

Введение

План лекции:

1. Содержание модуля
 2. Основные определения
 3. Важность данных для процессов цифровой трансформации
 4. Проблематика информации и данных
 5. Виды данных
 6. Пирамида данных
-

Содержание модуля Smart data

Лекции:

- Введение
- Большие данные и хранилища данных
- Графы знаний
- Основы машинного обучения
- Традиционные БД
- Семантические данные и Semantic Metamining

Лабораторные работы / практические занятия

Исследовательская работа

Основные определения

Данные:

- Интерпретируемое формализованным способом представление информации, пригодное для коммуникации, интерпретации или обработки. [ИСО/МЭК 2382-1] [ГОСТ Р 52292-2004]
- Информация, представленная в виде, пригодном для обработки автоматическими средствами при возможном участии человека [ГОСТ 15971-90] [ГОСТ Р 50304-92] [ОСТ 45.127-99]
- Дáнные — зарегистрированная информация: представление фактов, понятий или инструкций в форме, приемлемой для общения, интерпретации, или обработки человеком или с помощью автоматических средств (ISO/IEC/IEEE 24765-2010).

В информатике и информационных технологиях:

Данные — поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи или обработки (ISO/IEC 2382:2015).

Данные — формы представления информации, с которыми имеют дело информационные системы и их пользователи (ISO/IEC 10746-2:1996)

Большие данные:

По данным Gartner -

«Большие данные» - это объемные, быстро доступные и разнообразные информационные ресурсы, которые требуют рентабельных, инновационных форм обработки информации для лучшего понимания и принятия решений.

Это определение намеренно субъективно и включает в себя гибкое определение того, насколько большим должен быть объем данных, чтобы считаться большими данными. Например, большими данными оперируют Amazon и Google и это сильно отличается от больших данных для компаний среднего размера в сфере страхования или телекоммуникаций.

Как следствие такой недосказанности, появилось множество разных определений больших данных, но в целом они говорят о таких данных, размер которых превышает способность типовых СУБД собирать, хранить, управлять и анализировать их. А также анализ и обработки которых может иметь большую ценность для бизнеса.

Smart data (интеллектуальные данные):

Интеллектуальные данные - это данные, которые предварительно обработаны таким образом, чтобы они могли быть обработаны непосредственно в точки из сбора перед отправкой на платформу аналитики для дальнейшей консолидации и анализа. Термин интеллектуальные данные часто ассоциируется с Интернетом вещей (IoT) и данными, которые производят интеллектуальные датчики, встроенные в физические объекты. Термин Smart напрямую связан с тем, что точка сбора данных достаточно интеллектуальна, чтобы принимать некоторые типы решений по входящим данным в режиме реального времени, без необходимости в вычислительных мощностях централизованных систем. В прошлом большая часть аналитики выполнялась с помощью пакетной обработки. Данные собирались по расписанию, преобразовывались в желаемое состояние, помещались в базу данных и обрабатывались периодически. Недостатком этого подхода является то, что к моменту анализа данных они уже устаревали. Напротив, программирование интеллектуальной аналитики данных (также называемое потоковой аналитикой) отслеживает данные непосредственно в их источнике, фиксирует события, которые являются исключениями, оценивает их, принимает решение и делится результатами - и все это в пределах определенного временного окна.

Важность данных для процессов цифровой трансформации

Цифровая трансформация - это культурные, организационные и операционные изменения в организации, отрасли или экосистеме посредством интеллектуальной интеграции цифровых технологий, процессов и компетенций на всех уровнях и функциях поэтапным и стратегическим способом.

Цифровая трансформация - это глубокая трансформация деловой и организационной деятельности, процессов, компетенций и моделей для полного использования изменений и возможностей сочетания цифровых технологий и их ускоряющегося воздействия на общество стратегическим и приоритетным образом с учетом нынешних и будущих изменений.

Области цифровой трансформации

Цифровая трансформация в том интегрированном и связанном смысле, которого она требует, может, среди прочего, затрагивать трансформацию:

- **Бизнес деятельность / функции** : маркетинг, операции, человеческие ресурсы, администрирование, обслуживание клиентов и т. Д.
- **Бизнес-процессы** : одна или несколько связанных операций, действий и наборов для достижения определенной бизнес-цели, при этом управление бизнес-процессами, оптимизация бизнес-процессов и автоматизация бизнес-процессов становятся очевидными (с новыми технологиями, такими как роботизированная автоматизация процессов) . Оптимизация бизнес-процессов имеет важное значение в стратегиях цифровой трансформации и сегодня в большинстве отраслей и случаев представляет собой сочетание ориентированных на клиентов целей и внутренних целей.
- **Бизнес-модели** : как функционирует бизнес, от подхода к выходу на рынок и ценностного предложения до способов, которыми он стремится зарабатывать деньги и эффективно трансформирует свой основной бизнес, используя новые источники дохода и подходы, иногда даже отказываясь от традиционного основного бизнеса после пока.
- **Бизнес-экосистемы** : сети партнеров и заинтересованных сторон, а также контекстные факторы, влияющие на бизнес, такие как нормативные или экономические приоритеты и эволюция. Новые экосистемы выстраиваются между компаниями с различным опытом на основе цифровой трансформации, информации, в результате чего данные и полезный интеллект становятся инновационными активами.
- **Управление бизнес-активами** : при этом основное внимание уделяется традиционным активам, но все в большей степени - менее «осязаемым» активам, таким как информация и клиенты (улучшение качества обслуживания клиентов является ведущей целью многих «проектов» цифровой трансформации, а информация является источником жизненной силы бизнеса, технологической эволюции и любых человеческих отношений). И клиенты, и информация должны рассматриваться как реальные активы со всех точек зрения.
- **Организационная культура** , в соответствии с которой должна быть четкая ориентированная на клиента, гибкая и гиперсознательная цель, которая достигается за счет приобретения основных компетенций по всем направлениям в таких областях, как цифровая зрелость, лидерство, разрозненность работников умственного труда и т. -доказательство. Культура также пересекается с процессами, бизнес-деятельностью, сотрудничеством и ИТ-стороной цифровой трансформации. Для более быстрого вывода приложений на рынок необходимы изменения. В этом суть

DevOps: разработка и эксплуатация. Чтобы заставить ИТ и ОТ работать вместе в компаниях / процессах / деятельности, также необходимы изменения (это не только информационные и операционные технологии, это процессы, культура, сотрудничество) . И т.п.

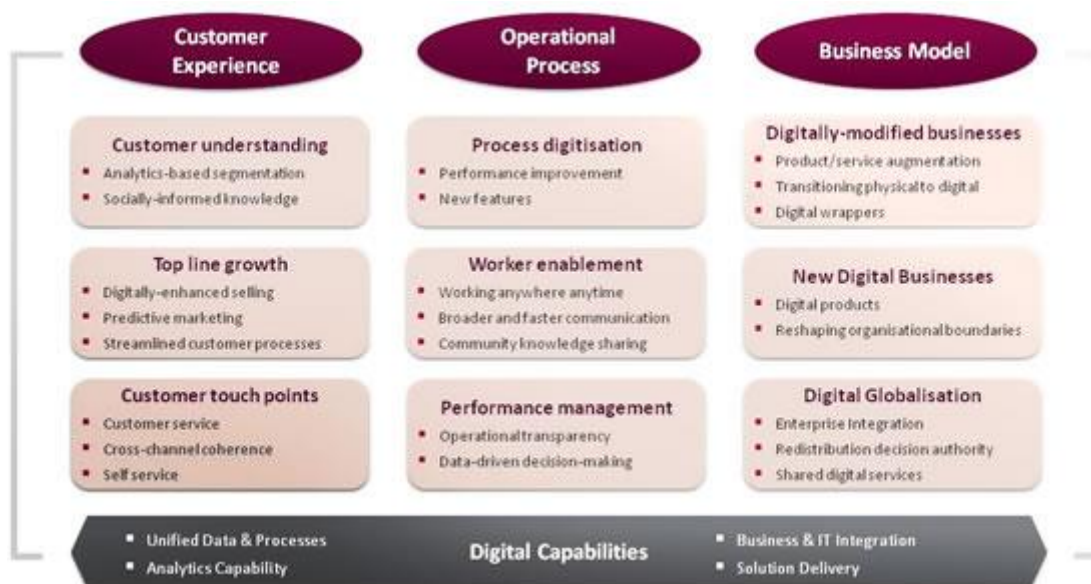
- **Экосистемные и партнерские модели** , среди прочего, рост кооперативных, совместных, совместных творческих и, наконец, совершенно новых подходов к бизнес-экосистеме, ведущих к новым бизнес-моделям и источникам доходов. Экосистемы будут играть ключевую роль в экономике как услуги и в достижении успеха цифровой трансформации.
- **Подходы к клиентам, работникам и партнерам.** Цифровая трансформация ставит людей и стратегию выше технологий. Изменяющееся поведение, ожидания и потребности любой заинтересованной стороны имеют решающее значение. Это выражается во многих подпроектах изменений, в которых клиентоориентированность, пользовательский опыт , расширение прав и возможностей сотрудников, новые модели рабочих мест, динамика изменения торговых партнеров и т. Д. (Могут) все присутствуют в картине. Важно отметить, что цифровые технологии никогда не являются единственным ответом на любой из этих человеческих аспектов, от удовлетворенности сотрудников до улучшения качества обслуживания клиентов. Люди вовлекают, уважают и расширяют возможности других людей, в первую очередь, технологии - это дополнительный инструмент и часть уравнения выбора и фундаментальных потребностей.

Этот список не является исчерпывающим, и де-факто несколько упомянутых аспектов связаны и пересекаются. Мы действительно смотрим на некоторые менее связанные с бизнесом явления «цифровой трансформации» и на так называемые сбои, но основное внимание уделяем бизнесу, что по определению означает целостный взгляд на цифровую трансформацию, в рамках которого такие аспекты, как клиентский опыт, технологическая эволюция и инновации Четкая цель, а не модное слово, являются ключевыми элементами.

Цифровая трансформация происходит повсюду

Итак, цифровая трансформация затрагивает каждую отрасль. Но это также может повлиять на все виды деятельности, подразделения, функции и процессы организации, поскольку может повлиять на саму бизнес-модель как таковую.

CapGemini Consulting была одной из первых, кто предложил концепцию цифровой трансформации и структуру цифровой трансформации, как вы можете видеть ниже. Компания сделала это в сотрудничестве с «Центром цифрового бизнеса Массачусетского технологического института» в ходе трехлетнего исследования, в ходе которого была определена эффективная программа цифровой трансформации как программа, учитывающая, что и как.



В приведенной ниже таблице McKinsey показаны только те аспекты, в которых может сыграть роль цифровая трансформация:

- **(Цифровой) клиентский опыт** (как уже говорилось, де-факто ключевой элемент многих цифровых преобразований, представляющих собой сочетание оптимизации клиентского опыта и улучшения процессов, а также экономии средств).
- **Иновации в продуктах и услугах**, где, например, могут использоваться модели совместного творчества.
- **Распределение, маркетинг и продажи**: еще один обычный подозреваемый и на практике область (наряду с обслуживанием клиентов), которая часто является одной из первых областей, претерпевающих цифровые преобразования.
- **Цифровое исполнение**, оптимизация рисков, усиленный корпоративный контроль и т. Д.

Другие, которые мы можем добавить, включают:

- Интеллектуальное управление информацией (при этом информация, данные и процессы, которые они используют, являются ключевыми и акцентируются на активации).
- Обслуживание клиентов, управление клиентским опытом и контакт-центры, управление взаимоотношениями с клиентами.
- Работа, человеческие ресурсы, новые способы сотрудничества, вовлечение и поддержка персонала (гибкая работа, социальное сотрудничество, сотрудничество на предприятии, унифицированные коммуникации и т. Д.).
- Обучение и образование.
- Закупки, управление цепочкой поставок (с цифровой цепочкой поставок) и отношения с поставщиками.
- И т.п.

