



Открытые данные в большом мире

Международное соглашение
СОКРАЩЕННАЯ ВЕРСИЯ

Данное соглашение было принято на форуме «Международная Наука 2015», первом из серии ежегодных конференций представителей четырех ОСНОВНЫХ международных научных организаций (Международного совета по науке — ICSU, Межакадемического партнерства — IAP, Всемирной академии наук — TWAS и Международного совета по социальным наукам — ISSC), созданных для участия мирового научного сообщества в формировании международной политики в сфере науки.

Соглашение определяет возможности и проблемы, связанные с революцией в области данных — преобладающей на сегодняшний день задачей глобальной политики в области развития науки. Оно закладывает основополагающие принципы, которые должны быть приняты в ответ на новые проблемы, и демонстрирует согласие научного сообщества с представителями правительств и межправительственных структур, которые рассматривают открытые данные в качестве фундаментальной предпосылки для совершенствования уровня научных исследований и извлечения максимальной пользы для общества от революции данных как в развитых, так и развивающихся странах.

Участники форума будут способствовать обсуждению и принятию этих принципов другими представительными научными организациями на национальном и международном уровнях.

Расширенная версия этого соглашения доступна, и включает в себя более глубокое обоснование положений данного документа и более подробно рассказывает о методах, направленных на поддержку развития открытых систем данных.

1. Мир больших данных

Цифровая революция последних десятилетий является мировым историческим событием, не менее глубоким и значительным, чем изобретение печатного станка. Она вызвала беспрецедентные изменения в процессе получения, хранения, обработки и мгновенной передачи огромных и сложных массивов данных, что должно повлечь за собой глубокие последствия для развития науки¹. Скорость изменений огромна. В 2003 году ученые объявили о завершении картирования генома человека. Для этого потребовалось более 10 лет и 1 млрд долларов. Сегодня этот процесс занимает считанные дни и стоит всего лишь малую часть этой суммы (тысячу долларов). «Большие данные», включающие в себя беспрецедентные по масштабам потоки данных, полученных с помощью компьютерных систем, а также «широкие базы данных», в которых многочисленные наборы данных могут быть семантически связаны, чтобы создать более глубокий смысл, являются двигателями этой революции, предлагая новые возможности для естественных, социальных и гуманитарных наук.

2. Возможности

Большие данные предоставляют возможность обнаружить явления и законы, которые до сих пор были вне нашего понимания; связывая различные аспекты систем, помогают лучше понять их поведение; оказывают помощь в описании сложных явлений и процессов; координируют описания состояния сложных систем и позволяют создать модели, прогнозирующие их динамическое поведение. Эти возможности глубоко актуальны для многих областей исследований, таких как прогнозирование погоды и климата; изучение головного мозга; определение

состояния мировой экономики; оценка производительности в сельском хозяйстве; демографические прогнозы; разгадка тайн истории; решение многих современных глобальных проблем, например, связанных с изменением окружающей среды, инфекционными заболеваниями и массовой миграцией, которые требуют сопряжения идей и данных в области различных научных дисциплин.

3. Вызовы и проблемы

Обретение этих возможностей создает серьезные проблемы для развития и организации науки. Открытые данные — это общедоступные связующие нити.

Императив открытых данных

Фундаментальная роль исследований, финансируемых государством, заключается в совершенствовании качества знаний, необходимых для формулирования суждений, внедрения инноваций и социального и личного благополучия. Технологии и процессы цифровой революции создают мощную среду, которая способствует повышению продуктивности научных исследований и росту потенциала научного творчества благодаря свободному перемещению потоков данных при помощи сетевого взаимодействия, объединяющего умы. Для реализации этой социальной революции в сфере науки необходимо согласиться с тем, что данные финансируемых государством исследований должны быть доступны для широкой общественности и повторного использования после завершения исследовательской работы, позволившей получить эти данные.

Самокоррекция

Общедоступная информация (данные), подтверждающая правильность научных гипотез, является основой научно-технического прогресса. Эти данные позволяют скрупулезно наблюдать за ходом аргументации и воспроизводить результаты наблюдений или экспериментов, чтобы принять или отвергнуть правомерность умозаключений исследователей. Необходимо, чтобы после публикации научной работы все приведенные в ней данные, связанные с ними метаданные, которые позволяют провести их повторный анализ, и коды, используемые в компьютерных операциях, автоматически становились открытыми для проверки, чтобы гарантировать осуществление жизненно важного процесса самокоррекции. Многочисленные примеры публикаций работ, результаты которых невозможно воспроизвести, подтверждают насущную необходимость открытия научных данных для мира больших данных. Однако недостаточно просто сделать данные открытыми. Процесс открытия данных должен быть разумным, а это означает, что данные должны быть доступны для анализа и повторного использования.

