

41. Data Access

An incredibly vast amount of **data** is now generated by our online lives, including all our uploaded documents, social media interactions, online errands and even GPS data from our cars. At the same time our ability to manage this data is becoming increasingly sophisticated.

to collect data	collecter des données
to store data on computers	stocker des informations dans des ordinateurs
a spreadsheet	tableur
data management	gestion de données

In Latin, '**data**' is the plural of '**datum**' and, historically and in specialized scientific fields, it is also treated as a plural in English, taking a plural verb, as in 'The data were collected and classified'. In modern nonscientific use, however, it is generally not treated as a plural. Instead, it is treated as a mass noun, similar to a word like information, which takes a singular verb. Sentences such as 'Data was collected over a number of years' are now widely accepted in standard English. Data can be defined as facts and statistics collected together for reference or analysis.

data mining	exploration/ exploitation de données
data capture	saisie/ récolte de données
to feed data into a computer	entrer des données dans un ordinateur
data processing	traitement de données
data science	science des données

digital marketing	marketing digital/ numérique
targeted advertising	publicité ciblée
phishing	hameçonnage
a cookie	témoin de connexion/ cookie

a sample survey	enquête par sondage/ échantillonnage
sample data	données d'échantillon
a census	recensement
scalability	évolutivité

Originally the term '**big data**' simply referred to the very large amounts of data being produced in the digital age, including data generated by emails, websites and social networking sites. The term is now also used to refer to specific datasets that are large both in size and complexity, with which new algorithmic techniques are required in order to extract useful information from them.

Although it has become a **buzzword**, it's a valid concept that centers on collecting, storing and using datasets generated both from structured data and unstructured data.

Doug Laney, an analyst, offered a definition for the notion back in 2001. He argued that Big Data has 3 primary components: **volume** (amount of data), **velocity** (speed at which the data is generated), and **variety** (breadth of data that exists).

a search engine	moteur de recherche
to enter a key term into a search engine	entrer un mot clé dans un moteur de recherche
to share information	partager des informations
a query term	terme de recherche
to submit a query	soumettre une requête
to access a site	accéder à un site
an IP address	adresse IP
to surf the web	surfer sur la toile

to rank a website

classer/ noter un site web

The 'Cloud' is a way of referring to a network of interconnected servers housed in data centers across the world which provide **a hub** for storing big data. It's called "cloud" because no one knows exactly where the data is physically stored.

to upload data to the cloud

to manage files

to run applications/ apps

to reboot a system

an outage

to purge data stores

to maximize storage space

data compression

data flow

télécharger des données vers le nuage

gérer des fichiers

exécuter des applications

redémarrer/ réinitialiser un système

panne informatique

vider les stockages de données

maximiser l'espace de stockage

compression de données

flux de données

ASCII (American Standard Code for Information Interchange) was the standard way of encoding text data so as it could be stored in a computer. Each character is allocated a decimal number (its ASCII code). Standard ASCII uses 8 bits (also defined as 1 **byte**) to store each character. It was gradually replaced by different standards among which **Unicode** for instance.

MOBILE AND WEB SERVICES

to tap into data

to detect patterns

to understand patterns

a dashboard

an algorithm

puiser dans les données

déceler des tendances

comprendre des schémas répétitifs

tableau de bord

algorithme

THE INTERNET OF THINGS

automation

electronic sensors

to manage a device remotely

a smart device

wi-fi (wireless fidelity)

a remote server

autonomous vehicle

automatisation

capteurs électroniques

contrôler un appareil à distance

appareil intelligent

wi-fi/ ASFI (accès sans fil à l'internet)

serveur distant

véhicule autonome

wearable technology

a handheld device

wireless technology

connected devices

state-of-the-art

an app/ an application

to access digital content

a remote user

technologie portable

appareil portatif

technologie sans fil

objets connectés

dernier cri

une appli/ application

accéder à des contenus numériques

utilisateur à distance

to sync documents

electronic boarding pass

barcode

a digital wallet

a fitbit wristband

synchroniser des documents

carte d'embarquement électronique

code barre

portefeuille électronique

bracelet fitbit

built-in electronics
to track movements
far-reaching consequences

éléments électroniques intégrés
suivre les déplacements
lourd de conséquences

VOCABULARY EXERCISES

EXERCISE 1: Find out what is hidden behind the following acronyms:

RFID UID IP NFC URL SQL

EXERCISE 2: Fill in the sentences with one of the compounds of data seen in this unit.