Цифровая трансформация и ключевая роль данных и информации

Даже если информация находится в самом центре цифровой трансформации, связь между управлением информацией и цифровой трансформацией проводится

недостаточно часто. В конце концов, независимо от формы оптимизации, коммуникации, сотрудничества, взаимодействия, опыта, инноваций и так далее: информация (или контент, или данные) является ключевым фактором успеха, чтобы это произошло (помимо человеческого фактора, лидерства, процессов, организации). и т.д., которые также требуют информации).

Информационный хаос и информация как средство реализации

По словам Джона Манчини из AIIM (ассоциации профессионалов в области управления информацией), существует четыре проблемы так называемого информационного хаоса :

- Как оптимизировать бизнес-процессы?
- Как мы можем получить представление о бизнесе на основе всей собираемой информации?
- Как мы используем информацию для лучшего взаимодействия с клиентами, сотрудниками и партнерами (также подумайте о всестороннем опыте)?
- Как мы управляем риском увеличения объемов и сложности контента?

Превращение этих проблем «информационного хаоса» в решения - вот в чем суть связи между цифровой трансформацией и управлением информацией.



Но это еще не все. Управление информацией играет роль фактора, способствующего цифровой трансформации, и на каждом этапе достижения целей цифровой трансформации, таких как повышение клиентоориентированности, привлечение эффективных интеллектуальных работников и повышение операционной эффективности.

Информация и управление информацией также являются ключевыми во всех трех частях хорошо известного треугольника: 1) люди, 2) процессы и 3) технологии / инструменты. Чтобы узнать больше, нажмите на ссылку ниже.

Цифровая трансформация и интеллектуальные данные (Smart data)

В контексте цифровой трансформации «управление информацией» и данными имеет решающее значение, но этого недостаточно. В сегодняшнем и завтрашнем бизнесе, основанном на информации и данных, важнее всего понимание, аналитика и действия: результаты.

Вот где вступают в игру контекст, семантика, искусственный интеллект и активация. С появлением Интернета вещей и Web 3.0 интеллектуальное измерение становится более важным в том, что касается понимания неструктурированной информации, автоматизации и подключенных устройств, а также использования информации в работе. Вот почему мы говорим об «интеллектуальной активации информации».

По мере того, как мы переходим к информационным организациям, и информация становится частью капитала и бизнес-активов предприятия, интеллектуальный подход к управлению информацией входит в зал заседаний.

В то же время деятельность и ценность данных рассматриваются с точки зрения вовлеченности, результатов и последней мили.

Обеспечение действенности данных, внедрение устройств (IoT) во все более сложный и растущий ландшафт данных, резкий рост неструктурированных данных, извлечение смысла и понимания из информации и ее использование в нужное время и в нужный момент для правильных причин и действий - все это критически важно .

От управления информацией до интеллектуальной активации информации

Это уже не просто «управление» информацией в традиционном смысле. Речь также идет не только о соединении систем и данных, но даже не о соединении посредством информации. С появлением Интернета вещей необходимость обеспечения качества данных и растущая потребность в их более быстром использовании и разблокировании, несмотря на огромный объем, добавили несколько элементов в уравнение информации и преобразования.

Среди них:

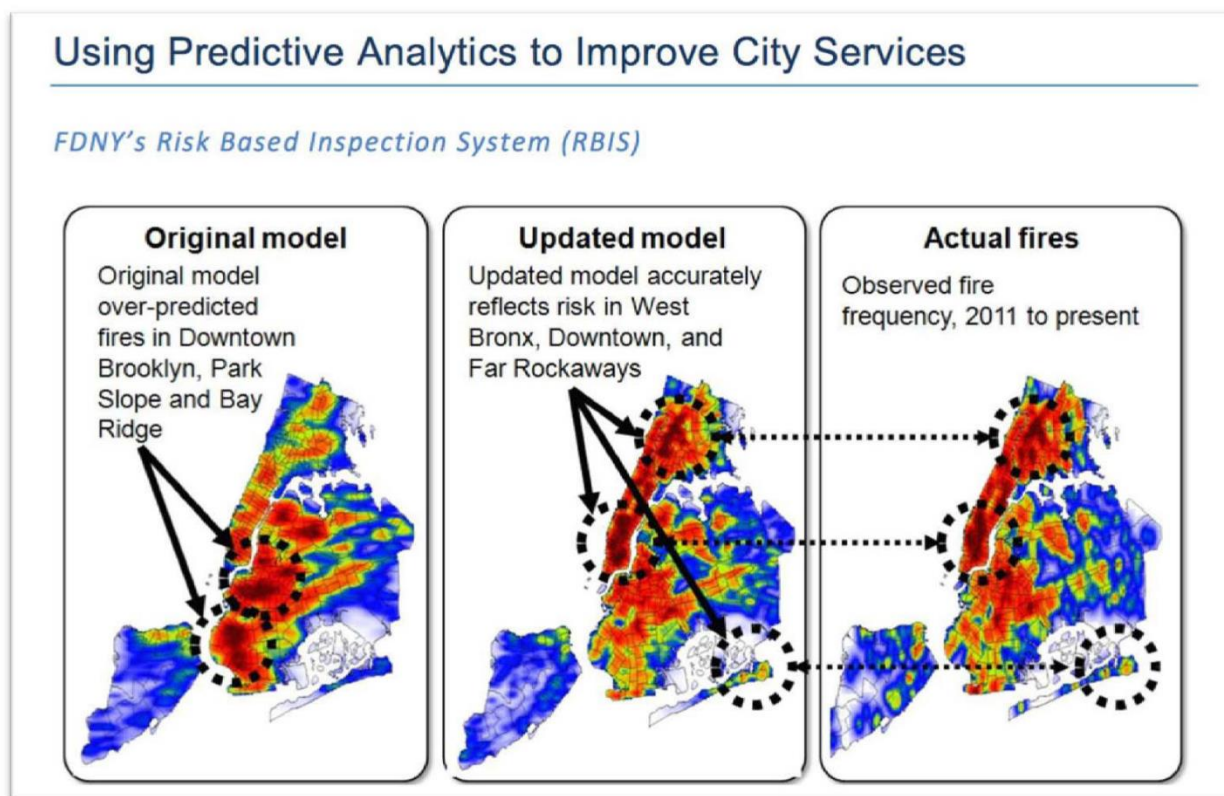
- интеллект (например, искусственный интеллект как единственный способ добавлять и извлекать смысл из еще большего количества данных и как единственный способ использовать информацию и данные в контексте Интернета вещей и между устройствами),

- скорость (скорость - это качество обслуживания клиентов и даже конкурентное преимущество),
- целостный подход к безопасности (с информацией и данными в качестве активов),
- необходимость оцифровывать и фиксировать бумажные данные (цифровое преобразование требует оцифровки и, следовательно, сканирования) ближе к источнику, владельцу и процессу, чтобы перейти на безбумажный (бумага замедляет цифровое преобразование),
- и повышенное внимание к точности, качеству и результатам.

Что все это значит и как это будет развиваться? Помимо наличия систем записей и систем взаимодействия, которые необходимы, мы переходим к системам интеллекта и интеллектуальной автоматизации и оптимизации, экосистемам кода, алгоритмам, когнитивным вычислениям (понимание и за его пределами) и быстрым / интеллектуальным данным. как способы добиться успеха в цифровой трансформации и, наоборот, информационные вызовы как движущие силы трансформации.

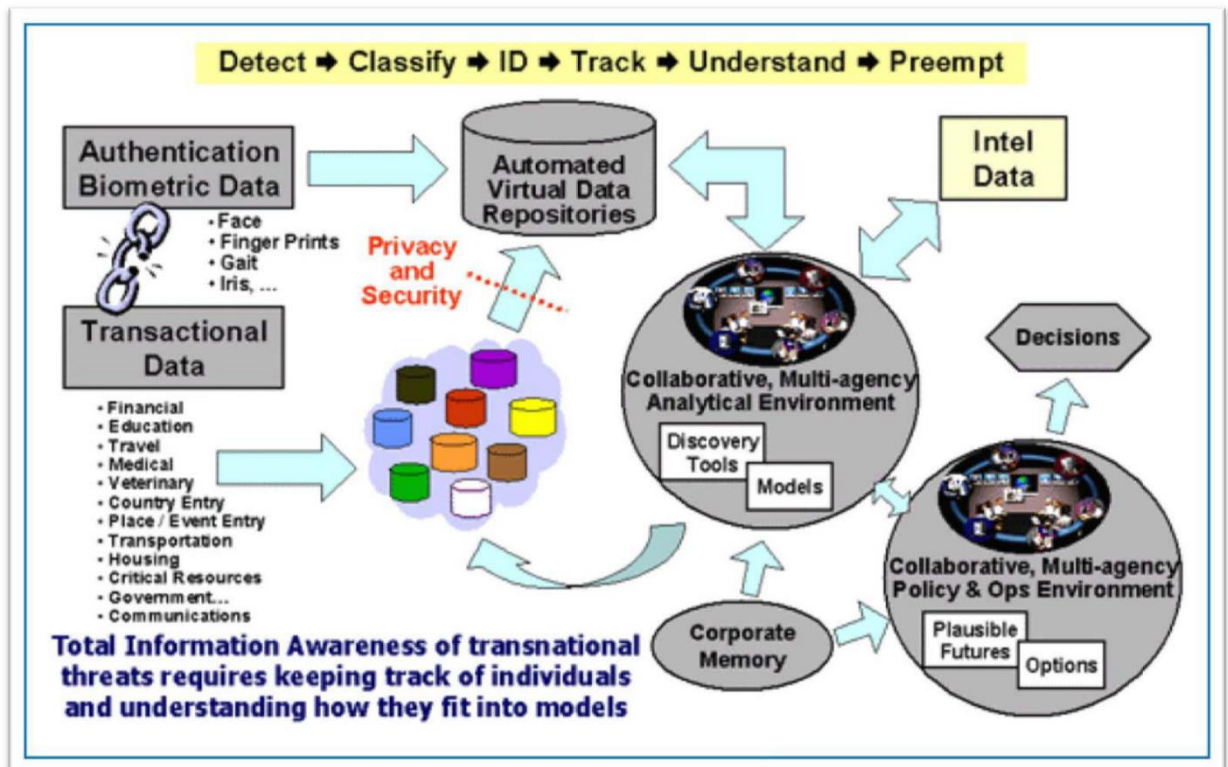
Проблематика информации и данных

Безопасность



New York City

Courtesy of Mayor's Office of Data Analytics. Used with permission.



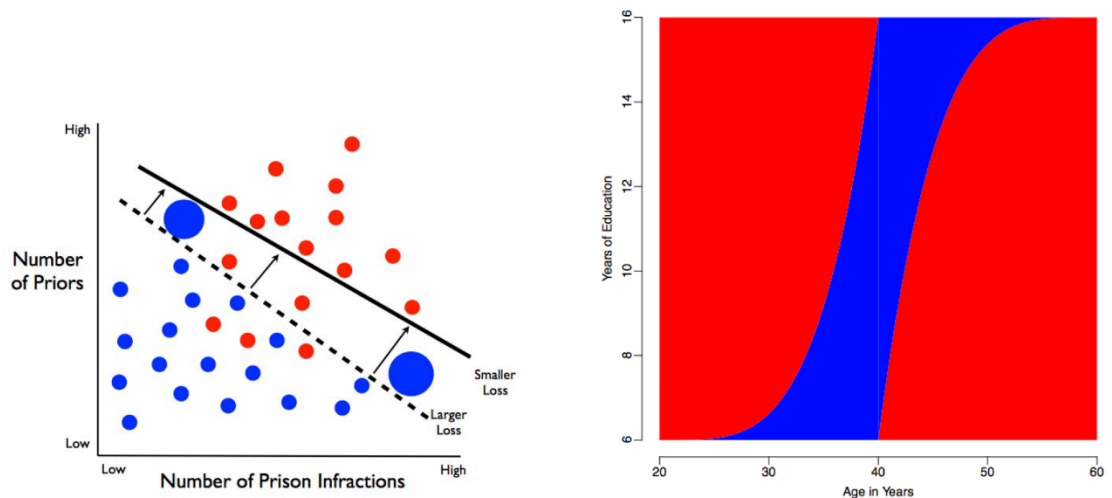
Total Information Awareness program

This image is in the public domain.



