрометров, но их список периодически пополняется.

Размер и надежность хранилища обеспечиваются за счет аппаратного RAID-6 массива, состоящего из 24 дисков, общим объемом 54 ТБ. Хранилище остается работоспособным при выходе из строя до 2 жестких дисков и поддерживает возможность их горячей замены.

Доступ к хранилищу данных для чтения и записи обеспечивается с помощью протокола sftp.

Таблица 1. Форматы данных спектрометров ИЯУ ИБР-2

№ канала	Спектрометр	Состав данных	Формат	Архив
2	ДИН-2ПИ	Файл с заголовком и гистограммами всех детекторов в виде таблицы	ascii	Нет
4	ЮМО	Файл с данными детекторов (последовательная запись измерений по числу итераций) Файл с метаданными	binary	Нет
5	ФДВР	Файлы со списком событий для отдельных МРD-контроллеров Гистограммы для отдельных детекторов Суммарные гистограммы для трех групп детекторов	xml binary binary binary	Нет
6а	РТД	Протокол длительности свипов (ASCII-файл)	ascii	Нет
6б	ДН-6	Гистограммы со всех детекторов (ПЧД и точечных)	binary	Нет
7а	ЭПСИЛОН	Данные в формате SONIX+	SONIX+	Нет
7б	СКАТ	Отдельные гистограммы для групп детекторов по числу итераций Отдельные гистограммы для всех детекторов по числу позиций (72 файла)	binary binary	Нет
7в	НЕРА-ПР	Гистограммы анализаторов (d, m, q, i), суммарные файлы для анализаторов (все итерации) Протокол температурного контроллера	binary ascii ascii	Нет
8	РЕМУР	Протокол эксперимента Данные по числу итераций для всех состояний флипперов или суммарные файлы	SONIX+	Да
9	РЕФЛЕКС	Данные в формате SONIX+ по числу итераций для всех состояний флипперов Суммарный файл InF-файл	SONIX+ SONIX+ ascii	Да
10	ГРЭИНС	Файлы с измеренными данными Данные по числу итераций для всех состояний флипперов и суммарные файлы	ascii SONIX+	Да
11	ФСД	Файл со списком событий	binary	Нет
12	ДН-12	Гистограммы для отдельных детекторов Данные в формате SONIX+	binary binary SONIX+	Нет
13	ФСС	Не определено до конца тестовых испытаний. Вероятно, аналогично ФСД или ФДВР	binary (?)	Нет
14		Радиографические снимки		Нет

ПРОГРАММА РЕГИСТРАЦИИ JOURNAL

В лаборатории разработана система автоматической регистрации проведенных измерений Journal [9]. Система внедрена на спектрометрах ЮМО, ФДВР и ГРЭИНС. Регистрируемая информация сохраняется в специализированной базе данных на одном из центральных серверов лаборатории. Программа поддерживает функции поиска измерений и составления отчетов.

Journal (рис. 1) состоит из следующих компонентов:

- интерфейс пользователей;
- база для хранения регистрационных данных;
- модуль, содержащий процедуры автоматической регистрации.

При разработке программы особое внимание уделялось возможности адаптации к пожеланиям пользователей разных спектрометров. Для унификации была разработана модель SQL-шаблонов. Часть списка регистрируемых параметров является общей для всех спектрометров. Остальные выбираются по согласованию с ответственным за установку. Интерфейс пользователя и функция регистрации адаптируются под шаблон, описывающий структуру данных в хранилище.

