

АЛЬТМЕТРИКИ БУДУЩЕГО

Исследовательская и инновационная деятельность по своей природе характеризуется большим массивом информации, распределенной по множеству различных баз данных и в Интернете. Их наиболее рациональное использование возможно с помощью цифровых технологий, способных связывать и анализировать наборы разрозненных сведений для продвижения в научной сфере и с целью формирования эффективной политики в области науки, технологий и инноваций (НТИ). В этой связи Организация экономического сотрудничества и развития начала реализацию проекта, направленного на изучение инициатив цифровой науки и инновационной политики (DSIP) в странах ОЭСР и государствах-партнерах. Его цель – оценка потенциала цифровых источников и инструментов для совершенствования научной сферы и ее результативного управления.

Данные, касающиеся науки и инноваций, в основном генерируются в профильных исследовательских организациях, вузах, финансовых агентствах, осуществляющих грантовую поддержку изысканий, а также в частных библиотеческих и патентных базах данных. Определенный объем информации собирается в Интернете, в том числе на веб-сайтах компаний и в социальных сетях. Формирование инфраструктуры DSIP имеет неоспоримые преимущества: увеличивает объем, степень детализации, гибкость и актуальность данных, способствует разработке новых показателей НТИ, оценке уровня инноваций, усилению технологического прогнозирования и выявлению ведущих экспертов и организаций в определенной сфере науки. Для отсле-

живания новых тем и технологий исследований, а также для поддержки инвестиционных решений в области НИОКР активно применяются новейшие цифровые инструменты, разрабатываются дополнительные услуги, благодаря чему данные интегрируются с внешними платформами и превращаются в мощные цифровые ресурсы, обеспечивающие всесторонний анализ сведений о результатах НТИ.

Во многих странах уже сформированы отдельные элементы современных систем цифровой науки. К примеру, в Бельгии в 2011 г. появилось Исследовательское информационное пространство Фландрии, представляющее собой ресурс для углубленного анализа научно-технических тенденций и разработки статистических показателей НТИ, а также отслеживания финансирования изысканий.

В Бразилии учреждена платформа Lattes Platform, основанная на интеграции различных цифровых данных бразильских государственных учреждений и высших учебных заведений. Она позволяет визуализировать информацию НТИ и разрабатывать дополнительные аналитические решения в области науки и инноваций.

В Польше запущена интегрированная информационная система POL-on, основной задачей которой является создание базы данных научных и образовательных учреждений страны. Собранные сведения используются Министерством науки и высшего образования для администрирования деятельности польских университетов и исследовательских подразделений. Определенный объем информации, собранной системой, доступен для общественности.

Концептуальный вид инициативы DSIP и ее основных компонентов представлен на *рисунке*.

ИСТОЧНИКИ ДАННЫХ

АДМИНИСТРАТИВНАЯ ИНФОРМАЦИЯ

- Налоговые органы
- Статистические ведомства
- Негосударственные организации
- Центры обработки данных

ФИРМЫ И НЕКОММЕРЧЕСКИЕ ОРГАНИЗАЦИИ

- Библиографические БД
- БД инновационных результатов

ОБРАЗОВАТЕЛЬНЫЕ И НАУЧНЫЕ УЧРЕЖДЕНИЯ

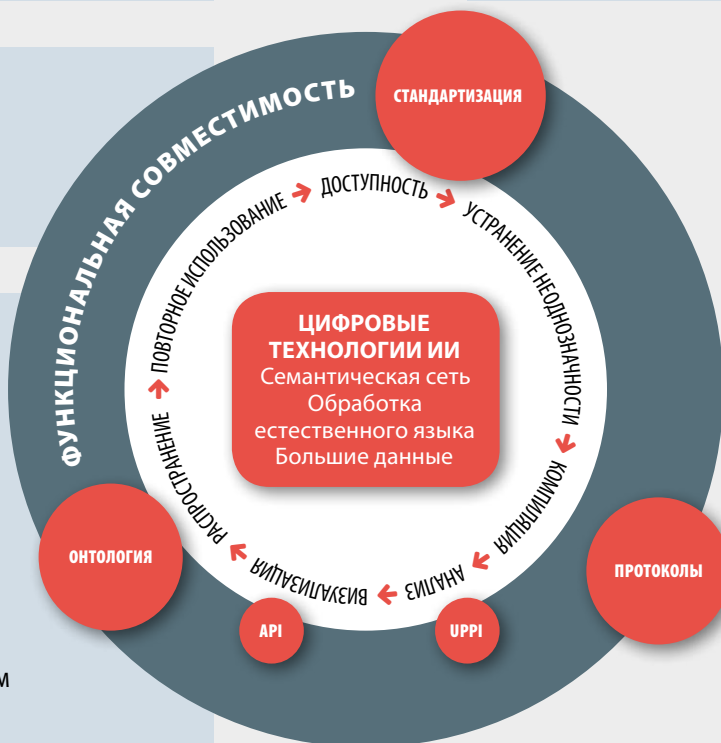
- Исследовательская информационная система
- Научные институциональные репозитории – электронный архив для длительного хранения, накопления и обеспечения долговременного и надежного открытого доступа к результатам научных исследований
- Персональные данные