Future Attribute Screening Technology

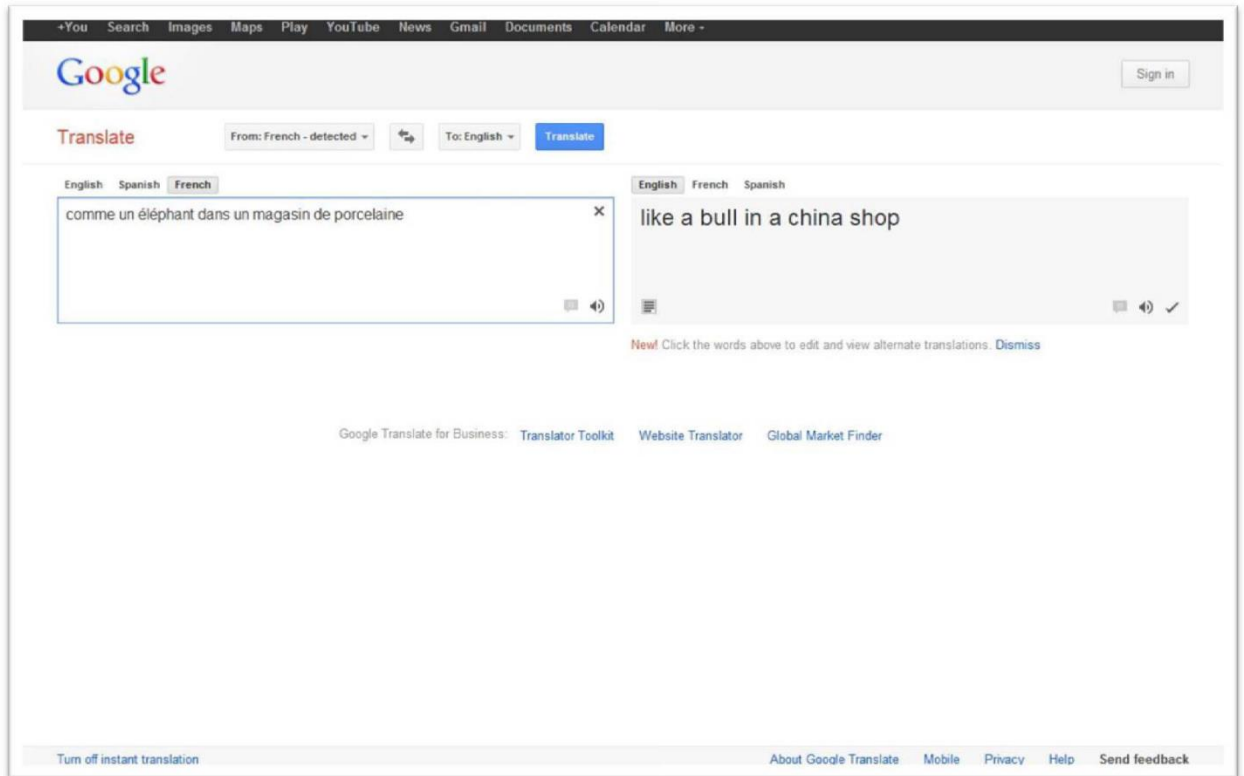
This image is in the public domain.



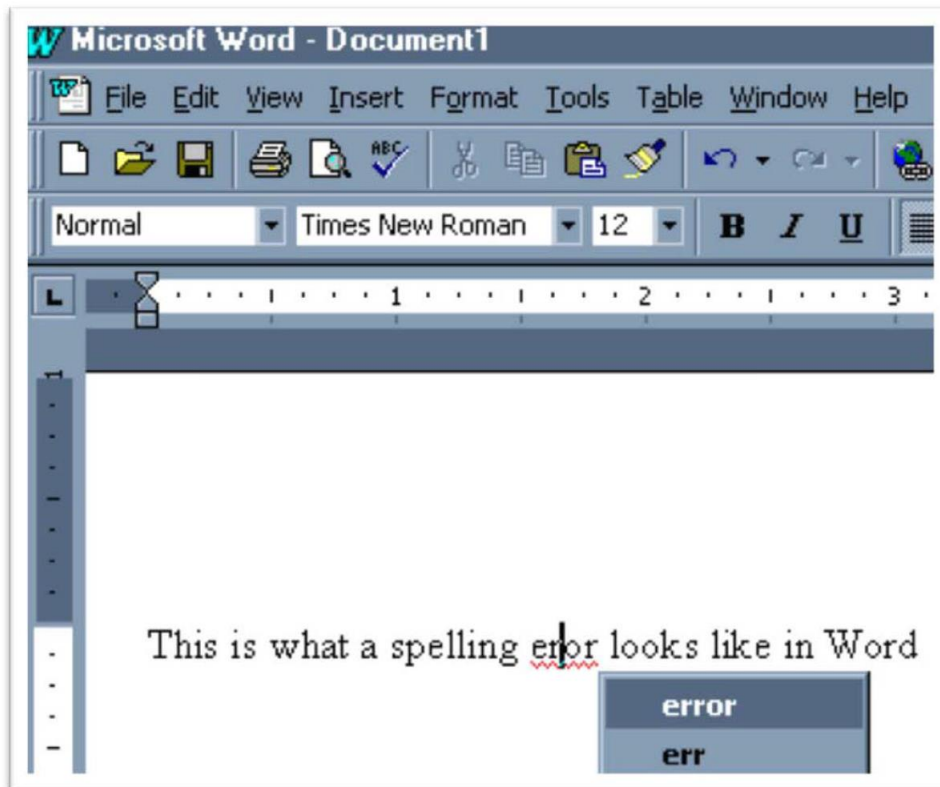
© John Wiley & Sons, Inc. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

Richard Berk's algorithm that predicts if a criminal released from jail will be involved in an homicide.

Сервисы



© Google. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.



© Microsoft. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

Ethical challenges

Potential consequences of Big data misuses and abuses (Bollier, 2010):

- imperil consumer freedom
- imperil civil security
- imperil personal privacy
- imperil of civil liberties

Dataveillance

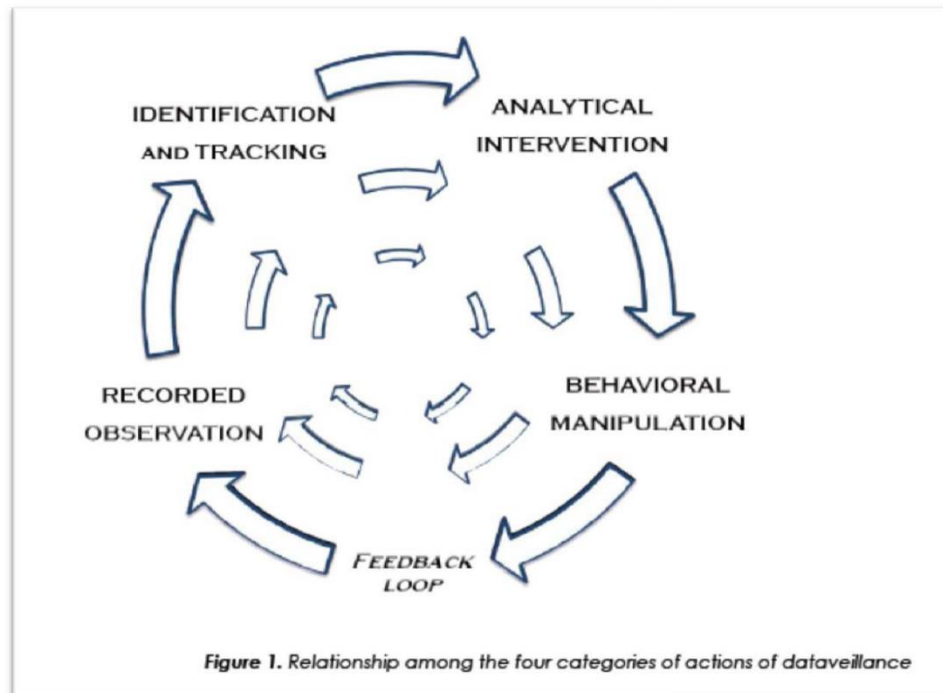
Definition: “systematic monitoring of people or groups, by means of personal data systems, in order to regulate or govern their behavior.”

(Degli Esposti, 2014)

Applications:

- Recruitment and retention
- Customer loyalty
- Supply chain efficiency
- Security and risk prevention

Dataveillance



© Surveillance & Society (Sara Degli Esposti). All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

(Ref. Degli Esposti, 2014)

Dataveillance

Main problems:

- unequal relationship between cie & customers
- happens without our knowledge (panopticon)

Civil liberties

People could eventually:

- pay more for health insurance
- be excluded from health coverage
- be denied a loan
- be subjected to profiling or extra scrutiny
- be kept in jail
- be arrested before committing a crime
- be found guilty by association

(Ref. Mayer-Schönberger and Cukier, 2014)

Civil liberties

Accusing a person of a possible behavior would:

- negate the very foundation of our justice (presumption of innocence and responsibility of people)
- deny human volition and free will
- exonerate us from any responsibility

(Ref. Mayer-Schönberger and Cukier, 2014)

Civil liberties

Should we accept to limit our individual freedom and rely on big data analytics for the common good?

Can big data analytics help us to avoid gender discrimination and racial profiling?

Privacy and civil liberties

Strategies currently in place:

- notice and individual consent
- opting-out
- anonymization
- hacking privacy violators

Potential solutions:

- shifting from 'privacy by consent' to 'privacy by accountability'
- defining new regulations (government or market?)
- redefining the notion of justice to preserve human agency
- opening the access to datasets and algorithms (transparency)
- mandating impartial experts to certify the algorithms and the interpretations
- blurring datasets (differential privacy)
- specifying how people can disprove a prediction about them

Potential solutions (Bollier, 2010):

- erasing data after a given period of time
- giving to people the right to take possession of their data
- limiting data collection to what is needed
- favouring self-discipline by introducing new social norms
- gaming the system by voluntarily altering our behavior (Nissenbaum)
- being proactive

Methodological implications

Paradigm shift – Big data requires researchers to:

- deal with the abundance and embrace messiness
- conduct research with no precise question in mind
- gain insights from data (induction) instead of testing theories by analyzing data (deduction)
- use hundreds of algorithms instead of selecting one
- stop obsessing with causality and be satisfied with correlations

(Ref. Kitchin, 2014)

Advantages of big data compared to traditional scientific methods:

- reduces the risk of error and bias associated with sampling (but can also face problems)
- provides the opportunity to create more dynamic and sophisticated models (but does not offer certainties)
- helps finding correlations that no theory can identify (but can lead to serious fallacies)
- allows to consider data once perceived as “noise”

TWO SCHOOLS OF THOUGHT

Empiricist (business):

- Big data can speak for themselves without the need of theories, models or hypothesis (fallacious)
- Big data analytics are free of human bias. They can be interpreted by anyone and their meanings transcend contexts (fallacious)

Data-driven science (academia)

- Use of existing theories and concepts to analyze the datasets
- Use of big data to identify good questions and formulate hypothesis
- Use of datasets for answering questions that these datasets were not designed to answer (problem)

Alternative path suggested by Kitchin (2014):

- Drawing on critical theory to frame how the research is conducted and how the results are interpreted.
- Admitting that research is never ideologically neutral and objective.
- Complementing big data studies with small data studies.

