



65th IFLA Council and General Conference

August 20-28, 1999

Code Number: 004-120-R
Division Number: VI
Professional Group: Statistics
Joint Meeting with: -
Meeting Number: 120
Simultaneous Interpretation: -

РАЦИОНАЛЬНЫЙ СБОР ДАННЫХ В ИНФОРМАЦИОННОЙ СРЕДЕ

Дж. Гормэн

Школа коммуникационного и информационного управления
Университет Виктории в Веллингтоне,
Новая Зеландия

(G. E. Gorman)

Abstract:

Библиотеки и информационные агентства в большой мере зависят от статистики при описании своих услуг, оценке своей деятельности и ее учета. При сборе данных, на которых основан статистический анализ, всегда имеют место предположения и неподдающиеся учету переменные величины, снижающие чистоту и объективность данных и поэтому затрудняющие анализ и интерпретирование данных. Этот доклад излагает некоторые из этих переменных величин для обращения внимания информационных руководителей на ловушки при сборе данных и содействия им при разработке средств контроля данных с целью увеличения эффективности использования статистики.

Статистика в центре внимания

Сбор и производство статистики - это важный вид деятельности в области управления в библиотеках конца XX века. Для большинства библиотечных руководителей "чем больше, тем лучше" - то есть увеличение количества данных приведет к увеличению количества полезной информации, в результате чего решения будут приниматься на расширенной информационной основе и, следовательно, руководство обслуживанием станет более адекватным.

Основополагающее предположение заключается в том, что данные о библиотечной деятельности можно преобразовать в полезную информацию и что информация станет знанием для управления.

Тогда понятно, что многие считают, что сбор данных - это основная деятельность процесса управления. Менее понятно то, что руководители в целом обычно рассматривают процесс сбора данных - интерпретации - использования - как что-то, что принимается без сомнения, и что многие просто действуют по этому образцу раз за разом, не помышляя о том, что может существовать лучший способ сбора и использования данных. Для ученого, работающего в области общественных наук и заинтересованного прежде всего в научной методике и чье главное занятие - это учить студентов проводить исследования в области библиотечной и информационной науки, такое положение, существующее уже десятки лет, тревожно.

Основная цель этого доклада - выдвижение предложения информационным работникам, что они могут с выгодой для себя рассмотреть использование не только количественных, но и качественных методов сбора и анализа данных для достижения повышенной надежности и более глубоких выводов в их исследованиях. Дополнительная цель - освещение некоторых опасностей, присущих неоспоримому принятию процесса сбора, и интерпретация данных.

Количественные переменные величины показывают, в какой мере конкретная характерная черта или признак присущи предмету, событию, личности, явлению. Примером может послужить число компьютеров, доступных студентам в университете. Качественные переменные величины классифицируют предметы, события, людей, явления по категориям с учетом признаков, по которым они отличаются друг от друга. Например, язык издания или журнала может быть английским, французским, ивритом или испанским [1]. Если библиотекари выйдут за пределы "сколько" и обратятся к признакам людей, вещей и деятельности, подлежащим учету, они, несомненно, глубже поймут свои организации и их работу.

Эта проблема не нова, да и решения, предложенные в этом докладе, не уникальны, но независимо от того, что было сказано прежде, проблема остается. Как представляется, стоит еще раз вернуться к реальности. Одна из моих коллег по Университету Виктории, Р.Каллен, поставила под сомнение опоры только на количественные данные в контексте своих исследований учета показателей. Итак, дискутируя с работами Пратта и Альтмана [2] и подразделением библиотечной и информационной статистики [3], она ставит вопрос надежности одной библиотечной статистики как прочного основания библиотечной деятельности и в особенности возможности того, могут ли эти данные пролить свет на соотношение вложенных ресурсов и результата. "В особых существенных вопросах удовлетворения читателей библиотечными информационными услугами затронуты лишь в малой части исследований, включенных здесь, хотя авторы в нескольких местах указывают, что дальнейший анализ возможен и в сущности желателен" [4].

