

Образ и подобие. О чем молчит научная статистика?

В последнее время при обсуждении проблем российской науки часто обращаются к статистике публикаций и к данным государственной статистики с целью оценить состояние науки и понять, что нужно для улучшения дел в этой области. Однако многие рассуждения по этому вопросу заставляют вспоминать, что «есть ложь, есть наглая ложь и есть статистика», то есть высказывания, правдоподобные по форме, но ложные по содержанию. Подобный дефект обусловлен не столько сущностью статистики как таковой, сколько ее неверным использованием.

Говоря о состоянии науки, часто приводят данные Росстата и информацию о научных публикациях, полученную из баз данных научных публикаций, таких как Web of Science Core Collection (Thomson Reuters) и Scopus (Elsevier). Государственная статистика науки во многих странах, в том числе и в России, опирается преимущественно на «Руководство Фраскати» – набор рекомендаций о сборе статистической информации в области научных исследований и разработок. Несмотря на наличие такого руководства, международные сопоставления крайне затруднительны, поскольку большинство стран придерживаются национальных статистических стандартов.

Между государственной статистикой и статистикой публикаций существует множество концептуальных разрывов, которые заложены уже на базовом уровне конструирования единиц наблюдения.

Государственная статистика действует в целях управления материальными и человеческими ресурсами, занятыми в науке. В силу этого используются такие единицы наблюдения, как «работники, занятые исследованиями и разработками», «сектор науки» (государственный, частный), «область науки» и т. д.

Статистика научных публикаций ведет свое происхождение от базы данных Института научной информации (Institute for Scientific Information – ISI), основанного Ю. Гарфилдом в 1960 г. Создавая свою базу данных, Ю. Гарфилд в первую очередь ставил перед собой задачу упорядочения больших объемов научной информации. Одним из решений было составление указателя цитирования, по аналогии с известным в США юридическим справочником ссылок Shepard's Citations. Английское название базы данных Science Citation Index на русский язык принято переводить как «Индекс научного цитирования», хотя правильнее было бы именовать ее «Указатель научного цитирования». Такой перевод привел к смещению смысла названия в русском языке от понятия индекса как «указателя», то есть инструмента поиска публикаций, к понятию индекса как «показателя», то есть инструмента оценивания публикаций.

Назначение базы данных Ю. Гарфилда состояло в том, чтобы ученые и библиотекари могли быстро найти нужную им информацию. Поэтому

информационную основу базы составляют единицы учета, принятые в библиографическом описании документов: «автор», «тип документа», «журнал» и т. п. Дополнительно в базе данных Ю. Гарфилда появились такие единицы, как «ссылка» и «аффилиация».

Различие между подходами, принятыми в государственной статистике науки и статистике публикаций, приводит к тому, что

- понятие «работник, занятый исследованиями и разработками» не тождественно понятию «автор»;
- понятие «аффилиация» слабо согласуется с принципами учета персонала в государственной статистике;
- классификаторы наук баз данных не совпадают с классификаторами государственной статистики.

Примеры концептуальных разрывов можно без труда продолжить. Таким образом, *de facto* сформировались два пространства статистических воззрений на науку, которые практически не пересекаются. Как правило, в исследованиях науки информация о публикациях и данные о «материальной» стороне дела существуют параллельно. Положение осложняется различиями между национальными традициями сбора статистической информации.

Статистика публикаций

Единых и однозначных стандартов сбора данных о публикациях в настоящее время не существует. Можно выделить лишь некоторые сходные приемы, сложившиеся в ходе работы с имеющимися источниками информации. Проводя исследования, специалисты в области наукометрии вынуждены каждый раз совершать серию последовательных выборов, что в конечном итоге влияет на результаты их работы.

Выбор базы данных

При сборе статистических данных о научных публикациях основными источниками информации служат международные базы Web of Science и Scopus. Поскольку структура информации в них различается, исследователи выбирают какую-то одну из них.