Виды данных

Независимо от того, являются ли данные предварительными и риторическими по своей природе или нет, очевидно, что данные различаются по своим характеристикам, которые в явном виде определяют, как с ними обращаться и что с ними можно сделать. В широком смысле данные различаются по форме (качественной или количественной), структуре (структурированная, полуструктурированная или неструктурированная), источнику (захваченный, производный, исчерпывающий, переходный), производителю (первичный, вторичный, третичный) и типу (индексный, атрибут, метаданные).

Количественные и качественные данные

Данные могут принимать множество материальных форм, включая числа, текст, символы, изображения, звук, электромагнитные волны или даже пустоту или тишину (пустое пространство само по себе является данными). Обычно они делятся на две большие категории. Количественные данные состоят из числовых записей. Как правило, такие данные обширны и относятся к физическим свойствам явлений (таким как длина, высота, расстояние, вес, площадь, объем) или являются репрезентативными и относятся к нефизическим характеристикам явлений (например, социальный класс, уровень образования, социальная депривация, рейтинг качества жизни). Количественные данные имеют четыре различных уровня

измерения, которые определяют способы их обработки и анализа (Kitchin and Tate 1999, см. Также Таблицу 1.1). Такие данные можно анализировать с помощью визуализаций, различных описательных и логических статистических данных и использовать в качестве входных данных для прогнозных и имитационных моделей.

Table 1.1 Levels of data measurement

Levels of measurement	Definition	Example
Nominal data	Categorical in nature, with observations recorded into discrete units.	Unmarried, married, divorced, widowed
Ordinal data	Observations that are placed in a rank order, where certain observations are greater than others.	Low, medium, high
Interval data	Measurements along a scale which possesses a fixed but arbitrary interval and an arbitrary origin. Addition or multiplication by a constant will not alter the interval nature of the observations. Data can either be continuous (e.g., time or length) or discrete (e.g., counts of a phenomenon) in nature.	Temperature along the Celsius scale
Ratio data	Similar to interval data except the scale possesses a true zero origin, and multiplication by a constant will not alter the ratio nature of the observations.	Exam marks on a scale of 0–100

Напротив, качественные данные не являются числовыми, например тексты, изображения, искусство, видео, звуки и музыка. Хотя качественные данные можно преобразовать в количественные, перевод включает в себя значительное сокращение и абстракцию, и большая часть богатства исходных данных теряется в результате такого процесса. Следовательно, качественный анализ данных обычно практикуется на исходных материалах, стремясь выявить и создать смысл и понимание, а не подвергать данные механическому использованию вычислительных методов. Однако значительный прогресс был достигнут в отношении обработки и анализа качественных данных с помощью вычислений с помощью таких методов, как машинное обучение и интеллектуальный анализ данных. Структурированные, полуструктурированные и неструктурированные данные.

Структурированные данные - это данные, которые можно легко организовать, сохранить и передать в определенной модели данных, например числа / текст, представленные в таблице или реляционной базе данных, которые имеют согласованный формат (например, имя, дату рождения, адрес, пол, так далее). Такие данные можно обрабатывать, искать, запрашивать, комбинировать и анализировать относительно просто, используя исчисления и алгоритмы, и их можно визуализировать с помощью различных форм графиков и карт и легко обрабатывать компьютерами. Полуструктурированные данные - это свободно структурированные данные, которые не имеют предопределенной модели / схемы данных и, следовательно, не могут храниться в реляционной базе данных. Их структура нерегулярная, неявная, гибкая и часто иерархически вложенная, но они

имеют достаточно согласованный набор полей, и данные помечаются, таким образом, разделяя контент семантически и предоставляя свободные, самоопределяющиеся метаданные контента и средства для сортировки, упорядочивания и структурирования данных. Примером таких данных являются веб-страницы с тегами XML (страницы, созданные с использованием расширяемого языка разметки [XML], который кодирует документы в формате, который является как человеком, так и машиночитаемым; Franks 2012; см. Связанные данные в главе 3). Напротив, неструктурированные данные не имеют определенной модели данных или общей идентифицируемой структуры. Каждый отдельный элемент, такой как повествовательный текст или фотография, может иметь определенную структуру или формат, но не все данные в наборе данных имеют одинаковую структуру. Таким образом, хотя их часто можно искать и запрашивать, их нелегко объединить или проанализировать с помощью вычислений. Такие неструктурированные данные обычно носят качественный характер, но часто могут быть преобразованы в структурированные данные с помощью классификации и категоризации. До относительно недавнего времени очень большие наборы данных обычно были структурированы по форме, потому что их, как правило, было намного проще обрабатывать, анализировать и хранить. В эпоху больших данных многие массивные наборы данных состоят из полу- или неструктурированных данных, таких как сообщения Facebook, твиты, загруженные изображения и видео, а также блоги, и по некоторым оценкам, такие данные растут в 15 раз быстрее, чем структурированные. (Zikopoulos et al. 2012), с достижениями в проектировании баз данных (например, базы данных NoSQL, которые не используют табличные модели реляционных баз данных, см. Главу 5) и методами машинного обучения (см. Главу 6), помогающими хранить и анализировать.

Собранные, отработанные, переходные и производные данные

Есть два основных способа создания данных. Во-первых, данные могут быть получены напрямую с помощью какой-либо формы измерения, такой как наблюдение, опросы, лабораторные и полевые эксперименты, ведение записей (например, заполнение форм или ведение дневника), камеры, сканеры и датчики. В этих случаях данные обычно являются преднамеренным продуктом измерения; то есть намерение состояло в том, чтобы получить полезные данные. Напротив, данные о выхлопных газах по своей сути производятся устройством или системой, но являются побочным продуктом основной функции, а не первичным выходом (Manuika et al. 2011). Например, электронная касса предназначена для суммирования закупленных товаров и обработки платежей, но она также производит данные, которые можно использовать для отслеживания запасов, производительности работников и покупок клиентов. Многие программные системы производят такие данные о выхлопных газах, многие из которых стали ценными источниками информации. В других случаях данные выхлопа носят временный характер; то есть они никогда не исследуются и не обрабатываются, а просто выбрасываются, потому что они слишком объемны или неструктурированы по своей природе, или потому что их обработка и хранение дорого обходятся, либо отсутствуют методы извлечения из них ценности, либо они малоэффективны. стратегическое или тактическое использование (Zikopoulos et al. 2012; Franks 2012). Например, Manuika et al. (2011: 3) сообщают, что «поставщики медицинских услуг ... отбрасывают 90 процентов данных, которые они генерируют (например, почти все видеопотоки в реальном времени, созданные во время операции)». Собранные и отработанные данные считаются «сырыми» в том смысле, что они не были преобразованы или объединены с другими данными. Напротив, производные данные производятся путем дополнительной обработки или анализа собранных

данных. Например, захваченные данные могут представлять собой индивидуальные подсчеты трафика через перекресток, а производные данные - общее количество отсчетов или отсчетов в час. Последние произошли от первых. Собранные данные часто являются входными данными в модель, а производные данные - выходными.

Например, данные подсчета трафика могут быть входными данными в транспортную модель с прогнозируемыми или смоделированными данными (такими как прогнозируемые подсчеты трафика в разное время или в разных условиях). В случае модели данные подсчета трафика, вероятно, были объединены с другими захваченными или производными данными (такими как тип транспортного средства, количество пассажиров и т. Д.) Для создания новых производных данных для ввода в модель. Производные данные создаются по ряду причин, в том числе для уменьшения объема данных до управляемого количества и для получения более полезных и значимых показателей. Иногда исходные захваченные данные могут обрабатываться с разными уровнями вывода в зависимости от их предполагаемого использования. Например, система наблюдения за Землей НАСА организует свои данные по шести уровням, начиная от необработанных захваченных данных, с возрастающей степенью обработки и анализа, и заканчивая выходными данными на основе анализа данных более низкого уровня (Borgman 2007; см. Таблицу 1.2).

Table 1.2 The six levels of data of NASA's Earth Observing System

Data level	Description
Level 0	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artefacts (e.g., synchronisation frames, communications headers, duplicate data) removed.
Level 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters computed and appended but not applied to Level 0 data.
Level 1B	Level 1A data that have been processed to sensor units.
Level 2	Derived geophysical variables at the same resolution and location as Level 1 source data.
Level 3	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
Level 4	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

Source: Adapted from <https://earthdata.nasa.gov/data/standards-and-references/processing-levels>

Source: Adapted from <https://earthdata.nasa.gov/data/standards-and-references/processing-levels>

Первичные, вторичные и третичные данные

Первичные данные генерируются исследователем и его инструментами в рамках разрабатываемого ими исследовательского проекта. Вторичные данные - это данные, созданные кем-то другим для повторного использования и анализа. Таким образом, первичные данные одного человека могут быть вторичными данными другого человека. Третичные данные - это форма производных данных, таких как подсчеты, категории и статистические результаты. Третичные данные часто

публикуются статистическими агентствами, а не вторичными, чтобы гарантировать конфиденциальность в отношении того, к кому относятся эти данные. Например, первичные данные ирландской переписи не могут быть опубликованы в качестве вторичных данных в течение 100 лет после генерации; вместо этого данные публикуются в виде сводных подсчетов и категориальных третичных данных. Многие исследователи и учреждения стремятся генерировать первичные данные, потому что они адаптированы к их конкретным потребностям и целям, тогда как эти варианты дизайна недоступны для тех, кто анализирует вторичные или третичные данные. Более того, те, кто использует вторичные и третичные данные в качестве исходных данных для своих собственных исследований, должны верить в достоверность оригинального исследования. Во многих случаях исследователи объединяют первичные данные со вторичными и третичными данными для получения более ценных производных данных. Например, розничный торговец может стремиться создать производный набор данных, который объединяет их первичные данные о продажах с третичными геодемографическими данными (данные о том, какие люди живут в разных районах, которые получены из переписи населения и других общедоступных и коммерческих данных), чтобы определить на какие места следует таргетировать маркетинговые материалы. Вторичные и третичные данные ценны, потому что они позволяют проводить репликационные исследования и создавать более крупные, богатые и сложные наборы данных. Позже они производят то, что Crampton et al. (2012) термин «усиление данных»; то есть данные, когда они объединены, дают возможность гораздо более глубокого понимания, выявляя ассоциации, отношения и закономерности, которые остаются скрытыми, если данные остаются изолированными. Как следствие, вторичный и третичный рынок данных представляет собой многомиллиардную отрасль (см. Главу 2).