В своем докладе Каллен далее показывает, что библиотека - это социальная конструкция и поэтому учет показателей - это тоже социальная конструкция. Тогда это значит, что нужно взглянуть на основную массу, включающую в себя значения, фокус и цель - три оси, важные для понимания библиотеки как социальной конструкции. По моему мнению, социальная конструкция - это средство рассмотреть библиотеки и информационные учреждения и в сфере социальных конструкций и неколичественные методы сбора и анализа данных становятся значительнее. Это особенно справедливо для трех областей: пользователей, фондов и услуг (или запросов); каждая из них обсуждается по очереди.

Можно ли учитывать пользователей разумным образом?

Сбор данных основан на предположении, что можно дать точное представление о предметах/населении, которые изучаются. В библиотеке такое предположение нужно поставить

под вопрос, когда речь идет о пользователях конкретной библиотечной услуги. Например, представим, что мы заинтересованы числом людей, пользующихся библиотекой. Как измерить или учесть это? Один простой способ - это считать входящих или выходящих механически или вручную. Много библиотек действует именно так - сколько годовых отчетов гордо показывают, что "в 199X" году XXXX читателей посетили библиотеку"? Но что это говорит нам? Это могут быть и случайные посетители, и серьезные ученые, интенсивно применяющие сложные поисковые услуги, и студенты, ищущие материалы по списку для чтения, и пожилые люди, использующие библиотеку как социальный центр, и родители, пользующиеся программами для своих детей... Иначе говоря, подсчет людей дает очень мало, ибо он не конкретизирует разные категории читателей и, следовательно, запросы, которые они, вероятно, ставят перед обслуживанием. Это важнейший пример того, что данные не могут давать целесообразной статистики или информации какой-либо ценности, так как принят столь грубый подход к пользователям. В основном предложении есть изъян, в данных имеются изъяны и, следовательно, в интерпретации также должны быть изъяны.

В действительности мы хотим знать профиль реальных пользователей библиотеки: кто они, что они ждут, когда входят в библиотеку, как они используют возможности и услуги, что они думают об оборудовании и услугах, почему они обращаются в библиотеку через электронные средства или сами приходят в библиотеку, и так далее. Эти очень полезные данные нельзя получить простым учетом читателей. Далее, никакой подсчет и даже самое сложное и подробное обследование пользователей не говорит ничего о потенциальных пользователях или не пользователях, но именно эти сведения руководители как раз и ищут: они хотят или как минимум нуждаются в том, чтобы знать потенциальный рынок своих услуг для того, чтобы можно было составить менеджмент-план для раскрытия этого резерва. Даже в такой стране, как Австралия, где библиотечное дело стоит на высоком уровне и где свыше 60% населения используют публичные библиотеки (что бы ни понималось под "использованием"), имеется очень большой контингент пользователей, которых нужно охватить нашим обслуживанием. Даже самый опытный ученый скажет, что сбор этих данных, которые полезнее и потому сложнее, о пользователях и не пользователях сопряжен с трудностями и что это требующее времени и дорогое занятие, но замены этому нет.

Какая польза от фондов?

Если подвергнуть сомнению ценность учета пользователей в отличие от ярко выраженных групп пользователей и выявления их оценки конкретного обслуживания, то можно ли перевести внимание от людей к предметам и в особенности к фондам (при любом их определении)?

В библиотеках считается почти библейской истиной, в особенности с добрых времен, когда была популярна формула Клаппа - Джордана, что Y размера книжных фондов дает сведения, которые важны в количественном и качественном смысле [5]. К тому же почти все годовые отчеты указывают, что размер фондов достиг уровня числа X книг, числа Y продолжающихся изданий, числа Z электронных ресурсов. А каково соотношение размера или количества фонда и его качества? Сей вопрос неизменно приводит статистиков в уныние, так как он подвергает сомнению ценность статистической работы. Но как и в случае с пользователями мы как информационные работники должны прежде всего заниматься величинами и смыслом независимо от того, изучаем ли мы пользователей или фонды.

