

САРАТОВСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ
им. Н. Г. Чернышевского

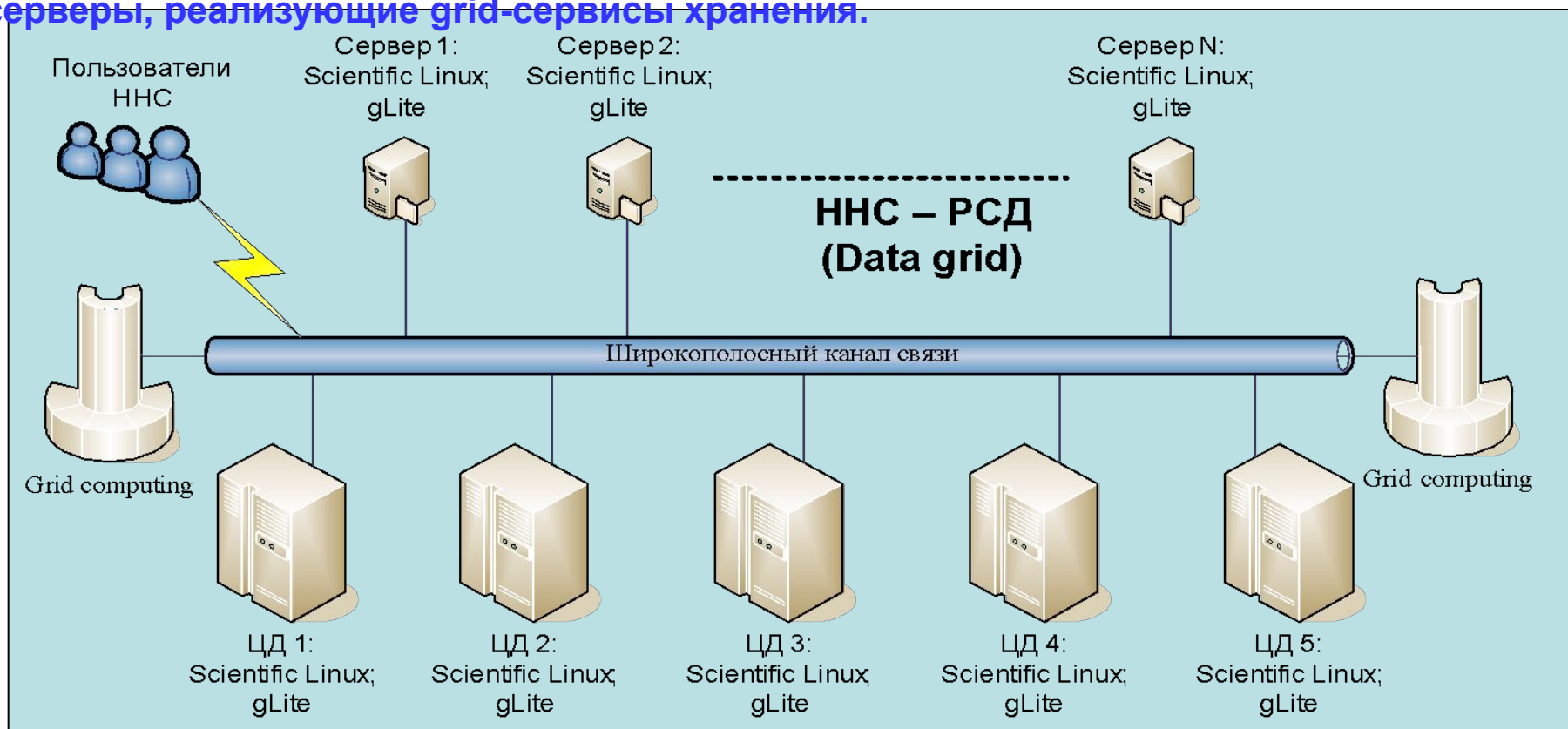
*Полномасштабное распределенное
хранилище данных национальной
нанотехнологической сети*

ПОВОЛЖСКИЙ
РЕГИОНАЛЬНЫЙ
ЦЕНТР
НОВЫХ
ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ

Соловьев Владимир Михайлович

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

В 2009 г. будет создана полномасштабная система распределенного сбора, хранения, обработки и управления потоками научных и технологических данных национальной нанотехнологической сети (ННС). Основу масштабируемой распределенной системы данных (РСД) составят пять онлайн-центров данных (ЦД) ННС, объединенных широкополосными каналами связи. В дальнейшем к этим ЦД, для наращивания объемов хранимой информации, могут подключаться географически удаленные ЦД и серверы, реализующие grid-сервисы хранения.



