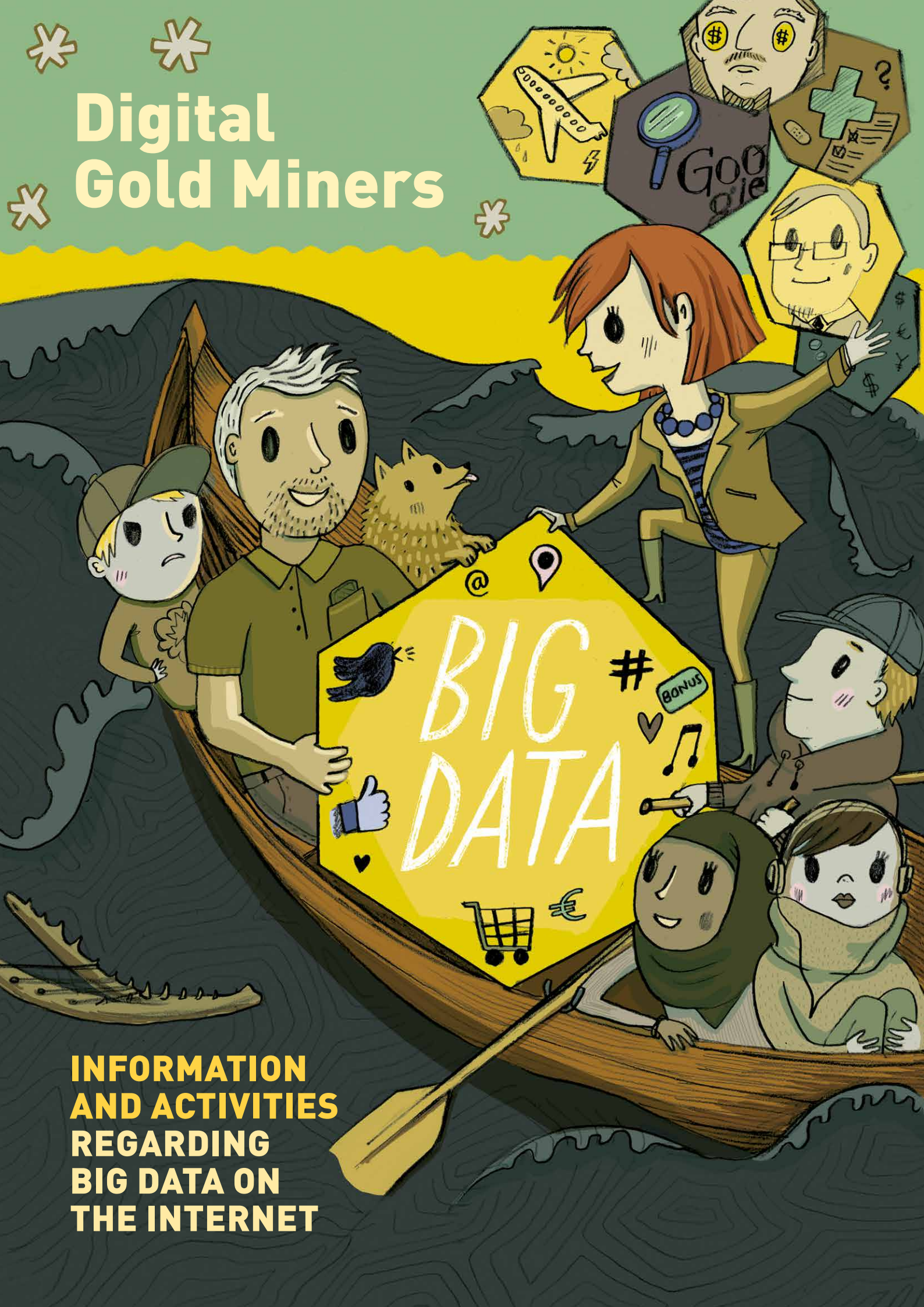
Digital Gold Miners



INFORMATION AND ACTIVITIES REGARDING BIG DATA ON THE INTERNET

Digital Gold Miners

Media Literacy Week

Contents

Introduction 3

What, who and how – three questions about big data 4

What is big data? 4

Who collects data? 5

How is data collected? 5

Mountain of information into a digital gold mine – how is big data utilised? 8

Valuable data 8

Doing business quickly and easily 9

Targeted advertising 9

Dangerous waters – is privacy under threat? 11

Big data and data protection 11

Hands off my identity! 12

Controlling your own data 13

Become active online 13

Application tips 14

KANSALLINEN AUDIOVISUAALINEN INSTITUUTTI
NATIONELLA AUDIOVISUELLA INSTITUTET
NATIONAL AUDIOVISUAL INSTITUTE

Co-funded by the European Union

This material has been produced by the National Audiovisual Institute, the Finnish Federation for Communications and Teleinformatics (FiCom), the Finnish Competition and Consumer Authority, Lastenlinkit.fi, the Finnish Lions Club, the Office of the Data Protection Ombudsman, the National Development Centre for Online Youth Work (Verke) and the Finnish Communications Regulatory Authority.

Illustration: Emmi Jormalainen

Introduction

How do Facebook ads know where I'd like to travel? How can individual search words help in predicting a flu epidemic? How can the information I share online help millions of others?

The information given by and collected from individual Internet users is cumulated into an enormous mass called *big data*. Analysing the information scattered across the Internet contains great possibilities, and it also raises many questions. Big data is like a digital gold mine of our time. The *Digital Gold Miners* guidebook contains basic information and assignments regarding big data. It focuses particularly on the collection of Internet users' information as well as the intended use of the collected information.