Адаптация научного рассуждения

Многие из сложных процессов и взаимосвязей, которые мы стремимся изучить и осознать при помощи использования широких или связанных данных, не поддаются анализу путем применения многих классических статистических методов. Они требуют более глубоких математических подходов, включая топологические методы, чтобы гарантировать правильность выводов, сделанных благодаря большим и широким данным. Требуемые переработки большого объема данных компьютерный анализ и машинное обучение становятся повсеместными, и имеют большое значение для научных открытий. Модели, которые машины способны идентифицировать, слишком сложны для человеческого

¹ Слово "наука" используется здесь в значении "системной организации знаний, которые могут быть рационально объяснены и надежно применены". Это понятие включает в себя (как и в большинстве языков, отличных от английского) гуманитарные и социальные науки, а также т.н. "STEM" (science, technology, engineering, medicine - естественные науки, технические науки, инженерное дело, медицина).

разума, что создает серьезные проблемы для создания человеко-машинного интерфейса и для работы исследователя в 21 веке.

Этические ограничения

Принцип открытости данных имеет этические последствия для исследователей и предметов исследования. Он может превалировать над индивидуальными интересами исследователей, которые генерируют данные. Таким образом, следует разрабатывать новые способы признания и поощрения их вклада. Конфиденциальность данных должна быть защищена. При режиме открытого обмена данными научных исследований существует опасность утраты контроля за их будущее использование, в то время как процедуры анонимизации не в состоянии гарантировать безопасность персональных данных.

Открытое глобальное участие

Большие данные и открытые данные имеют особое значение для наименее развитых стран (НРС). Как правило, НРС не располагают значительными ресурсами для проведения научных исследований. Если они не могут участвовать в исследованиях на основе больших и открытых данных, разрыв между ними и более развитыми странами в ближайшие годы может возрасти в геометрической прогрессии. Они не смогут собирать, хранить и обмениваться данными, участвовать в глобальных исследовательских проектах, не будут в состоянии внести свой вклад в качестве полноправных партнеров в глобальные усилия, направленные на исследования изменений климата, развитие здравоохранения и защиту природных ресурсов, а также в полной мере извлекать выгоду из тех видов деятельности, в которых глобальные решения могут быть достигнуты только при наличии глобального участия. Таким образом, и развивающиеся, и развитые страны имеют четкую и прямую заинтересованность в том, чтобы в полной мере мобилизовать научный потенциал НРС и тем самым внести свой вклад в достижение Целей Устойчивого Развития, сформулированных ООН.

Возможности

Эффективное использование открытых данных возможно только при условии системных действий на персональном, дисциплинарном, национальном и международном уровнях. Несмотря на то, что наука немислима без международной кооперации, научные исследования часто ведутся в рамках национальных систем и организаций, которые также должны воспользоваться предоставляемыми им возможностями. Финансирующие научные исследования организации и научно-исследовательские институты должны выделять средства и стремиться улучшить условия работы исследователей, в разумных пределах предоставлять открытые данные и поддерживать процессы открытия данных.

Все большее число научных сообществ открывают для себя преимущества совместного использования данных в различных областях, как, например, лингвистика, биоинформатика и химическая кристаллография, и с успехом извлекают большую пользу для своих исследований в результате международного сотрудничества, направленного на обеспечение широкого доступа к открытым данным.

Международные научные организации — Комитет по данным (CODATA) Международного совета по науке (ICSU), Мировая система данных (WDS) и Союз по научно-исследовательским данным (RDA) принимают на себя обязательство содействовать и оказывать всемерную поддержку систем и процедур, которые будут обеспечивать международный доступ к данным и совместимость данных.

Открытая наука и общедоступные знания

Идея "открытой науки" разработана с учетом необходимости укрепления диалога и взаимодействия ученых и общественности в решении многих актуальных проблем, связанных с реализацией проектов и внедрением результатов научных исследований. Существуют и юридические лимиты открытости: необходимость обеспечения безопасности, конфиденциальности и интересов собственности. Существуют и противоположные компенсационные тенденции: стремление к приватизации знаний, противоречащее этике научного поиска и основной потребности человечества свободно использовать различные идеи. Если мировое научное сообщество не хочет допустить того, чтобы здание мировой науки обрушилось под давлением вышеперечисленных негативных тенденций, оно должно быть привержено политике открытых данных, открытой информации и открытого знания.

4. Принципы открытых данных

В связи с увеличением масштабов проблем, связанных с революцией данных, «Международная наука 2015» предлагает следующие принципы открытых данных.

