

## ПОСЛЕ EGI – WGI?

В.П. Шириков

По существу, данный обзор можно считать продолжением тех вводных, что были сделаны автором на наших конференциях в Абрау в 2004 и 2007 годах ( см.[1, 2] ). При этом частично используется материал, нашедший отображение в докладе [3] на юбилейной конференции по электронным библиотекам и коллекциям в 2008 году ( см.[3]), а также в авторских обзорах в периодических изданиях Информационных бюллетеней ЛИТ ОИЯИ [4, 5, 6].

Речь шла и идёт о том, как и в какой степени реализуются и развиваются идеи основоположников GRID – тематики ( К.Кессельмана, Я.Фостера..), ставших ключевыми фигурами для объединений Globus Alliance и Globus Grid Forum, занявшихся организацией проработки и реализации систем типа Grid: их технической и программной основы, т.е. тех наборов программных средств ( Globus Toolkits, GT ), с помощью которых можно создавать эксплуатационные варианты систем. Исходной целью было сравняться по масштабу и общедоступности с реализацией « Всемирной информационной паутины World Wide Web », созданной на основе идей и программного задела Тима Бернерс- Ли почти 20 лет назад. К сожалению, несмотря на то, что указанные выше объединения начали разработку универсальных пакетов программой поддержки подобных структур более 10 лет назад – единой вычислительной структуры не получилось, а история как-то изложена в указанных выше обзорах. Получилось своеобразное « лоскутное одеяло » использования вычислительных ресурсов : в Европе своё, ( с применением версий пакетов GT стали строить « локальные гридики » в рамках локальных сетей организаций или стран ( как NorduGrid для северных стран ), в Америке своё.. Реализация проекта EGEE ( European Grid for E-science), в рамках которого до 2010-го года работали в основном те, кто был связан с обработкой данных с ускорителя LHC ( и не только), вынудила ответственных за программное обеспечение своих GRID-структур организовывать программные системные мосты для перехода к использованию EGEE ( эта ситуация охарактеризована в обзоре [ 4 ] ); возникла проблема обеспечения интероперабельности средств EGEE и Американского OSG ( Open Science Grid)... Наконец, в рамках расширения возможностей EGEE и унификации его использования по крайней мере для Европейских стран была запущена реализация панЕвропейского проекта EGI (European Grid Initiative) как преемника EGEE . Целью было и укрепление общей компьютерной ресурсной базы (например, включением в состав совместно используемого странами-участницами оборудования суперкомпьютерных центров из 15 европейских стран) плюс унификация использования того программного системного обеспечения, которое необходимо для доступа и использования объединённого Европейского Грид. Как указывалось в обзоре [5], всеми организационными и финансовыми вопросами занялся Совет EGI Council, куда входят и представители от России и Белоруссии : в их ответственность входит и предоставить для общего использования : например, грид-инфраструктуру RDIG ( Russian Data IntensiveGrid) и оборудование федерации суперкомпьютерных центров «Скиф –полигон» ( в которую вошли суперкомпьютерные центры ряда университетов и институтов России).

Ситуация с расширением рамок EGI за пределы Европы ( скажем , объединением с Американскими Грид- структурами и не только, что позволило бы говорить о проекте WGI ( Worldwide Grid Initiative)) , не очевидная, хотя, казалось бы, общей системной программной основой начала работ по созданию всех грид-структур были упомянутые выше пакеты Globus Toolkits и их развитие.. Так, в статье по адресу <http://x-com.parallel.ru/about.html> авторами из МГУ под руководством В.В.Воеводина отмечается : " Направление создания универсальных средств по созданию глобальных полигонов, объединяющих в рамках высокоскоростных сетей значительные распределённые ресурсы – интересное, однако реальные системы крайне тяжелы в установке, администрировании и сопровождении; организация расчётов на доступных компьютерах требует привилегированных административных полномочий, многие компьютерные платформы вообще не поддерживаются, тиражирование крайне затруднено. Примером работ в этом направлении является инфраструктура EGEE... " . Правда, в рамках проекта EGI усилия по преодолению указанных трудностей предпринимаются, но всё же..Для ряда прикладных задач типа той , которая описана в статье " Grids for Experimental Science: The Virtual Control Room " (см. [http://www.globus.org/alliance/publications/papers/clade\\_submitted\\_corrected.pdf](http://www.globus.org/alliance/publications/papers/clade_submitted_corrected.pdf) ) , авторам вполне достаточно было взаимодействия с системой Access Grid, когда для контроля и интерпретации результатов в проведении экспериментов по термоядерному синтезу на установке Токамак требовалось оперативное привлечение вычислительного ресурса...

