

Healthgrid Белая книга

Винсент Бретон, Кевин Дин и Тони Соломонидес,
редакторы по поручению создателей Белой книги Healthgrid'a

The Healthgrid White Paper

Vincent Breton, CNRS-IN2P3, LPC

Kevin Dean, Internet Business Solutions Group, Cisco Systems

Tony Solomonides, CEMS, University of the West of England, Bristol BS16 1QY, UK,

Editors on behalf of the Healthgrid White Paper collaboration

<http://clrwww.in2p3.fr/PCSV/IMG/pdf/WP%20IOS%20complete.pdf>

Аннотация

В течение последних четырех лет в сообществе исследователей, работающих над грид-технологиями и технологиями высокопроизводительного компьютеринга, велись дискуссии о препятствиях, с которыми столкнутся грид-технологии в области здравоохранения, и возможностях, предоставляемых ими для разработки медицинских приложений. Итогом дискуссий стала первая конференция по Healthgrid'у, прошедшая в Лионе, Франция, 16-17 января 2003 года. Конференция должна была расширить представления участников о возможностях и преимуществах применения грид-технологий для создания европейской / международной грид-инфраструктуры для здравоохранения.

Вопросы, обсуждавшиеся на этой конференции, сошлись с позицией отделения Еврокомиссии по е-здоровоохранению, чьим предписанием, принятым на встрече в Лиссабоне, было "Создать интеллектуальную среду, которая сделает возможным повсеместное наблюдение за состоянием здоровья граждан и поможет профессионалам здравоохранения справиться с некоторыми главными проблемами, учитывать факторы риска и вводить в клиническую практику новые медицинские знания". В этом контексте понятие "Охрана здоровья" включает не только клинические процедуры, но работу со всем спектром информации – от молекулярного уровня (генетическая и протеомическая информация), затрагивающего клетки и ткани, до уровня отдельного человека и населения в целом (социальное здравоохранение). Грид-технологии предлагают возможность

создания общей работоспособной основы для всех различных членов этого большого "семейства охраны здоровья" и должны привести к расширению знаний и взаимодействию различных дисциплин.

Следствием первой конференции по HealthGrid'у было создание HealthGrid-ассоциации, некоммерческой исследовательской организации, юридически зарегистрированной во Франции, но включающей широкий круг исследователей и институтов, объединяющих опыт создания гридов для здравоохранения.

После второй конференции, прошедшей в Клермон-Ферране 29-30 января 2004 года, возникла потребность в создании "Белой книги", отражающей текущее состояние дел и будущее гридов для здравоохранения. В подготовке этого документа приняли участие свыше пятидесяти экспертов из разных областей грид-технологий, приложений e-здравоохранения и медицины.

1. ОТ ГРИДА К HEALTHGRID'у : ПЛАНЫ И ТРЕБОВАНИЯ

1.1. ОБОСНОВАНИЕ

Доказательная медицина требует, чтобы медицинские решения принимались на базе достоверных знаний о пациенте в сочетании с подтвержденными экспертами научными доказательствами, а не на интуиции и искусстве врачей. К тому же очевидно, что существует насущная необходимость перехода от работы с бумажным представлением информации о пациенте к её электронному представлению. Страны ЕС вкладывают значительные средства в создание систем электронных историй болезни пациентов. При этом возникает проблема стандартизации и создания такого интерфейса через общие "языки", который сделает информацию разделяемой. Сложной и дорогостоящей может быть также технология защиты информации. Кроме того, ключевыми требованиями для обеспечения высокого качества обслуживания в здравоохранении являются доступ к множеству различных источников медицинских данных, обычно географически разделенных, и наличие компьютерных средств, позволяющих извлекать знания из этих данных.

Исследовательские проекты по интеграции биомедицинских знаний, достижения в области обработки медицинских изображений, развитие новых вычислительных средств и использование этих технологий в диагностике и лечении позволяют предположить, что базирующиеся на гриде системы смогут внести значительный вклад в достижение этой цели. Усовершенствованный доступ не просто улучшает интеграцию, предлагаемую гридом, но поднимает на новый уровень выгоды от такого доступа.

Грид-технология была объявлена одной из ключевых технологий в 'Области европейских исследований' ('European Research Area'). Основная трудность – вывести эту технологию из лабораторий во внешний мир за пределы одной только e-науки и сделать её доступной для e-бизнеса, e-правительства и e-здравоохранения. Были наглядно доказаны выгоды от использования грид-технологий в областях, связанных с длительными процессами моделирования большого числа экспериментов. Одной из главных областей применения грид-технологий является физика высоких энергий [1, 2, 3]. Хотя потенциальные возможности грид-технологий очевидны для многих приложений, требующих больших вычислительных мощностей и хранилищ данных большого объема, работающих с географически распределенной информацией или требующих

повсеместного доступа, продвижение грида происходит медленно. Объясняется это отсутствием адекватной инфраструктуры, недоверием пользователей и, чаще всего, небольшим числом приложений.

HealthGrid должен быть средой, где можно хранить данные, представляющие интерес для медицины, обрабатывать их и делать их легко доступными для различных участников системы здравоохранения: исследователей, врачей, центров здравоохранения, администрации и, в долгосрочной перспективе, граждан. Если бы такая инфраструктура могла гарантировать защиту данных, соблюдение этических норм и обязательных правил, она позволила бы объединить постгеномную информацию и медицинские данные и открыть возможность для индивидуализированного здравоохранения.

