http://dx.doi.org/10.15576/GLL/2019.3.53

GLL
Geomatics, Landmanagement and Landscape No. 3 • 2019, 53–66

GEOINFORMATION IN THE INVISIBLE RESOURCES OF THE INTERNET

Karol Król

Summary

The article presents an introduction to deliberations on the type and scope of information, including geoinformation, made available in the Deep Web and Dark Web. It has been shown that geoinformation is present in the Surface Web, although only a small fragment is available in the search results. In the search indexes, the main pages of specialised geoinformation portals, Internet maps and databases are typically made available. Most geoinformation is available on the Deep Web, which requires the use of specialized search engines or exploration of thematic maps. It was also pointed out that geoinformation contradicts the assumptions of the Dark Web. The Tor network, which is the basis of the Dark Web, was created to ensure anonymity and prevent location in space.

Keywords

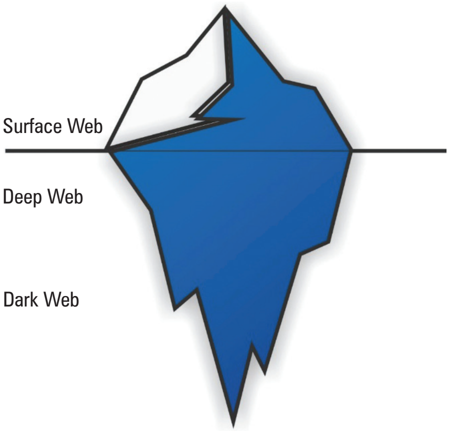
Surface Web • Deep Web • Dark Web • geoinformation • indexable Web

1. Introduction

Not all Internet resources are available at your fingertips. To reach many of them, we need to go deeper. It is often necessary to know where to look, and how to search [Ehney and Shorter 2016]. The visibility of an online resource depends on whether it can be accessed via search tools. It is worth noting, however, that each search engine creates its own invisible Internet, depending on the indexing algorithm that it applies [Devine and Egger-Sider 2009]. Invisible resources are included in the category of "unconscious resources". This peculiar "unconscious" quality is a perverse definition, which is meant to draw attention to the lack of knowledge, on the part of the users, about resources that are not available through widely used search engines. This kind of resources also includes the growing body of geo-data resources. The purpose of the present work is to analyse selected geo-data sources available in the deep resources of the Polish Internet.

The Surface Web, which Internet users use every day, consists of resources that the search engines can find, and subsequently offer in response to a user's query. However, the resources available via the search engines are just the tip of the iceberg (Fig. 1) –

a traditional search engine indexes only a small portion of all data. Other content is immersed in the so-called Deep Web.



Source: Author's own study

Fig. 1. Schematic depiction of the multi-layered accessibility of Internet resources

2. Surface Web

The most common and open area of the web that is publicly available can be called a "Surface Web" or "Web's surface". When a user searches for specific information on the Internet, he usually uses one search engine or another. Search results constitute the response to a query consisting of specific Keywords. Sites (resources) are included in the search results thanks to the indexing crawlers (bots) that search the Internet's resources and record the links they found. In the indexing process, each resource is categorized. This means that if a website does not have links leading thereto, which are placed on another website, then the search engine will not find it [Sherman and Price 2003]. The Visible Web consists of websites that are accessible to the general public, and which are usually indexed by search engines. This area includes all sites that can be found using search engines such as Google, Yahoo or Bing.

3. Deep Web

The first mention of the "Deep Web" or "invisible web" appeared already 20 years ago, when the phenomenon was relatively new, little known and still surprising for many Internet users. Since then, the indexing bots have overcome many technical barriers that had formerly prevented them from finding "hidden" Internet resources [UC Berkeley 2010]. Invisible Internet is a fraction of structured or partially structured contents [Pederson 2013]. Invisible Web, also known as "undernet" or "hidden Web" (Table 1), consists of resources that are difficult or impossible to find through typical search engines and catalogues. Despite the fact that these resources are beyond the reach of traditional search tools, they constitute an integral, natural part of the web [Ehney and Shorter 2016].

