



**Российская Академия Наук
Сибирское отделение
Институт вычислительных технологий**



***«Технологии создания и
поддержки распределенных
информационно-вычислительных
ресурсов в СО РАН»***



Информационные технологии

Проблемы создания информационных ресурсов и продуктов являются приоритетными направлениями развития многих стран, в том числе и в России, где эти проблемы отнесены к критическим направлениям развития общества.

«Информация является важнейшим стратегическим ресурсом и наибольший экономический и социальный успех сегодня сопутствует тем странам, которые активно используют современные средства компьютерных коммуникаций и сетей, информационных технологий и систем управления информационными ресурсами»

Жискар д'Эстен





Информационное общество

- Наблюдаются переход к исключительно распределенной схеме создания, поддержания, хранения ресурсов и, в то же время, – стремление к виртуальному единству посредством предоставления свободного доступа к любым ресурсам сети через ограниченное число «точек доступа».
- Таким образом, в современном информационном обществе все больший приоритет получают технологии использования распределенных информационно-вычислительных ресурсов, называемые в последнее время «GRID-технологиями»





Проблемы эксплуатации информационных ресурсов

- Стремительное развитие глобальных информационных и вычислительных сетей ведет к изменению фундаментальных **парадигм** обработки данных, которые можно охарактеризовать как переход к Распределенным Информационно-Вычислительным Ресурсам (РИВР) и создание инфраструктуры для свободного доступа к ним.



Эффективная эксплуатация информационных ресурсов возможна только в том случае, когда они постоянно поддерживаются авторами, т.е. на основе технологий использования информационно-вычислительных ресурсов – GRID-технологии.



Сибирское отделение РАН

Sci-Potential of Siberia - Microsoft Internet Explorer
Address: http://www-sbras.nsc.ru/win/sbras/siberia/soran_siberia.html

Общая картина академической деятельности в Сибири (СО РАН), а также в высших учебных заведениях, важнейшими из которых являются Сибирские Университеты. Общее представление о состоянии науки в Сибири можно получить, взглянув на список академических Институтов, используя информацию, представленную ниже.

Legend:

- Научные организации СО РАН
- Научные организации СО РАМН
- Научные организации СО РАНХ
- Университеты
- Международные научные центры

Нажав на имя города на карте, можно получить дополнительную информацию.



Сибирское отделение РАН

- В настоящее время в Сибирском отделении РАН имеются разнообразные базы данных и других ресурсов, содержащих информацию о различных аспектах научной, образовательной и хозяйственной деятельности сообщества.
- Потребности сегодняшнего дня вызвали к жизни принципиально новые виды информационных ресурсов – информационные системы, электронные публикации и коллекции, облеченные в форму электронных библиотек.





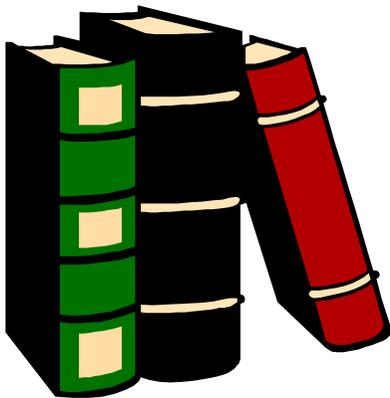
Сибирское отделение РАН

- Электронные библиотеки – особая форма организации информации, доступ к которой осуществляется через единый (возможно распределенный) каталог.
- Унификация каталогизации на основе международных рекомендаций обеспечивает интероперабельность ресурсов.
- Основная задача, технологии GRID, состоит в объединении всех (доступных) ресурсов сети в единую распределенную интегрированную среду.





Электронные библиотеки



- Создание распределенной информационно-вычислительной среды СО РАН в нашем подходе опирается на идею электронных библиотек.
- В рамках этого подхода электронные библиотеки рассматриваются как отдельная конкретная технология работы с **метаинформацией**, основанная на унифицированной каталогизации всех ресурсов (аналоги – профили LDAP, Z39.50).



Экологические системы

- Необходимость разработки механизмов, обеспечивающих как функционирование общей информационно-аналитической рабочей среды, так и доступ к научным ресурсам, и их сохранность, является приоритетом в научных исследованиях.
- Эти вопросы приобретают особую важность для исследований экологических систем, когда различные группы исследователей (вследствие особенностей решаемых проблем, а также природы вопросов, связанных с окружающей средой), разделенные географически, должны осуществлять совместную работу, обмен данными и знаниями, и координировать свои действия с целью оптимизации использования информационно-вычислительных ресурсов, сервисов и приложений.





Экологические системы

- Следует отметить, что в настоящее время сбор данных об окружающей среде и корректное использование наборов данных приобретает все большее значение.
- Данные об окружающей среде часто незаменимы; они всегда уникальны, хотя бы по времени их получения. Их сбор часто обходится очень дорого.
- По этим причинам огромную важность имеет задача получения максимальной пользы от данных, полученных в результате каждого исследования.





Интеграционные проекты

- 1. Создание распределенной информационно-аналитической среды для исследований экологических систем (проект № 34 - д.ф-м.н. Гордов Е.П., чл.-к. РАН Колчанов Н.А., чл.-к. РАН Федотов А.М.)**
- 2. Создание средств спутникового экологического мониторинга Сибири и Дальнего Востока на основе новых информационных и телекоммуникационных методов и технологий (проект № 86 - чл.-к. РАН Левин В.А., чл.-к. РАН Шайдуров В.В., д.б.н. Харук В.И.)**



GRID - Мета-Компьютер

Формирование в ресурсах «Сети» единого, математически однородного поля компьютерной информации, способного стать универсальным и машиннезависимым носителем данных, унифицированных программ и глобально распределённых вычислительных процессов.

Отличительная особенность данного направления в том, что оно основано на фундаментальных свойствах и законах развития компьютерной информации.