1. Because of the rise of the '____ data' era, demand for talent in data _____ is growing rapidly.
2. Companies need to develop a data _____ program that includes standard operating procedures as well as monitoring and quality control measures.
3. How can we talk of understanding and modelling global processes if the _____ data sets used refer almost exclusively to the developed world?
4. Data _____ techniques are used to uncover patterns from the data, which can then be used for various purposes.
5. A data transfer system should allow a smooth and possibly automated data _____ from individual operators to the public authorities and to a publicly accessible website.

EXERCISE 3: Compounds of 'click'.

Choose the right phrase from the list to fill in the blanks. Check their meanings if you don't know them.

pay-per-click advertising clickbot clickstream analysis to generate clicks
clickwrap agreement

1. On a Web site _____ is the process of collecting, analyzing and reporting aggregate data about which pages a website visitor visits -- and in what order.
2. LinkedIn is working on a series of changes to sponsored content that are designed to _____.
3. A _____ is a software robot that clicks on ads to help an attacker conduct click fraud. Some _____ can be purchased, while others are malware that spread as such and are part of larger botnets.
4. If you have ever noticed the advertisements that appear alongside search results on Google and other search engines, you are already familiar with _____.
5. A _____ is an online contract that confirms a user's consent to a company's terms and conditions. This type of contract is a substitute for a real signature and is often used to enforce software licenses or authorize online transactions.

EXERCISE 4: Mind your spelling! Translate the following words into English.

un algorithme
un serveur (informatique)
un code barre
l'automatisation
une adresse IP

CORRECTION OF THE VOCABULARY EXERCISES

EXERCISE 1: Find out what is hidden behind the following abbreviations:

RFID Radio Frequency Identification
UID Unique Identification Number
IP Internet Protocol
NFC Near-Field Communication
URL Uniform Resource Locator

SQL Structured Query Language

More acronyms you may find when dealing with IT:

ARPANET	Advanced Research Projects Agency NETwork
B to B	Business to Business
BPR	Business Process Re-engineering
CMMI	Capability Maturity Model Integration
COBIT	Control Objectives for Information and related Technology
EDI	Electronic Data Interchange
EDM	Engineering Data Management.
ERP	Enterprise Resource Planning
IaaS	Infrastructure As A Service
ISO	International Organization for Standardization
ITIL	Information Technology Infrastructure Library
KM	Knowledge Management
LAN	Local Area Network
OCR	Optical Character Recognition
PaaS	Platform As A Service
PPL	Pay Per Licence
PPU	Pay Per Use
QoS	Quality of Service
RAD	Rapid Application Development
SaaS	Software As A Service
SLA	Service Level Agreement
TCO	Total Cost of Ownership
VPN	Virtual Private Network
VM	Virtual Machine
XML	eXtensible Markup Language

EXERCISE 2: Fill in the sentences with one of the compounds of data seen in this unit.

1. Because of the rise of the big data era, demand for talent in data science is growing rapidly.
2. Companies need to develop a data management program that includes standard operating procedures as well as monitoring and quality control measures.
3. How can we talk of understanding and modelling global processes if the sample data sets used refer almost exclusively to the developed world?
4. Data mining techniques are used to uncover patterns from the data, which can then be used for various purposes.
5. A data transfer system should allow a smooth and possibly automated data flow from individual operators to the public authorities and to a publicly accessible website.

EXERCISE 3: Compounds of 'click'.

Choose the right phrase from the list to fill in the blanks. Check their meanings if you don't know them.

pay-per-click advertising clickbot clickstream analysis to generate clicks
clickwrap agreement

1. On a Web site, clickstream analysis (also called clickstream analytics) is the process of collecting, analyzing and reporting aggregate data about which pages a website visitor visits -- and in what order.
2. LinkedIn is working on a series of changes to sponsored content that are designed to generate more actionable clicks.
3. A clickbot is a software robot that clicks on ads to help an attacker conduct click fraud. Some clickbots can be purchased, while others are malware that spread as such and are part of larger botnets.
4. If you have ever noticed the advertisements that appear alongside search results on Google and other search engines, you are already familiar with pay-per-click advertising.
5. A clickwrap agreement is an online contract that confirms a user's consent to a company's terms and conditions. This type of contract is a substitute for a real signature and is often used to enforce software licenses or authorize online transactions.