Отдельной проблемой можно считать и проблему создания информационных систем и коллекций, которые называют "Digital Libraries" (DL) и VDL ("Virtual Digital Libraries"). Речь не идёт в основном о библиотеках в традиционном смысле, к этому понятию относят цифровые коллекции разного типа – например, коллекцию фотографий или снимков событий в экспериментах , дополненную средствами поиска через Web интересующей фотографии ( снимка) по определённым признакам. Для реализации таких средств должна быть

предварительно проведена обработка каждого элемента коллекции, что может потребовать значительных вычислительных ресурсов. В своём авторском обзорном докладе на конференции RCDL'2008 (Десятой Всероссийской конференции по тематике электронных библиотек и коллекций) я приводил пример реализации проекта DILIGENT ( Digital Library Infrastructure on Grid Enabled Technology) и его предвидевшемся развитии в последующие годы в рамках проекта D4Science ( сейчас он представлен на сайте по адресу <http://www.d4science.eu> ). Одной из первых прикладных целей проекта DILIGENT было создание сервисов для проекта SAPIR ( Search in Audio Visual Content Using Peer-to-Peer IR) как части проекта Chogus, т.е. для задачи создания в интересах этих проектов нового типа представления и поиска данных, отсутствовавших в традиционно используемых поисковых системах типа Google и Yandex. Указанным проектом DILIGENT авторов из CNR-ISTI ( Пиза, Италия) заинтересовался ЦЕРН и помог выделением компьютерных мощностей из ресурсов EGEE для создания и формализованного описания информационных объектов: с применением сервисов "gCube on top of gLite"(см. <http://www.gcube-system.org> ), разработанных авторами проекта, был проведён на инфраструктуре EGEE 16-недельный прогон ( data challenge) по обработке 37 млн. фотографий из on-line базы данных Flickr (известного модифицированного Web-приложения для поиска и обмена фотографиями), сгенерировано около 112 млн. текстовых и image-объектов...

Может быть, полезно ещё раз вспомнить и старую статью 2002-го года " The Semantic Grid: a Future e-Science Infrastructure" ( <http://www.semanticgrid.org/documents/semgrid-journal/semgrid-journal.pdf> ), где авторы предсказывали, что программная среда компьютеризованной науки и все Grids должны будут включать в себя трёхуровневую систему сервисов:

1. Data/Computation Services, средства размещения данных и их транспортировки между обрабатывающими программами, обеспечение вычислительных и сетевых ресурсов;
2. Information Services, средства представления, запоминания и доступа к информации, управления ею;
3. Knowledge Services, средства накопления, представления, обновления, "публикации" (сетевое распространения) знаний для помощи учёному в его исследовательском процессе.

Все положения демонстрировались детальным формализованным примером цикла полной автоматизации обработки экспериментальных данных в сетевой компьютерной среде ( от начала поступления данных на анализ до подведения итогов результата обработки научным сообществом) с применением конкретного перечня сервисов каждого из указанных уровней ; подчеркивалась роль семиуровневой системы онтологий ( аппарата формализованного представления информации) для нормального функционирования всей клиент-сервисной структуры приведённого примера.

Когда-то, комментируя эту статью в обзорном докладе на конференции "Научный сервис в сети ИНТЕРНЕТ" в 2003 году (см.[7]), я отмечал следующее ( на основе её авторских определений):

Разделение понятий "информация" и "знание" сделано просто: информация – это какие-то данные и их значения, определение, смысл ( "данное целое число относится к температуре во время реакции", "эта строка – имя человека"), а знание – это информация, побуждающая к действию ("данное значение температуры критическое, необходима остановка реакции"). Соответственно "сервис" можно определить как программный процесс реализации какого-то действия из набора служебных и прикладных программ в какой-то научной предметной области или в междисциплинарных сферах : например, сервис автоматического уведомления учёных, заинтересованных в результатах проведённой другими сервисами обработки какого-то набора данных. Агенты в этой схеме – своеобразные "брокеры" на бирже ( рынке) программных услуг-сервисов, программные инициаторы процессов : агент по своей инициативе или поручению от другого агента организует поиск нужного сервиса в каком-то репозитории, сверяет полномочия поручителя с указаниями в описании сервиса, запускает сервис в работу и предпринимает какие-то действия по концу его работы. Что касается упомянутой системы онтологий ( документов или файлов с метаданными, которые формально определяют классы, типы и свойства объектов, понятий, терминов.., а также отношения между ними за счёт использования описаний свойств классов и подклассов и логических правил вывода), то в упомянутой статье отмечается, что проблемы аннотирования контента ( содержания коллекций информации разных типов) и сервисов определяют необходимость порождения аппаратом онтологий следующих типов метаданных:

Domain ontologies : описания (концептуализация) важных объектов, их свойств и отношений между ними ( согласованный набор аннотаций, понятий, определений в предметной области...);

Task ontologies : описания задач и процессов, их свойств и отношений ( например, набора характеристик фаз процесса химического анализа..);

Quality ontologies: описание атрибутов знания ( например, аннотации к тому, могут ли результаты, полученные какими-то средствами, быть более эффективно получены и расширены более совершенными средствами) ;

Value ontologies : характеристика тех атрибутов, которые относятся к установлению значимости ( важности) контента ( " стоимость" полученных в эксперименте физических данных, например);

Argumentation ontologies : широкий набор аннотаций, имеющих отношений к описанию причин – почему контент был накоплен ( например, данные с какого-то эксперимента), почему он был использован тем

или иным способом, кто его одобряет или не признаёт..

Понятно, что в реализации такой архитектуры накопления, обработки и использования её результатов в значительной степени замешаны и понятие семантического Grid, и понятие семантического Web.. В этом смысле интересен доклад Хорошевского В.Ф. из ВЦ РАН "Онтологические модели и Semantic Web : откуда и куда мы идём ? ", опубликованный в трудах упомянутой выше конференции RCDL`2008 ...

Должен отметить, что многие работы по рассматриваемой в обзоре тематике рассматривались на четырёх международных конференциях "Распределённые вычисления и грид-технологии в науке и образовании" в ЛИТ ОИЯИ: последняя прошла в 2010 году: тезисы докладов опубликованы ( ISBN 978-5-9530-0253-0), предполагается публикация полных текстов. Впрочем, скажем, полный текст работы " Mediation Based Semantic Grid" сотрудников из ИПИ РАН ( соучастников реализации и развития международного проекта AstroGrid ) на русском языке и сейчас доступен по адресу <http://synthesis.ipi.ac.ru/synthesis/publications/10semgrid/10semgrid.pdf>.

Наконец, в заключение можно продолжить разговор по модной теме, которой посвящался заключительный раздел редакторского обзора [6] : о совместном использовании грид-технологии и технологии "облачной обработки данных" ( Cloud computing). Будет ли общий всемирный грид WGI или попрежнему будет многогридовая структура – от указанной темы не уйти. В этом смысле интересующимся можно рекомендовать материалы Европейского исследовательского консорциума по информатике и математике ( ERCIM ), подготовившего в октябре прошлого года специальный выпуск по этой теме ( см. <http://ercim-news.ercim.eu/en83> , где в принципе со страницы по этому адресу можно организовать скачивание 64-х страниц общим объёмом в 17 мегабайт ( файл в pdf-формате).

#### ЛИТЕРАТУРА:

1. В.П.Шириков. Программное обеспечение Grid: переоценка ценностей.//Труды Всероссийской научной конференции " Научный сервис в сети Интернет", Новороссийск, 2004, стр.142-144.
2. В.П.Шириков. Системное обеспечение "бесшовной" структуры и средств использования" Computational/Data Grid of Grids" для разных областей деятельности : достижения, нерешённые проблемы, виды на реализацию. //Труды Всероссийской научной конференции "Научный сервис в сети Интернет", Новороссийск, 2007, стр.10-13.
3. В.П. Шириков. RCDL`1999 – RCDL`2008: DL, VDL, Semantic WEB/GRID//Труды Десятой Всероссийской научной конференции, Дубна, 2008, стр.24-27.
4. В.П. Шириков. Программное обеспечение Grid: состояние и перспективы. [http://lit.jinr.ru/Inf\\_Bul\\_3/bullet.htm#\\_Toc98590864](http://lit.jinr.ru/Inf_Bul_3/bullet.htm#_Toc98590864).
5. В.П.Шириков. Обеспечение "бесшовной" структуры и средств использования "Computational /Data Grid of Grids" .[http://lit.jinr.ru/Inf\\_Bul\\_4/bullet\\_6.htm#\\_Toc190687952](http://lit.jinr.ru/Inf_Bul_4/bullet_6.htm#_Toc190687952).
6. В.П. Шириков. О новом проекте общеевропейской GRID-инфраструктуры. [http://Inf\\_Bul\\_5/bullet\\_8.htm](http://Inf_Bul_5/bullet_8.htm)
7. В.П. Шириков. Как у нас с интеллектом в Web и Grid для создания полноценного научного сервиса? //Труды Всероссийской научной конференции "Научный сервис в сети Интернет", Новороссийск, 2002, стр.33-38.