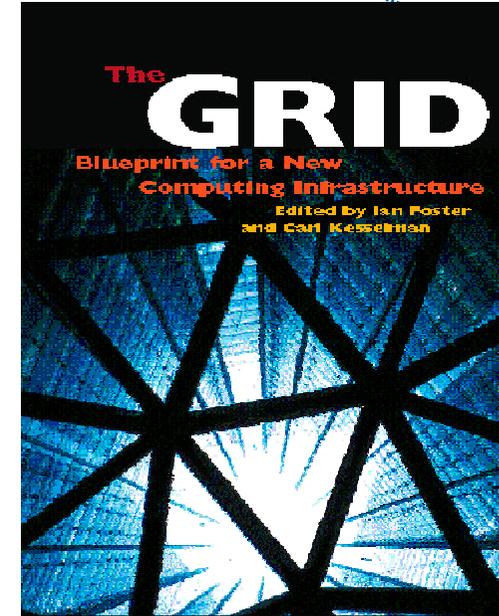




GRID

Концепция метакомпьютера

- + Коммуникации
- + Виртуальные организации
- + Middleware технологии (GT)
- Коммерческий грид (DCGrid,...,Oracle10g)
- Общественный грид (SETI@home,...)





GRID-технологии

Проблемы безопасности

Аутентификация, в т ч серверов

Авторизация

Конфиденциальность и целостность данных

Биллинг, в т ч контента

Аудит

Строгое выполнение обязательств

Дополнительная функциональность

Участники взаимодействия получают полный контроль над ресурсами

Не P2P платформа - виртуализация ресурсов

Не клиент - серверная платформа
OGSA - сервис - агентная архитектура

Распределенная платформа

Отсутствие единого центра управления

Отсутствие единого диспетчера
(его цель - выполнять задачи VO)

Сервисы по требованию (фабрики)

Виртуализация (доступ не к серверам а к сервисам)

Гетерогенность ресурсов

Диспетчеризация, а не распараллеливание

Есть круг задач, которые grid позволяет эффективно решать (с приемлемыми затратами)

Высокая размерность задачи

Для отказоустойчивости и приемлемой производительности тиражирование данных и миграция служб или процессов с одного ресурса на др.

Интеграция ресурсов

WSFR - симбиоз с Web
VO, в т.ч. динамические

Доверительные отношения разного вида



Проект распределенной информационно-вычислительной среды для исследования окружающей среды

Цели:

Интегрированная система предоставления фактографической информации и предоставления в объединенное пользование вычислительного и поддерживающего оборудования и программного обеспечения, необходимого для эффективного доступа к данным.

В рамках проекта предполагается создание пилотной модели распределенной базы данных и метаданных, которые доступны всем участникам проекта. Доступ к банкам данных, моделей и программ анализа данных.

В результате выполнения Проекта должна быть построена основа для создания распределенной информационной системы с единой точкой доступа (порталом**).**



Функционирование информации в распределенной среде

- 1) Интероперабельность БД (информационных хранилищ) и стандартизация хранилищ данных – интеграция БД.
- 2) Извлечение данных – построение знаний – интеллектуальные агенты.
- 3) Доступ к библиотекам алгоритмов для обработки данных и организация систем моделирования.
- 4) Настройка данных на алгоритмы и алгоритмов на данные.
- 5) Распределение задач по различным порталам (распределенная обработка данных).



Функционирование информации в распределенной среде

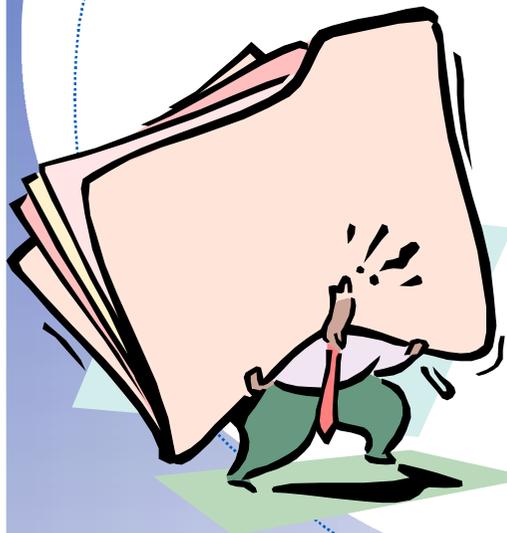
При этом информационный портал должен выступать как

- *Центр* аккумуляции информационных ресурсов, обеспечивающий их централизованное хранение и оперативную обработку.
- *Центр* регистрации распределенных ресурсов
- Точка доступа к распределенным ресурсам
- *Центр* администрирования ресурсов
- *Центр* сбора и обработки статистики использования информационных ресурсов.



РИБР → GRID

- Использование распределенных информационно-вычислительных ресурсов в интегрированной среде опирается на унификацию систем метаданных, описывающих как сами данные, так и правила доступа к ним.
- Для разработки распределенных информационных систем сегодня используются различные технологии (RPC, DCOM, RMI, ODBC, JDBC, CORBA, SOAP, ebXML, WSDL, WSFL, UDDI и др.).



Большая часть этих технологий по своей сути ориентирована не на работу с информационными ресурсами, а на сетевое взаимодействие программ и распределенные вычисления в гетерогенных средах.



META-DATA

- Однако ни одна из упомянутых технологий не обеспечивает универсальных способов работы с информационными ресурсами, т.е. необходимого уровня абстрагирования от конкретных систем и платформ при сохранении высокой степени функциональности и предоставления механизма доступа к данным.





META-DATA

- Именно такие требования присущи технологиям, основанным на использовании международных стандартов взаимодействия открытых систем (OSI). В качестве примеров здесь можно привести протокол ANSI Z39.50 (стандарт ISO-23950) или протокол LDAP (рекомендации ITU X.500).





META-DATA

Протоколы семейства OSI, связанные с интеграцией ресурсов, основаны на концепции метаданных – формализованных знаний о внутренней структуре и поведении информационных и вычислительных ресурсов (или документов).

Метаданные содержат структурированные сведения о ресурсе, представляющие его свойства (атрибуты) и функции – информацию, предназначенную для анализа, дизайна, развития, обработки и использования ресурсов.



META-DATA

Понятие метаданных уже давно и успешно используют при разработке и эксплуатации информационных хранилищ и систем аналитической обработки данных, для электронного документооборота и управления потоками работ, управления знаниями.