The material can be used to give thought to your own role as a provider and user of information. The goal is to support active and responsible citizenship and offer tools for the management of information and life in the digital age.

Instructions for the user of the material

This material is suitable for supporting the media education of young people of 13 years or older, for example, by schools, libraries, youth work and leisure time media education.

The guidebook consists of four different sections that shed light on big data as a phenomenon from an Internet user's point of view.

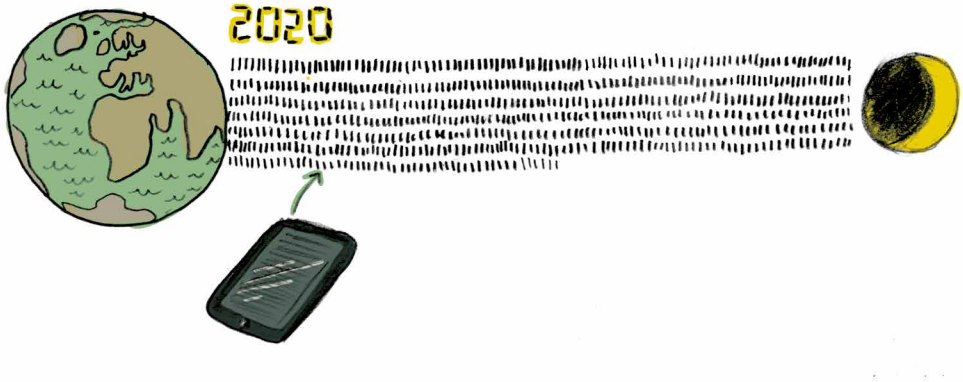
- What, who and how – three questions about big data
- Mountain of information into a digital gold mine – how is data utilised?
- Dangerous waters – is privacy under threat?
- Controlling your own data

At the end of each section, there are operational ideas aimed at pupils or students. The ideas can either be used as such or applied to suit your group. Some of the assignments require a computer and Internet connection.

What, who and how – three questions about big data

WHAT IS BIG DATA?

Big data refers to the collection, storing and analysis of very large, unorganised masses of information in different forms and coming from different sources by utilising the Internet and modern technology. Big data is characterised by information being generated and spread very rapidly. Our digital universe – all the data we create and share – is growing constantly by 40 percent a year. International Data Corporation (IDC) estimates that the amount of data accumulated by 2020 will be so large that storing it would require such an enormous number of tablet computers that, when stacked on top of each other, they would reach the Moon more than 6.5 times.