Данная Белая книга представляет обзор технологий HealthGrid'a, содержащий описание текущего состояния грида и е-здравоохранения и анализ промежуточных разработок и возможностей. Существуют многочисленные движущие силы, стимулирующие внедрение и эксплуатацию средств защищенного, всеобъемлющего, повсеместного и прозрачного доступа к информации и вычислительным ресурсам, который могут обеспечить грид-технологии. Многие технические проблемы, возникающие в е-здравоохранении (стандартизация данных, интеграция баз данных, извлечение знаний из контента, управление персональными данными ...) могут быть решены с использованием грид-технологий. Требуется, однако, преодолеть много барьеров. Мы рассматриваем применение грида в таких областях как физика высоких энергий или численное моделирование и обсуждаем их отличие по отношению к здравоохранению с тем, чтобы наметить путь к успешному применению грид-технологий в е-здравоохранении и, в конечном итоге, к созданию HealthGrid'a.

1.2. ВВЕДЕНИЕ В HEALTHGRID

1.2.1. Европейский сектор здравоохранения

Е-здравоохранение использует информационно-коммуникационные технологии (ИКТ) для создания интеллектуальной среды, которая дает возможность повсеместно следить за состоянием здоровья граждан, помогает профессионалам в области здравоохранения справляться с трудными проблемами или интегрировать новые знания в клиническую практику.

В е-здравоохранении было разработано много приложений для управления информацией и решения технологически сложных задач в современном здравоохранении. Е-здравоохранение – это не только хорошая стратегия для повышения качества здравоохранения, но и хороший бизнес. Е-здравоохранение, или сектор телематики¹, становится третьей промышленной опорой, на которых держится здравоохранение, после фармацевтической промышленности и производства оборудования для получения медицинских изображений. Подсчитано, что затраты на ИКТ-системы возрастут к 2010 году с 1% до 5% [10], сегодня в Интернете насчитывается более 1500 сайтов, посвященных охране здоровья, и предсказывается увеличение доходов в е-здравоохранении к 2004 году с \$22 миллиардов до \$348 миллиардов. Охрана здоровья является второй наиболее часто упоминаемой темой в Интернете [11].

¹ telematics - интегрированные средства обработки и передачи информации.

Важным с точки зрения бизнеса является использование в е-здравоохранении приложений, базирующихся на службах. Например, хостинг ASP (Application Service Provision) позволяет провайдерам служб специализироваться на установке и сопровождении приложений и служб для своих клиентов. ASP перекладывает заботу об аппаратной инфраструктуре на провайдеров, так что клиентам для доступа к программному обеспечению достаточно только браузера Интернета. Основные преимущества ASP, такие как специализация персонала и ресурсов, хороший спрос на рынке и растущие инвестиции, усиливаются благодаря ситуации в секторе здравоохранения: рынок в здравоохранении фрагментирован, поскольку многие пользуются собственными системами; работа с ними часто скучна и утомительна и могла бы быть рационализирована; организации располагают довольно устаревшими компьютерными системами и меньшим, чем в других секторах, персоналом, знакомым с ИКТ. Существуют, однако, и некоторые неблагоприятные обстоятельства. Так, соблюдение права собственности является гораздо более критичным в здравоохранении, чем в других областях. Кроме того, медицинские записи появляются с длительными перерывами во времени и требуют долговременного хранения, а сохранение конфиденциальности информации является обязательством перед законом. К тому же, требования к высокому качеству обслуживания гораздо более критичны в здравоохранении, и при работе с медицинской информацией могут потребоваться линии связи с высокой пропускной способностью для получения минимального времени ответа. Тем не менее, электронная обработка медицинских данных открыла широкие возможности для решения таких задач как диагностика, планирование в хирургии или терапии, причем как в клинической практике, так и в клинических исследованиях.

Связывание баз данных, содержащих медицинскую информацию, необходимо, но это только часть решения. Должна производиться дальнейшая обработка информации с целью извлечения знаний, поскольку полный объем медицинской информации делает невозможной её непосредственную обработку. Извлечение знаний может обеспечить средства для анализа релевантной информации и проведения исследований, направленных на охрану здоровья населения.

Электронная обработка медицинских данных находится сейчас на разных стадиях развития в разных центрах обработки и даже в разных медицинских учреждениях. Широко используются больничные информационные системы (Hospital Information Systems) для наблюдения за стационарными больными. Лаборатории и отделения диагностики, работающие с медицинскими изображениями, в значительной степени пользуются электронными данными пациентов. В первичной медицинской помощи эти технологии менее продвинуты. Однако, главная проблема – это так называемая электронная история болезни пациента (EPR, Electronic Patient Record). EPR обещает согласованный, понятный доступ к полной информации о пациенте и управление этой информацией, как для отдельного человека, так и для групп населения. Для достижения этой цели уже предпринимаются значительные усилия, включая и работы по стандартизации.

Самый важный момент – это защита данных. Персональные данные (любая часть информации, по которой её владелец может быть идентифицирован непосредственно или в сочетании с информацией, которая является доступной или может быть таковой) – это конфиденциальная информация, поэтому доступ к ней должны иметь только авторизованные и аутентифицированные лица, а сами данные должны быть закодированы, чтобы гарантировать их конфиденциальность и целостность. Более того,

электронное архивирование персональных данных строго регулируется европейскими и национальными законами. Другими важными аспектами являются повсеместно распространенный доступ и отказоустойчивость, поскольку медицинское обслуживание должно быть доступно круглосуточно.