Если принять то, что между качеством и количеством есть взаимосвязь, хотя я не разделяю этого мнения, то нужно поставить под вопрос ценность данных о фондах в случае предположения какой-либо связи между размером и уровнем обслуживания. Для того, чтобы компенсировать это, много библиотек считает выдачу по абонементу или число выданных справочных материалов, журналов, CD-ROM и так далее. Все же любой учет выдачи по абонементу или числа использования легко искажается в том случае, если какой-нибудь другой ученый, работающий над уникальным проектом, абонент, увлекшийся особой темой и та далее, запрашивают необычную литературу. В эпоху повышенного внимания к коммерческой стороне

и услуг с добавленной стоимостью можно также спросить, являются ли выдача по абонементу или число использования библиотечного материала действенным показателем общей картины.

Конечно, можно подкрепить данные о выдаче по абонементу и числе выдач вводом в фонды каких-то качественных показателей. Это обычно означает какую-то иерархическую систему, по которой книги местной коллекции подводятся под какую-то внешнюю мерку. В Новой Зеландии это может значить, что, например, те же книги, которые хранятся в публичной библиотеке Нью-Йорка и в городской библиотеке Веллингтона, в Веллингтоне имеют повышенную ценность. Или университетская библиотека может поставить издания престижных академических издательств и агентств ООН выше популярных романов или изданий местной администрации. Но должна ли библиотека или информационная служба реагировать на потребности конкретного населенного пункта или она занимается самоизмерением по национальным или международным критериям? То есть, для городской библиотеки Веллингтона это могут или даже должны быть материалы, уникальные для ее фонда, а не те, которые имеют это значение для публичной библиотеки Нью-Йорка, которые окажутся наиболее подходящими для удовлетворения нужд местных пользователей.

Иначе говоря, учет фондов как в книжной, так и в любой иной форме, не является мерой спроса, а подсчет числа выдач не служит мерой уровня или качества использования; это лишь свидетельство того, что кто-то обратился к фондам электронным образом, возможно, из-за того, что нет ничего "лучше". Но разве речь идет об этом вопросе? Для слишком многих библиотекарей важен не уровень спроса на услуги, а размер фондов или книговыдачи.

Являются ли запросы заменой подсчета пользователей или фондов?

Подсчет пользователей и фондов может дать полезные данные, хотя бы и ограниченной ценности, но нужно повторить, что за такими обрубленными числами кроется слишком много библиотечных услуг и что к этому нужно отнестись как к замене осмысленного анализа данных. Некоторые учреждения приняли такую альтернативу. Они считают читательские запросы (через сотрудников, электронные системы или путем иных методов вопроса-ответа). Некоторые отличные примеры этого можно найти в труде "Библиотеки на рабочем месте". Это великолепный отчет Д.Спиллера и LISU. Какое число разысканий (как для конечного потребителя, так и через посредников) было по вашему мнению произведено из библиотечно-информационного центра? На какое число разысканий был по вашему мнению дан ответ из библиотечно-информационного центра? [6]

Интересуясь числом запросов, библиотеки обычно регистрируют число запросов за данный отрезок времени или наблюдают отмеченное взаимодействие потребителей и неодушевленных информационных ресурсов. При сборе данных всегда складывается такая ситуация, когда тот, кто их собирает, довольно легко может исказить их или отнестись к ним тенденциозно, ведь обычно это работник библиотеки. Вполне вероятно, что эта процедура может устроить его и поэтому он может преувеличить данные с целью приукрасить справочную службу. Например, работник может намеренно изменить числа и преувеличить действительное число запросов. Чаще бывает так, что простой вопрос "Как попасть туда?" может быть сочтен как запрос, хотя в реальности персонал должен считать только информационные запросы.

Как и в отношении оборота фонда, мы хотим узнать что-то об уровне запросов. Все ли запросы равны? Нет. Требуют ли некоторые много времени и сил? Конечно. Так почему не задать вопросы о времени и деталях ответов на запросы? Подумать только, насколько богаче станут данные, если задать следующий вопрос. Как вы считаете, процентов из общего числа запросов, на которые ответил библиотечно-информационный центр, приходится на работу, которая длится следующие отрезки времени (затем нужно перечислить отрезки времени от 1 до 10 минут и так далее)? Или же если спросить информацию о типе запроса: развлекательный, информационный, исследователей? Улучшат ли такие вопросы понимание природы, глубины и качества предоставляемых услуг?