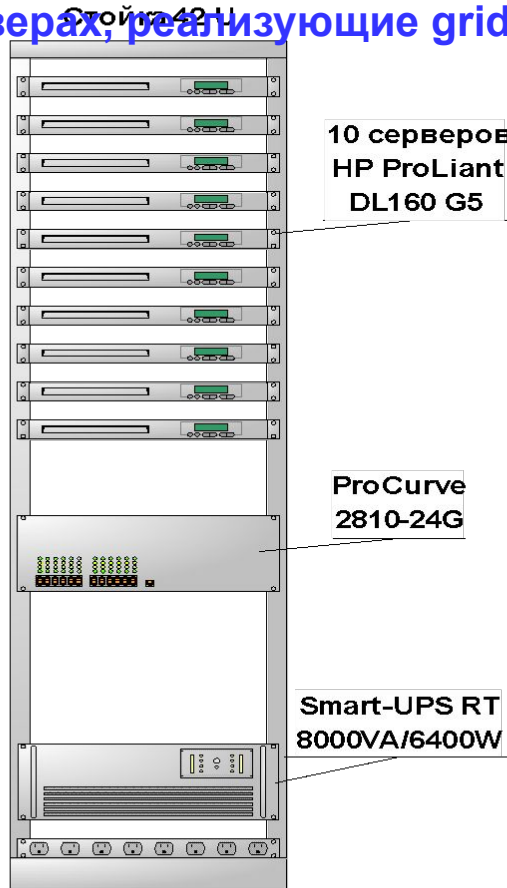
Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Центр данных (ЦД) - это data grid cluster, выполненный на десяти серверах, подключенных через коммутатор к ННС. Конструктив ЦД представляет собой автономную стойку, питающуюся через источник бесперебойного питания. На серверах установлена операционная система Scientific Linux CERN 4.7, поддерживающая распределенную файловую систему. Для реализации grid-сервисов хранения на серверах установлен программный комплекс gLite 3.1. Аналогичное ПО до...



Кластер установленный в РНЦ «Курчатовский институт»

08/14/2023



Кластер установленный в Саратовском государственном университете

3

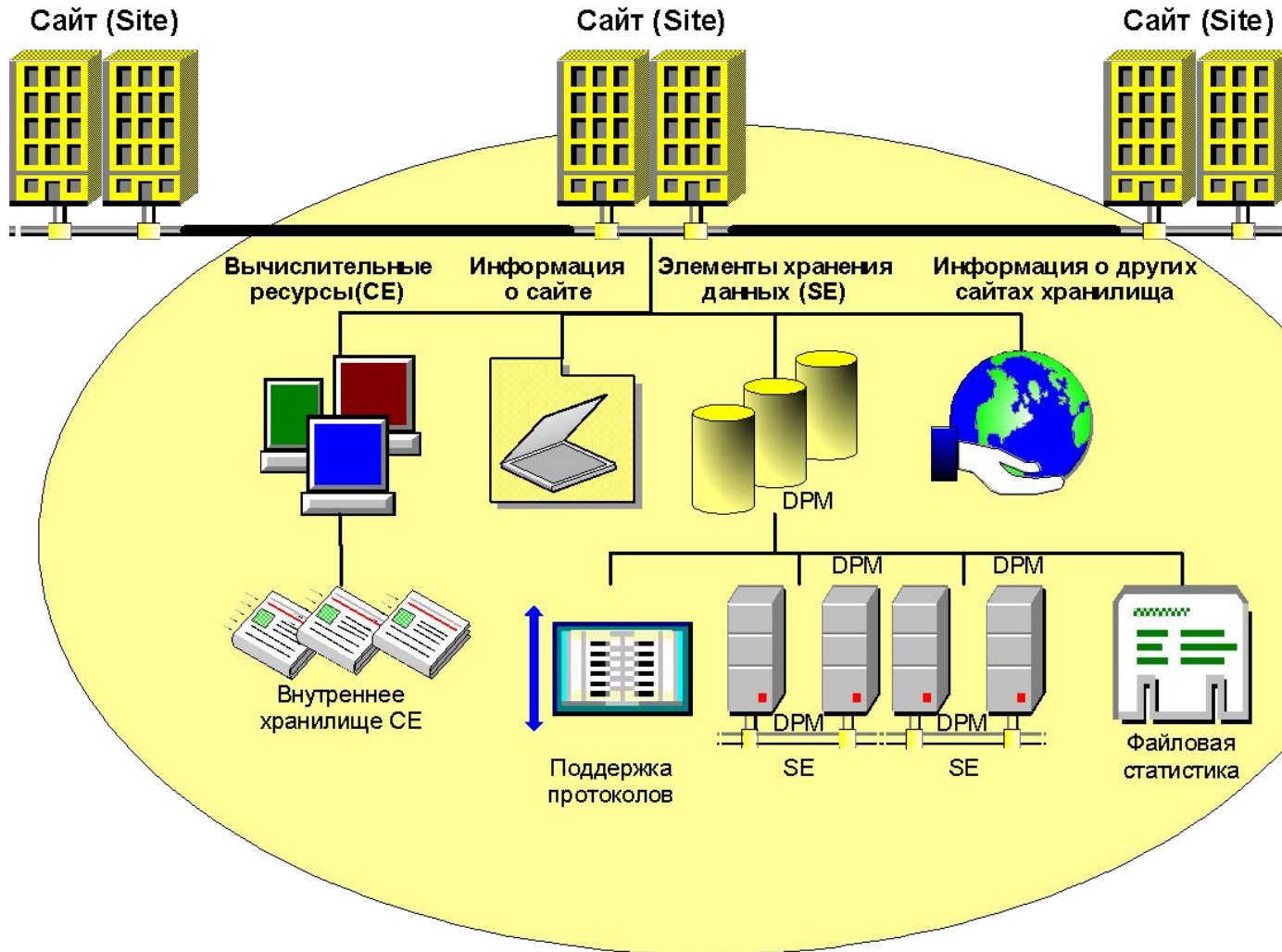
Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