Объем медицинской информации огромен, а сама она рассредоточена. Требуются большие ресурсы для хранения данных о пациентах, медицинских изображений, биосигналов, нешифрованных текстов, видеоматериалов, фотографий и других форм цифровых данных. К тому же, здравоохранение имеет распределенную структуру, а информация, которой владеют больницы, пункты первичной медицинской помощи и травматологические отделения, не консолидирована. Для связывания интегрированных баз данных требуются новые достижения в компьютеринге и сложные структуры. Медицинская информация далека от 'стандарта'. Часто она представляется во взаимно не совместимых форматах, а имеющиеся стандарты не полны и не универсальны. Даже использование стандартного протокола не дает уверенности, что независимо полученные данные, представляющие 'одну и ту же' часть информации, будут идентичны. Окончательные результаты зависят от налаженности и качества оборудования и квалификации персонала.

В среднесрочной перспективе разумно ожидать, что большинство служб здравоохранения будет использовать компьютерные ресурсы для хранения, обработки и совместного использования информации о пациентах. Отрабатываются нужные технологии, и уже устанавливаются коммуникационные сети с высокой пропускной способностью между центрами охраны здоровья по всей Европе. Новой ключевой работоспособной технологией является грид.

1.2.2. Введение в грид

Грид-компьютеринг нацелен на обеспечение глобальной ИКТ-инфраструктуры, которая сделает возможным для динамичных и географически рассредоточенных организаций и сообществ (виртуальных организаций) скоординированное, гибкое и безопасное совместное использование разнообразных ресурсов, включающих компьютеры, приложения, данные, устройства памяти, сети и научный инструментарий. Применение грид-технологий организациями изменит способы и средства решения сложных задач, поскольку грид предлагает беспрецедентные возможности для разделения ресурсов и совместной работы. Подобно тому, как World Wide Web трансформировал для нас способ обмена информацией, концепция грида поднимает на новый уровень параллельный и распределенный компьютеринг, предлагая унифицированную, отказоустойчивую, прозрачную инфраструктуру, доступную по запросу (on demand), для решения все более сложных задач.

Гриды можно разделить на вычислительные гриды, гриды данных /информации/ знаний и коллективные гриды. Назначение вычислительного грида – создание виртуального суперкомпьютера, который динамически соединяет в одно целое производительность большого числа отдельных компьютеров, чтобы обеспечить платформу для современных высокоэффективных и/или высокопроизводительных приложений, которые не могут быть выполнены одной системой. С другой стороны, гриды данных нацелены на совместное использование огромных объемов данных. Гриды информации и гриды знаний расширяют возможности гридов данных тем, что обеспечивают поддержку категоризации данных, обнаружения информации, онтологий,

разделения знаний и их повторного использования. Коллективные гриды устанавливают виртуальную среду, позволяющую географически рассредоточенным пользователям или группам пользователей кооперироваться для достижения общих целей. Технологии коллективного грида дают возможность для реализации виртуальных лабораторий или удаленного контроля и управления оборудованием, датчиками и приборами.

От начальных экспериментов по исследованию возможностей, предлагаемых широкополосными сетями, грид-технологии вышли теперь на такой уровень, когда стало возможным их промышленное использование, например, появились Information Power Grid в NASA, DataGrid в ЦЕРНе, TeraGrid в NSF. Однако, во многих областях широкомасштабное разделение ресурсов еще не стало реальностью. Это объясняется, главным образом, недостатком общепринятых стандартов и разнородностью и раздробленностью имеющегося программного обеспечения грида, его средств и предоставляемых услуг. Основным органом, занимающимся глобальной стандартизацией грид-служб, протоколов и интерфейсов, является Global Grid Forum (GGF), куда входят представители промышленности, исследовательских и академических сообществ.

В соответствии с недавним обзором двадцати европейских грид-проектов наиболее широко используемым промежуточным программным обеспечением является Globus toolkit, за ним следует Unicore. Однако, за последние два года Globus toolkit, который первоначально был разработан в академическом сообществе для удовлетворения потребностей в разделении ресурсов для высокопроизводительного компьютеринга (HPC – High-Performance Computing), значительно уступил концепции ориентированности на грид-службы и растущей поддержке использования коммерческих Веб-служб. Первой попыткой свести вместе грид-технологии и Веб-службы стала Открытая архитектура грид-служб (OGSA, Open Grid Services Architecture). Недавно GGF объявил о намерении реализовывать OGSA на базе структуры WSRF (Web Services Resource Framework), стандарты для которой разрабатываются в настоящее время группой OASIS. Это позволит использовать для грид-компьютинга стандартные технологии Веб-служб, имеющие широкую промышленную поддержку.

Характерным для дальнейшего развития грид-технологий будет полное принятие концепции ориентированности на службы и технологий Веб-служб, полная виртуализация ресурсов и служб и растущее использование семантической информации и онтологий. Потребуется значительные усилия для обеспечения соответствующих средств высокого уровня и создания среды, которая спрячет сложность и сократит расходы на создание грид-приложений. Предпосылками для широкомасштабного распространения грид-технологий послужат принятие развитых стандартов безопасности, поддержка высокого качества обслуживания и создание связанных с гридом бизнес-моделей и процессов.

1.2.3. HealthGrid'ы

HealthGrid'ы – это грид-инфраструктуры, состоящие из приложений, служб или компонент промежуточного программного обеспечения, применяемых для решения специфических задач, возникающих при обработке биомедицинских данных. Ресурсами в HealthGrid'ах являются базы данных, вычислительные мощности, медицинский опыт и даже медицинское оборудование. Таким образом, HealthGrid'ы тесно связаны с здравоохранением.