Table 1. Selected terms linked to Deep Web and Dark Web

Term	Source
Regular Web, clearnet	Avarikioti et al. 2018
Clear Web	Gehl 2016, Gollnick and Wilson 2016
Deep Web content	Symanovich 2019
Darknet, Dark Web	Chertoff and Simon 2015, Gehl 2016, Avarikioti et al. 2018
Dark websites, Dark Web	Maddox et al. 2015, Gehl 2016
Deep Web sites	He et al. 2007
Deep Web	Bergman 2001, Bradbury 2014, Dalins et al. 2018
Deep Net, Invisible Web	Lewandowski and Mayr 2006, He et al. 2007, Chertoff and Simon 2015, Weimann 2016
Hidden Web, Indexable Web, Indexed Web	Raghavan and Garcia-Molina 2000, Bergman 2001, He et al. 2007
Hidden value on the Web	Bergman 2001
Surface Web	Bergman 2001, He et al. 2007
Undernet	Chertoff and Simon 2015, Gehl 2016
The dark side of the Internet	Kim et al. 2011
Dark-web places	Maddox et al. 2015
Hidden services	Hyperion Gray 2019

Source: Author's own study

The Internet can be seen as a huge repository of various kinds of information, or an encyclopaedia with entries on every topic [Lin and Chen 2002]. However, standard search engines gain access to a small fraction of all information available on the web. Deep Web, in a sense, is the content of databases and other Internet services, which for various reasons are not indexed by typical search engines. Search engine robots (so-called "spiders" or "crawlers") do not collect information generated in real time, which are ephemeral in nature, such as data displayed on thematic layers of online maps, search results for specialized search engines, stock exchange quotes, current weather at a specific location, or flight schedules of airlines. Search engine programs are not able to fill in forms required to generate specific information. All such resources form part of the Deep (Invisible) Web [Avarikioti et al. 2018].

While the Surface Web connects billions of static HTML pages, it is believed that much more information is "hidden" in the Deep Web, whereas access to that information requires database queries. Such information is usually not available at a static URL – instead, it is "folded" into websites as a form of response to queries sent via the database query interface, e.g. created using CGI (Common Gateway Interface), HTML forms, or JavaScript [He et al. 2007]. Databases cannot be searched or indexed by traditional search engines, because they do not have a static URL that could be included in the search index. Since search engines cannot currently effectively index databases, it is believed that such data is "invisible" and therefore it remains largely "hidden". Nevertheless, these databases constitute a goldmine of information, as many of them contain detailed and specific data that is not available in other parts of the web [Lin and Chen 2002]. In addition, a large part of the network is hidden from indexing crawlers: websites can be excluded from indexing using a robots.txt file. Many resources are available after logging in. All these technical limitations exclude huge amounts of information from the search results [Devine and Egger-Sider 2004]. These phenomena gave rise to characteristic, metaphorical phrases defining the availability of information on the "surface of the Web" or in the "visible web" versus the "deeply hidden" web, the "invisible web", or the web underneath. In various articles, the authors write about "drilling" deep in the database, "collecting", "taming" or "extracting" information [Devine and Egger-Sider 2004]. These metaphors refer to mining and resource extraction. Others, in turn, refer to the depths of the sea. Michael K. Bergman [2001] compared searching the Internet to dragging nets across the ocean: many of the content can be caught online, but there is also much of it that is only available at greater depth.

The content of geoinformation databases or the contents of the public library catalogue is not available from the level of a web browser – to find a book in a digital rental library, it is not enough to enter its title in the Google search bar. The search engine will not return the result associated with a specific library. Such kinds of information are found in the Deep Web. In such cases, it is necessary to use the industry search engine, available on the library's website, which searches library databases. Almost every time a search takes place within a given site, access to "deep web content" is obtained [Symanovich 2019].

Many users believe that the Invisible Web is not worth their attention because it is full of spam and various ephemeral content. Conversely, it is worth considering exploration, if only because of the multitude and diversity of resources it contains. These resources are usually of high quality, they are unique, specialized in nature, and they are collected in clearly defined thematic areas. The resources available in the Deep Web include: private web, disconnected pages (unlinked content – websites that are not linked), contextual Web, dynamic content (content that is dynamically generated), limited access content (content available after logging in) or non-HTML content [Chertoff and Simon 2015, Hurlburt 2017].