The information masses of big data are accumulated from sources such as GPS data, information searches, social media contents and data collected by various detectors and sensors. Different types of data from different sources, such as photographs, location information and social media discussions, can be combined in many different ways to be analysed and refined. Big data has no value in and of itself, but the data can be refined to suit a wide variety of uses. This is why big data can be considered a kind of digital gold mine.

WHO COLLECTS DATA?

The enormous mass of information is growing often unbeknownst to the Internet user. Every tap of a key, loading of a page and reading of online news can increase our digital footprint. An increasing number of everyday items is being connected to the Internet. Thus, the information forms a massive pile, but it is actively collected as well.

Information is collected by, for example:

- social media service providers (e.g. user profiles)
- administrators of search engines and online services (search histories, information logs)
- businesses (transactions, customer registries, online behaviour of customers)
- public administration (e.g. health care, authority registries, traffic design and monitoring)
- navigation service providers (e.g. time and location information, positioning of people and vehicles)

HOW IS DATA COLLECTED?

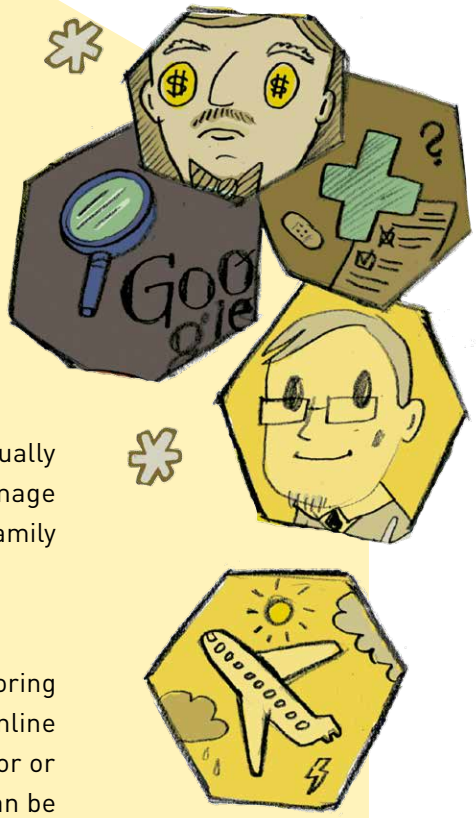
Collecting user information online and via mobile applications is carried out through a variety of processes.

Information is obtained in situations such as when people create and use profiles in social media services or provide customer information on a business' website when ordering products.

A search engine will remember your previous searches by retaining your search history, and an online TV service is able to recommend an interesting movie or TV series based on your viewing history. The functionality of travel instructions and navigation services is based both on location information from individual users and choices made by several users while travelling.

A device connected to the Internet can be identified by IP address. Cookies enable websites or applications, along with their administrators, to collect information about a person's online behaviour. Technical information can also be collected directly from a browser, for example: the more the user has modified the browser by installing add-ons, the better the user can be identified.

INTERNET GLOSSARY



Web browser (often called just a browser) is a computer program that enables its user to view and submit text, images and other information found on www pages. Browsers include Firefox, Chrome, Internet Explorer and Safari.

Application refers to a program installed on a computer or smartphone, a game or even an add-on installed on a web browser. Applications are usually tools for a limited purpose, such as file transfer, image sharing, hiding your IP address or communication with family and friends.

Administrator is a person or a team monitoring the functionality and users of an online service. Most online services that can be registered to have an administrator or at least an owner of the domain. The administration can be contacted with questions regarding the use of the online service. You can check the contact information of the owners of Finnish online services ending in .fi at domain.fi, in case the service does not provide contact information of the administration.

WHAT ARE "COOKIES"?

Websites may use cookies. A cookie is a file sent to and stored on the user's computer.