Индексные и атрибутные данные и метаданные

Данные также различаются по типу. Индексные данные - это те данные, которые обеспечивают идентификацию и связывание, и включают уникальные идентификаторы, такие как номера паспорта и социального страхования, номера кредитных карт, серийные номера производителя, идентификаторы цифровых объектов, IP- и MAC-адреса, номера заказа и доставки, а также имена, адреса и почтовые индексы. Индексированные данные важны, потому что они позволяют связывать большие объемы неиндексированных данных и отслеживать их с помощью общих идентификаторов, а также позволяют различать, комбинировать, дезагрегировать и повторно агрегировать, искать и другие формы обработки и анализа. Как обсуждалось в главе 4, индексированные данные становятся все более распространенными и детализированными, повышая релевантность наборов данных. Атрибутные данные - это данные, которые представляют аспекты явления, но не являются индексными по своей природе. Например, в отношении человека индексные данные могут быть отпечатком пальца или последовательностью ДНК, со связанными атрибутивными данными, такими как возраст, пол, рост, вес, цвет глаз, группа крови и т. Д. Подавляющая часть данных, которые генерируются и хранятся в системах, являются атрибутивными данными. Метаданные - это данные о данных. Метаданные могут относиться либо к содержимому данных, либо ко всему набору данных. Метаданные о содержимом включают имена и описания конкретных полей (например, заголовки столбцов в электронной таблице) и определения данных. Эти метаданные помогают пользователю набора данных понять его состав и то, как его следует использовать и интерпретировать, а также облегчают объединение наборов данных, совместимость и обнаруживаемость, а также судить об их происхождении и

происхождении. Метаданные, относящиеся к набору данных в целом, имеют три различных формы (NISO 2004). Описательные метаданные касаются идентификации и обнаружения и включают такие элементы, как заголовок, автор, издатель, тема и описание. Структурные метаданные относятся к организации и охвату набора данных. Административные метаданные касаются того, когда и как был создан набор данных, деталей технических аспектов данных, таких как формат файла, а также того, кто владеет и может использовать данные. Распространенным стандартом метаданных для наборов данных, который объединяет эти три типа метаданных, является Dublin Core (<http://dublincore.org/>). Этот стандарт требует, чтобы наборы данных имели 15 сопутствующих полей метаданных: заголовок, создатель, тема, описание, издатель, участник, дата, тип, формат, идентификатор, источник, язык, отношение, охват и права. Метаданные являются важными компонентами всех наборов данных, хотя они часто игнорируются при курировании данных, особенно среди исследователей, которые собирают первичные данные для собственного использования, а не для совместного использования.

Данные кадрирования

До сих пор в этой главе данные уже начали концептуально оформляться с точки зрения обсуждения онтологии данных (что такое данные), их различных форм и того, как они вписываются в пирамиду знаний. Существует множество других способов осмысления и понимания данных, например, с технической точки зрения, касающейся качества, достоверности, надежности, подлинности и удобства использования данных, а также способов их обработки, структурирования, совместного использования и анализа; или этический взгляд на причины, по которым создаются данные, и способы использования данных; или политическая или экономическая перспектива, которая рассматривает, как данные нормативно рассматриваются и оспариваются как общественные блага, политический капитал, интеллектуальная собственность или товар, а также как они регулируются и продаются; или пространственная и / или временная перспектива, которая рассматривает, как технические, этические, политические и экономические режимы, касающиеся производства данных и их использования, развиваются и видоизменяются в пространстве и времени; или философская перспектива, которая рассматривает различные и оспариваемые онтологии и эпистемологии данных. Многие из вопросов, кратко обсуждаемых в этом разделе, возвращаются на протяжении всей книги.

Технически

По всем дисциплинам данные рассматриваются с нормативной, технической точки зрения. На карту поставлено то, в какой степени методы сбора и измерения генерируют определенные, чистые и точные данные, и как такие данные могут и должны обрабатываться, структурироваться, совместно использоваться и анализироваться таким образом, чтобы поддерживать их целостность, обеспечивая тем самым надежность и достоверность. Из них можно сделать обоснованные выводы. Всегда есть сомнения в достоверности данных, потому что они по своей сути абстрагируются, обобщаются и аппроксимируются в процессе их производства (Goodchild 2009). Таким образом, значительное внимание уделяется таким вопросам, как репрезентативность данных, неопределенность, надежность, ошибка, систематическая ошибка и калибровка в рамках разработки и реализации исследования, при этом эта информация записывается в виде метаданных. Учитывая, что данные являются суррогатом некоторого аспекта явления - свет, представляющий звезду, физические характеристики, представляющие растение, слова, представляющие мысли человека, - репрезентативность связана с тем,

насколько хорошо данные отражают явление, которое они стремятся представить, и насколько хорошо образец полученные данные представляют собой всю совокупность. Что касается первого, то ключевой вопрос заключается в том, насколько мы можем быть уверены в том, что научные методы точно охватывают рассматриваемое явление. Это была особая проблема в социальных и гуманитарных науках, и ее оказалось трудно решить. Например, хорошо известно, что то, что люди говорят, что они будут делать, и то, что они делают, часто совершенно разные, и то, что люди делают, часто не то, что они намеревались. Поэтому возникает вопрос, насколько хорошо данные интервью отражают человеческое поведение или насколько хорошо поведение отражает сознательное мышление. Точно так же существуют опасения по поводу того, в какой степени ключевые индикаторы адекватно отражают и отражают работу домена. Например, в какой степени такие показатели, как количество цитирований, индекс Хирша и зарегистрированные патенты, указывают на высокое качество работы сотрудников университета (в отношении факультета гуманитарных наук они считаются очень плохими показателями)? Решение состояло в том, чтобы попытаться разработать все более и более сложные исследовательские проекты, которые противодействовали бы недостаткам различных методов или в значительной степени игнорировали недостатки.

Что касается того, насколько хорошо выборка представляет популяцию, мы могли бы решить сгенерировать подробные, продольные, астрономические данные по 50 звездам, чтобы лучше понять их природу, но в какой степени мы можем быть уверены, что эти 50 звезд представляют качества из миллиардов существующих звезд? Даже в эпоху больших данных, которые стремятся быть исчерпывающими, а не выборочными при генерации данных (см. Главу 4), данные по своей сути являются выборкой (не все люди используют социальные сети или совершают покупки с помощью кредитных карт, и действительно многие люди в разных странах). в мире нет доступа к телефонам или компьютерам), что означает, что данные являются репрезентативными для группы людей, даже если эта группа очень велика. Опять же, решение заключалось в разработке ряда методов выборки, которые стремятся обеспечить репрезентативность в различных условиях (часто в зависимости от случайности выборки), и статистических методов, которые вычисляют степень, в которой мы можем быть уверены, что выборка представляет собой совокупность. (Китчин и Тейт, 1999). Надежность касается повторяемости или последовательности в получении одного и того же результата при использовании инструмента исследования. Голледж и Стимсон (1997) описывают три вида надежности: (1) донкихотская надежность, когда единственный метод наблюдения постоянно дает неизменное измерение; (2) диахроническая надежность, стабильность наблюдения во времени; и (3) синхронная надежность, схожесть наблюдений за один и тот же период времени. Надежность важна, потому что общепризнано, что чем более последовательным является показатель при производстве данных, тем больше доверия к нему можно отнести. Ошибка - это разница между измеренным и реальным значением и может включать в себя отсутствие (недостающие данные), ошибки (например, неправильное кодирование или неправильную классификацию или неправильное применение метода) и недопонимание. Смещение - это особый вид ошибки, при которой данные искажаются из-за постоянной последовательности ошибок. Предвзятость обычно вызвана тем, что метод, инструмент или техника выборки, использованные для создания данных, оказывают чрезмерное влияние на то, какие данные производятся, или могут быть внесены из-за идеологической позиции или стремлений исследователя, часто непреднамеренно (Китчин 1996). Такие методы обработки, как агрегирование, могут вносить систематическую ошибку за счет

уменьшения дисперсии в наборе данных, приводящей к ошибкам экологической ошибки, то есть при условии, что агрегированное значение точно представляет агрегированных лиц (например, если бы у нас было два человека весом 50 кг и двое весом 150 кг). их средний совокупный вес будет составлять 100 килограммов, но никто в наборе не весит это количество) (Kitchin and Fotheringham 1997). Неопределенность касается того, в какой степени

Этично

Этика касается мысли и практики, связанных с такими ценностными концепциями, как справедливость, равенство, справедливость, честность, уважение, права, права и забота. Каждое общество действует, руководствуясь сочетанием здравых, неформальных и само собой разумеющихся моральных ценностей, а также строго кодифицированных этических позиций, закрепленных в правилах, принципах, политике, лицензиях и законах, подлежащих исполнению со стороны государства и других агентств. Эти этические позиции часто оспариваются, когда разные группы придерживаются противоположных взглядов на сами ценности и степень, в которой этические позиции должны быть закреплены в законодательстве, и их обсуждение является упражнением в моральной философии. Такое оспаривание также существует в отношении данных, особенно в отношении того, какие данные генерируются и средства производства, как данные распространяются, продаются и защищаются и для каких целей используются данные.

В то время как некоторые данные считаются относительно безвредными, например измерения, относящиеся к погоде, другие данные считаются высокочувствительными, например, данные, относящиеся к отдельным лицам, которые могут использоваться для создания подробной картины их жизни и регулирования этих жизней. В некоторых случаях сбор данных может нанести вред, например, опрос лиц, переживших военные преступления, может вызвать психологический стресс. Здесь возникают вопросы о том, в какой степени создание данных и различные формы наблюдения за данными (наблюдение, осуществляемое посредством обработки и анализа записей данных) и анализ данных нарушают конфиденциальность и другие права человека и могут использоваться для активной социальной сортировки людей (предоставляют дифференцированные услуги в зависимости от их характеристик) (Graham 2005). Эти опасения усугубляются с учетом того, что цифровые данные можно легко объединять, обмениваться и обмениваться, и мы живем в эпоху повсеместного инвазивного сбора данных и надзора. Возможно, неудивительно, что агентства, финансирующие академические исследования и высшие учебные заведения, теперь регулярно оценивают этические аспекты исследовательских проектов с точки зрения их потенциальных более широких последствий, а страны приняли законы, такие как законы о защите данных и конфиденциальности, чтобы попытаться предотвратить неправомерное использование и злоупотребление данными. Эти и связанные с ними вопросы более подробно обсуждаются в главе 10.

Политически и экономически

Рассмотрение этики данных начинает раскрывать способы, которыми данные оформляются более широкими политическими и экономическими проблемами. Какие данные генерируются, как они обрабатываются, анализируются и используются, контекстуализируется с учетом: того, как они нормативно воспринимаются населением и используются штатами, а также понятий того, как они должны регулироваться и регулироваться законодательством; дискурсы, используемые в дискурсивных режимах, которые поддерживают или противостоят

их возникновению и применению; принятие решений о финансировании и инвестировании в данные; развитие капитализма и способы использования данных для управления усилиями и увеличения стоимости и прибыли; и торгуются как товар с появлением многомиллиардного рынка данных, состоящего из разнообразного набора игроков (производителей, агрегаторов, продавцов, аналитиков и потребителей; см. главу 2). Те, кто производит данные, должны учитывать общественное и политическое мнение, этические соображения, нормативно-правовую среду, доступное финансирование и надежность своих инвестиций по сравнению с ресурсами. Лица, отвечающие за законодательную и финансовую сферы, должны обдумывать и принимать решения о том, как формировать ландшафт, в котором работают производители и пользователи данных, а также учитывать свои собственные режимы данных и то, что они рассказывают о своих программах, приоритетах и способы управления и управляемость (Lauriault 2012). В обоих случаях играет роль разнообразный набор политических и экономических рациональностей, при этом конкурирующие голоса стремятся повлиять на мнение и более широкую область данных. Движение открытых данных, например, рассматривает данные как общественное достояние, которое должно составлять общественное достояние и быть свободно доступным (за исключением конфиденциальных личных данных) или быть доступным через соглашения о добросовестном использовании. Напротив, бизнес рассматривает данные как ценный товар, который, с одной стороны, необходимо защищать с помощью режимов интеллектуальной собственности (авторское право, патенты, права собственности), а с другой - не быть настолько ограниченным этическими соображениями, что они не могут использоваться для прироста капитала. Для сообществ и государств данные - это средство, с помощью которого политические программы и работа могут быть узаконены, реализованы и оспорены путем создания основанных на фактах нарративов и контрдискурсов, которые имеют большую риторическую ценность, чем анекдоты или настроения (Wilson 2011; Garvey 2013). Другими словами, данные составляют в терминах Фуко (1981) форму власти / знания; средство, с помощью которого можно наблюдать и регулировать людей, явления и территорию (Lauriault 2012). Эти альтернативные интересы часто могут согласовываться парадоксальным образом, хотя у них могут быть совершенно разные программы, например, поддержка крупным бизнесом движения за открытые данные в отношении общедоступных данных (см. Главу 3). Другими словами, данные проявляются и расположены в сложных и спорных политических экономиках и, в то же время, они используются для формирования таких дискуссий и режимов. Более того, данные представляют собой экономический ресурс, который является ключевым компонентом следующего этапа экономики знаний, изменяя способ производства на управляемый данными (см. Главу 7). С конца 1980-х годов такие ученые, как Кастельс (1988, 1996), утверждали, что последний цикл капитализма опирается на производство знаний, которые создают новые продукты и формы труда, облегчают реструктуризацию экономики и повышают производительность, конкурентоспособность, эффективность и т.д. устойчивость и накопление капитала. В частности, большие данные являются последним достижением в углублении и продвижении этого цикла, обеспечивая множество доказательств, которые используются компаниями, с одной стороны, для мониторинга и оценки деятельности компании в режиме реального времени, сокращения потерь и мошенничества, а также совершенствовать корпоративную стратегию, планирование и процесс принятия решений, а с другой - разрабатывать новые товары, выявлять новые рынки и ориентироваться на них, внедрять динамическое ценообразование, реализовывать неиспользованный потенциал и получать конкурентные преимущества (Manyika et al. 2011; Zikopoulos et al. 2012).