ИНТЕРНЕТ

- Веб-ресурсы исследовательских учреждений
- Социальные сети
- База данных научных результатов
- Общедоступные универсальные интернет-энциклопедии
- Сайты научных организаций

ФУНКЦИИ И ЦЕЛИ DSIP

- Оптимизация рабочих процессов
- Эффективный мониторинг и оценка научной политики
- Обоснованное принятие решений



ПОЛЬЗОВАТЕЛИ DSIP

- Административные органы
- Негосударственные организации
- Высшие учебные заведения
- Научно-исследовательские учреждения
- Компании
- Гражданское общество

Рисунок. Концептуальное представление инициативы DSIP

Примечание: API – интерфейс прикладного программирования; UPPI – уникальный, постоянный и универсальный идентификатор

Основные элементы системы состоят из различных источников, которые поступают в цикл данных, разрешенный стандартами взаимодействия, включая уникальные, постоянные и унифицированные идентификаторы (UPPI).

ЦЕЛИ ИНИЦИАТИВЫ DSIP

Системы DSIP выполняют ряд функций и сулят множество преимуществ. Среди них можно отметить следующие.

- *Оптимизация рабочих процессов.* Цифровые инструменты значительно упрощают рутинные административные процедуры, повышают эффективность работы научных организаций, а также облегчают получение грантов. Ученые, претендующие на них, могут использовать идентификаторы функциональной совместимости, чтобы связать свои исследовательские профили с заявками на гранты. Такой инструмент применяется в качестве цифрового входа в эстонскую исследовательскую систему ETIS.

- *Детальный и своевременный анализ данных для принятия решений в области НТИ.* Цифровизация предлагает наиболее оптимальную интеграцию сведений, собранных разными агентствами, обеспечение прироста контента и, соответственно, возможности для разработки интегрированной межведомственной политики на уровне исследовательской или инновационной системы, касающейся в том числе и более рационального распределения финансирования. К примеру, Министерство образования, культуры, спорта, науки и технологий Японии и Национальный институт политики в области науки и технологий запустили платформу SciREX, содержащую перечень научно-технических запросов, требующих инвестиционной поддержки.
- *Мониторинг исследований и управление научной деятельностью.* Системы DSIP предлагают возможность оценки текущих изысканий, финансируемых государством, позволяют более гибко корректировать их направления и в перспективе сделать этот процесс открытым и непрерывным. Так, в Колумбии создана Технологическая платформа SCIENTI, которая благодаря разработанным индикаторам и показателям НТИ способна в реальном времени отслеживать результаты научной работы. Для достижения таких целей будут использоваться новейшие цифровые ресурсы, такие как искусственный интеллект, аналитика больших данных и другие, которые выявляют закономерности, необходимые для развития новых областей исследований, технологий, отраслей и управления ими. Цифровые технологии также могут поддерживать краткосрочное прогнозирование НТИ, осуществлять стратегическое планирование, определять спрос на рабочую силу в конкретных областях и устранять потенциальные несоответствия со стороны предложения на рынке труда. В Российской Федерации, к примеру, Институт статистических исследований и экономики знаний НИУ ВШЭ разработал систему iFORA для поддержки форсайт-исследований.
- *Накопление информации.* Помимо обширных баз данных системы DSIP часто содержат данные аудита входов и выходов пользователей на тот или иной ресурс. Подобные кейсы важны для выявления ведущих экспертов в конкретной области науки, в частности определения рецензентов проектных пред-

ложений, для фиксации центров передового опыта и установления партнерских контактов. Например, Министерство бизнеса, инноваций и занятости Новой Зеландии разработало Новозеландскую исследовательскую информационную систему, предназначенную для улучшения поиска данных по вопросам, связанным с исследованиями и инновациями. Здесь представлены сведения об объемах государственных инвестиций в научные изыскания, потенциальных сферах взаимодействия науки и бизнеса, а также о наиболее авторитетных ученых и организациях. При этом основная цель платформы – максимальная коммерциализация разработок и укрепление отношений между научными кругами и промышленностью.