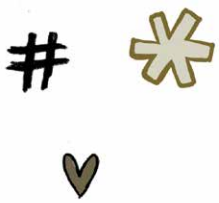
Cookies make it possible to store an Internet user's information temporarily while the web browser is open, or for a longer time, whereupon items such as passwords or contents of an online shopping cart will remain even after the website and the browser have been closed.

Cookies often make using the Internet easier, as the user information has already been stored on the website's memory. On the other hand, cookies also make it possible to track Internet users, as a cookie can reveal, among other things, what website the user browsed previously and for how long. The administrator of a website can also provide the user's information to other parties, such as advertisers or businesses specialising in analysing.

The browser's settings allow cookies to be disabled or limited. However, some services, such as online banks, require cookies in order to function properly. For more about disabling cookies, see your browser's settings and instructions.

Ideas for activities 1.

First, make your own list detailing what you use the Internet for. Examine the list and think about all the ways the devices, websites and applications you use gather information about you. Examine the answers in a group and think about what kind of information may have been collected from each group member.



Tip: While analysing the assignment, the participants can look into an Australian blogger's listing "What does Facebook know about you?" which lists things that Facebook can deduce about its users.

It is the year 2030. Think about what services and items could collect information about you in the future. Choose a life situation or event in which you think the information collected about you would benefit you and bring additional value or comfort into your life. Draw a picture or tell a story about yourself at that moment.

Mountain of information into a digital gold mine – how is big data utilised?

VALUABLE DATA

Utilising enormous masses of information requires sophisticated analysis and combining single items of data with other available information. The accumulated data has direct monetary value as well, as businesses specialising in analysing big data sell Internet users' information to other businesses and operators.

In addition to businesses and advertisers, the refined information can also benefit individual Internet users or online service developers, for example. Furthermore, analysing large amounts of data may contribute to scientific research and help public administrations and businesses to improve matters such as the fluency and safety of transport.

Online search engines can provide interesting and necessary information about national and global phenomena. Recording people's different search results and comparing their frequency can provide information about matters such as epidemics, consumer habits or economic fluctuations. The search results mainly indicate real-time events and phenomena, but they can sometimes be used to predict the future as well.

A search engine may predict an influenza wave

When someone contracts a high fever and sore throat, there will be online searches regarding these symptoms. An influenza site maintained by Google collects this information and thus predicts budding disease waves before they become actual large epidemics.

So far, health care authorities have often not received reliable information about epidemic outbreaks until several days later, when people have sought medical treatment for the symptoms. The real-time information of search engines is hoped to become a tool for detecting epidemics at an earlier stage and preventing them from spreading. Epidemics predicted from search results and actual influenza waves have been reported to match well, but as of yet, the prediction model based on search results cannot be considered completely reliable.

DOING BUSINESS QUICKLY AND EASILY

For the common Internet user, the refining of big data can be observed as rapidly developing services and increased user-friendliness. By monitoring Internet users' movements and behaviour, service providers gain awareness of issues or favourite features regarding their websites, for example.

Purchasing, banking and agency services, as well as many other forms of business, have become partially or even completely Internet-based. For example, real-time travel instructions, a saved online shopping cart and the retention of user information are possible due to data being collected, stored and utilised in service development.

TARGETED ADVERTISING

Online forms are one example or information being collected for targeted advertising. While the website is helping you find a product and outlining related alternatives you might like, the administrators receive information about you and your interests.

Laura wants to go somewhere warm

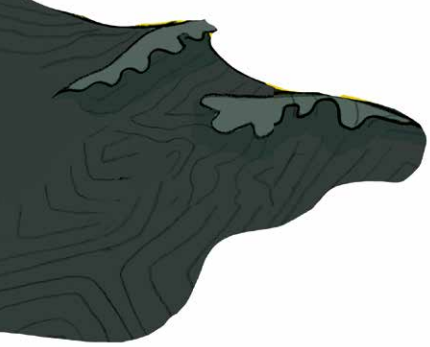
The weather outside is miserable and I still have a few vacation days left. I feel like going somewhere warm. I'm looking for last minute flights on zeebookers.com. Should I go to Spain, I wonder if it's already warm in Greece... I'm filling in dates and searching for nice destinations. The website also offers affordable hotels and I'm ticking the search form to determine what features I want my accommodation to have. Should I book this holiday right away?