Поступая таким образом, производство и анализ данных позволяет компаниям работать более разумно в отношении того, как они организованы и работают, способствуя гибкости и инновациям, сокращая риски, затраты и операционные потери, улучшая качество обслуживания клиентов и максимизируя рентабельность инвестиций и прибыль. Стимулируя накопление капитала, большие данные способствуют новому разделению труда и следующему витку неравномерного развития. Таким образом, данные можно рассматривать как агент интересов капитала.

Во времени и в пространстве

Данные имеют как временность, так и пространственность. Какие данные производятся и способы их обработки, анализа, хранения или удаления варьируются во времени и пространстве; данные и окружающие их сообщества имеют историю и географию. То, как данные обрабатываются и анализируются, со временем меняется под влиянием организационных изменений и улучшений в регистрации и администрировании, новых законов, касающихся обработки и защиты данных, новых технологий, новых методов сортировки и анализа данных, различных статистических географических регионов (например, новых локальных районов или границы округов) и новые статистические методы. Более того, наборы данных, действующие в одной юрисдикции, могут сильно отличаться от другой. Даже в пределах юрисдикции способы производства и обработки данных одним предприятием могут варьироваться в зависимости от институциональных или личных капризов. Рассмотрим переписи населения. Перепись представляет собой всестороннее обследование местности и ее населения, которое обычно проводится каждые десять лет. Цель состоит в том, чтобы собрать ключевую информацию о том, кто живет в данном районе, и их характеристиках (например, возраст, пол, семейное положение, состав домохозяйства, религия, раса, социальный класс и т. Д.) И аспектах их жизни (их работа, жилье, так далее.). Чтобы можно было измерить изменения, переписи требуют преемственности в отношении задаваемых вопросов и методов их проведения. В то же время, для сбора новых данных, представляющих интерес, которые отражают более широкие изменения в обществе, требуется преобразование, например, новые или измененные вопросы (см. Рисунок 1.2: обратите внимание, что даже когда вопросы сохранялись во время переписей, их формулировка была часто совсем разные). Кроме того, то, как впоследствии проводится перепись, определяется институциональными, политическими и экономическими силами и новыми техническими достижениями: см. Linehan (1991) для истории ирландской переписи 1821–1991 гг. И Lauriault (2012) для анализа канадской переписи. 1871–2011 гг. Более того, построение переписи оспаривается и обсуждается, поскольку заинтересованные стороны конкурируют за включение, изменение или удаление вопросов. В некоторых случаях изменения могут быть довольно радикальными, например, решение в Германии прекратить свою перепись в 1980-х годах (см. Hannah 2011). Как следствие, национальная перепись всегда находится в противоречии между преемственностью и изменением, но, тем не менее, со временем развивается и имеет различные географические регионы. На сегодняшний день, однако, было немного историй и географических географических объединений данных (хотя см. Alder 2002; Desrosières 1988; Hannah 2011; Hewitt 2010; Lauriault 2012; Poovey 1998; Porter 1995).

Philosophically

PARTICULARS FOR INDIVIDUALS ON FORM A AT EACH CENSUS

- = Not collected; Y = precoded; E = entered by Enumerator; y = reply menu given in notes

Y	Year of Census	1841	1851	1861	1871	1881	1891	1901	1911	1926	1936	1945	1951	1958	1961	1966	1971	1979	1981	1988	1991	Y	
M	Month	6	3	4	4	4	4	3	4	4	4	5	4	4	4	4	4	4	4	4	4	4	M
D	Date	6	30	7	2	3	5	31	2	18	26	12	8	8	9	17	18	1	5	13	21	D	
C	Capacity of Form-Number of Persons	20	15	15	12	15	15	15	15	10	10	9	10	11	10	10	10	10	6	8	8	B	C
1	Name and surname	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	1
2	Sex	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	2
3	Relationship to head of household	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	3
4	Age [A]; [B] Years, Months; [C] Date of Birth	A	A	A	A	A	A	A	A	B	B	C	C	C	C	C	C	C	C	C	C	C	4
5	Marital Status [5 shows separated]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Y	Y	Y	Y	Y	5
6	Place of birth (County, City, Country)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	6
7	Absent family members (separate table)	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7
8	Deaths in family members since last Census (separate table)	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8
9	Literacy (Read, write); y = notes give reply menu	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	9
10	Occupation: * pre 1926 no separate "industry"	↑X	↑X	↑X	↑X	↑X	↑X	↑X	↑X	X	X	X	X	X	X	X	X	X	X	X	X	X	10
11	Marriage date "D" or duration "d"; w women only	D	D	D	D	-	-	-	-	dw	dw	-	Dw	-	-	Dw	-	Dw	-	Dw	-	-	11
12	Irish: * as addendum to Literacy questions; y as 9	-	*y	*y	*y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	12
13	Incapacity (D=deaf, dumb, blind; L= Lunatic)	-	D	D	D	DL	DL	DL	DL	-	-	-	-	-	-	-	-	-	-	-	-	-	13
14	Religion	-	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	14
15	Live births to present marriage for married women	-	-	-	-	-	-	-	-	X	X	-	X	-	X	-	X	-	X	-	X	-	15
16	Children of present marriage still living	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	16
17	Orphanhood (for children)	-	-	-	-	-	-	-	-	X	X	X	-	-	-	-	-	-	-	-	-	-	17
18	Dependents = no. <16 for married men and all widowed	-	-	-	-	-	-	-	-	X	X	X	-	-	-	-	-	-	-	-	-	-	18
19	Industry (Question "Employer and Employers' Business")	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X	X	X	X	X	X	X	19
20	Area, val, agric. holdings; H=household; I=individual	-	-	-	-	-	-	-	-	H	H	H	H	-	H	I	H	-	I	-	-	-	20
21	Period of unemployment previous year; 3 causes	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-	21
22	Period of residence for immigrants	-	-	-	-	-	-	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	22
23	Home address of visitors	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-	-	23
24	Employ, status-sep. ques. re employee/own account	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	-	Y	Y	Y	24	
25	Subsidiary occupation	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	25
26	Age at which fulltime education ceased	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	-	X	X	X	26	
27	Types of school etc. attended fulltime: 1- duration	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	-	Y	-	-	-	27
28	Scientific or technological qualifications	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	-	X	X	X	28	
29	Usual residence now	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	-	X	X	X	29	
30	Usual residence one year ago	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	-	X	X	X	30	
31	Means of travel to work, school etc.	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	-	Y	Y	Y	31	
32	Distance to work, school etc.	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	-	X	X	X	32	
33	Became resident within past year: Yes/No	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	33
34	Address of place of work, school or college	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	34	
35	Present status - sep. ques. re labour force status	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	Y	Y	35	
36	Lived outside > - 1 year; when, whence came	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	36	
37	Highest level of education completed - (for no. 27)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	37	
38	Is person farming - principal or subsidiary?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	38

Figure 1.2 Questions concerning individuals on the Irish census 1841–1991

Источник: Воспроизведено из Linehan 1991 г.

Для некоторых на онтологическом уровне данные безвредны. Данные - это просто данные, важные элементы, которые нейтральным и объективным образом извлекаются из мира с учетом технических ограничений. Они «не имеют никакого внутреннего значения, не обязательно представляют какие-либо интерпретации или мнения и не несут ориентировочных характеристик, которые могут раскрыть их важность или актуальность» (Pérez-Montoro и Díaz Nafria 2010). Они преданалитические и предфактуальные. С этой точки зрения сенсор не имеет политики или повестки дня. Он просто измеряет свет, тепло, влажность и т. Д., Производя показания, которые отражают правду о мире, если только они не испорчены техническими сбоями. Другими словами, датчик создает объективный, реалистичный взгляд на мир, показывая вещи такими, какие они есть на самом деле, причем реальность измеряемого объекта не зависит от процесса измерения (Desrosières 1998). В процессах измерения, в которых люди играют центральную роль - в лаборатории или при проведении опроса или интервью - используется форма механической объективности, которая придерживается определенных правил и строгого систематического метода для получения отдаленных, отстраненных, беспристрастных и прозрачных данных, которые являются свободна от предубеждений и предпочтений исследователя и, таким образом, не зависит от местных обычаев, культуры, знаний и контекста (Porter 1995). Таким образом, наука, когда она практикуется должным образом, не имеет политики или скрытых планов, и тогда данные можно принимать за чистую монету. Действительно, термины, обычно используемые для подробного описания того, как обрабатываются данные, предполагают мягкие технические процессы: «собран», «введен», «скомпилирован», «сохранен», «обработан» и «добыт» (Gitelman and Jackson 2013). Политическим является только использование данных. Другими словами, это люди, а не наука, искажают данные и передвигают их в своих целях. Для других такая точка зрения несостоятельна. То, как мы воспринимаем данные, как мы их измеряем и что мы с ними делаем, активно определяет природу данных. Для них данные не существуют до их поколения; они не возникают из ниоткуда.