Метаданные позволяют описывать классы объектов, представляющие типы информационных документов, атрибуты и отношения между различными типами документов и ресурсов

Объединенные информационные ресурсы



Welcome to Z39.50 Gateway Novosibirsk - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Mail Stop

Address <http://z3950.uiggm.nsc.ru/zgwk/> Go Links >>

Z39.50



Home Page

Помощь

О шлюзе

Добро пожаловать в Распределенную Информационную Систему Сибирского Отделения РАН

Выберите нужную область поиска:

- [Информационные ресурсы СО РАН](#)
- [Распределенный электронный каталог библиотек г.Новосибирска](#)
- [Библиографическая научно-техническая информация](#)
- [Труды сотрудников СО РАН](#)
- [Сотрудники СО РАН](#)
- [Электронные коллекции СО РАН](#)
- [Геоинформационные ресурсы СО РАН](#)

- [Тезаурусы и классификационные схемы](#)
- [Информация о серверах Z39.50 \(Explain\)](#)

При поддержке [Сибирского Отделения РАН](#) Copyright © ОИГГМ СО РАН, 2000-2001

Internet



ИРИС СО РАН

- Система метаданных описывающая:
 - Организации
 - Основные направления деятельности
 - Персоны
 - Директории пользователей и сертификаты
 - Проекты
 - Публикации
 - Разработки и инновационные предложения
 - Электронные коллекции и научные базы
 - Библиографические базы и каталоги библиотек
 - Виртуальный музей истории СО РАН



ИРИС СО РАН

- Интеграция сетевых сервисов и информационных ресурсов в корпоративных распределенных информационных системах с множеством управляющих серверов требует создания единой корпоративной распределенной справочной системы о конфигурации и содержании самой системы, в основу которой может быть положена уже хорошо зарекомендовавшая себя в сети служба каталогов (X.500 - LDAP).



ИРИС СО РАН

Распределенная информационная система СО РАН - Microsoft Internet Explorer
http://ldap.nsc.ru:210/
LDAP
СО РАН
Глобальный каталог: Организации СО РАН
Организации | Сотрудники | Ресурсы | Сервисы
Новости
Каталог LDAP
Научные центры СО РАН
Организации СО РАН
Сотрудники СО РАН
Информационные ресурсы
Сетевые сервисы
Каталог библиотек
СО РАН, Новосибирск
Н-ск-БФП, корпорация
Томск, Омск, Москва
Смело-от-ИИИ-Омск
Омск-от-ИИИ-Омск
Базы данных ИИИ
Сотрудники
Труды сотрудников
Электронные публикации
Геоинформационные ресурсы
Рубрикаторы
Телефоны
Экстренка
СО РАН
Полное название: Институт вычислительных технологий СО РАН
Город: Новосибирск
Описание: Области исследования: разработка информационно-технологических решений в задачах проектирования, математическое моделирование и вычислительные технологии в области механики сплошной среды, физики, энергетики и экологии.
Идентификатор: ИТ
Короткое название: ИИТ СО РАН
Директор: ГМ. ЗДЕСЬ
URL: http://www.ict.nsc.ru
Дата изменения: 201603160424572
dn: ou=ИИТ СО РАН, ou=ИИИ, ou=ИИИ, ou=ИИИ, ou=ИИИ, ou=ИИИ

LDAP каталог СО РАН

Томск - 02.07.2006

Информационные ресурсы
Сетевые сервисы
Каталог библиотек
СО РАН, Новосибирск
Н-ск-БФП, корпорация
Томск, Омск, Москва
Смело-от-ИИИ-Омск
Омск-от-ИИИ-Омск
Базы данных ИИИ
Сотрудники
Труды сотрудников
Электронные публикации
Геоинформационные ресурсы
Рубрикаторы
Телефоны
Экстренка
СО РАН
Шокин Юрий Иванович
ИИТ СО РАН
Привлечение математиков
Телефакс: +7-383-3316150
Телефакс ИИ: +7-383-3316150
Город: Новосибирск
Почтовый индекс: 630030
Эл. почта: shokin@ict.nsc.ru
URL: http://www.ict.nsc.ru/shokin/



ИРИС СО РАН

Основными предпосылками внедрения справочной системы являются:

- Необходимость учета и поддержания в актуальном состоянии информации о сервисах и ресурсах
- Необходимость организации единой политики доступа к ресурсам на основе единых правил
- Необходимость единого учета пользователей
- Необходимость однократной авторизации пользователей
- Необходимость децентрализованного формирования специальных справочников и обеспечения доступа к ним через различные интерфейсы. Эти справочники должны содержать:
 - Информацию о доступных в системе сервисах (FTP, SMTP, WWW и т.п.) с адресами обслуживающих их серверов
 - Информацию о доступных в системе информационных ресурсах с указанием точки и метода доступа к ним
 - Информацию об администраторах
 - Информацию об организациях, участвующих в создании и поддержке системе
 - Контактную информацию о персонах
 - И т.д. и т.п.



ИРИВС СО РАН

Системный уровень обеспечивает доступ к распределенному хранилищу данных (провайдеры данных) – основанных на использование протоколов Z39.50, трансляция SQL-запросов, LDAP.

Доступ обеспечивается путем **регистрации и именованния** наборов данных

Регистрация - данных имеющих различный технологический способ хранения: файловые структуры и иерархические БД (LDAP, Z39.50)

Именованние обеспечивает унификацию запросов:

- Именованный SQL (реляционные БД)
- Абстрактный SQL (type 104)
- Абстрактный RPN



Meta-DataGrid

Реализация программного обеспечения, позволяющего конечному пользователю работать в тематическом веб-портале, функционирующем в качестве системы управления и поддержки исследований по окружающей среде Сибири в соответствии со следующими требованиями:

- **Уровень услуг** – пользователь имеет доступ к одним и тем же услугам независимо от своего местоположения.
- **Актуальность** – должна предоставляться максимально свежая информация –обеспечивается децентрализованным управлением: каждый сервер должен отвечать только за свою локальную часть базы, чтобы обновление данных и сопровождение можно было выполнять немедленно
- **Единое пространство имен**, позволяющее представлять информацию как единый логический каталог.
- **Схемы данных**, допускающие локальные расширения.
- **Единый протокол доступа**. Приложения, нуждающиеся в ресурсах справочника, должны производить запросы, используя стандартизированный протокол.