Хотя конечной целью для е-здравоохранения в Европе было бы создание единого HealthGrid'a, т.е. грида, объединяющего все ресурсы е-здравоохранения и, естественно, включающего средства защиты и авторизации для поддержания субсидиарности¹ независимых узлов HealthGrid'a, развитие, с большой вероятностью, будет идти по пути создания отдельных HealthGrid'ов, может быть, с зачатками интероперабельности.

При дальнейшем развитии грид-технологий[13] будут более точно учитываться задачи, присущие именно е-здравоохранению. Приложения HealthGrid'a ориентированы как на индивидуальную охрану здоровья, так и на эпидемиологический анализ. Индивидуализированная охрана здоровья улучшается посредством эффективного и надежного сочетания незамедлительной доступности персональной клинической информации и наличия развитых услуг в диагностике и терапии. Эпидемиологические HealthGrid'ы объединяют информацию, собранную среди многочисленных групп населения, с целью извлечения знаний, которые могут привести к открытию новых взаимосвязей между симптомами, болезнями, наследственными признаками и другими клиническими данными.

Следующие задачи являются ключевыми для HealthGrid'ов:

- HealthGrid'ы тесно связаны с данными, но многие больницы очень неохотно идут на передачу информации за пределы больницы. Для широкомасштабного развертывания HealthGrid'ов и их привлекательности для бизнеса очень важно поднять защиту данных до такого уровня конфиденциальности, который сделал бы возможным общий доступ к данным извне (см. ниже). Хотя за хранение данных отвечают больницы, разделение данных и обрабатывающие приложения могут создать возможности для бизнеса. Интеграция баз данных требует дальнейшей работы в компьютеринге и создания сложных структур данных;
- управление распределенными базами данных и возможности для извлечения данных являются важными средствами для многих биомедицинских приложений в таких областях как эпидемиология, создание лекарств или даже диагностика. Экспертные системы, запускаемые на гриде, должны уметь 'рыться' в больших распределенных базах данных, извлекая знания, которые могут привести к раннему распознаванию новых источников болезней, к определению групп риска среди населения, к наблюдению за развитием болезней или к поиску подходящих протеинов для борьбы с определенными болезнями;
- защита данных в грид-инфраструктурах достаточна для исследовательских работ, но должна быть усилена в будущем, чтобы гарантировать конфиденциальность данных. Передача и хранение данных в зашифрованном виде недостаточны, должны быть обеспечены сохранность данных и автоматическая псевдо-анонимизация, чтобы гарантировать полноту и надежность данных и предотвратить утечку информации при неконтролируемом доступе к ресурсам. С биомедицинской информацией нужно обращаться очень аккуратно, чтобы избежать утечки конфиденциальных данных. Нарушение конфиденциальности наносит непоправимый ущерб, поскольку нет способа вернуться к предыдущей ситуации. Защита передачи данных должна сочетаться с защитой хранения, при

¹ **Принцип субсидиарности** предполагает процедуру распределения и перераспределения полномочий между управленческими уровнями государственной власти, разграничение компетенций.

этом должны использоваться такие механизмы защиты, которые позволят предотвратить неавторизованный доступ к гриду злонамеренных пользователей с целью дешифровки и визуализации персональных данных. Необходима автоматическая псевдо-анонимизация;

- устойчивость гридов по отношению к ошибкам и отказам очень хорошо подходит для медицинских приложений, которые всегда должны быть готовы к исполнению ('always on'). Грид-технологии могут облегчить доступ к реплицированным ресурсам и информации, требуя только постоянного подключения пользователей к Интернету;
- исследовательским сообществам в области биокомпьютинга и биомоделирования очень нужны ресурсы, которые они могут получить через грид. Для доступа к большим базам данных необходима согласованность с медицинскими стандартами. Имеется много уже существующих и появляющихся новых стандартов, которые должны быть приняты во внимание. Очевидно также, что в гриде должно быть обращено особое внимание на сложную и мультимедийную информацию, такую как медицинские изображения, сигналы, видео и т.д. и на то, что эта информация более чувствительна к форматам данных;
- наконец, необходима определенная гибкость для управления виртуальными организациями на высоком уровне. Управление ресурсами должно быть более точным и динамичным, зависящим от многих критериев, таких как срочность, авторизация пользователей и другая административная политика.

На сегодняшний день большинство грид-приложений для медицины следует классическому подходу ориентированности на высокую производительность. Численное моделирование органов на основе данных, полученных от пациентов [14, 15], обеспечивает информацию для расширения знаний о болезнях или для разработки медицинского оборудования. Обращение к подходам, направленных на нужды пациентов, можно найти на исследовательском уровне в таких областях, как радиотерапия, черепно-лицевая хирургия или нейрохирургия.

Приложения в других областях имеют дело с обработкой информации большого объема, такой как медицинские изображения. Работе со снимками рака молочной железы были посвящены несколько успешных грид-проектов [16, 17] и имеются проекты здравоохранения, подходящие для их передачи в грид [18]. Основное внимание было уделено интеграции данных на федеральном уровне и их совместному использованию и реализации полуавтоматических средств обработки, которые могли бы улучшить специфичность и чувствительность программ скрининга¹ рака молочной железы. Много усилий было затрачено на предварительную обработку данных и обеспечению конфиденциальности.

Рассматриваются также такие подходы к применению грида в медицине, когда центром внимания является человек [19]. Главная цель такого подхода – сделать

¹ **Скрининг** - массовое предварительное обследование, использующее относительно недорогие и нетрудоёмкие технологии.