Растет автоматизация справочных систем и поэтому первое, что приходит в голову - это монтаж регистрируемых механизмов в электронные системы, которые позволят вести поиск данных в масштабе запросов, количестве полученных данных и так далее. Говоря об автоматизированных системах, нужно упомянуть вопрос подсчета числа взаимодействия пользователей и машин. Эти данные труднее исказить или как минимум легче устранить неинформационные запросы из подсчета.

Другая крайность - это сбор данных о замеченных случаях взаимодействия пользователей, что, как известно, ненадежно из-за зависимости от удаленного, ненавязчивого наблюдения. Этот метод сбора данных особо подвержен тенденциозному подходу и прежде всего в библиотечной среде с дешевой рабочей силой (т.е. студенты-наблюдатели). Это может привести к "... избирательной регистрации наблюдаемых данных. Наблюдатели будут, вероятно, регистрировать некоторые предметы и отношения с разной степенью интереса, предвзятости и под влиянием своих собственных симпатий [7]. Иначе говоря, важны навыки наблюдения и в той мере, в какой в этих навыках присутствуют изъяны, в той же мере эти изъяны будут наличествовать и в данных. В отличной работе А.Келлехера о наблюдении имеется ряд предупреждений о методах сбора данных, которые коротко можно изложить так: наблюдатель должен иметь навыки наблюдения и он не должен вкладывать никаких мотивов в наблюдаемое взаимодействие и поведение. В любой информационной обстановке естественная тенденция - это выдвижение предположения, что взаимодействие в какой-то мере связано с задачей (пользователь ищет информацию для конкретной цели) и это может привести к выдвижению мотива, которого и нет.

Проблема непредвзятости

Конечно, с учеными, которые выступают за расширение числа методов сбора данных, имеется, как это известно всем практикам, одна загвоздка: они живут в башнях из слоновой кости, страшно удаленных от "реального мира". Нужно признать, что в этом "реальном мире" непредвзятые стороны, для которых ведется много сбора данных и их анализа, просто не желают многих деталей, они не хотят глубоких раздумий о данных, а хотят лишь простой таблицы, которая показывает, насколько учреждение X лучше заведения Y ("лучше" значит увеличенный бюджет, повышенная справочная деятельность, более многочисленные фонды). То есть нужно признать, что сбором данных в существенной мере заняты те, кому подотчетны практики, а те, кому подчинены мы, нередко мелочны по натуре. Являются ли вышестоящие инстанции административными, руководящими, политическими или финансовыми, важно признать, что у них есть полномочия диктовать, какие данные собирать, как их использовать и представлять. Каждая библиотека или информационная служба подотчетна кому-то еще, ибо они зависят от этого кого-то в отношении финансирования и вообще обоснования самого своего существования. Этот "кто-то еще" должен понимать информационные потребности библиотек. Если позволить внешним инстанциям диктовать условия сбора информации и нормы представления, то тогда вполне можно ждать, что они приспособят это к своим интересам, а не к интересам библиотеки. А почему бы им и не сделать так?

Усложнение авторизованных библиотечных систем и облегчение сбора количественной информации о пользователях, фондах, поддержках, сделках значит, что мы уделяем все более внимания простому определению количества как средству оценки. По мере того, как это происходит, крепнет убеждение руководящих инстанций, что данные можно собирать исключительно просто: здесь нажать на кнопку, там дать команду. Следовательно, уменьшается вероятность того, что удастся вырваться за рамки сконцентрированного на числах шаблона, так как наши инспекторы по-прежнему считают это наиболее эффективным способом оценки нашей работы. Нужно признать также, что в программном обеспечении, которое должно помочь в анализе качественных данных (которые нелегко анализировать) просто нет благосклонного отношения к пользователям и легкой интерпретации, что необходимо при анализе данных. Вопреки положительной оценке программного обеспечения анализа качественных данных такими специалистами, как Майлз и Хуберман [8], скептическое

восприятие большинства пакетов, находящихся в продаже, остается. В конце концов компьютерные программы используют технические методы обработки качественных данных, которые по сути соответствуют другим методам, требующим много времени. Нужно провести четкое различие между эффективностью (минимальная стоимость чего-то на единицу) и эффектностью (успешное завершение задачи или командировки). Руководящие инстанции почти неизменно настроены на эффективность, а технология, которая укрепит сбор и анализ данных, укрепит, конечно, эффективность (и только эффективность). В отличие от этого информационные работники - это работники индустрии обслуживания, где успешное завершение нашей задачи - эффективность - должно иметь первостепенное значение.