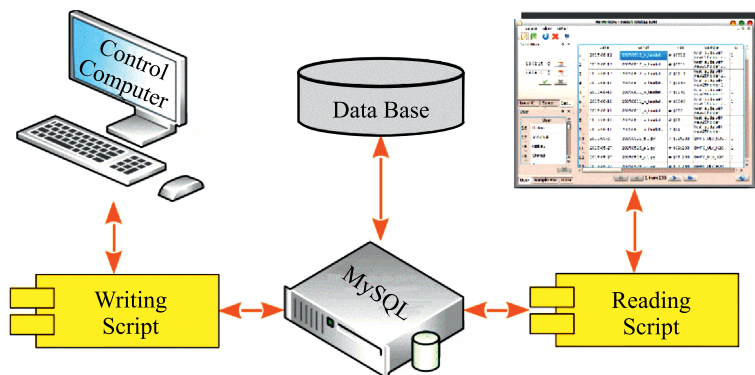


Рис. 1. Структура программы Journal

ОЦЕНКА НЕОБХОДИМОГО ОБЪЕМА ХРАНИЛИЩА

В табл. 2 приведена оценка текущей потребности в дисковом пространстве хранилища. Наибольшая потребность отмечена для фурье-спектрометров и рефлектометров. На январь 2018 г. суммарная годовая потребность ФСД составляет около 6 ТБ, а ФДВР — 4 ТБ. Потребность всех рефлектометров составляет около 3 ТБ. Вклад остальных спектрометров составляет менее 100 ГБ в год, на данном этапе им можно пренебречь. Таким образом, суммарная годовая потребность составляет около 13–14 ТБ.

Таблица 2. Потребность в объеме сохраняемых данных спектрометров по оценке пользователей

Номер канала	Спектрометр	Данные за год (ТБ)	Тип спектрометра
2	ДИН-2ПИ		
4	ЮМО		
5	ФДВР	4	Фурье
6а	РТД		
6б	ДН-6	0,005	
7а	ЭПСИЛОН		
7б	СКАТ	0,0015	
7в	НЕРА-ПР	0,002	
8	РЕМУР	1	Рефлектометр
9	РЕФЛЕКС	1	Рефлектометр
10	ГРЭИНС	1	Рефлектометр
11	ФСД	6	Фурье
12	ДН-12	0,001	
13	ФСС		Фурье
14		1	

В будущем с учетом роста потока данных на фурье-спектрометрах до 6 ТБ в год, введения в эксплуатацию нового фурье-спектрометра ФСС годовая потребность увеличится до 21 ТБ. Кроме того, применение нового контроллера DeLiDaq-2 на рефлектометрах также способно повлиять на результат.

В целом существующий объем дискового пространства хранилища достаточен для сохранения экспериментальных данных спектрометров ИЯУ ИБР-2 как минимум на 2–3 года.

ПОЖЕЛАНИЯ ПОЛЬЗОВАТЕЛЕЙ ПО СОСТАВУ И ДИСЦИПЛИНЕ РАБОТЫ С ДАННЫМИ В ХРАНИЛИЩЕ

По результатам проведенного опроса, а также обсуждения на семинаре по методике и технике нейтронного эксперимента НЭОКС ИБР-2 ОНИРКС ЛНФ пользователи ожидают получить автоматическое сохранение:

- спектрометрических данных;
- протоколов измерения;
- протоколов устройств окружения образца;
- другой информации (например, фотографий образца).

Желательно, чтобы структура файлов в хранилище совпадала со структурой файлов на управляющем компьютере спектрометра. В ряде случаев требуется ручная запись данных в хранилище.

ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

Программное обеспечение для работы с хранилищем. Общая структура хранилища и сервиса для работы с ними приведена на рис. 2.

Data Storage организует хранение и доступ к экспериментальным данным. Auto Upload автоматизирует резервное копирование данных от управляющих компьютеров. Система Journal выполняет регистрацию измерений и организацию каталога данных. GUI — графический интерфейс пользователя.

С учетом разницы в форматах и составе сохраняемой информации решение должно быть адаптировано к требованиям спектрометров. Как отмечено в п. «Влияние форматов данных на сервис пользователя», предполагаемый сервис не может быть единым для всех случаев. Выбор унифицированного формата данных (SONIX+, NeXus) позволяет организовать развитый сервис с поиском по любому параметру, регистрируемому в системе управления измерением, и визуализацию данных непосредственно из хранилища.