- *Инклюзивность в области НТИ.* DSIP играет важную роль в обеспечении долгосрочного устойчивого финансирования исследований и инноваций и укреплении доверия населения к достижениям науки. С помощью системы возможно их обсуждение с заинтересованными сторонами в доступной форме, например посредством интерактивной визуализации данных, предоставления гражданам знаний о значимости и необходимости научных разработок. Коста-Рика, например, запустила интегрированную платформу Hipatia, созданную на основе различных административных сведений, чтобы помочь населению страны лучше понять национальный научный потенциал и его влияние на развитие экономики.

СОВМЕСТИМОСТЬ

Систематизация разрозненной информации из внешних источников требует разработки общих форматов данных и других средств взаимодействия, включая, помимо прочего, интерфейсы прикладного программирования (API), онтологии, протоколы и UPPI.

Интегрированная и функционально совместимая система облегчает отслеживание результатов исследований и инновационной деятельности в разных странах, позволяет значительно снизить административную нагрузку, а также более эффективно анализировать данные. К тому же концентрация оперативных и подробных сведений на одном ресурсе дает возможность разрабатывать более гибкую и адаптированную политику управления наукой. Однако для полноценного функционирования системы в международном

масштабе, по мнению экспертов ОЭСР, необходимо преодолеть несколько серьезных барьеров, среди которых обеспечение качества данных, их защита и доступ к ним, оптимальная интеграция, устойчивое финансирование, наличие цифровых навыков и доверие к новым технологиям.

На техническом уровне нужно определить вид цифровой системы, обеспечивающей совместимость данных, на семантическом – решить проблему их неоднородности, на управленческом – согласовать функционал. Не стоит забывать об эффективности стандартов информации, особенно в смешанной экосистеме, включающей как старые, так и новые форматы. В связи с этим в некоторых DSIP используются национальные идентификаторы, например регистрационные номера компаний и социального страхования, а также идентификаторы исследователей в конкретных странах (таблица).

В последние годы были предприняты попытки разработать общие стандарты и словари для улучшения международной совместимости инфраструктур DSIP. К ним относятся UPPI, которые присваивают стандартизированный код, уникальный для каждого исследовательского объекта, сохраняющийся длительное время и распространяющийся на различные наборы данных. Некоторые UPPI существуют как неотъемлемая часть или поддержка коммерческих продуктов, таких как базы данных публикаций/цитирования, исследовательские информационные системы, службы управления цепочками поставок и т.д. Другие представляют собой исключительно системы идентификаторов, к примеру ORCID, целью которого является устранение двусмысленности имен в научных исследованиях путем разработки уникальных идентификаторов для отдельных исследователей. Эти системы обеспечивают простой реестр UPPI и базовую связанную информацию (например, имя и место работы для физических лиц, название и расположение для организаций). Кроме того, они часто непосредственно включают или содержат ссылки на широкий спектр дополнительных сведений, например об образовании ученых, их занятости и др.

ХАРАКТЕРИСТИКИ UPPI

По мнению аналитиков проекта, UPPI должны соответствовать определенным требованиям. Среди них можно отметить следующие:

- *точность* – соответствие формальному шаблону (регулярному выражению), устанавливающему набор назначаемых идентификаторов, что облегчает их проверку и применение;
- *стабильность* – неподверженность изменениям, исключение ненужных деталей или информации, которая может быть трансформирована путем использования случайного буквенно-цифрового кода фиксированной длины и структуры;
- *принадлежность* не более чем к одному локальному объекту во избежание двусмысленности. Например, буквенно-цифровой идентификатор должен включать либо цифру ноль, либо букву «о», поскольку их путают;
- *уникальность* – связанность объекта не более чем с одним идентификатором и исключение его «переработки» для другого объекта;
- *регистрация* – четкая задокументированная фиксация схемы идентификатора, его работы и любых изменений, происходящих с ним;
- *веб-совместимость* – отмена использования символов, выполняющих определенные функции HTML, и форматов обмена (например, XML), таких как «:», «/», «.»;
- *веб-разрешаемость* – возможность преобразовываться в веб-адрес, по которому можно получить доступ к данным или информацию об объекте. На практике это означает, что идентификатор должен состоять из шаблона URI – универсального идентификатора ресурса, например <http://orcid.org/>, и локального идентификатора, относящегося к конкретной записи, например 0000–0002–2040–1464. При совместном использовании URI и локальный идентификатор создают разрешимый веб-адрес (например, <http://orcid.org/0000-0002-2040-1464>). Это позволяет легко проверить идентификатор, чтобы убедиться, что он относится к фактической записи, а она в свою очередь к правильному объекту;
- *доступность* – использование метаданных, в частности имени объекта, к которому относится идентификатор, для прозрачной ссылки и действий в общедоступном поиске [2].

Инициаторы проекта считают, что каждый дополнительный регистрант системы UPPI увеличивает ее ценность и формирует так называемый сетевой эффект, в рамках которого в будущем UPPI станет общепринятым способом идентификации научных объектов. Еще одним отраслевым стандартом для интеграции данных

является API, обеспечивающий автоматизированный протокол обмена данными. Он активно внедряется многими странами на всех государственных цифровых ресурсах и базах данных, в том числе и научно-исследовательских. Улучшение доступа к наборам административной информации положительно влияет на функциональность и надежность результатов анализа, предоставляемых системами DSIP.

ИСПОЛЬЗОВАНИЕ DSIP В ОЦЕНКЕ ИССЛЕДОВАНИЙ

В последнее время наука все чаще сталкивается с требованием продемонстрировать эффективность осуществленных исследований, подтвердить их экономический эффект и отчитаться о затраченных средствах. В этой связи решающее значение приобретает оценка научных достижений, основанная на цифровых фактических данных. Эта процедура имеет ряд преимуществ, обусловленных технической возможностью работы с большими связанными базами, с тонким уровнем детализации информации и ее агрегированием. Однако существуют значительные риски того, что такая оценка не всегда будет точна, поскольку данные доступны в количественном, явно сжатом виде, а для объективности картины они должны охватывать все нюансы интересующего явления. Более того, они должны быть доступны и совместимы с другими источниками. Следует учитывать и разные уровни анализа:

с точки зрения работы отдельных ученых, их учреждений, страны или научной области в целом. Наиболее распространенные сегодня библиометрические показатели, в том числе количество цитирований, являются, по мнению аналитиков, суррогатной мерой оценки результатов научных исследований. Поэтому необходимо предпринять усилия по созданию открытой, надежной и последовательной системы измерения всех видов научной деятельности. Обеспечить такой процесс способны инфраструктуры DSIP, благодаря которым большинство важнейших измерений на этом поле будет представлено в цифровом виде.

Подобные альтернативы предназначены для оценки таких нетрадиционных научных продуктов, как наборы данных и программное обеспечение, не учитываемых в индексах цитирования. К тому же эти показатели, в отличие от представляемых в малоизвестных научных журналах, будут более доступны широкой аудитории, специалистам-практикам или общественности в целом. В качестве более широкого и своевременного источника оценки исследований, чем цитирование, предлагаются метрики, созданные на платформах социальных сетей. Они усовершенствованы для содействия открытой науке и помощи в фильтрации быстрорастущих объемов информации.

Одной из главных движущих сил цифровизации науки и инновационной политики становятся решения, разработанные бизнес-компаниями и некоммерческими организациями. Они ока-

Тип	Примеры
UPPI для участников научной деятельности	<ul style="list-style-type: none"> Открытый идентификатор исследователя и участника (ORCID) Цифровой идентификатор объекта (DOI) Глобальная база данных идентификаторов исследований (GRID) Международный стандартный идентификатор имени (ISNI) Идентификатор Ringgold
Идентификаторы авторов, созданные издателями/индексаторами	<ul style="list-style-type: none"> Идентификатор исследователя Идентификатор автора Scopus
Стандарты управления данными о результатах НТИ	<ul style="list-style-type: none"> Общеввропейский формат исследовательской информации (CERIF) Словарь Consortia Advancing Standards in Research Administration Information (CASRAI) Онтология VIVO
Протоколы	<ul style="list-style-type: none"> Протокол инициативы открытых архивов для сбора метаданных (OAI-PMH)