Invisible resources include content that conventional search engines such as Google, Bing or Yahoo! have no access to [Devine and Egger-Sider 2004]. Minkle [2002] called such resources the "buried treasure of the web". The deep part of the web is not completely invisible – it is invisible only to users of conventional search engines [Ford and Mansourian 2006]. Pedley [2002] drew attention to the size and quality of information that is found in this area of the web. Sherman and Price [2003] grouped the invisible web into four main categories according to the reasons behind the "invisibility of resources": (1) the opaque web, (2) the private web, (3) the proprietary web, and (4) the truly invisible web [Ford and Mansourian 2006]. Sherman and Price [2001] noted that the invisible web is huge and that it grows faster than Surface Web.

True size of the Internet

Currently, there are over 1.5 billion websites on the Internet. Of these, less than 200 million are active [Stats 2019]. There are around 3.5 billion users of the Internet, which is almost 45% of the over 7-billion world population [Hurlburt 2017]. Visible World Wide Web resources account for only 6 to 10% of the entire Internet. The remaining 90–94% represents content that has not been indexed. Search engine robots do not reach most of the resources found in the deep web, although 95% of these constitute publicly available information. Internet resources that are not indexed are growing rapidly, and they usually take the form of databases – more than half of the invisible web is found in specialized databases.

Deep Web is characterized by growing scale, domain diversity, and numerous structured databases [Khare et al. 2010]. It grows so fast that an effective estimation of its size can be difficult or even impossible [Lu 2008]. Gulli and Signorini [2005] estimated the size of the Surface Web (public indexable web) at about 11.5 billion websites. Out of this set, 9.36 billion websites were available in the indexes of the four major search engines, including Google. It is estimated that the size of the Invisible Web is about 500 times larger than the Surface Web. In 2001, Bergman estimated that the Invisible Web contains nearly 550 billion individual documents, while the Surface Web contains only one billion. These studies revealed that the Deep Web was about 400–500 times larger than the Surface Web [Bergman 2001]. However, a few years later, it was shown that these estimates were highly debatable [Lewandowski and Mayr 2006]. According to other studies, the Deep Web consisted of approximately 307,000 websites, 450,000 databases and 1,258,000 interfaces, and it continued to grow rapidly, with a 3-7-fold increase in 2000–2004 [He et al. 2007]. Today's Internet is much larger – it is estimated that 555 million domains contain thousands or millions of unique websites. With the development of the network, the content of the Deep Web will also grow [Pederson 2013].

The Deep Web is sometimes associated with the Dark Web (or Dark Net), so the terms are sometimes used interchangeably. While useful information can be found in the Deep Web, the Dark Web can be a space for illegal and dishonest activities.

4. Dark Web

The Internet and Web technologies have originally been developed assuming a perfect world where all users are honorable. However, the dark side has emerged and bedeviled the world [Kim et al. 2011]. The term Deep Web is used to describe content posted on the Internet, which for various reasons is not indexed by search engines. The Dark Web is part of the Deep Web, which has been deliberately hidden, and is not available through standard web browsers. The Dark Web is part of the Internet that mainstream software cannot access [Gehl 2016]. Dark Net, colloquially, refers to a distinct network supporting cryptographically hidden sites [Moore and Rid 2016].

For many users, the Internet is a great place, a tool for work, communication and entertainment. However, there are nooks and crannies of the network that seem to be distant, and many of these are dark and shady. The very term Dark Web brings to mind the images of dark alleys, gloomy dead ends, dangerous people, and socially harmful activities. The Dark Web is a secretive, anonymous place where murky types offer access to illegal goods and services [Bradbury 2014].

The Dark Web is a mysterious place. Most Internet users do not know about it and will never go there [Gollnick and Wilson 2016]. Dark Web is part of the Internet, which most people probably do not know how to get access to, and probably most of them would not like to explore the content that is shared there [Jardine 2015]. The Dark Net is a place where illegal activities are undertaken – "cybercrime is like cancer, spreading from the Dark Web to the rest of the Internet" [Hurlburt 2017].