EXERCISE 4: Mind your spelling! Translate the following words into English.

un algorithme	an algorithm
un serveur (informatique)	a server
un code barre	a bar code/ barcode
l'automatisation	automation
une adresse IP	IP address

UNDERSTANDING

EXERCISE 1: Read the following text and answer the pertaining questions.

Wearable technology creeps into the workplace

"Physically it was like getting a vaccination; a pain in the hand that was over very quickly," explains Hannes Sjoblad, describing the moment a piercing specialist implanted a microchip under his skin. The NFC chip allows the Swede to swipe into his office, set the alarm system, register loyalty points at nearby retailers and access his gym.

Around 15 to 20 per cent of the 250 people working at the Epicenter co-working space in Stockholm where Sjoblad is "chief disruption officer" have opted into the program, which eliminates the need for key-fobs or electronic entry cards. Since announcing it earlier this year, Sjoblad has been flooded with inquiries from companies looking to adopt a similar system. "Security companies, office operators, real estate companies and even military organisations want to see how this technology works," Sjoblad says.

It's all part of a trend toward using technology - usually wearable devices like smartglasses, wristbands, smartwatches and badges rather than implantable ones - to monitor employee movements and improve productivity. The promise of data-driven efficiency can be alluring to the board room, but it comes at a cost: the employee's right to privacy.

"It started with big data discussions around gathering business insights and not having the human accounted for in that data puzzle. Wearable technology can help make the workforce visible in that," says Chris Brauer, director of innovation at Goldsmiths, University of London. Devices must be paired with a powerful back-end system, however.

"Wearables are not useful on their own," adds Guillaume Roques, head of developer relations EMEA at Salesforce. "They have to be part of the move toward a system of intelligence, which combines big data, the cloud and analytics. Connecting them all together is a big challenge.

A key trend is companies using wearable devices to track employee health - giving staff fitness monitors to keep tabs on their activity levels as part of "wellness" programs. This data can be tied into health insurance policy premiums or other incentive programs to reduce healthcare costs. Activity data collected from FitBit wristbands can be tied into employee incentive programs.

Oil giant BP, for example, has distributed more than 24,500 Fitbit fitness trackers to staff of its North American business in 2015 alone using such an incentive program.

"These programs are often strongly linked to companies negotiating lower rates on collective insurance policies. Underwriters are more trusting of these devices than the self-reporting of employees," explains Brauer.

According to Gartner around 2,000 companies worldwide offered their staff fitness trackers in 2013, rising to 10,000 in 2014. The firm predicts that by 2016 most companies with more than 500 employees - will offer fitness trackers.

In industries with high-risk roles - such as mining and oil and gas - wearables can play a critical role in safety.

Truck drivers at Rio Tinto's coal mines in Hunter Valley, Australia, for example have been using a device called "SmartCap", which looks like a regular baseball cap but has sensors to detect the alertness. It provides an early warning for when a driver is approaching a "microsleep," designed to reduce fatigue-related accidents.

Meanwhile XOEye has developed a set of industrial smartglasses that can capture HD video of complex problems encountered in construction, manufacturing or field services.

Two-way communication means that a remote viewer - be it a manager or a technical specialist - can guide or train the wearer from afar. APX Labs is working with Salesforce on a similar system, called Skylight.

One of the best established applications for body-worn devices in the workplace is to help streamline logistics. For example, UK supermarket chain Tesco gives armbands to staff in a distribution centre in Ireland. These can track the goods being transported across 9.6 miles of shelving, eliminating the need to mark clipboards and giving managers estimated completion times.

Similarly the "pickers" who work in Amazon warehouses wear GPS tags and have a handheld scanner that tells them the most efficient route to take to collect an item for delivery.