информацию доступной для всего медицинского сообщества (пациент, родственники, врачи, детские ясли и сады...) с учетом, конечно, прав доступа и языковых ограничений.

Биоинформатика – это та область, где больше всего начинают применяться грид-технологии. Основная проблема, стоящая перед биоинформатикой, это развитие и сопровождение инфраструктуры для хранения и передачи биомедицинской информации, доступа к этой информации и моделирования биомедицинских процессов. Текущие работы по биокомпьютингу [19.А] согласуются с целями грид-технологий. Работа по интеграции клинической и генетической распределенной информации и развитие словарей стандартов облегчат разделение данных и ресурсов.

1.3 ЧЕГО НЕ ХВАТАЕТ, ЧТО ВОЗМОЖНО И ЧТО ТРЕБУЕТСЯ ПРОМЫШЛЕННОСТИ

Грид-технология – это ‘движущаяся цель’. Бурное развитие платформ и версий создает значительные трудности для доведения приложений до стадии производства. Главный интерес промышленности состоит в определении и эксплуатации бизнес-моделей на гриде. Но промышленность нуждается в большей стабильности и стандартизации грид-инфраструктур прежде чем начнется разработка жизнеспособных бизнес-моделей. Действительно, имеющемуся промежуточному программному обеспечению грида не хватает нескольких компонент, обязательных для бизнес-приложений:

- программному обеспечению грида недостает надежных и полных служб учета, которые могут четко идентифицировать поставщиков, потребителей и использование ресурсов в сценарии, в котором задействован широкий круг разнородных ресурсов, обладающих различными свойствами. Экономические Вопросы эксплуатации грида еще требуют дальнейшей разработки;
- текущие работы, касающиеся устойчивости по отношению к ошибкам и отказам, безусловно повысили надежность и производительность программного обеспечения грида, что является ключевым моментом при эксплуатации приложений в здравоохранении, но еще не дошли до производственного уровня;
- модели защиты и секретности в гриде не адекватны приложениям, которые могут быть сертифицированы конечным пользователем и полномочными органами в здравоохранении;
- должны быть проведены надежные испытания, чтобы удостовериться в качестве обслуживания и устойчивости к ошибкам у всех компонент, что необходимо для приложений в здравоохранении. Еще важнее сертификация промежуточного программного обеспечения для приложений здравоохранения, т.к. они могут оказать влияние на здоровье пациентов и затрагивать правовые и этические факторы;
- важная проблема, которая возникает при эксплуатации грида, это использование лицензий на базовое программное обеспечение;

- в настоящее время эти лицензии обычно не допускают использования программного обеспечения в среде грида, где компьютеры и пользователи не определены четко. С развитием новых бизнес-моделей могут появиться и новые модели лицензирования. До тех пор успешным приложениям лучше использовать собственное или лицензионное программное обеспечение;
- прежде чем приступать к разработке приложений, необходимо создать производственную инфраструктуру (production infrastructure), в которой эти приложения будут работать. В такой инфраструктуре можно реализовать многие службы, тестировать их и получить подтверждение готовности. На такой платформе можно выполнить тестирование и подтверждение готовности для любого медицинского приложения, хотя окончательная эксплуатация может проводиться на самостоятельных ресурсах.

Существуют, по крайней мере, три сценария, согласно которым грид-технологии, применяемые в здравоохранении, могут быть успешными с точки зрения промышленности:

- консолидация ресурсов: необходимы объединенные решения для приложений, данных и ресурсов в центре и регионах. (Существующие в настоящее время технологии распределенных баз данных не предлагают нужного уровня интероперабельности или возможности предоставления других ресурсов, кроме данных);
- усиление эффективности: идеальными с точки зрения бизнеса являются такие приложения, для которых требуется максимальная нагрузка на ресурсы, следующая за неактивными периодами;
- сокращение производственных затрат для приложений, где отдача от вложений низкая, но социальное влияние может быть значительным. Создание общественно-частного консорциума может способствовать достижению целей, которые представляют основной интерес для большинства населения, но приносят небольшую экономическую выгоду. Предоставление ресурсов для проведения экспериментов *in silico* может стимулировать открытие легких в применении, доступных, эффективных лекарств для болезней, которым не уделялось должного внимания.

Предстоит пройти долгий путь, прежде чем HealthGrid'ы войдут в эксплуатацию, и промышленность должна оказывать помощь в проведении исследований, чтобы извлечь выгоду из надежных и интересных результатов.

1.3.1. Фармацевтическая промышленность

Сближение биотехнологии и ИКТ открывает пути к новым методам создания лекарств, что требует для фармацевтической промышленности огромных вычислительных мощностей для моделирования, конструирования и проверки взаимодействия лекарств с рецепторами, чтобы на основании результатов принять решение о синтезировании и опробовании лекарства.

Чтобы лекарство могло появиться на рынке, требуется несколько лет исследований. Необходимо сокращать время выхода на рынок новых лекарств, и этого можно добиться путем увеличения объема вычислений, выполняемых для докинга¹. Вычисления для виртуальных соединений порождают информацию такого объема, анализ которой требует и большого времени, и больших затрат. Результаты анализа нужно сохранять для дальнейшей обработки и использования, что требует создания механизмов для защиты и сохранения конфиденциальности этой информации при её использовании в интегрированных на федеральном уровне базах данных.