Источники данных

Система получения данных

Пользователи



Atlas "Atmospheric aerosols of Siberia" - Microsoft Internet Explorer

Файл Правка Вид Избранное Сервис Справка

Назад Поиск Избранное

Адрес: http://web.ict.nsc.ru/aerosol/data_elem.rdf?id=11&lang=en

Atlas "Atmospheric aerosols of Siberia"

Main Data Publication Utils Login

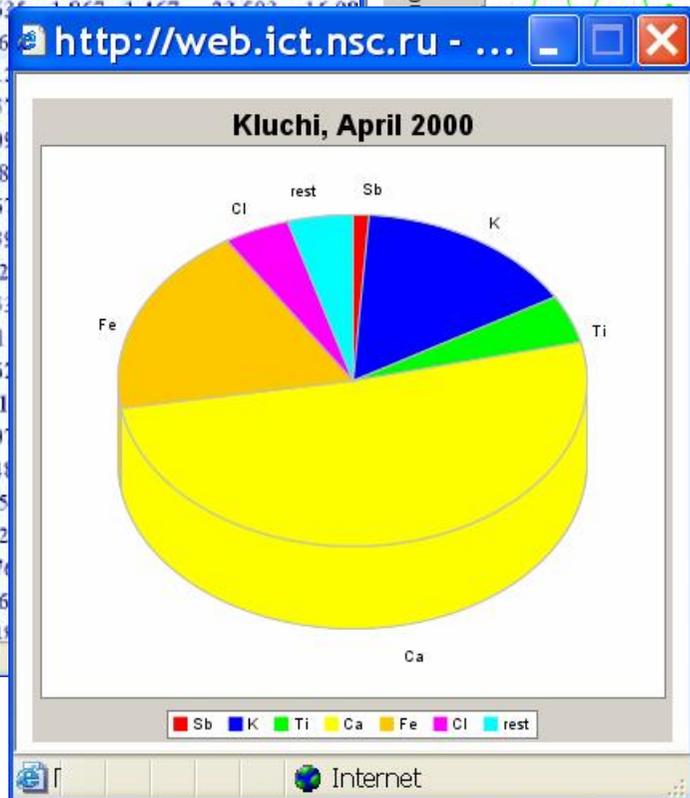
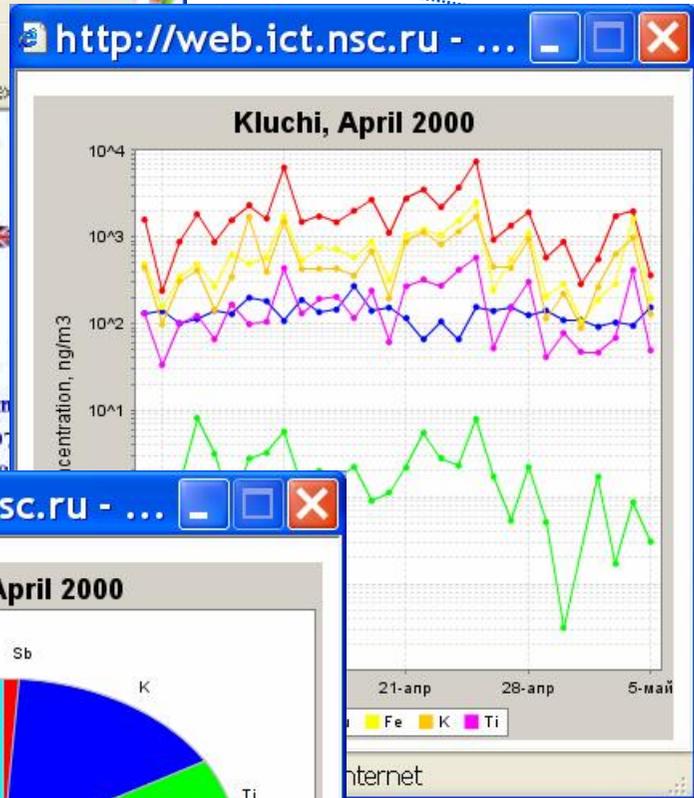
Data > Klyuchi > Element composition

Expedition: Klyuchi, April 2000
 Start date: 06.04.2000
 Final date: 05.05.2000

Element composition:
 Graphics: @L
 @E
 - Data for expedition
 - Data for station
 - Diagram
 Correlation
 Factors
 Daily measurements
 Ion composition:
 Carbon composition:
 Combined functions:

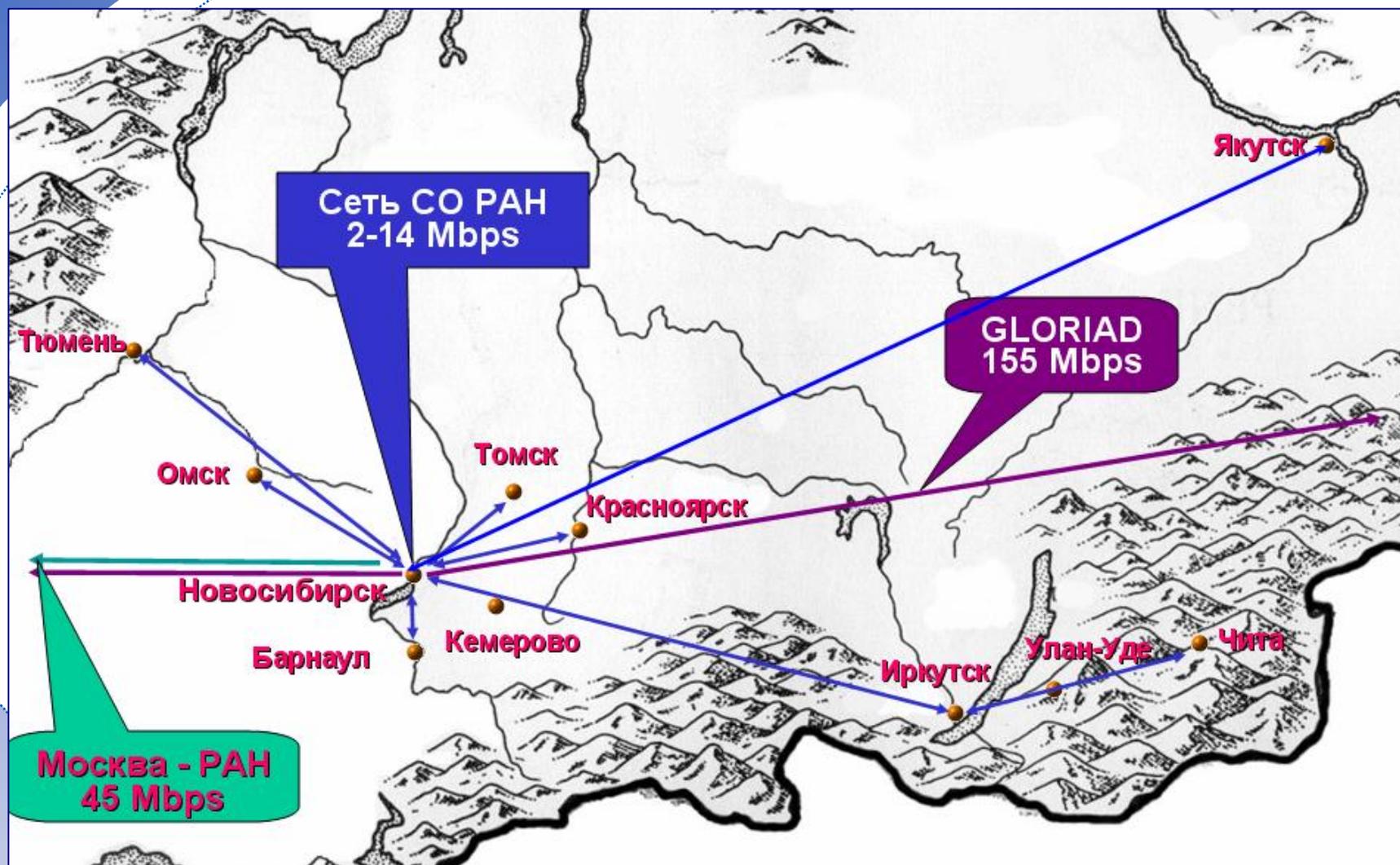
Comments: Hv
 Remember it

	< x > _g	σ _g	< x > _a	#days	EF _a	EF _g	EF _{og}	min	max	max/m
<input type="checkbox"/> As	1.796	3.104	3.421	28	156.577	70.026	3.478	0.221	20.109	90.97
<input type="checkbox"/> Ba	9.66	2.05	11.772	25	1.97	1.625	1.867	1.167	23.593	15.89
<input type="checkbox"/> Bi	0.08	1.866	0.099	15	60.539	43.6				
<input type="checkbox"/> Br	3.327	1.644	3.763	30	125.02	102.3				
<input checked="" type="checkbox"/> Ca	1444.592	2.203	1927.603	30	3.756	3.5				
<input type="checkbox"/> Ce	3.78	2.167	4.849	15	3.324	3.09				
<input checked="" type="checkbox"/> Cl	127.964	1.338	133.497	30	72.235	51.8				
<input type="checkbox"/> Co	2.335	2.142	3.057	30	9.759	9.67				
<input type="checkbox"/> Cr	1.898	3.43	3.437	24	1.871	1.39				
<input type="checkbox"/> Cs	4.387	1.398	4.613	13	121.231	96.2				
<input checked="" type="checkbox"/> Cu	1.403	3.137	2.246	29	3.788	2.5				
<input type="checkbox"/> Ga	0.299	3.083	0.456	30	1.87	1.62				
<input type="checkbox"/> Gd	2.435	2.335	3.416	30	33.985	33.1				
<input type="checkbox"/> Ge	0.028	4.07	0.062	15	1.972	1.07				
<input checked="" type="checkbox"/> K	435.415	2.284	594.373	30	1.66	1.48				
<input type="checkbox"/> La	5.522	2.532	8	26	11.288	10.5				
<input type="checkbox"/> Mn	22.03	2.264	30.329	30	2.24	2.2				
<input type="checkbox"/> Mo	0.106	1.601	0.12	30	8.377	6.74				
<input type="checkbox"/> Nb	0.171	2.095	0.223	28	0.687	0.6				
<input type="checkbox"/> Ni	1.080	2.741	1.470	30	1.681	1.45				





Регион деятельности Сети передачи данных СО РАН





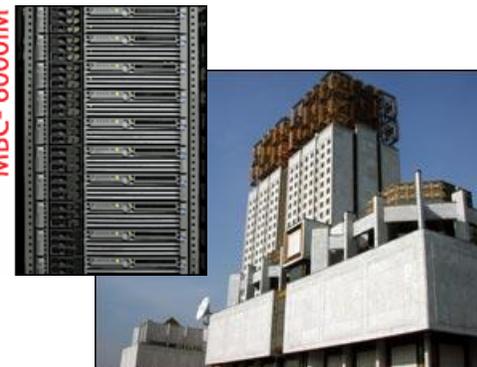
Объединенные вычислительные ресурсы:

**Межведомственный
суперкомпьютерный центр РАН**



**Иркутский суперкомпьютерный
центр СО РАН**

MBC-6000IM



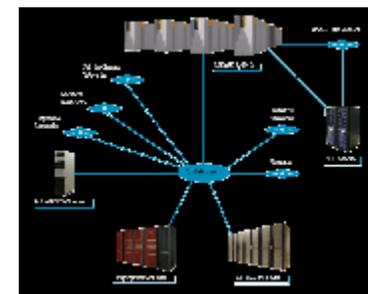
МЦЦ / www.jscc.ru

**Красноярский суперкомпьютерный
центр СО РАН**



**Новосибирский
суперкомпьютерный центр СО РАН**

**Центр высокопроизводительных
вычислений (HLRS), Германия,
Штутгарт**



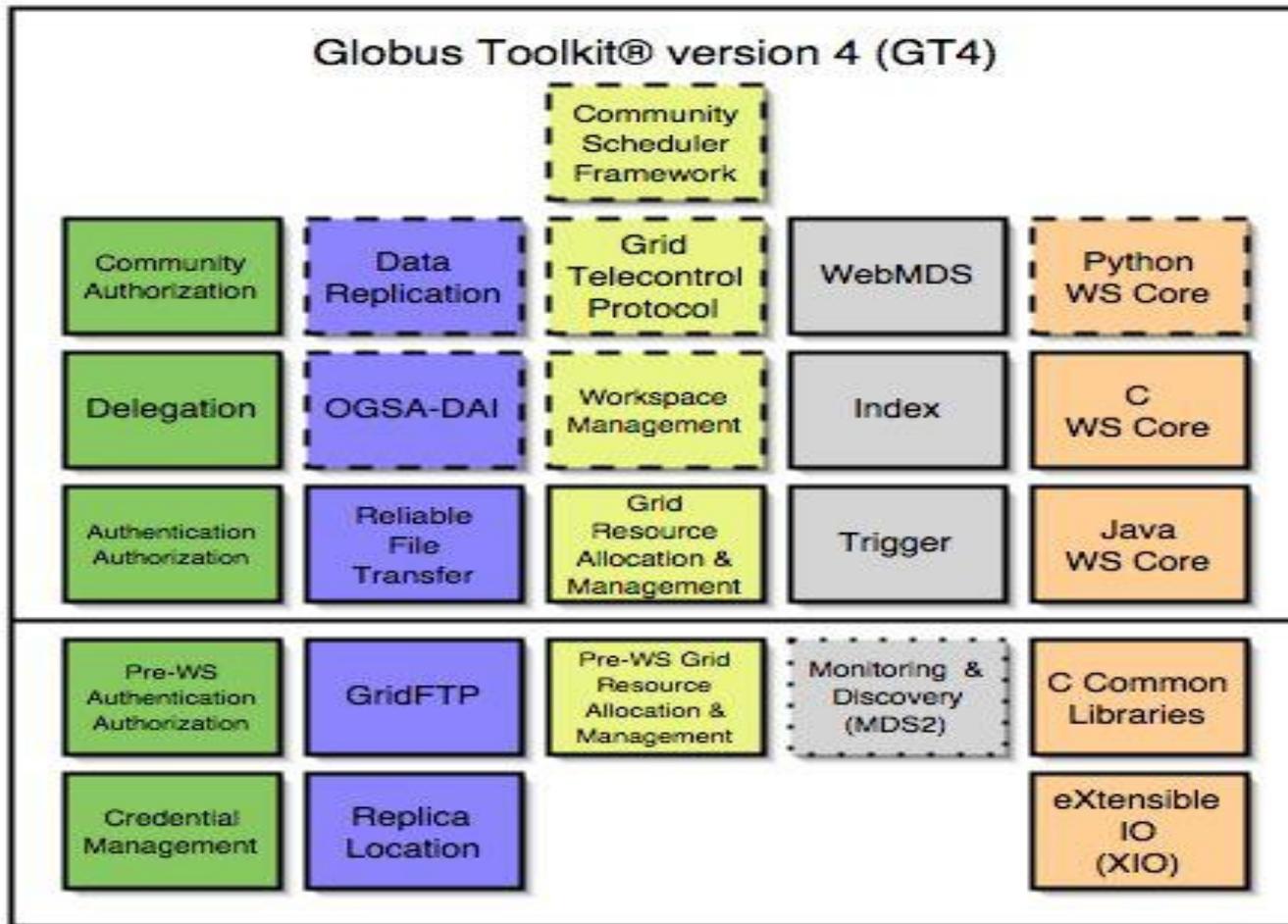


Элементы концепции

виртуальные организации



информационные объекты



репозиторий

сервисы

базовые
сервисы





Высокоуровневые грид-сервисы

Сервис репозитория ИО в стандарте **METS**

(Metadata Encoding and Transmission Standard)

Стандарт для кодирования описательных, административных и структурных метаданных для ИО в составе ЭБ, основывающийся на языке **XML** schema, в рамках инициативы федерации электронных библиотек (Digital Library Federation). Поддерживается подразделением Библиотеки Конгресса Network Development and MARC Standards Office.

ü семантические модели понятийной сети и модели вычислительных процессов

ü описание собственных и внешних информационных ресурсов и часть ресурсов

(регламент получения и обновления информации; формат данных и способ доступа к ним; состав ИО, описанный в терминах модели и способ ее хранения в источнике; способы идентификации ИО в источнике; первоисточник информации об объекте или отдельных его атрибутах и др.)

ü описание компонентов самого метахранилища

ü каталог Grid/Web-сервисов [**WSDL** (Web Services Description Language)]

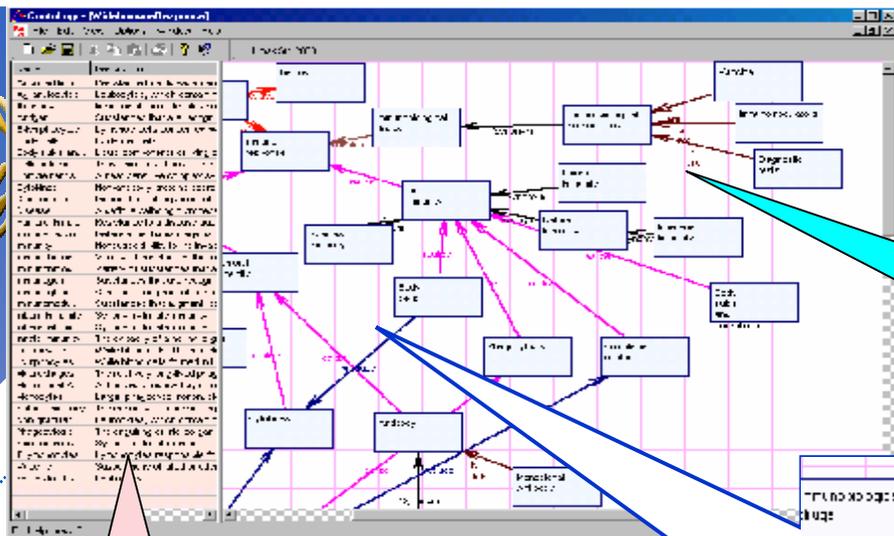
Library Bibliogrid Proxy Service + ZING (New Z39.50)

GDQS

(The Grid Distributed Query Service) высокоуровневый интеграционный сервис от OGSA-DAI, поддерживающий **OQL** в качестве сквозного языка запросов.