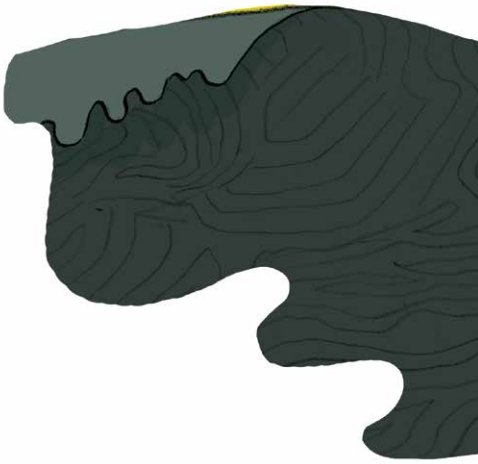
The next day, the sidebar on Facebook features an ad by flyrealcheap.com regarding travel packages to Greece, Spain and Bulgaria. I wonder how Facebook knew I need a holiday...

The administrator of a website may release and sell collected information to so-called third parties, such as advertisers, online services or a business conducting analyses and profiling Internet users. One visit to a website and ticking a few boxes is not enough to provide the advertiser with reliable information about your consumer habits, but the more websites you visit and the more searches you do, the more the advertisers are able to target you.

Ideas for activities 2.

Form groups of four, for example. One pair of the group uses a search engine to find texts and news stories with the search words "big data benefits", while the other pair uses the words "big data threat". Examine the texts for about half an hour and make notes about the threats and benefits you find. After that, return to the groups, compare your results and discuss which seem more significant, the threats or the benefits. You may also arrange a debate exercise, in which one group defends and one opposes the big data phenomenon.





Dangerous waters – is privacy under threat?

From a common Internet user's perspective, big data raises some questions especially regarding privacy protection and misuse of the information. Do you ever read the terms of use of websites or online applications? Or have you given thought to whether someone could misuse your information or use it for malicious purposes?

BIG DATA AND DATA PROTECTION

Giving your personal information online and having your personal information collected by service providers have quickly become common. It is easy to miss how widely an online service collects information with cookies, for example. The big data accumulated on the Internet tells a lot about Internet users and makes it possible to form increasingly detailed profiles about people. In fact, big data has been feared to undermine people's privacy and data protection. Many countries have legislation governing the collection, use and releasing of personal information.

Online services often release a privacy statement and terms of use. Familiarise yourself with them before giving your personal information and using the service so that you know what you are committing to and giving your information for. You will also find out who it is you are doing business with and in which country the service provider resides. Legislation governing Internet users' rights varies from country to country and region to region.

Complicated terms of use?

Terms of use and privacy protection clauses can sometimes feel long or tiresome, but you should pay close attention to them especially on the websites you use frequently. The fundamental principle of websites and online applications is often that by using the service, you accept the related terms and conditions, as the example below illustrates.

"By using Website, you commit to complying with the currently valid terms of use. As the owner of Website may make changes to the terms of use from time to time, we advise that you regularly check these terms. If you decline the terms of use, we kindly ask you not to use Website."

Online services also change their terms of use relatively often, so you should make sure that your privacy settings are up to date. It is also good to remember that all personal information posted online may one day become public.

HANDS OFF MY IDENTITY!

Keeping your personal information on online databases and having them displayed on social media, for example, increase the risk of various kinds of misuse. One form of online scamming that has become common in recent years is online identity thefts. The scammer may, for example, create a fake profile of another Internet user on social media or order products from an online store with another person's information.

The more detailed information you give about yourself, the easier it is to carry out an identity theft. Your phone number and birth date alone are enough to enable a scammer to do many kinds of harm. So familiarise yourself with the privacy settings of the online services you use and check how detailed information you publish about yourself.