Фактически, эта информация избыточна, но это вызывается недостатком интероперабельности между различными приложениями и источниками данных. Имеющиеся средства не позволяют ни эффективно обрабатывать результаты исследований, ни извлекать достаточно знаний. Это означает, что значительная часть информации оказывается ненужной, а результаты не используются. Для успеха фармацевтической промышленности крайне необходимо сотрудничество ученых и исследователей, работающих в промышленности.

Следующий шаг в разработке лекарств – это интеграция фенотипа с генотипной информацией и факторами окружающей среды, что должно привести к созданию ‘персональных’ лекарств, а это потребует анализа по запросу (on-demand) и больше ресурсов и программных средств.

1.3.2. Отрасль медицинских информационных технологий

Самая важная задача в медицинских информационных технологиях – это достижение максимальной степени интероперабельности, безшовного доступа и обработки распределенной электронной медицинской информации. Решение этой задачи, базирующееся на электронных историях болезни пациентов, требует взаимодействия промышленности, исследователей и организаций по выработке стандартов.

Для достижения этих целей недостаточно только объединить распределенные базы данных. Во-первых, не вся информация в этих базах совместима, и не только по формату, но и из-за различающихся процедур, медицинского оборудования, человеческого вмешательства и других факторов. Чтобы интероперабельность могла стать реальностью, федерация данных должна быть достигнута на семантическом уровне. Во-вторых, много медицинской информации в настоящее время не обрабатывается электронным образом. Основные показатели состояния организма (vital sign), проверка восприятия (perception tests), лабораторные анализы хотя обычно и записываются и даже в цифровом виде, но недоступны для дальнейшей обработки из-за отсутствия коммуникаций или несовместимости данных. Интерфейсы для оборудования и форматы данных согласовываются и стандартизируются, но спрос на них растет медленно.

Ведение полных электронных историй болезни пациентов потребует значительных ресурсов для их хранения и обработки, так что клинические учреждения, несомненно, вынуждены будут использовать услуги компьютеринга. Настоятельной необходимостью будет взаимодействие различных устройств. Такая инфраструктура сможет предоставить новые услуги, такие как компьютерная диагностика, обработка медицинских

¹ **Докинг** - это одна из важнейших стадий процесса компьютерного конструирования лекарства. Задачей докинга является построение модели структуры комплекса молекулы лиганда (биологически активного вещества) и молекулы рецептора (биомишени).

изображений, извлечение основных показателей состояния организма, оценка клинических результатов или даже моделирование.

1.4. ЧЕГО НЕ ХВАТАЕТ, ЧТО ВОЗМОЖНО И ЧТО ТРЕБУЕТСЯ ДЛЯ ЗДРАВООХРАНЕНИЯ И МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ

Ситуация в “обычном” здравоохранении сильно отличается от ситуации в медицинских исследованиях. Главная цель гридов, ориентированных на здравоохранение, это надежный и эффективный доступ к большим объемам данных. Медицинские же исследования имеют дело с более широким кругом вопросов. Для медицинских приложений существенным является наличие компьютерных ресурсов, возможность извлекать знания из очень больших баз данных и средства для решения суперсложных задач (grand-challenge problems).

Медицинские приложения в биокомпьютинге для биомедицинских исследований с помощью грида принадлежат к семейству “киллер-приложений”. Высокий уровень развития генетики и биомедицинских технологий приближает их к медицине, а решение суперсложных задач биокомпьютинга возлагается на грид [19.A]. Другой важной областью для грид-приложений является биомедицинское моделирование и имитация. Биомедицинские модели сильно связаны друг с другом, затрагивают сложные физические процессы и требуют интенсивных вычислений. Связывание моделей важно для получения реалистичной имитации (simulation), что могло бы вызвать ответную реакцию медицинской науки и технологии медицинских инструментов. Задачу получения в перспективе “виртуального человека” технологически можно осуществить только при помощи очень больших вычислительных ресурсов. Инфраструктур национальной е-науки для достижения такой грандиозной цели недостаточно.

Гриды для здравоохранения должны быть обеспечены соответствующими службами для запроса, хранения и извлечения мультимедийных медицинских данных из гридов данных. Должны быть предусмотрены способы сохранения секретности данных, чтобы доступ к медицинским базам данным был возможен не только для их владельцев, но и извне.

Чтобы избежать дублирования работ и гарантировать применимость результатов, нужна координация с инициативами по электронным историям болезни пациентов. Для получения доступа к данным очень важно будет установить связь с медицинскими информационными системами, такими как больничные информационные системы (Hospital Information Systems), системы компьютерного архивирования медицинских изображений (Picture Archiving Computer Systems), систем радиологических и лабораторных данных и данных первичной медицинской помощи (Radiology, Laboratory and Information Systems), а создание библиотек служб облегчит разработку медицинских приложений.

Наконец, последнее, но не менее важное: грид – это одно из важнейших средств распространения знаний в развивающихся странах. Совместное использование медицинских данных, процедур, служб и накопленного опыта позволит исследователям в этих странах, где такие средства не существуют, повысить качество медицинского обслуживания и накопить нужный опыт в медицине.

1.4.1. Обработка медицинской информации

Конечной целью биомедицинской и медицинской информатики является поддержка непрерывной индивидуализированной охраны здоровья – от предупреждения заболевания до реабилитации. Однако, внедрение информатики и технологических средств в клиническую практику развивается гораздо медленнее, чем ожидалось, и коммуникационный зазор между клиницистами и специалистами по информатике все еще существует.