While wearable technology can bring huge benefits, they also bring challenges, particularly as devices start to gather more and more personal and biometric data. Consumer-grade gadgets don't always have rigorous encryption and other protections to safeguard personal data, which could leave companies exposed to data leaks or theft.

"We hear about data breaches every week - and it's naive to think that the same won't happen with these miniaturised devices," says technology lawyer Paul Lanois.

There's also the risk of inadvertently creating an oppressive working environment that damages staff morale. "It can be seen as an intrusive surveillance tool rather than something that improves productivity or performance," explains Brauer.

UK-based data science consultancy Profusion found this out the hard way. It ran a study to see what data employers could glean from wearable devices 24 hours per day, seeking to improve the wellbeing of the workforce. The research involved tracking 171 different metrics including heart rate, activity levels, location and other data taken from smartphone applications.

"One participant found the idea of continuously checking his heart rate made him nervous," says Profusion chief executive Mike Weston. Another feared that her line manager would be keeping track of her self-reported stress levels. "She felt uncomfortable being under the microscope."

Weston says this shows how careful companies need to be when implementing wearable technology program. "If there's a creepiness factor around what you're doing, you probably shouldn't be doing it."

From *Bloomberg Business*, by Olivia Solon on August 7 2015

1. Explain what wearable technology is.
2. Give 3 examples of wearable technology from the text and say which companies use them.
3. Mention 4 advantages of wearable technology in the workplace.
4. Find 4 drawbacks of wearable technology in the workplace.

EXERCISE 2: Read the following text and answer the pertaining questions.

Stitch Fix's CEO on Selling Personal Style to the Mass Market

At Stitch Fix our business model is simple: We send you clothing and accessories we think you'll like; you keep the items you want and send the others back. We leverage data science to deliver

personalization at scale, transcending traditional brick-and-mortar and e-commerce retail experiences. Customers enjoy having an expert stylist do the shopping for them and appreciate the convenience and simplicity of the service.

Of course, making something seem simple and convenient to consumers while working profitably and at scale is complex. It's even more complex in the fashion retail industry, which is crowded, fickle, and rapidly changing. Other apparel retailers attempt to differentiate themselves through the lowest price or the fastest shipping; we differentiate ourselves through personalization. Each Fix shipment, as we call it, is a box containing five clothing and accessory items we've chosen just for you. Those choices are based on information you and millions of others have given us—first in an extensive questionnaire you fill out when you sign up, and then in feedback you provide after each shipment. Stitch Fix sold \$730 million worth of clothing in 2016 and \$977 million worth in 2017. One hundred percent of our revenue results directly from our recommendations, which are the core of our business. We have more than 2 million active clients in the United States, and we carry more than 700 brands. We're not upselling you belts that match that blouse you just added to your cart, or touting a certain brand because you've bought it before, or using browsing patterns to intuit that you might be shopping for a little black dress—all activities that have low conversion rates. Instead we make unique and personal selections by combining data and machine learning with expert human judgment.

Data science isn't woven into our culture; it is our culture. We started with it at the heart of the business, rather than adding it to a traditional organizational structure, and built the company's algorithms around our clients and their needs. We employ more than 80 data scientists, the majority of whom have PhDs in quantitative fields such as math, neuroscience, statistics, and astrophysics. Data science reports directly to me, and Stitch Fix wouldn't exist without data science. It's that simple.

Don't forget the people.

The analytical part of me loves our algorithmic approach. But shopping is inherently a personal and human activity. That's why we insist on combining data with a human stylist who can alter or override the product assortment our styling algorithm has delivered. Our stylists come from a range of design and retail backgrounds, but they all have an appreciation for the data and feel love and empathy for our clients. Humans are much better than machines at some things—and they are likely to stay that way for a long time.

For example, when a client writes in with a very specific request, such as "I need a dress for an outdoor wedding in July," our stylists immediately know what dress options might work for that event. In addition, our clients often share intimate details of a pregnancy, a major weight loss, or a new job opportunity—all occasions whose importance a machine can't fully understand. But our stylists know exactly how special such life moments are and can go above and beyond to curate the right look (...). That creates incredible brand loyalty.