В nanoиндустрии при проведении уникальных и дорогостоящих экспериментов требуется долговременно хранить очень большие объемы «сырых» необработанных экспериментальных данных в десятки и сотни Тбайт.



Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Ресурсный уровень хранилища данных (РСД)



В РСД совокупность вычислительных ресурсов (Computing Element, CE) и элементов хранения данных (Storage Elements, SE) сосредоточена на сайте (Site)

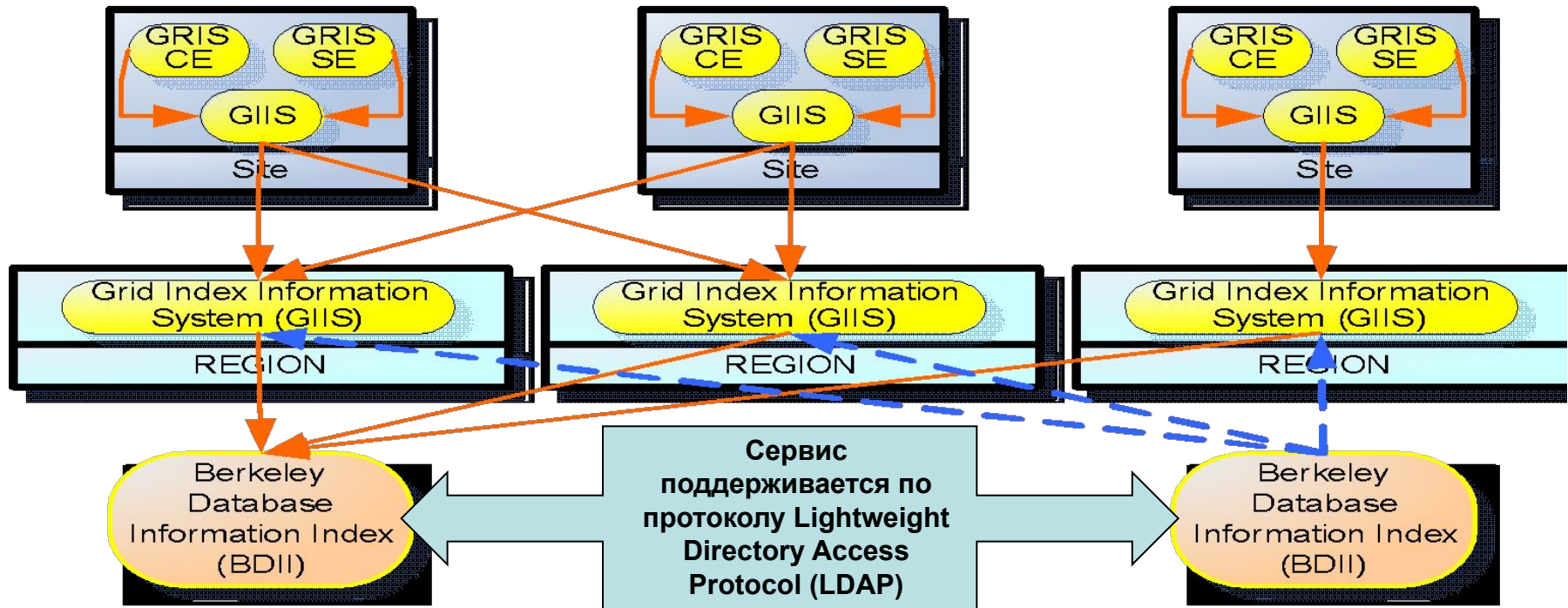
Для объединения всех элементов хранения данных (SE) в единое хранилище, используется grid-сервис DPM (Disk Pool Manager)

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Ресурсный уровень хранилища данных (РСД продолжение)

Иерархическая модель предоставления информации о grid-сервисах в РСД:

CE, SE □ GRIS □ GIIS □ BDII



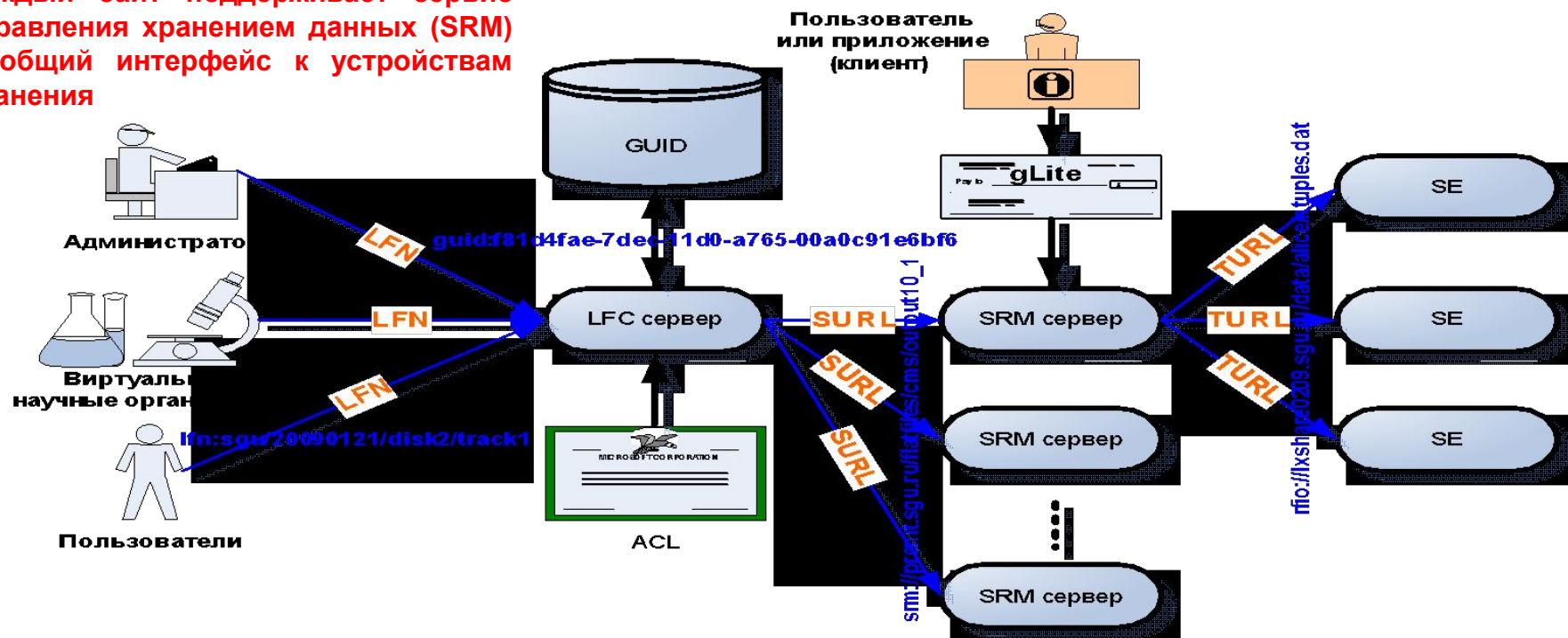
На каждом сайте информационные провайдеры GRIS (Grid Resource Information Server) собирают информацию о grid-сервисах CE и SE. Эта информация аккумулируется в коллекторах информации сайтов GIIS (Grid Index Information System) и сервисах BDII (Berkeley Database Information Index). На самом верхнем уровне находится grid-сервис top-BDII, имеющий информацию со всех сайтов.

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Ресурсный уровень хранилища данных (РСД продолжение)

Получение данных в РСД с помощью сервиса SRM (Storage Resource Manager)

Каждый сайт поддерживает сервис управления хранением данных (SRM) – общий интерфейс к устройствам хранения



- **LFN** - Logical File Name, логическое имя файла, создаваемое пользователем для его идентификации в РСД.
- **LFC** - Large Hadron Collider Computing Grid File Catalogue, сервис определяющий местоположения файлов в РСД.
- **GUID** - Globally Unique Identifier, внутренний (машинный) идентификатор элемента данных, соответствующий LFN.
- **ACL** - Access Control List, списки управления доступом к данным.
- **SURL** - Site Uniform Resource Locator, определитель физического местоположения реплики элемента данных (SE) в РСД.
- **TURL** - Transfer Uniform Resource Locator, идентификатор транспорта, определяющий место и протокол, получения файла.