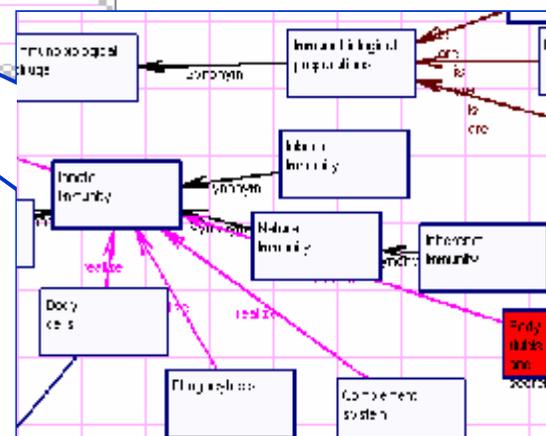
G-ONTOLOGY

Грид – технология семантического связывания разнородных ИО, основанная на стандартах построения онтологий и языке **OWL** (Web Ontology Language)



Грид -онтологии

Понятия и связи –
создание, просмотр,
модификация

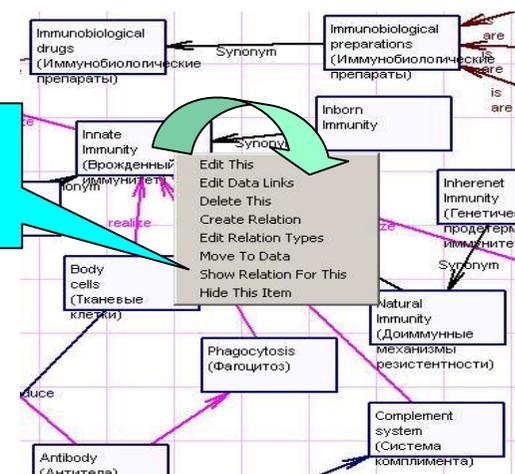


B-lymphocytes	Lymphoid cells concerned with humoral immunity
Vaccine	Suspensions of killed or attenuated microorganisms
Species immunity	Synonym - Innate Immunity
Innate Immunity	The capacity of a normal organism to remain resistant to a disease agent
Inherent Immunity	Synonym - Innate Immunity
Humoral Immunity	Resistance to a disease agent resulting from antibodies
Diagnostic tests	Diagnostic techniques involving the demonstration of specific antigens
Natural Immunity	The capacity of a normal organism to remain resistant to a disease agent
Disease	A definite pathologic process with a characteristic clinical course
Immunity	Nonsusceptibility to the invasive or pathogenic action of a disease agent

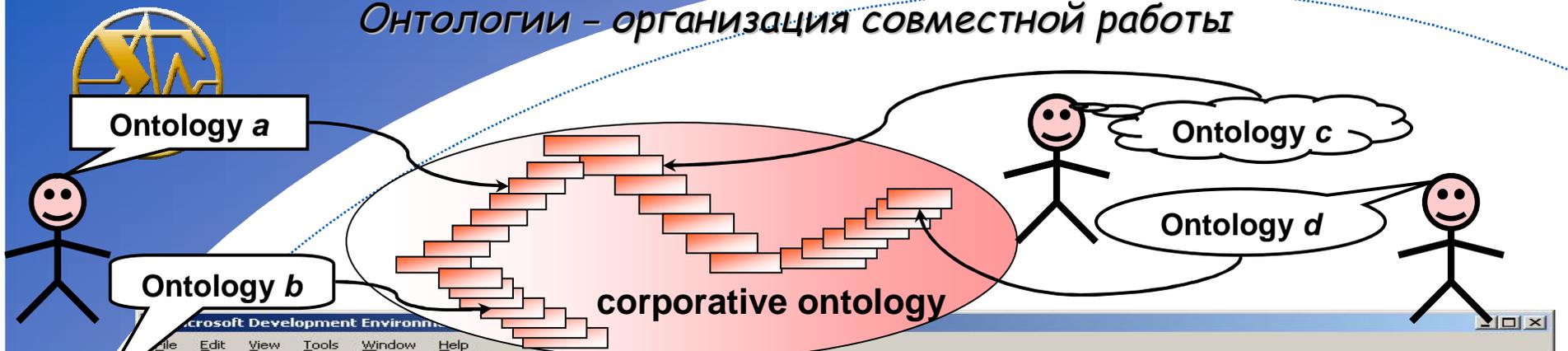
Список
ПОНЯТИЙ

Описание
ПОНЯТИЙ

Понятия и связи -
свойства



Онтологии – организация совместной работы



Библиотека онтологий

Ontology a

Ontology b

Интеграция ресурсов



Заключение

ЭБ необходимы в Grid среде, среде VO, ориентированной прежде всего на высокопроизводительные вычисления, где ЭБ рассматриваются прежде всего как среда обработки данных