Выбор типов документов

Анализируя статистику научных публикаций, можно говорить о трех ипостасях их существования:

- как средства научной коммуникации,
- как документа базы данных,
- как единицы статистического учета.

В каждом из этих случаев предполагается своя классификация.

Журналы классифицируют публикации по рубрикам, исходя из сложившихся традиций. Базы данных, обрабатывающие множество журналов, учитывают эти традиции, однако вынуждены переводить различные журнальные классификации публикаций в единую систему типов документов.

При этом базы данных трансформируют изначальные принципы классификации, двигаясь в русле внутренней логики своего развития.

Статистика научных публикаций использует унифицированную классификацию баз данных. На ее основе анализируются те документы, которые, по мнению исследователей, характеризуют состояние науки. Однако консенсуса по этой проблеме нет (например, далеко не все исследователи включают в статистику публикаций труды конференций).

Выбор временного периода при подсчете цитирования

Известно, что традиции цитирования в различных дисциплинах сильно различаются. В естественных науках больше принято цитировать самые «свежие» работы, тогда как в социально-гуманитарных дисциплинах активно цитируются, в том числе, довольно старые публикации. Это приводит к тому, что, как и в случае выбора типа публикаций, не может быть единственного обоснованного решения о том, за какой временной период учитывать цитирование. В зависимости от целей и задач может анализироваться «накопленное» цитирование (общее число ссылок на документ с момента его публикации до момента запроса к базе данных), например цитирование за трехлетний период или за более длительное время.

Проблема классификаторов

Классификация публикаций по областям науки обычно реализуется исходя из тематики журналов: статья считается принадлежащей к области X, потому что журнал, в котором она опубликована, относится к данной области. Недостаток такого способа классификации в том, что журналы могут публиковать статьи, относящиеся к разным, хотя и близким, научным специальностям. Сегодня «журнальный» способ классификации публикаций ввиду своей простоты является наиболее распространенным.

К этому стоит добавить, что классификаторы публикаций не согласуются с классификаторами областей науки и научных специальностей, принятыми в государственной статистике. Это обстоятельство не дает возможности однозначно соотнести информацию о публикациях в той или иной дисциплине с другими статистическими показателями.

Учет персонала

Конструирование относительных показателей продуктивности, таких как отношение числа статей к числу ученых в стране, на сегодняшний день крайне затруднительно. В государственной статистике науки используются категории «работники, занятые исследованиями и разработками» и «профессорско-преподавательский состав», что проводит четкую границу между исследователями и преподавателями. В свою очередь, в базах данных, собирающих информацию на основе библиографических описаний, используется категория «автор». Автором статьи может быть преподаватель, исследователь, аспирант или даже сотрудник, занимающий техническую

должность. Иными словами, базы данных публикаций позволяют определить только числитель (количество публикаций). Чтобы оценить знаменатель, то есть общее число работников, вовлеченных в производство научного знания и способных стать авторами публикаций, необходимо обратиться к государственной статистике, которая, как было указано выше, оперирует совсем другими категориями.

Построение относительных показателей продуктивности в науке требует большой работы по гармонизации категорий государственной статистики и категорий статистики публикаций. Пока же в большинстве науковедческих исследований информация о публикациях выносится в отдельный раздел и никак не соотносится с другими статистическими показателями.

Мы попытались очертить проблему отношения государственной статистики науки и статистики публикаций. Факты говорят о том, что образ науки, формируемый государственной статисткой, плохо согласуется с образом, конструируемым на основе библиометрической информации. Напрашивается вывод, что некритическое комбинирование разнородных описаний науки, которое зачастую практикуется в публицистике и политических дискуссиях, может привести к неожиданным последствиям **(Маркова Ю. Образ и подобие. О чем молчит научная статистика? // Поиск (<http://goo.gl/xXhfEB>). – 2015. – 1.05).**