It's simple: A good person plus a good algorithm is far superior to the best person or the best algorithm alone. We aren't pitting people and data against each other. We need them to work together. We're not training machines to behave like humans, and we're certainly not training humans to behave like machines. And we all need to acknowledge that we're fallible—the stylist, the data scientist, me. We're all wrong sometimes—even the algorithm. The important thing is that we keep learning from that.

From *HBR's 10 Must Reads on AI, Analytics, and the New Machine Age* by Katrina Lake, May-June 2018

1. What is Stitch Fix's line of work?
2. What is Stitch Fix's comparative advantage/ competitive edge?
3. How can Stitch Fix personalize each box?

GOING FURTHER

1. In text 2, Katrina Lake writes: "A good person plus a good algorithm is far superior to the best person or the best algorithm alone. We aren't pitting people and data against each other. We need them to work together." Can you think of other examples to illustrate that collaboration?

2. Have you heard of any other companies that developed the same idea as Stitch Fix?

CORRECTION of the UNDERSTANDING part

EXERCISE 1: Read the following text and answer the pertaining questions.

1. Explain what wearable technology is.

Wearable technology is a category of electronic devices that can be worn as accessories, embedded in clothing, implanted in the user's body, or even tattooed on the skin. The devices are hands-free gadgets with practical uses, powered by microprocessors and enhanced with the ability to send and receive data via the Internet.

The rapid adoption of such devices has placed wearable technology at the forefront of the Internet of things (IoT).

2. Give 3 examples of wearable technology from the text and say which companies use them.

- a microchip under one's skin which allows people working at the Epicenter co-working space in Stockholm to swipe into their offices, set the alarm system, register loyalty points at nearby retailers and access their gyms.
- Oil giant BP distributed Fitbit fitness trackers to staff of its North American business in 2015
- Truck drivers at Rio Tinto's coal mines in Hunter Valley, Australia, have been using a device called "SmartCap", which has sensors to detect the alertness. It provides an early warning for when a driver is approaching a "microsleep".
- the pickers who work in Amazon warehouses wear GPS tags and have a handheld scanner that tells them the most efficient route to take to collect an item for delivery.

3. Mention 4 advantages of wearable technology in the workplace.

The different examples illustrate how wearable technology can ensure better security by reducing the risk of accident, guarantee better health by encouraging people to exercise more, streamline logistics and make companies save money.

4. Find 4 drawbacks of wearable technology in the workplace.

Most of these gadgets lack protections to safeguard personal data, which could leave companies exposed to data leaks or theft.

There's also the risk of putting pressure on the staff that may feel they are constantly monitored.

The gathered information may be misused: for instance health-insurance premiums may increase for those people who are considered as leading an unhealthy life.

Using wearables to ease the work of employees may actually make them dependent on those technologies which may lead to accidents.

EXERCISE 2: Read the following text and answer the pertaining questions.

1. What is Stitch Fix's line of work?

Stitch Fix is a company where an expert stylist along with an algorithm do the shopping for the customers who subscribed to their service and send them clothing and accessories they think they'll like; customers keep the items they want and send the others back.

2. What is Stitch Fix's comparative advantage/ competitive edge?

Stitch Fix is an apparel retailer. But while others try to differentiate themselves through the lowest price or the fastest shipping, they do it through personalization/ customization. Indeed, each shipment is a box containing five clothing and accessory items tailored to each customer's tastes.

3. How can Stitch Fix personalize each box?

Stitch Fix's choices are based on the information they collected from their subscribers in an extensive questionnaire they fill out when they sign up, and then in feedback they provide after each shipment. Stitch Fix combine data and machine learning with expert human judgment: the human stylist can alter or override the product assortment the styling algorithm has delivered.

GOING FURTHER

1. In text 2, Katrina Lake writes: “A good person plus a good algorithm is far superior to the best person or the best algorithm alone. We aren’t pitting people and data against each other. We need them to work together.” Can you think of other examples to illustrate that collaboration?

Personal answers are expected here. Maybe in the medical field with AI being able to diagnose diseases before doctors can but having humans doing any procedures or psychological follow-up. Or the necessary combined use of human brains and GPS to avoid accidents.

2. Have you heard of any other companies that developed the same idea as Stitch Fix?

Have a look at: <https://lookiero.co.uk>