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Интерфейс пользователя для управления данными (клиент РСД)

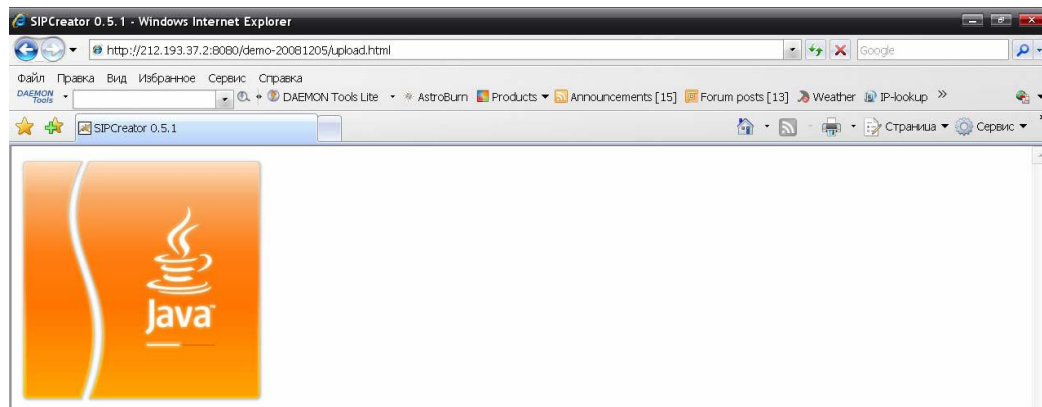
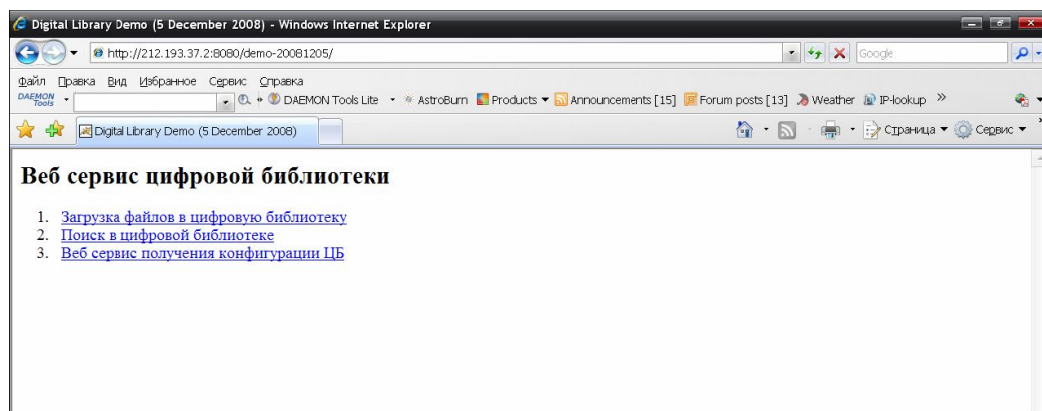
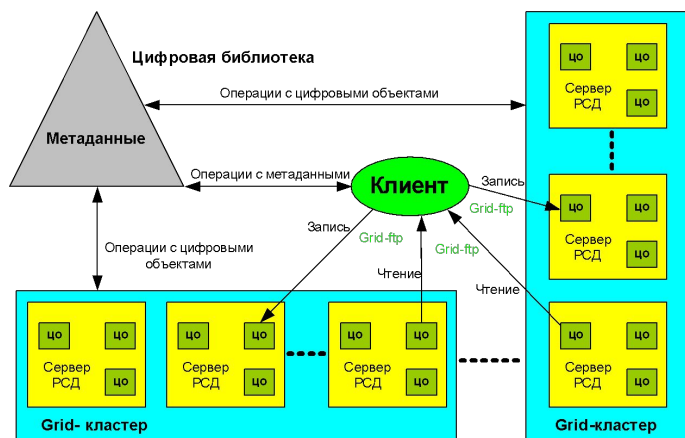
Screen shots приложения клиента РСД при работе с данными

Client implements the following main services:

- organizes interaction of the user of NNC with the data storage by the grid-ftp protocol;
- organizes access to grid-services on the basis of digital certificates;
- organizes interaction with the digital library for the creation of digital objects, search of data and management of data flows in the RSD.

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Интерфейс пользователя для управления данными (клиент РСД продолжение)



Клиент предназначен для организации web-доступа к РСД. В шести разработанных версиях реализованы различные возможности: от выполнения всех сервисов на стороне клиента до выполнения всех сервисов на сервере.

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Управление цифровыми объектами в РСД

Screen shots применения клиента РСД при работе с цифровой библиотекой

The screenshots illustrate the workflow in the Fedora Commons client:

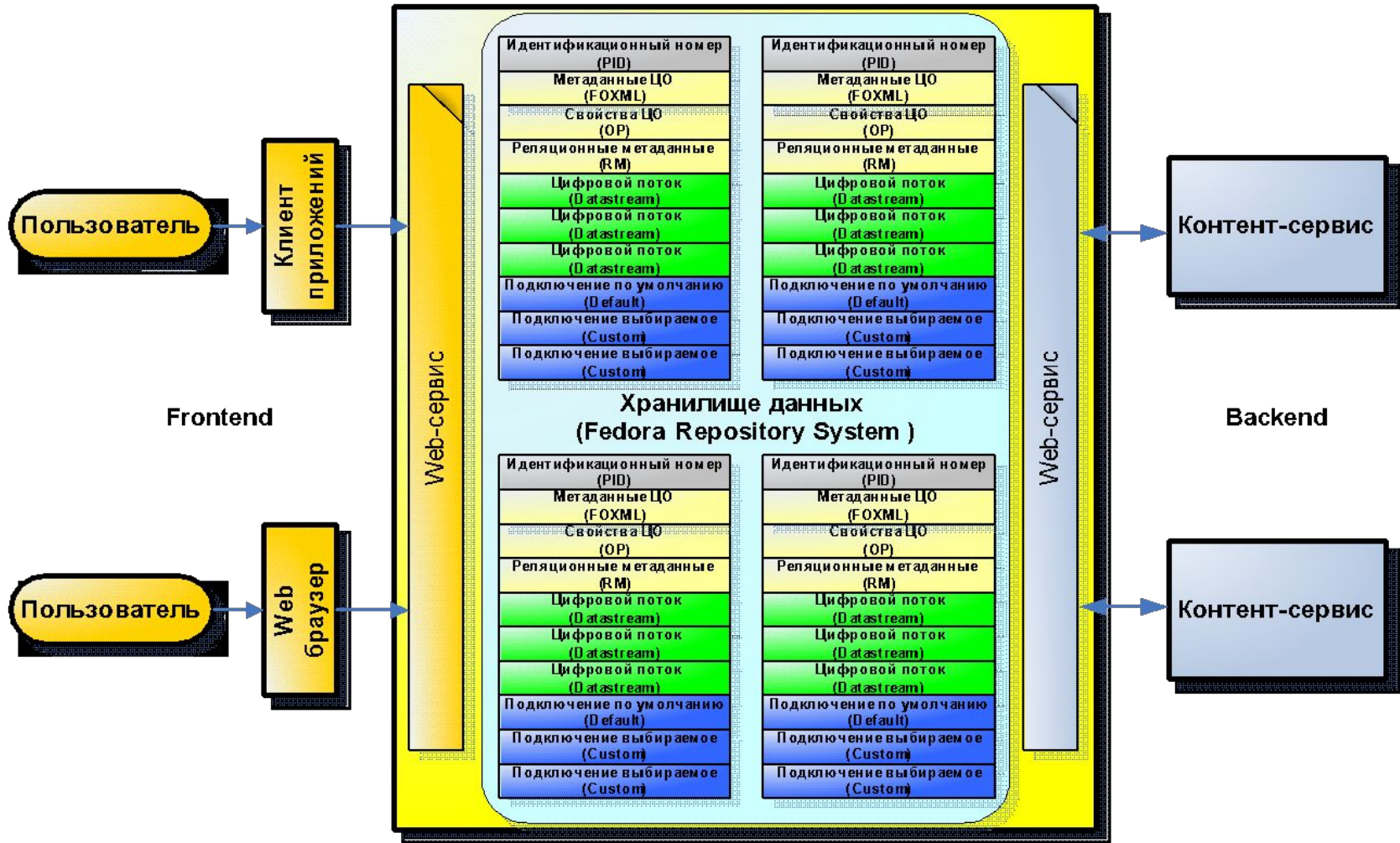
- 1. Search Repository page with the search bar and navigation options.
- 2. 'Fields to edit' page showing a list of fields like pid, label, state, etc.
- 3. Search results page showing a list of objects with their PIDs.
- 4. Object Profile HTML Presentation page showing the details of a specific object.
- 5. Object Items HTML Presentation page showing a table of item descriptions for a specific object.

Item ID	Item Description	MIME Type
CONTENT	описание.jpg	text/xml
DC	Dublin Core Record for this object	text/xml

Для работы с цифровыми объектами в РСД разработана подробная инструкция для пользователей

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Управление цифровыми объектами в РСД (продолжение)