Понятие эксперимента. Для удобства группировки данных в хранилище введем понятие «эксперимент». Он может состоять из одного или последовательности измерений, например итераций, которые имеют общую цель. Содержание эксперимента, а также состав сохраняемых данных для каждого спектрометра определяются по согласованию с ответственным за установку и программируются в библиотеке операций спектрометра (см. приложение 1).

Эксперименты различаются по идентификатору, сохраняемому в базе данных (БД). Идентификатор вырабатывается автоматически и служит для организации поиска в хранилище. Для каждого спектрометра при сохранении данных в хранилище создается структура каталогов по правилам, принятым на спектрометре.

Дисциплина работы с данными в хранилище. Копирование данных в хранилище автоматизировано. Удаление данных осуществляется по просьбе ответственных за спектрометры исключительно администратором хранилища.

Если процесс измерения организован итеративно, то данные по итерациям передаются по мере их появления. Протокол эксперимента (см. приложение 3), а также суммарные файлы передаются в хранилище после завершения эксперимента. Перезапись информации в хранилище запрещена. При сохранении данных проводится автоматическая проверка имени файла. При попытке сохранить файл с существующим именем это имя корректируется автоматически.

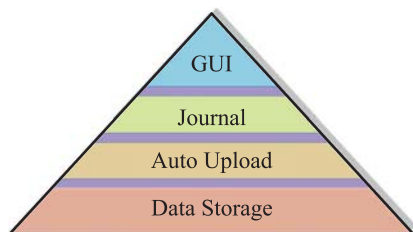


Рис. 2. Общая структура хранилища и сервиса для работы с ними

В некоторых случаях ответственные за спектрометр имеют возможность записывать данные вручную, а именно:

- на установке радиографии (14-й канал), где автоматизация отсутствует;
- в случае сбоя автоматического копирования данных;
- для ручного резервного копирования большого (более 1 ТБ) объема данных.

Во время цикла данную деятельность по ряду причин целесообразно ограничить.

Влияние форматов данных на сервис пользователя. Формат SONIX+ образован совокупностью двух файлов: файла со спектрометрической информацией и файла с так называемым снимком БД Varman [10]. По сравнению с другими форматами, используемыми в ЛНФ, последний имеет три важных преимущества:

- снимок БД содержит всю зарегистрированную в системе управления информацию об измерении (так называемые метаданные), включая конкретные значения параметров всех используемых устройств, имена файлов и т. д. Полнота и актуальность информации позволяют организовать максимально содержательную обработку измерений;

- помимо значений, снимок БД содержит описания структур сохраняемых параметров. Следовательно, при изменении этих структур (что происходит время от времени) возможность правильного извлечения параметров сохраняется;

- есть однозначное соответствие между данными (спектры) и значениями параметров (снимок БД). Следовательно, можно организовать поиск, просматривая снимки БД непосредственно. Это более затратно, чем в случае использования программы Journal, но зато список параметров поиска максимально полон.

Такими же преимуществами обладает и формат NeXus. Он регламентирует создание контейнера для всех данных, связанных с измерением на спектрометре, т. е. полную информацию об измерении.

В остальных случаях:

- метаданные приведены либо частично (inf-файлы), либо отсутствуют;
- способ записи метаданных не регламентирован;
- конкретные значения параметров извлечь затруднительно, так как они сохранены различным образом в разных файлах. Ассоциация между спектром и детектором часто произвольно кодируется в имени.

Таким образом, наиболее развитый сервис, включающий поиск по значениям параметров и визуализацию спектров, можно организовать для спектрометров, сохраняющих файлы в формате SONIX или NeXus. В остальных случаях визуализация затруднена либо ограничена. Без сохранения снимков БД возможности поиска с помощью системы Journal лимитированы списком сохраняемых в ее БД параметров.