Что можно сделать?

В предшествующем изложении имеется ряд обстоятельств. Они касаются того, что можно сделать для изменения положения от сбора данных, ориентированного на рентабельность, в котором во главу угла поставлены числа, к разумным методам сбора и анализа, при которых роль контекста повысится. Все они предлагаются не как альтернативы, а как средства укрепления стандартных статистических измерений, применяемых повсеместно в информационном секторе.

- Нужно обратить серьезное внимание на подлинные недостатки количественных методов сбора и анализа данных и постараться включить качественные методы, позволяющие глубже понять пользователей, фонды, обслуживание.
- Не нужно обращать так много внимания на пользователя как на родовое понятие, а уделять внимание конкретным категориям пользователей и характеристике их пожеланий и потребностей.
- Нужно уделять меньше внимания числовым характеристикам фондов, а больше внимания уделять приемлемым показателям качества фондов.
- Не нужно обращать так много внимания на простые запросы пользователя, нужно уделять внимание содержанию и уровню этих вопросов.
- Нужно применять методы сбора качественных данных с полным учетом проблем, связанных с обеспечением использования этих методов без получения заданных сведений.
- Нужно объяснять руководящим инстанциям, что рентабельность и эффективность - это не эквивалентные понятия и что эффективность полезнее рентабельности в информационных учреждениях.
- Нужно работать с разработчиками программного обеспечения при создании высококачественных информационных компьютерных программ, более приемлемых в смысле благосклонного отношения к пользователям и аналитических способностей.

Заключение

В докладе Доула и Хьюрика [9] обсуждены "новые измерения" для библиотечной оценки и в особенности относительно электронных ресурсов. Авторы дали превосходный обзор обычных измерений и четко изложили современное положение. Приятно видеть, что теперь рассматриваются измерения, ориентированные на пользователя. Жаль, что они составляют лишь очень малую часть обычных измерений, основанных на расходах, времени, мероприятиях. Если это - будущее сбора данных в библиотеках, то я не уверен, что произойдет много изменений в лучшую сторону в этой, по моему мнению, удручающей ситуации.

Неплохие результаты может дать кое-какая работа, которая ведется под покровительством американской "Коалиции для сетевой информации" и в особенности работа Ч.Макклуре. В работе "Оценка академической сетевой обстановки: стратегии и возможности выбора" он и С.Лопата представляют учебник сетевой оценки, подход к которой в значительной мере носит качественный характер и которая выдвигает веские аргументы в пользу применения качественных методов при оценке академических сетей [10]. Все же, как представляется, это не

вызвало общего одобрения, и, конечно, не оказало особого воздействия на тех, кто занят сбором данных.

В конечном счете мы выступаем за обращение внимания библиотечных работников на то, что осмысленные данные были связаны с контекстом, и что смысл зависит от интерпретации, что они получаются на основе переменных величин, измерять которые сложно и трудно, что понимание - это индуктивный процесс. Это отличается, но не обязательно противоречит традиционному количественному подходу статистиков, которые предполагают возможность выявления и измерения переменных величин сравнительно прямолинейным образом, и что нормы и консенсус можно получить из данных путем простой дедукции [11]. И то, и другое имеет место в информационной работе, но не нужно вытягивать одно за счет другого - или скорее продолжать подчеркивать одну (количественную) сторону за счет другой (качественной) стороны.

Нужно вспомнить классическую работу Уэбба и др. о ненавязчивых измерениях там, в главе 8, содержится страстный призыв статистиков к ученым: "Используйте все имеющиеся средства нападения" [12]. Больше чем 30 лет спустя настало время информационным профессионалам обратить внимание на этот призыв и выглянуть за рамки чисел, на источники потенциально углубленного содержания.

Ссылки

См. библиографию в оригинальном тексте на английском языке.