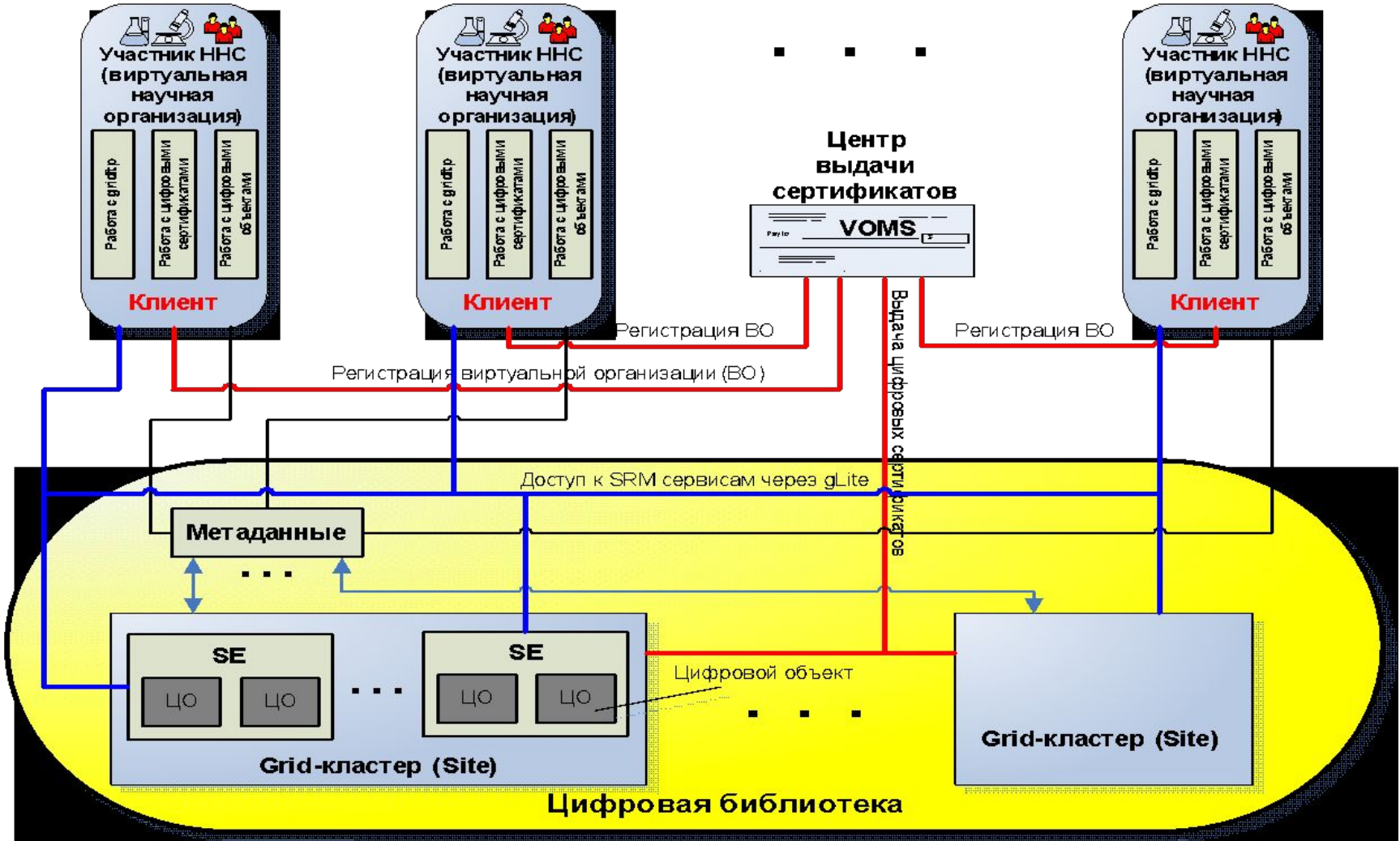


Интеграция клиента и Web-сервисов

Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Управление цифровыми объектами в РСД (продолжение)