Ideas for activities 3.

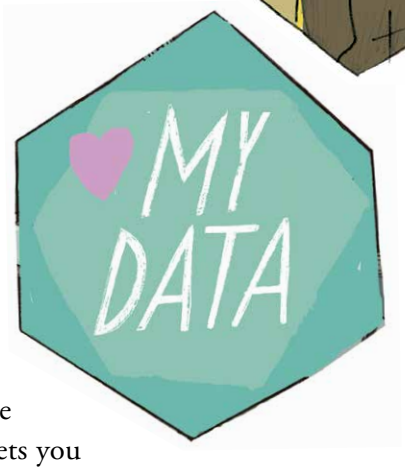
Are you familiar with the terms of use of the social media service you most frequently use? Examine the terms of use of various services together. Did you find any surprises? Do you find the terms clear or difficult to understand?

Find the data privacy statement of an online magazine, a gaming website or social media service. Research and think about the following together: What information does the web page collect from its readers? What information does the web page collect without the reader actively giving his or her information? What is the information needed for? What information are you willing to give to the website? Discuss: how careful are you with your privacy?

Controlling your own data

You cannot stop the accumulation and active collecting of data unless you give up modern technology entirely. However, you do have control over the information collected from you and your actions.

You can make your data profitable or offer it voluntarily to be used for scientific research, for example.



BECOME ACTIVE ONLINE

One way of utilising big data is the flexible pricing of products based on demand, time and consumer behaviour. Flexible, i.e. dynamic, pricing is utilised in the sales of concert and flight tickets, for example. When purchasing tickets, the prices may vary significantly depending on the date and even the time. When you purchase tickets, make use of online watches that inform you when the tickets you want are on a special sale.

Releasing your user information can also be considered voluntary and unpaid work for businesses. If businesses generate revenue with your user information, should you be entitled to a share? Some services may pay you for access to your user information and when selling the information anonymously to third parties. However, the revenue shares paid to individual users are currently no greater than a few dozen euros per year.

In addition to sharing information about their online behaviour, private persons may share the processing power of their computers to contribute to science, for example. Stanford University is seeking treatments to cure conditions such as degenerative memory diseases, and the University of Oxford is seeking a solution to climate change.

http://folding.stanford.edu/

http://www.climateprediction.net/

You may remove the cookies monitoring your online behaviour with these instructions. If you do not want your computer to store cookies, you can remove them regularly or use a private browsing mode.

Chrome: https://support.google.com/chrome/answer/95647?hl=en-GB&vid=1-635760859144753372-1423426481

Firefox: https://support.mozilla.org/en-US/kb/delete-cookies-remove-info-websites-stored

Explorer: http://windows.microsoft.com/en-GB/internet-explorer/delete-manage-cookies#ie=ie-11

APPLICATION TIPS

You can activate a **"Do Not Track" feature** on your browser, which sends websites a request that they stop monitoring your actions. However, as of yet, most websites do not acknowledge the browser's request in any way and collect visitors' information regardless.

The **incognito or private browsing modes** of browsers do not mean anonymous browsing on the Internet. The websites you visit can still see your IP address and your browser's individual features. However, the incognito mode prevents your browsing history from being stored on your computer, as well as web pages' cookies. This can prove practical when you are looking for a surprise present on the family computer, but it does not protect your privacy significantly when online.

There are add-ons for some browsers (such as Firefox and Chrome) that can be used to monitor how your information is spreading online or limit it. Add-ons such as **Lightbeam and Collusion** show the third parties in the background of a website. **Adblock Plus**, on the other hand, prevents advertisements from showing up on the browser and videos.

There are also online classification systems for websites and online stores based on users' own reviews (such as Web of Trust). Applications such as this do not affect the information collected from users, but they can help individual online store customers assess whether they should give their personal information to a certain store.

Ideas for activities 4.

Search for information online both with the incognito mode and the default settings. Does the search yield similar results from Google, for example? Does YouTube recommend the same videos with both settings? Examine the results in a group or in pairs.