В последние годы велось активное обсуждение причин медленного внедрения результатов исследований. Некоторые причины вызываются как самими исследованиями, так и их реализацией и связаны с трудностями, присущими именно медицинской информатике, такими как сложность информации и организации, влияние человеческого фактора, различия в финансовой культуре и культуре производства и др. Например, алгоритмы, разработанные и успешно примененные для очень узкого класса специальных случаев, должны пройти проверку на их соответствие потребностям здравоохранения. Вкладом в развитие технологий в мире медицины могла бы стать объединенная платформа биомедицинской и медицинской информатики, которая сделала бы возможным взаимодействие и интеграцию ресурсов, что, в свою очередь, послужило бы поддержкой доказательной медицины и подтверждения правильности результатов исследований.

Ключевой момент в медицинской информатике – это управление медицинской информацией, эффективная и качественная сертификация информации и обмен знаниями между всеми участниками процесса оказания медицинской помощи. Предварительно полученное знание должно быть сохранено и представлено в структурированной форме, чтобы его можно было извлекать в правильном контексте и в организованном виде для использования в образовательных и исследовательских целях и, в то же время, для получения новых медицинских диагнозов и генерации новых медицинских знаний.

Рассматривались также основная стратегия и область действия медицинской информатики в контексте её взаимосвязи с биоинформатикой. Основой для сотрудничества двух этих дисциплин могут стать такие задачи, как понимание причин заболевания на молекулярном уровне, эффективное наблюдение за заболеваниями хронических больных и интеграция клинических и генетических данных. Интересную перспективу открывает сочетание всеобщей компьютеризации, облегчающей передачу и накопление биологических данных в реальном времени вне клиники, с биомаркерами и другими индикаторами, что может привести к новому этапу для систем помощи на дому.

В завершение нужно сказать, что возникает потребность в соблюдающих этические нормы приложениях для обмена знаниями и их синтеза – внутри сложных систем взаимодействия между исследователями и пользователями, в междисциплинарной среде – с тем, чтобы ускорить получение выгоды от исследований через усовершенствование здравоохранения, развитие более эффективных услуг и препаратов и укрепление системы здравоохранения. Этим требованиям отвечают грид-технологии, которые обеспечивают функциональную и архитектурную структуру, способствующую проявлению синергии и решающую вопросы этики и конфиденциальности.

1.4.2. Биомедицинское моделирование

Исследования физической природы биомедицинских процессов, происходящих в человеке, начались совсем недавно. Объединение точных и полных средств моделирования инженерных процессов способствовало разработке биомедицинских моделей структурной динамики, динамики жидкостей, химических процессов или распространения электрического напряжения, которые описывали с большой степенью точности физику многих органов и тканей.

Все эти модели обычно применяются к ограниченным малым областям или не дают желаемой точности, поскольку требуют большой памяти. Кроме того, сложность биомедицинских моделей человека вызывается высокой степенью связанности между собой химических, структурных, магнитных и электрических процессов. Эта сложность требует усовершенствования биомедицинских моделей и наличия беспрецедентно огромного объема вычислительных ресурсов и ресурсов памяти.

Таким образом, движение к модели “виртуального человека” – это долгосрочная цель биомедицинского компьютеринга. Решение такой задачи требует тесного сотрудничества многих организаций, разделяемых вычислительных ресурсов, моделей и данных. К точным медицинским моделям нет свободного доступа, обычно они представляют собой наиболее ценный капитал исследовательских центров. Необходимы средства для кооперации, не нарушающие права интеллектуальной собственности.

1.4.3. Геномика¹

Проекты секвенирования генома (Genome-wide sequencing)² были выполнены для многих организмов, включая *Homo Sapiens* [4] и *Mus Musculus* [5]. Это коренным образом меняет традиционный подход к биомедицинскому открытию, в котором понимание некоторой биологической функции требует идентификации (и секвенирования) одного или больше генов, ответственных за эту функцию: в настоящее время ситуация такова, что были секвенированы тысячи генов, но они ждут, чтобы им была приписана какая-то функциональная информация.

Тот факт, что гены с неизвестными функциями составляют свыше 70% всех генов, наводит на мысль, что в настоящее время понимание большинства биологических и патологических процессов еще далеко от завершения. Вследствие этого с невероятной высокой скоростью развиваются новые технологические платформы, которые используют информацию о секвенировании генома, чтобы систематическим образом исследовать функции генов. Применение технологии микрочипов [6] к более трансляционным (translational) областям исследования, таким, как исследование рака, раскрыло её огромный потенциал при установлении диагноза. Недавние работы показали, что можно использовать анализ профиля экспрессии генов из образцов опухолей для определения

¹ Геномика - недавно возникшее направление науки, объектом изучения которой являются геномы

² Секвенирование –методы расшифровки нуклеотидной последовательности нуклеиновых кислот, определение последовательности ДНК в геноме

набора генов (сигнатур), чья экспрессия коррелирует, положительно или отрицательно, со специфическими клиническими признаками, такими, как продолжительность жизни больного без возникновения метастазов при раке молочной железы [8], и для влияния на терапию [7]. Другие типы огромных массивов данных, генерируемых в проектах по геномике, содержат: уровни экспрессии белков, измеряемые при помощи протеомического скрининга¹; данные о взаимодействии белок-белок в различных организмах; данные о структуре белков; секвенирование геномов дополнительных организмов и сравнение геномов; последовательность полиморфизмов в человеческой популяции, мутационный анализ рака и наследственных болезней у человека; анализ потери функции в различных организмах при помощи РНК [9].