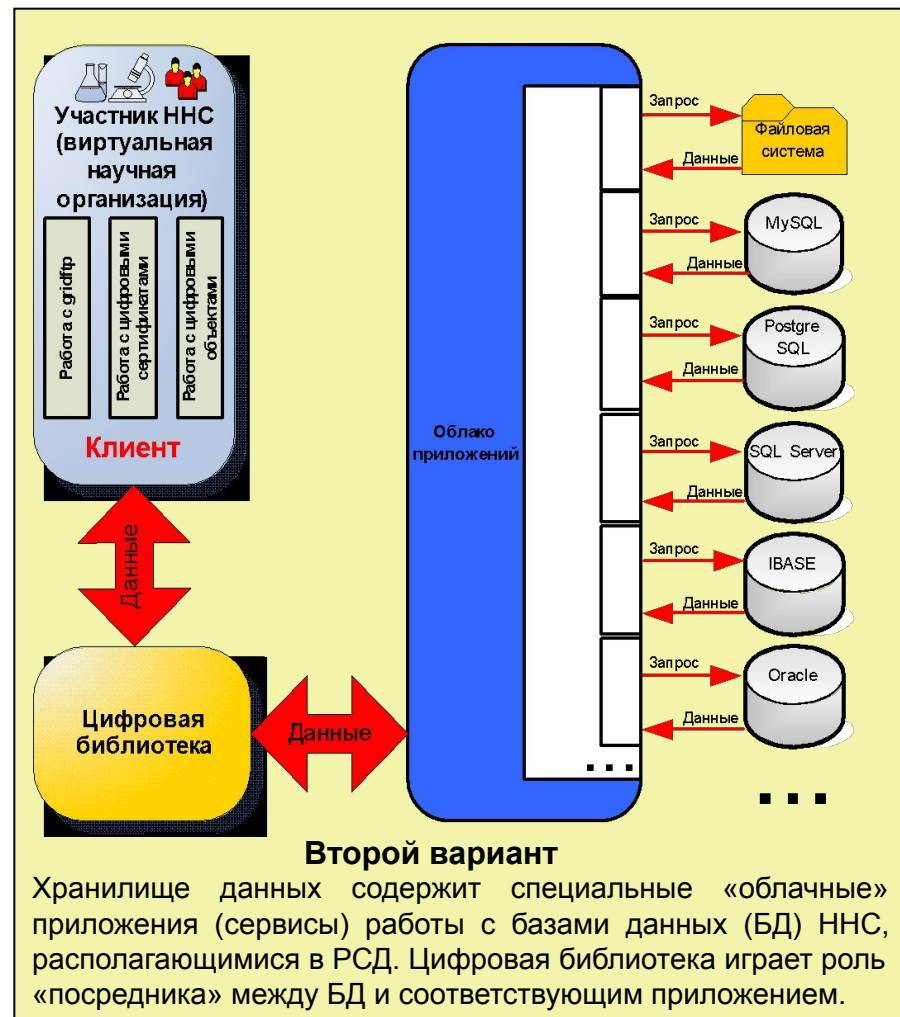
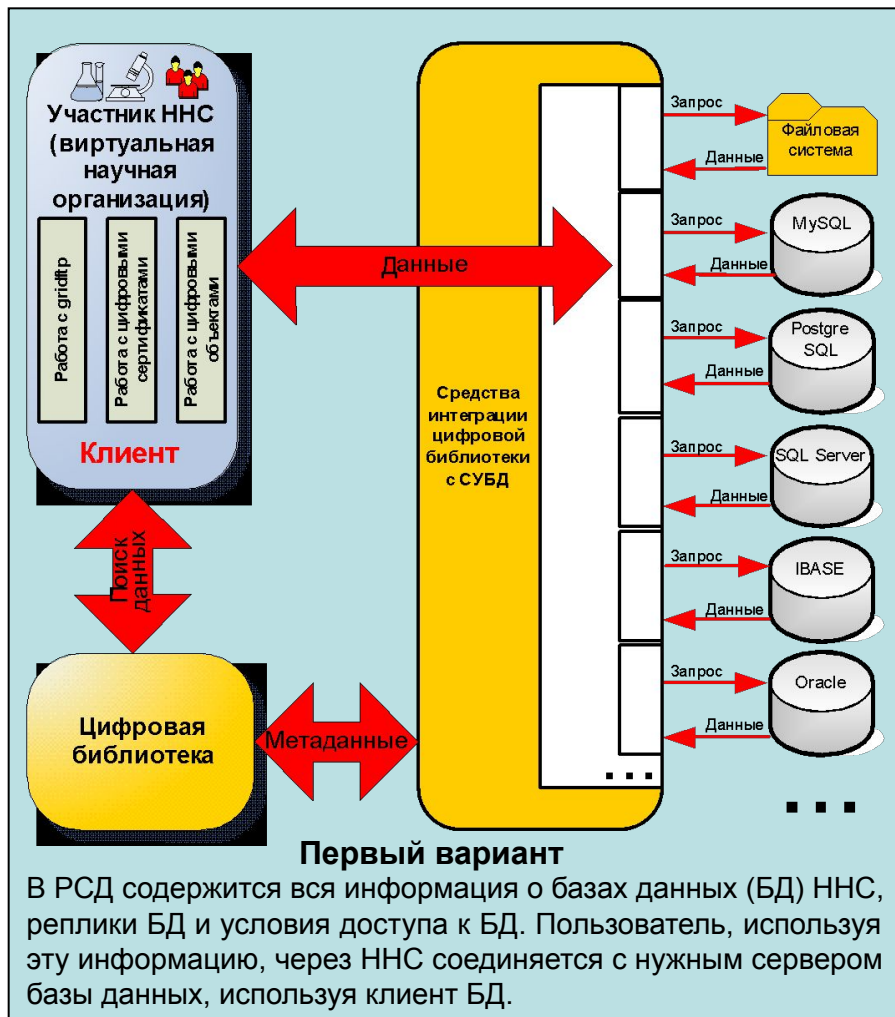
Работа с цифровой библиотекой



Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Управление цифровыми объектами в РСД (продолжение)

Работа с базами данных в РСД



Полномасштабное распределенное хранилище данных национальной нанотехнологической сети

Результаты работы

В результате выполнения работы пользователь получает масштабируемое прозрачное хранилище гетерогенных данных с гарантированным качеством сервиса. В полномасштабном распределенном хранилище данных обеспечивается требуемый уровень защиты, сохранность данных, необходимое количество реплик, высокая скорость доступа к данным и унифицированный механизм обмена данными разного типа (файлы, таблицы, массивы, базы данных и т.д.). Хранилище совместимо с вычислительными сервисами grid-инфраструктуры и поддерживает долговременное хранение, поиск и удобный доступ к данным. Полномасштабное распределенное хранилище управляет жизненным циклом данных, включая создание материалов, передачу, сохранение и доступ ко всем цифровым материалам.

Спасибо за внимание

Докладчик – Владимир Михайлович Соловьев

E-Mail: svm@sgu.ru

Телефон: (8452) 210 660