Данные производятся с помощью методов измерения, абстрагирования и обобщения, которые были задуманы для выполнения задачи и записываются в формы и меры, которые соответствуют стандартам, изобретенным людьми (например, метрическая система). Они представляют собой эпистемологические единицы, имеющие репрезентативную форму, позволяющую проводить эпистемологическую работу, а данные об одних и тех же явлениях можно измерять и записывать различными способами, каждый из которых предоставляет различный набор данных, которые можно анализировать и интерпретировать различными способами (Poovey 1998). То, как генерируются данные, не является неизбежным: протоколы, организационные процессы, шкалы измерений, категории и стандарты разрабатываются, обсуждаются и оспариваются, и существует определенная путаница в генерации данных. Возьмем случай измерения численности населения страны: необходимо принять решение о том, кто учитывается, а кто не учитывается (например, включать посетителей, законных и нелегальных иностранцев, тех, кто сознательно или не принимал участие, и т. Д.) И где их следует вести подсчет (например, где они находятся в ночь переписи или где они обычно проживают); установлены всевозможные правила и процедуры, но по-прежнему существует неопределенность и различия между счетчиками (Porter 1995). Следовательно, онтологическое определение и разграничение данных - это не нейтральный технический процесс, а нормативный, политический и этический процесс, который часто оспаривается и имеет последствия для последующего анализа, интерпретации и действий (Bowker and Star 1999). Тем не менее, однажды созданные онтологии данных работают, чтобы исключить изменчивость и определить, что будет видимым и невидимым в наборе данных, хотя этот процесс конвергенции и стабильности всегда открыт для сопротивления и переделки из-за множества действующих лиц, вещей и процессов. и противоречивость данных, которые нелегко вписываются в систему (Bowker and Star 1999). Более того, как только данные созданы, их можно отсортировать, объединить и разделить на кусочки любым количеством способов по различным категориям. Другими словами, данные не являются независимыми от системы мышления и инструментов, лежащих в основе их производства (Bowker and Star 1999). И такие системы мышления пронизаны философскими предпосылками и убеждениями и практикуются по-разному. В самом деле, как отмечает Боргман (2007: 38), хотя наука пытается изобразить себя универсальной, «их практики являются локальными и сильно различаются». Таким образом, данные не просто представляют реальность мира; они представляют собой конструкции о мире (Desrosières 1998; Poovey 1998). С такой точки зрения «научное знание создается, а не невинно обнаруживается» (Гительман и Джексон 2013: 4). Таким образом, [d] ата трудно отделить от программного обеспечения, оборудования, документации и знаний, необходимых для их использования. Например, если данные производятся инструментом, таким как сеть датчиков, интерпретация этих данных требует понимания инструмента - например, что датчики обнаруживают, при каких условиях, с какой частотой запоминающих устройств, например, изношенная компьютерная лента и поврежденные жесткие диски (Boellstorff 2013). В самом деле, данные не только готовятся, но и открыты для «незапланированного, неожиданного и случайного», трансформируемого парачеловеческим, сложным материальным и возникающим во времени способами, которые не всегда следуют заранее заданному алгоритмическому «рецепту» (Boellstorff 2013). Учитывая социальную и материальную природу данных, мы активно переделываем «наши материальные, технологические, географические, организационные и социальные миры в такую среду, в которой данные могут процветать ... вступая в симбиотические отношения с данными» (Ribes и Джексон 2013: 152). Таким

образом, в то время как «[d] ата рассматривается как нечто существующее снаружи - что-то, что связано с реальным» (Shah 2013, исходный курсив), они более продуктивно понимаются и как компонент реального, и как производитель реального. Данные - это не просто абстракция и репрезентативность, они являются конститутивными, и их создание, анализ и интерпретация имеют последствия. Как утверждают Гительман и Джексон (2013: 2): «если данные каким-то образом подчиняются нам, мы также подчиняемся данным». Данные собираются из мира, но, в свою очередь, работают в мире. Данные не являются и никогда не могут быть благоприятными (Shah 2013). Вместо этого «[d] ата ... необходимо понимать как обрамление и обрамление» (Гительман и Джексон 2013: 5). Другими словами, концептуализация данных - это гораздо больше, чем обычно признают наука и бизнес.

Пирамида данных

Общие сведения.

Многое было написано о пирамиде знаний или данных, информации, знаниях, мудрости, DIKW, иерархии и ее использовании в управлении знаниями, KM. В данной статье мы продолжаем обсуждение, но придерживаемся другой точки зрения. Утверждается, что пирамида знаний является артефактом процессов управления знаниями, а не артефактом реальности. Это другой подход, поскольку несколько авторов (Акофф, Шарма, Бейтс, Фрике) рассматривают пирамиду как естественное выражение отношений между DIKW и логическим развитием для генерации IKW. В некоторых статьях утверждается, что естественные отношения между DIKW на самом деле представляют собой перевернутую пирамиду, а пирамида знаний - это искусственно построенный артефакт, представляющий отношения между DIKW в контексте управления знаниями организации. Это концептуальный документ, который призван способствовать обсуждению и пониманию исследователями природы УЗ и ее взаимосвязи с общими процессами обучения. Ожидается, что с помощью этого обсуждения можно будет по-новому взглянуть на процессы управления знаниями, такие как процессы фильтрации и отбора. Другая цель этого документа - создать модель, которая объединяет KM и Business Intelligence, BI и / или другие «интеллектуальные возможности», такие как Customer Intelligence, CI. Также ожидается, что это обсуждение поможет лучше понять процессы управления знаниями, чтобы можно было разработать более совершенные системы управления знаниями (KMS). Наконец, в этой статье будут даны определения терминов, важных для УЗ, включая определения данных, информации, знаний и интеллекта. Есть надежда, что они могут быть приняты как рабочие, согласованные определения, но признается, что эти термины имеют философский характер и могут обсуждаться сколько угодно долго. Эта дискуссия приветствуется, но не поощряется. В качестве дисциплины нам необходимо разрешить дебаты, но мы также должны объединиться в консенсусный набор рабочих определений. Надеемся, что этот документ будет способствовать достижению консенсуса.

Традиционная пирамида знаний

Ссылки на иерархию знаний можно найти в популярной литературе, но, как правило, Акофф считается первой академической публикацией. На рисунке 1 показана пирамида традиционных знаний, первоначально предложенная

Акоффом. Из рисунка следует, что данные порождают информацию, знания порождают мудрость. Дополнительный вывод заключается в том, что данных больше, чем информации, больше информации, чем знаний, и больше знаний, чем мудрости. Эта модель использовалась в бесчисленных презентациях HICSS, а также в других презентациях УЗ, и утверждается как данность, что это общепринятая модель, показывающая иерархию DIKW. Модель не дает философского определения данных, информации, знаний или мудрости, и это не является целью данной статьи, поскольку уже существует множество источников, которые приводят аргументы в пользу различных определений. Однако цель данной статьи - предложить согласованные рабочие определения, сфокусированные на УЗ. Пирамида традиционных знаний использует следующие обобщенные основные определения: Данные основные, дискретные, объективные факты, такие как кто, что, когда, где, о чем-то. Информация - данные, которые связаны друг с другом через контекст, так что они предоставляют полезную историю, например, связь того, кто, что, когда, где данные для описания конкретного человека в определенное время. Знание - информация, которая была понята культурно таким образом, что она объясняет, как и почему о чем-либо или обеспечивает понимание и понимание чего-то. Мудрость - размещение знаний в рамках или номологической сети, которая позволяет применять знания в различных и не обязательно интуитивных ситуациях. Эти определения подразумевают, что существует взаимосвязь между данными, информацией, знаниями и мудростью, и именно пирамида знаний обеспечивает графическое представление этих взаимосвязей в виде свернутой иерархии данных, ведущих к информации, ведущей к знаниям и, наконец, ведущим к мудрости (в дальнейшем это будет называться процессом создания DIKW). Хьюстон и Хармон описывают это с помощью суммирования: $I = \Sigma (D)$, $K = \Sigma (I) = \Sigma \Sigma (D)$ и $W = \Sigma (K) = \Sigma \Sigma \Sigma (D)$. Пирамида обычно используется в инструкциях по базам данных и концепциям УЗ, чтобы отразить, что «данные», «информация», «знания» и «мудрость» - это разные концепции и все более высокие уровни абстракции. В этой статье утверждается, что пирамида DIKW является неверным представлением взаимосвязей между этими концепциями для целей исследования информационных систем. Сохранение нечеткости этих понятий в дисциплине информационных систем приводит к путанице. Нам нужно лучше определить эти конструкции, чтобы начать устранять некоторые двусмысленности в нашем исследовании.



Figure 1, The Knowledge Pyramid [1]

Противоположность пирамиде традиционных знаний дает Туоми [link]. Туоми [link] утверждает, что данные не являются строительным блоком для информации, знаний и мудрости, поскольку данные не наблюдаются, не собираются и не записываются в вакууме. Скорее, наше понимание мира через нашу мудрость и знания побуждает нас собирать конкретную информацию и данные для поддержки использования наших знаний и мудрости. С этой точки зрения иерархия движется вниз по пирамиде, а не вверх по пирамиде, и данные не существуют как набор несвязанных фактов, поскольку все собранные факты связаны с нашими базовыми знаниями и мудростью. Хотя это усовершенствование пирамиды знаний, оно все же приводит к путанице.

Пересмотренная пирамида знаний

Пересмотренная пирамида знаний пытается поместить иерархию знаний в контекст природного или реального мира. На рисунке 2 показана пересмотренная пирамида. Это показывает, что данные, информация, знания и мудрость существуют в глобальном контексте, то есть люди постоянно собирают и обрабатывают данные в информацию, знания и мудрость. Однако собранные и обработанные данные - это еще не все, что доступно, и они ограничены способностями датчиков обнаруживать, интерпретировать и захватывать данные. Датчики могут быть нашими человеческими чувствами, другими человеческими чувствами или механическими, где механическими датчиками являются все, что не является человеческим чувством, например световой детектор, детектор радиоволн, измеритель давления, введенный в протокол транзакции и т. Д. Они отражают что по мере того, как датчики улучшают нашу способность собирать больше данных, улучшается. Человеческий пример улучшения датчиков - использование лазиков для улучшения зрения, позволяя человеку «видеть» гораздо больше, обогащая захватываемое зрение. Пунктирные стрелки отражают преобразование в информацию, знания и мудрость с использованием процессов понимания, анализа и осмысления. Эти линии являются двунаправленными, что указывает на то, что генерация информации, знаний и / или мудрости может продвигаться вверх по иерархии или обратная связь вниз по иерархии, поскольку пользователь понимает больше из того, что они изучают (это признает встречное мнение Туоми [link]). Линии между слоями отражают социальные сети, используемые для передачи различным пользователям. Социальные сети используются в этом контексте свободно и относятся к любым формальным или неформальным, прямым или косвенным методам, используемым для передачи данных, информации, знаний и мудрости между пользователями. Примеры включают настройки класса, молву, опубликованные статьи, презентации на конференциях, электронную почту и т. Д. В конечном итоге это представление иерархии знаний и общего процесса обучения людей и обществ, в результате которого генерируются и используются несколько больших массивов знаний. . Конечные точки ведут к обучению. Зачем учиться? У обучения есть множество определений, но для этой модели обучение определяется как приобретение DIWK, которое приводит к изменению поведения или ожиданий внутри человека или группы, которые проводят обучение. Почему пирамида знаний перевернута, т.е. почему информации больше, чем данных, больше знаний, чем информации, и больше мудрости, чем знаний? Одна из причин чисто математическая: если информация представляет собой структурирование данных в значимые комбинации, то количество возможных комбинаций для

количества x данных минимально $x!$ подразумевая, что возможно больший объем информации, чем исходный объем данных. Учитывая, что пользователи данных могут иметь разные системы отсчета для обработки данных в разных областях мышления (например, бухгалтеры против маркетологов или инженеры против биологов), вполне вероятно, что объем генерируемой информации превышает исходный объем данных. Этот же аргумент можно использовать для генерации знаний и мудрости, особенно если также учесть, что у пользователей могут быть разные этические, религиозные или культурные системы убеждений, которые могут заставить их интерпретировать информацию, а затем и знания по-разному (например, христиане могут генерировать отличная мудрость от того же знания, чем буддийское, или свободные общества, отличные от тоталитарных обществ). Это согласуется с Нонака [link], Дженнексом [link] и другими, которые утверждают, что все знания зависят от контекста. Примером этого является передача технологии, которая берет DIKW из одной дисциплины и пытается применить ее к другой, например, используя знания, используемые для создания наушников с шумоподавлением, и применяя их для идентификации радиоактивных веществ (оба используют алгоритмы обучения для выявления закономерностей).