Таблица. Средства обеспечения функциональной совместимости в DSIP и родственных системах. Источник: [1]

зывают значительное влияние на развитие инициатив DSIP, использующих их цифровые продукты и услуги в качестве строительных блоков: технологические архитектуры, инструменты для управления данными, консалтинговые услуги, связанные с запуском и обслуживанием цифровых инфраструктур. Следует отметить, по мнению аналитиков проекта, положительную тенденцию укрепления сотрудничества между государственным и частным секторами не только в формате покупки готовых решений, но и в разработке новых. Особенно активны крупные академические издательства Elsevier и Holtzbrinck Publishing Group вместе с аналитической фирмой Clarivate Analytics. Они создают и объединяют набор данных и услуг в платформы, которые имитируют многие функции полноценных систем DSIP. Несколько продуктов, разработанных этими компаниями, включая библиографические базы данных, уникальные идентификаторы и организационную CRIS, часто выступают в качестве ключевых компонентов государственных инициатив DSIP.

Кроме того, цифровые гиганты Alphabet и Microsoft, а также национальные технологические корпорации, такие как Baidu (Китайская Народная Республика) и Naver (Корея), разработали платформы для поиска научных результатов. И хотя влияние этих компаний на цифровизацию науки и инновационной политики ограничено, объем информации, накапливаемой ими, может стать ключевой базой национальных систем DSIP. Например, API Microsoft Academic Graph позволяет получать контент о публикациях, цитировании, авторах, учреждениях, областях исследований, журналах и конференциях, который может использоваться в системах DSIP. Академические поисковые системы (Google Scholar, Microsoft Academic, Baidu Scholar и Naver Academic), собирающие данные о научных публикациях и цитировании, способны анализировать уровень научных работ и проводивших их ученых, а также определять рейтинги университетов.

К потенциальным участникам инициатив DSIP относятся организации, реализующие консультационные проекты для поддержки науки и финансирования научных разработок. Примером может служить Science-Metrix – канадская исследовательская компания, специализирующаяся на оценке научно-технологической деятельности с использованием библиометрических данных, которой в 2018 г. было поручено разработать

методы и показатели уровня исследований и инноваций для Национального научного фонда США.

Некоммерческие организации также вносят весомый вклад в формирование DSIP путем финансирования и разработки решений для этой инициативы. Например, Фонд Альфреда П. Слоуна поддержал проекты по сбору систематических данных об эффективности финансируемых государством исследований (ETOILE, UMETRICS), по предоставлению свободного доступа и обмена научными результатами (arXiv.org, FORCE11, Impactstory) и устранению неоднородности данных (улучшение цитирований, разработка уникальных идентификаторов). Австралийское некоммерческое социальное предприятие Cambia в сотрудничестве с Квинслендским технологическим университетом запустило открытую платформу Lens PatCite, объединяющую базы данных нескольких национальных и международных патентных ведомств, а также научных учреждений, включая PubMed, Crossref и Microsoft Academic, для обеспечения открытого доступа к связанной патентной информации.

Благодаря высокому уровню функциональности цифровые решения, разработанные некоммерческими организациями, широко используются государственными органами, а также бизнес-структурами. Часто они выступают важными элементами DSIP, улучшая их совместимость. Таким образом, для активного формирования подобных экосистем необходимо стратегическое сотрудничество частного и некоммерческого секторов, межведомственная координация, использование общих ресурсов для документирования метаданных, разработки стандартов для них и обеспечения наиболее продуктивного взаимодействия программных и аппаратных средств. Инициатива DSIP должна основываться на принципах совместного проектирования, совместного творчества и совместного управления. Эти слагаемые обеспечат необходимые условия для внедрения новых технологий и включения неиспользуемых источников данных в системы DSIP. ■

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Обзор науки, технологий и инноваций ОЭСР, 2018 г.: Адаптация к технологическим и социальным изменениям // https://doi.org/10.1787/sti_in_outlook-2018-en.
2. Плохие идентификаторы – это выбоины на информационной супермагистрали: полезные уроки для исследователей // <https://biologue.plos.org/2017/07/06/bad-identifiers-are-the-potholes-of-the-information-superhighway-take-home-lessons-for-researchers/>.

Наталья МИНАКОВА