Ответственность

Ученые

i. Ученые-исследователи, получающие государственное финансирование, обязаны внести свой вклад в создание общественного блага путем получения и передачи новых знаний, неотъемлемой частью которых являются связанные с ними данные. Они должны сделать такие данные открытыми и доступными для других пользователей в кратчайший срок после их создания всеми способами, обеспечивающими их повторное использование для выполнения других исследовательских проектов.

ii. Данные, которые подтверждают опубликованные научные гипотезы, должны находиться в открытом доступе в разумных и допустимых пределах². Это должно обеспечить связь между различными системами данных. Данные должны храниться в хорошо управляемых и надежных хранилищах с низкими барьерами доступа.

iii. **Научно-исследовательские институты и университеты** несут ответственность за создание благоприятных условий для открытых данных. Это включает в себя обеспечение подготовки кадров в области управления данными, их хранение и анализ и соответствующую техническую инфраструктуру, в том числе библиотеки и службы управления данными. Научные организации и финансирующие органы должны разработать соответствующие стимулы и поощрения для сотрудников, участвующих в процессах, направленных на открытие данных. Необходимо достижения консенсуса по таким критериям на национальном уровне, а в идеале и на международной арене. Принимая во внимание современную тенденцию к интернационализации, университеты и другие научные учреждения в развитых странах должны сотрудничать со своими партнерами в развивающихся странах в целях интенсификации процесса создания открытых данных.

iv. Издатели

отвечают за то, чтобы обеспечить доступ к данным для рецензентов во время подготовки публикации, требуют открытого доступа к данным одновременно с публикацией, в которой приводятся эти данные, а также ссылок на цитируемые данные. Издатели также несут ответственность за

то, чтобы публикация была доступна для последующего анализа путем открытого предоставления метаданных, текста публикации и возможности осуществлять поиск данных.

v. Финансирующие учреждения

должны предусмотреть затраты на открытие данных в исследовательском проекте, сделав их неотъемлемой частью затрат на ведение исследований, и должны обеспечивать достаточные ресурсы и надлежащее содержание инфраструктуры и хранилищ. При оценке результатов научных исследований следует принимать во внимание наукометрические показатели, индекс цитируемости и должным образом учитывать вклад создателей данных.

vi. Профессиональные ассоциации, научные общества и академии

должны разработать руководящие принципы и политику в области открытых данных, оповещать о новых предоставляемых возможностях, соблюдая при этом эпистемические нормы и общепринятые правила.

vii. Библиотеки, архивы и хранилища

несут ответственность за предоставление услуг и разработку технических стандартов работы с данными, чтобы гарантировать доступность данных для пользователей в течение длительного времени.

Границы открытости

viii. Принцип открытости данных должен быть обязательным для научных организаций, получающих государственное финансирование. Исключения должны ограничиваться вопросами конфиденциальности, безопасности, охраны и коммерческого использования в интересах общества. Предлагаемые исключения должны быть подтверждены в каждом индивидуальном случае, но ни в коей мере не становятся общепринятыми.

Реализация принципов

ix. Цитирование и источники

При использовании данных в научных публикациях, полученных другими исследователями, эти данные должны быть приведены с ссылкой на автора, указанием источника и постоянного цифрового идентификатора.

x. Функциональная совместимость

Данные научного исследования и метаданных, обеспечивающие их оценку и повторное использование, должны быть совместимы в максимально возможной степени.

xi. Неограниченное повторное использование

Если данные исследования еще не находятся в свободном доступе, они должны быть сделаны доступными для их повторного использования путем отказа от прав на них или предоставлением соответствующего разрешения, с указанием, что они могут быть повторно использованы без каких-либо ограничений, кроме обязательной ссылки на исследователя, впервые получившего эти данные.

xii. Связанность

Открытые данные должны быть, как можно чаще, связаны с другими данными, на основании их содержания и контекста, в целях максимального увеличения их смыслового значения.

Документ был подготовлен рабочей группой ICSU-IAP-ISSC-TWAS в составе:

- **Джоффри Боултон**,
Эдинбургский университет, Президент КОДАТА, председатель
- **Доминик Бабини**,
Университет Буэнос-Айреса и CLACSO (представитель ISSC)
- **Саймон Ходсон**,
Исполнительный директор КОДАТА (представитель ICSU)
- **Жианхуи Ли**,
Академия наук КНР, CNIC (представитель IAP)
- **Цилидзи Марвала**,
Университет Йоханнесбурга (представитель TWAS)
- **Мария Г. Н. Мусоке**,
Университет Макерере, Уганда (представитель IAP)
- **Пол Ф. Юлир**,
Академия наук США (представитель IAP);
Независимый консультант по управлению данными
- **Салли Уатт**,
Маастрихтский университет, KNAW (представитель ISSC)

Полная версия соглашения на сайте
<http://www.science-international.org>

Печатную версию можно получить в
Международном совете по науке (ICSU), по адресу:
5 rue Auguste Vacquerie, 75116 Paris, France.



www.icsu.org
www.interacademies.net
www.worldsocialscience.org
www.twas.org