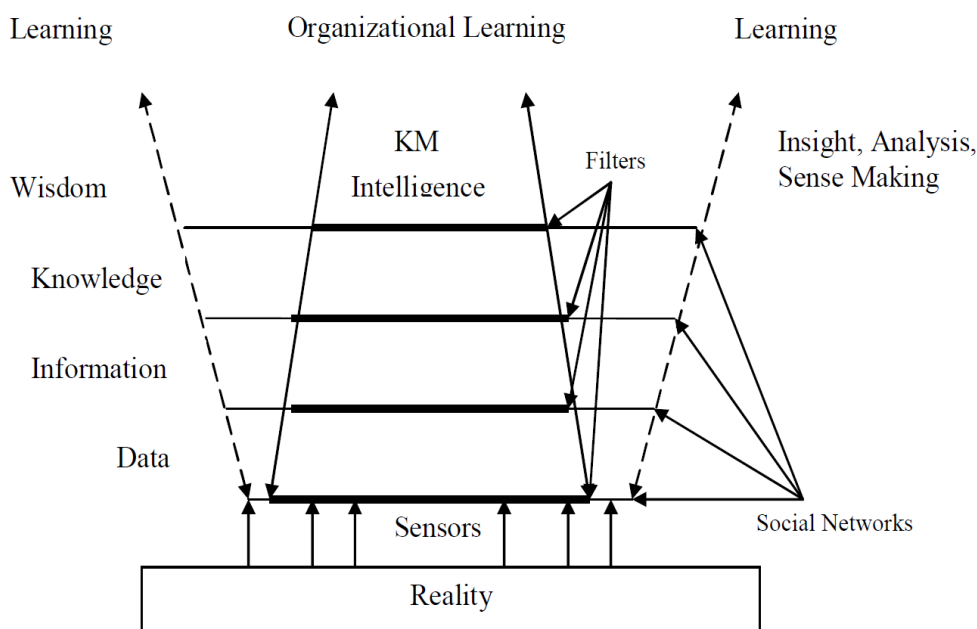


Figure 2, Revised Knowledge Pyramid

Сплошные вертикальные стрелки представляют собой применение управления знаниями в пересмотренной пирамиде знаний. Jennex [link] определил управление знаниями как «практику выборочного применения знаний из предыдущего опыта принятия решений к текущей и будущей деятельности по принятию решений с явной целью повышения эффективности организации». Это означает, что управление знаниями не пытается уловить все знания или мудрость. Скорее, управление знаниями нацелено на конкретные знания и мудрость, необходимые организации для выполнения конкретных задач. Подразумевается, что пирамида знаний УЗ является подмножеством пересмотренной пирамиды знаний, как

показано на рисунке 2. Есть и другие отличия. УЗ, как правило, не использует мудрость, но начинает применять интеллектуальные концепции и различать знания, необходимые для принятия решения, и конкретные практические знания, необходимые для принятия конкретного решения в конкретном контексте. Например, маркетинговые знания необходимы для создания маркетинговых кампаний, но конкретные знания клиентов необходимы для принятия решений о том, как продавать конкретных клиентов. Некоторых пользователей это сбивает с толку, поскольку это означает, что для КМ требуются разные типы знаний. Однако это, как правило, согласуется с концепцией мудрости Туоми [link]. Эта путаница может быть движущей силой для тех, кто занимается бизнес-аналитикой, бизнес-аналитикой или клиентской аналитикой, CI и т. Д., Поскольку они видят необходимость различать общую информацию и знания и конкретную информацию о решениях и знаниях (BI определяется как сбор, анализ, и представление бизнес-информации для принятия решений [link], в то время как CI то же самое только для информации о клиентах [link]). На рисунке 2 это показано с помощью термина «интеллект», а не «мудрость», где «интеллект» относится к очень конкретным практическим знаниям. Термин «разведка» взят из области разведки, той области, которая стремится генерировать практические знания, которые будут использоваться при формулировании стратегий и тактик для достижения конкретной цели, например, победы над противником во время войны или определения того, какие действия следует планировать. Этот термин выбран нелегко, поскольку он вписывается в иерархию DIKW, означающую, что это интерпретация данных, информации и знаний для создания курса действий или принятия конкретных решений. Кроме того, стрелки КМ показаны как двуглавые, поскольку иерархия может двигаться в любом направлении. Фактически, ожидается, что, хотя процесс обучения будет, как правило, идти снизу вверх, что означает, что он начинается с интерпретации данных, организационное обучение и УЗ обычно будут процессом сверху вниз, когда сначала определяется, какие действия или решения должны иметь место или быть приняты, и исходя из этого определить, какой интеллект, затем знания, затем информация, затем данные необходимы для поддержки принятия конкретных действий или принятия конкретных решений. Пирамида знаний об УЗ заканчивается организационным обучением, а не обучением, потому что УЗ, как правило, ориентировано на организацию (признано, что существует личный УЗ, но во многих случаях личный УЗ также определяется потребностями организации). Организационное обучение определяется как поддающееся количественной оценке улучшение деятельности, расширение имеющихся знаний для принятия решений или устойчивое конкурентное преимущество [link]. Хубер, Давенпорт и Кинг [link] считают, что организация узнает, если в результате обработки DIKW ее потенциальное поведение изменится. Это соответствует прекращению существования пересмотренной пирамиды знаний только в организационном контексте. Другое важное отличие заключается в применении фильтров. В то время как общий процесс обучения направлен на то, чтобы донести данные, информацию, знания и мудрость до всех, кто хочет их использовать, КМ этого не делает. Как указывалось ранее, управление знаниями стремится поддерживать принятие конкретных решений и, следовательно, требует определенных данных, информации, знаний и интеллекта. Кроме того, КМ стремится поделиться этим с нужными людьми в нужное время [link]. Это означает, что КМ фильтрует данные, информацию и

знания, чтобы генерировать конкретную полезную информацию, которая предоставляется конкретным, ограниченным пользователям. Фильтры размещаются в социальных сетях, чтобы ограничить доступ, а также отделить и уловить то, что нужно, от того, чего нет. Это довольно новый термин для УЗ, поскольку в литературе по УЗ часто используется термин стратегия УЗ. Позиция данной статьи заключается в том, что фильтры УЗ - это реализация стратегии УЗ. О важности фильтров свидетельствуют Дженнекс и Олфман, которые обнаружили, что стратегия управления знаниями является ключевым фактором успеха. Последнее различие между пирамидой традиционных знаний и пересмотренной пирамидой знаний заключается в удалении вершины. Это было сделано для устранения путаницы, поскольку вершина обычно подразумевает наличие конечной точки, такой как окончательная ключевая мудрость для пирамиды традиционных знаний или единый ключевой элемент данных для пересмотренной пирамиды знаний. Хотя это несколько удовлетворяет точность наблюдения и с каким типом калибровки? (Borgman 2007: 183). Тем не менее, наука часто пытается разделить данные бесплатно и независимо от такого контекстуального понимания, организуя и разделяя данные через базы данных, в которых беспорядок их создания улучшается, а пользователи `` защищены " от необходимости знать, как данные были собраны и систематизированы (Гительман и Джексон, 2013). Таким образом, Рибс и Джексон (2013: 165) утверждают, что научные концепции данных как нейтральных и объективных - это фикции, которые `` предполагают или проецируют мир, в котором данные плавают без своего происхождения, теряя форму, сущность и историю, и тем самым становятся свободными для восприятия. путешествовать по миру как недифференцированная и универсальная валюта ». Напротив, они утверждают, что «данные более надежны, чем это». Следовательно, никакие данные не являются преаналитическими или объективными и независимыми элементами. Как говорят Гительман и Джексон (2013: 2, после Bowker (2005)), «сырые данные - это оксюморон»; «Данные всегда уже готовые »и никогда не являются полностью « сырыми »». «Данные необходимо представить как данные, чтобы они существовали и функционировали как таковые» (Гительман и Джексон 2013: 3). Данные являются как социальными, поскольку они расположены в контексте, так и материальными, поскольку они имеют форму (в виде битов, символов, чисел и т. Д.), Хранящуюся на бумаге, магнитных лентах, жестких дисках и т. Д. (Wilson 2011; Гительман, Джексон 2013). Оба активно формируют структуру данных. Например, большие данные зависят от дискурсивных, политических и экономических дискурсов, которые поддерживают их развертывание (см. Главу 7), а также от баз данных, компьютеров, серверов и сетей, которые обеспечивают их производство, обработку, совместное использование, анализ и хранение. (см. главу 5). Такие носители облегчают гниение данных, неправильное размещение, забывание или удаление данных или их эрозию из-за битовой гнили (поломка

Обсуждение

Пересмотренная пирамида знаний признает Акоффа [link] и Туоми [link] и обнаруживает, что они оба не имеют недостатка. Акофф [link] предполагает, что данные генерируют информацию, которая генерирует знания, которые генерируют мудрость. Туоми [link] считает, что все данные и информация собираются на основе имеющихся знаний и мудрости. Пирамида Акоффа течет вверх, пирамида Туоми течет вниз, и реальность, наблюдаемая и реализованная в пересмотренной

пирамиде знаний, признает, что поток идет в обоих направлениях. Туоми [link] ошибается, полагая, что поток идет только вниз по пирамиде. Фундаментальные исследования часто включают сбор данных о том, чего у нас нет знаний или мудрости. Примером могут служить исследования в области физики в конце 1800-х годов. Ученые не понимали и не знали об атомной структуре, субатомных частицах и радиации. Данные были собраны и наблюдались учеными, которые отметили, что фотопленка затуманивалась при приближении к определенным материалам. Эти данные были собраны, а затем помещены в контекст как информация о том, что указанные конкретные материалы вызвали запотевание. Рассмотрение этой информации привело к формированию теорий, объясняющих, почему это произошло, генерируя потенциальные знания. Анализ этих знаний привел к экспериментам, призванным доказать правильность этого потенциального знания. Этот пример показывает, что и Акофф [link], и Туоми [link] правы, и оба ошибаются. Поток творения идет в обоих направлениях пирамиды. Другой пример, подтверждающий этот аргумент, - использование интеллектуального анализа данных для создания информации путем обнаружения закономерностей в данных. Обычно называемое открытием знаний, это анализ этих новых обнаруженных закономерностей или информации, а также применение контекста, культуры и других знаний, которые создают знания; и позиционирование этих знаний в контексте существующих проблем или конкурентных преимуществ, что создает действенные знания или интеллект. В этом примере преобладающим направлением творческого потока является восходящая пирамида, но также можно утверждать, что предыдущие знания о бизнес-потребностях, процессах и поведении повлияли на интерпретацию обнаруженных закономерностей в знания и интеллект, так что творческий поток фактически идет в обоих направлениях. Последний пример - из инженерии. Инженеры обычно используют все, что могут, при проведении тестирования, даже если они не уверены в назначении данных. Это иллюстрирует мой прошлый опыт разрушения сосуда под высоким давлением. Испытание требовало испытаний на разрыв при давлении в сосуде 2500 фунтов на квадратный дюйм с использованием воды. Использовались стандартные приборы для измерения давления и температуры, добавлен видеонаблюдение. Не было реальной необходимости в сборе видеоданных, за исключением того, что, как считалось, было бы интересно посмотреть. Поскольку вода, несжимаемая жидкость, использовалась в качестве среды под давлением, не ожидалось, что это будет серьезный отказ (насилие в этом случае представлено очень большим и очень быстрым высвобождением энергии), поскольку ожидалось, что судно трещина, и вода вытечет или, возможно, разбрызгивается. Команда испытателей была очень удивлена, когда испытание привело к впечатляющему высвобождению энергии. В основе сюжета лежали видеоданные, и было интересно увидеть, как испытательная установка распалась во время первого испытания, а затем увидеть, что испытательная установка была окружена мешками с песком и другими подпорными стенками для второго испытания. Суть этого примера заключается в том, что существует множество ситуаций, в которых мы собираем данные без реальной причины, кроме как мы можем, и именно последующий анализ этих данных приводит к генерации неожиданной информации, знаний и разведанных.