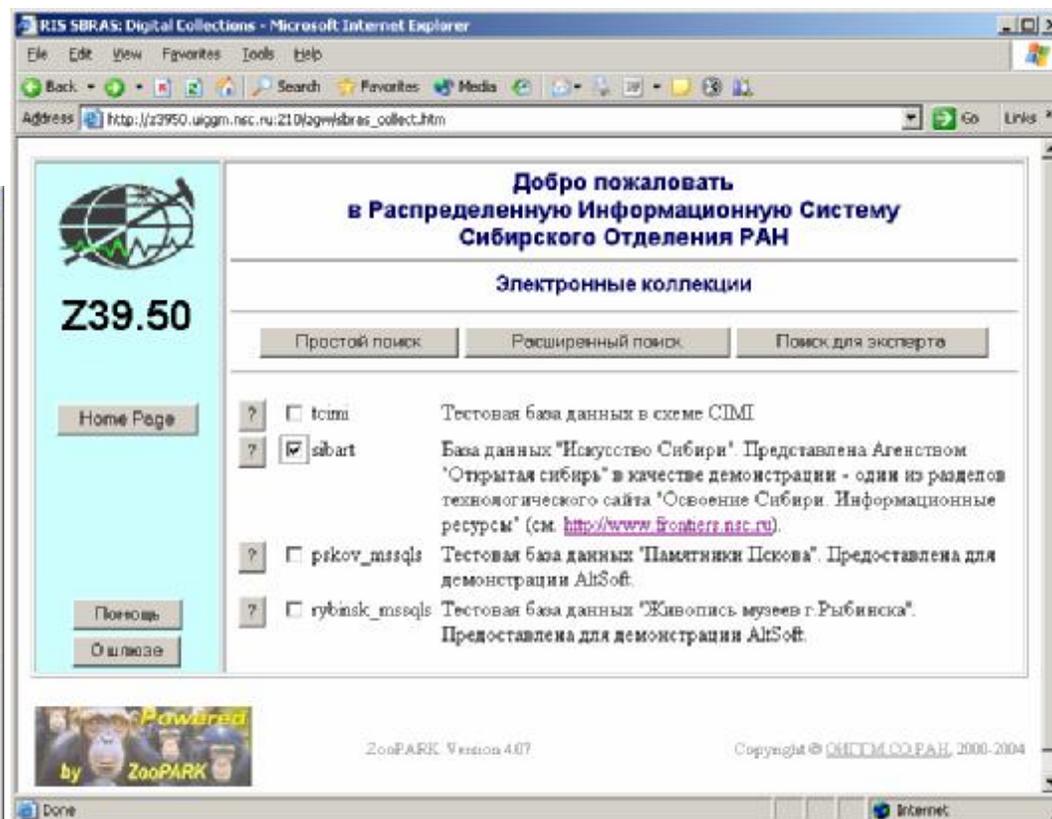
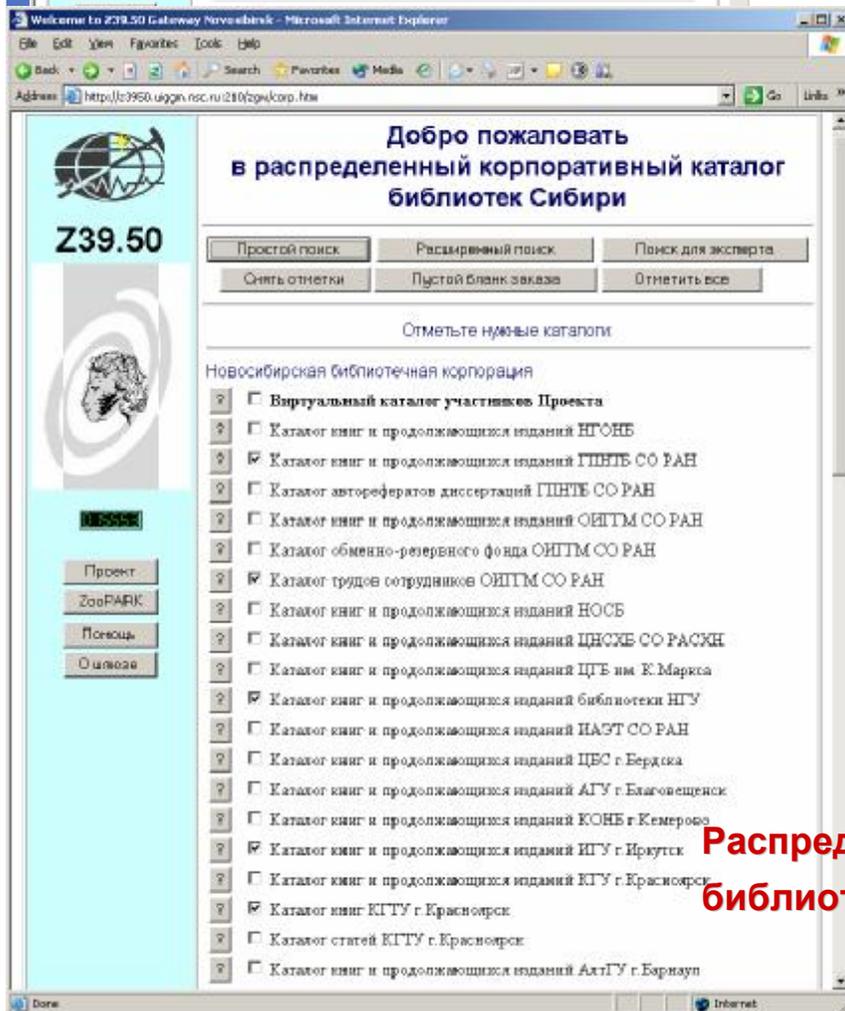
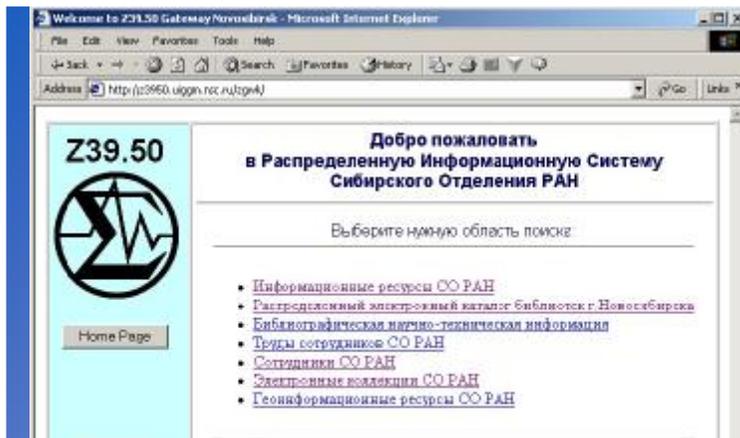
Виртуально, распределенно, гетерогенно, федеративно

Использование большого числа готовых решений базового middleware для построения ЭБ

Работы в рамках данного проекта можно рассматривать и как создание и развитие среды программирования для создания ЭБ

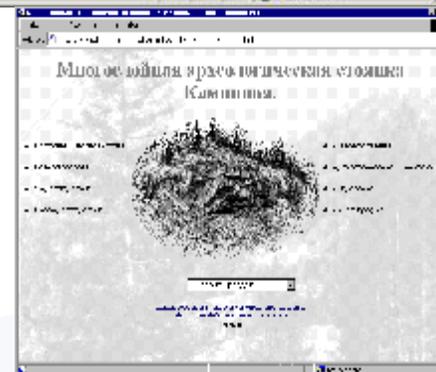
Полигон для выбора, создания и апробации информационных сервисов, в том числе работающих в БД и коллекциях, не имеющих публичного доступа

Шлюз Z39.50



Распределенная
библиотечная система

Интеграция ресурсов





Интеграция ресурсов

Благодарю
за
внимание!



"Разработка и использование программно- аппаратных средств ГРИД-технологий и перспективных суперкомпьютерных вычислительных систем семейства "СКИФ" в совет министров Союзного государства

- Основная идея технологии Grid, по словам директора Российского научного центра "Курчатовский институт" по научному развитию Алексея Солдатова, заключается в использовании каждым входящим в систему компьютером логических и физических мощностей всей сети. В целом это похоже на локальную сеть с тем отличием, что помимо возможности доступа к данным на других компьютерах новая технология делает возможным объединение ресурсов находящихся на расстоянии компьютеров для решения одной задачи. Солдатов убежден, что в ситуации с Grid Россия не сможет отказаться от использования нового решения, иначе проиграет технологическую конкуренцию