Проведение исследований в геномике приводит к тому, что биомедицинские базы данных постоянно и экспоненциально увеличиваются в числе и размерах, и параллельно с этим развиваются и средства извлечения информации из этих баз данных. Главные исследовательские лаборатории (например, NCBI в США и EBI в Европе) занимаются сбором и регулярным обновлением информации. Эти данные могут быть проанализированы с использованием Веб-интерфейса наряду с группой хорошо известных приложений (главным образом, с программами извлечения данных), которые интенсивно используют процессор и требуют больших объемов ввода/вывода.

Часто процесс анализа данных требует конвейерной передачи результатов через разные приложения. Выборка результатов, полученных базирующимися на Вебе приложениями - неудобная и чреватая ошибками процедура, требующая "ползания по экрану". Она становится еще более трудной в связи с изменениями в Веб-интерфейсе. Несмотря на то, что вычислительные ресурсы, выделяемые для каждого из исследователей по отдельности, ограничены, одновременный доступ к Веб-приложениям приводит к перегрузке главного ресурсного центра. В результате биологи предпочитают выгружать файлы из баз данных и проводить их обработку локально.

Это имеет два главных последствия: каждый отдельный исследователь должен отслеживать изменения в базе данных и копировать для себя актуальную версию данных; массивная выгрузка громадных объемов данных ухудшает производительность Веб-сайта и приложений.

Еще один важный аспект – недостаток стандартизации в публикуемых базах данных: перекрестные ссылки данных затруднены (если вообще возможны) из-за избыточности и несвязности, нет ни стандартного языка запросов, ни центрального управления данными и, наконец, различные приложения требуют разных форматов для одних и тех же данных. Контроль за качеством данных и, соответственно, за конфиденциальностью полученных результатов очень слаб.

Ожидается, что грид-инфраструктура преодолеет многие из этих препятствий, возникающих при использовании базирующихся на Вебе подходов к работе с данными о геномах, предоставляя новые услуги, такие как прозрачный доступ к компьютерным ресурсам для проведения интенсивных вычислений, требующихся для решения биомедицинских задач. Другая важная задача – это создание разделяемых связанных реляционных баз данных и управление этими базами для устранения несвязности и

¹ протеомика - наука, занимающаяся инвентаризацией (определением и классификацией) белков клетки

противоречивости в актуальных базах данных и для обеспечения инфраструктуры, которая позволит собирать данные, полученные в результате геномных экспериментов, и предоставит средства для управления копиями данных и координации их обновления.

Наконец, в качестве главного требования нужно рассматривать защиту баз данных (все аспекты, касающиеся конфиденциальности данных), шифрование данных при передаче и, последнее, но не менее важное, аутентификацию и авторизацию пользователей.

1.5. ССЫЛКИ

1. "LHC Computing Grid Project", <http://lcg.web.cern.ch/LCG/>
2. "The CrossGrid Project", Technical Annex <http://www.lip.pt/computing>
3. "Project Presentation" The DATAGrid project, <http://www.eu-dataGrid.org>
4. "Human Genome Resources", <http://www.ncbi.nlm.nih.gov/>
5. "Mouse Genome Resources", <http://www.ncbi.nlm.nih.gov/>
6. Z.G. Goldsmith and N. Dhanasekaran "The Microrevolution: Applications and impacts of microarray technology on molecular biology and medicine (Review)". Int J Mol Med. 13:483-495, 2004.
7. "Current Progress in the Prediction of Chemosensitivity for Breast Cancer," Daisuke Shimizu, et. al. Breast Cancer 11:42-48, 2004.
8. M. J. van de Vijver and Others "A Gene-Expression Signature as a Predictor of Survival in Breast Cancer", N Engl J Med. 347: 1999-2009, 2002.
9. Derek M. Dykxhoorn, Carl D. Novina & Phillip A. Sharp, "Killing the Messenger: Short Rnas that Silence Gene Expression", Nat Rev Mol Cell Biol 4: 457-467, 2003.
10. "The Emerging European Health Telematics Industry". Deloitte & Touche, Feb, 2000. Health Information Society Technology Based Industry Study – Reference C13.25533.
11. "Medical practice websites enhance patient care" Ehealthcoach, December 2002.
12. M. Schmidt, G. Zahlmann "What are the benefits of Grid Technology for a health care solutions provider?", Siemens Medical Solutions, Proc. HealthGrid conf. January 2003.
13. "Next Generation Grid(s)", European Grid Research 2005 – 2010 Expert Group Report, 16th June 2003,
14. J.Fingberg, et. Al. "Bio-numerical simulations with SimBio", Physikalische Methoden der Laser- und Medizintechnik, pp. 114-120, VDI Verlag, 2003.
15. "Grid-Enabled Medical Simulation Services" GEMSS.

16. R. McClatchey, et. Al. "The MammoGrid Project Grids Architecture" CHEP'03, San Diego March 24th 2003.

17. "The eDiamond Project", <http://www.ediamond.ox.ac.uk/>

18. "Magnetic resonance imaging for breast screening (MARIBS)", Official Web site of <http://www.icr.ac.uk>

19. "myGrid: personalised bioinformatics on the information Grid", Robert D. Stevens, Alan J. Robinson and Carole A. Goble - Bioinformatics Vol. 19 Suppl. 1 2003.

20. A. Sousa Pereira, V. Maojo, F. Martin-Sanchez, A. Babic, S. Goes, "The Infogenmed Project", ICBME 2002: December 2002.

21. S. Nørager, Y. Paindaveine, "HealthGrid Terms of Reference", Version 1.0, 20th September 2002