Интерфейс пользователя для работы с хранилищем. Интерфейс пользователя для работы с хранилищем данных предполагает наличие следующих функций:

- просмотр структуры каталогов и загрузка результатов измерений из файлового хранилища. Можно воспользоваться готовыми приложениями TotalComander, FileZilla и другими;
- поиск результатов измерений в файловом хранилище по параметрам проведенного измерения (параметры определяются ответственными за установку);
- визуализация спектров в формате SONIX+;
- отображение снимка БД;
- вывод протокола эксперимента или протоколов устройств окружения образца и т. п.;
- вывод скрипта измерения;
- вывод изображений, записанных в результатах измерений.

Авторы выражают глубокую признательность всем коллегам, принявшим активное участие в дискуссии по вопросам, связанным с концепцией, за полезные идеи и конструктивную критику.

Приложение 1. СОСТАВ ДАННЫХ ЭКСПЕРИМЕНТА

В состав сохраняемых данных по согласованию с ответственным за спектрометр входят:

- файлы со спектрами в формате, принятом на спектрометре;
- снимок БД;
- протокол эксперимента;
- скрипт эксперимента;
- другая информация по согласованию с ответственным за установку, например фотографии, inf-файлы и т. д.

Приложение 2. СПИСОК ОБЯЗАТЕЛЬНЫХ ПАРАМЕТРОВ В СНИМКЕ БАЗЫ ДАННЫХ

Снимок БД, помимо прочего, должен содержать следующие параметры:

- идентификатор эксперимента;
- время начала эксперимента;
- время начала измерения;
- текущее время;
- длительность экспозиции;
- имя образца;
- имя пользователя;
- имя ответственного за установку;

- имя эксперимента;
- набор уникальных параметров, задаваемых пользователями, индивидуальный для каждого спектрометра.

Приложение 3. СОСТАВ ПРОТОКОЛА ЭКСПЕРИМЕНТА

В файле протокола записываются:

- имя файла протокола;
- идентификатор эксперимента;
- имя ответственного за установку;
- имя пользователя;
- имя образца;
- время начала эксперимента;
- время окончания эксперимента;
- длительность суммарной экспозиции;
- комментарий;
- отметки о событиях и проведенных измерениях.

ЛИТЕРАТУРА

1. *Kirilov A. S.* Instrument Control Software at the IBR-2: Directions of Development. JINR Communication E10-2017-89. Dubna, 2017.
2. <http://www.nobugsconference.org/>
3. <https://indico.esss.lu.se/event/357/>
4. <https://icatproject.org/>
5. <http://www.nexusformat.org/>
6. *Кирилов А. С.* Современное состояние и перспективы развития программного обеспечения комплекса спектрометров реактора ИБР-2М. Препринт ОИЯИ Р10-2011-101. Дубна, 2011.
7. <https://sonix.jinr.ru/wiki/doku.php?id=:ru:index>
8. *Морковников И. А., Кашунин И. А., Сухомлинов Г. А.* Организация передачи данных с Sonix+ на файловый сервер Nfserv-b. Сообщение ОИЯИ Р10-2014-101. Дубна, 2014.
9. *Морковников И. А., Иванов О. И., Кирилов А. С.* Программа ведения рабочего журнала на спектрометре ЮМО. Сообщение ОИЯИ Р10-2012-134. Дубна, 2012.
10. *Кирилов А. С., Юдин В. Е.* Реализация базы данных реального времени для управления экспериментом в среде MS Windows. Сообщение ОИЯИ Р13-2003-11. Дубна, 2003.

Получено 26 апреля 2018 г.

Редактор *А. И. Петровская*

Подписано в печать 19.06.2018.

Формат 60 × 90/16. Бумага офсетная. Печать офсетная.

Усл. печ. л. 0,68. Уч.-изд. л. 0,77. Тираж 200 экз. Заказ № 59428.

Издательский отдел Объединенного института ядерных исследований

141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.

E-mail: publish@jinr.ru

www.jinr.ru/publish/