Paradoxically, Dark Web is unique because it is not particularly user friendly to Internet users. The use of public services such as Facebook or Twitter requires the creation of an account, which takes only a few minutes. Access to the resources placed in the Dark Net is not so simple. Using the Dark Web often requires encryption and decryption of messages and the ability to use a relatively esoteric, virtual (crypto) currency. Obviously, none of the sites in the Dark Web is advertised, except that many of them are demonized in the media [Maddox et al. 2016]. Because of all this, the Tor – which is the browser that allows access to the Dark Net – is used by relatively few (specialized) users. Studies have shown that in the United States, in 2015, there were 55 Tor bridge users per 100,000 Internet users, while in Canada there were 79 users per 100,000 Internet users [Jardine 2015].

The Dark Web uses the Onion Router hidden service protocol. Tor and other similar networks enable the users to stay in the network almost completely anonymously by encrypting data packets and sending them through several network nodes (onion routers) [Chertoff and Simon 2015]. The Onion Router (Tor) is both free software and an open network that helps users remain anonymous and protect themselves against traffic analysis [Cardullo 2015]. Tor in its entirety originated as a collaborative project between the US Naval Research Laboratory and the non-profit organisation called Free Haven Project. The underlying purpose was to create a distributed, anonymous, easily deployable and encrypted network to be used by those who needed it [Moore and Rid 2016].

The Dark Web is often portrayed as a place where illegal content is hidden, and at the same time as a place that provides complete freedom (of expression) [Gehl 2016]. Tor can be used to bypass content filtering mechanisms, censorship, and other communication restrictions. In addition, an important feature of Tor is the ability to host websites anonymously, which provides a certain amount of immunity. Publishers who run "dark websites" (ending in .onion) are able to hide their identity and location from Internet users [Dingledine et al. 2004]. In most cases, users with sites that end with .onion do not know the identity of the host, and the host does not know the identity of the users. This distinguishes Tor from the typical Internet at large, in which websites are linked to a company or a location, and visitors are often identified and monitored using various tracking technologies, such as cookies, account registration, IP addresses, or geolocation [Gehl 2016].

Tor is essentially a neutral tool that can be used in just cause as well as for criminal purposes. Some users value Tor's anonymity because it makes it difficult to censor sites or content that may be stored elsewhere in the world [Owen and Savage 2015]. However, the same open-source tools that provide privacy protection and allow users to bypass censorship can also become a space for black markets, and they may fuel criminal activity [Hurlburt 2017]. Tor is typically used for criminal purposes in liberal countries, while its "virtuous use" is found in countries where political repression prevails [Jardine 2015]. For example, the Russian state offered 110,000 USD (GBP 65,000) to a person or an organization that would break Tor's encryption and anonymity [BBC 2014].

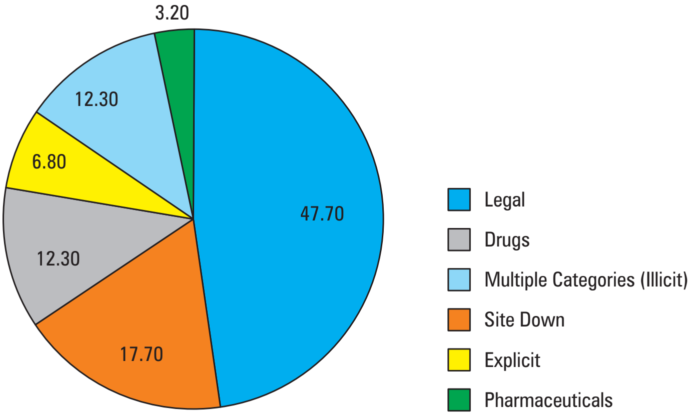
5. Dark Web is infamous for illegal content. Is this a justified belief?

Although the Dark Web also contains harmless content (such as websites of people who think that freedom of speech is at risk), it is infamous for illegal content. The Dark Net is a space where the digital black market is developing, enabling the purchase of sensitive (personal) data, services, items, products and substances with limited availability, such as for instance drugs (an obvious example of that is the Silk Road crypto-market, the place of illegal drug trafficking) [Maddox et al. 2016]).

The Dark Web can be a source of threats to both private individuals and enterprises. This dark side of the Internet is a space where stolen data is a marketable product. Counterfeit and stolen documents have long been sold in the Dark Web [Sixgill 2018]. Much stolen information, including credit card numbers, financial documents, login data, proprietary source code, tax documents, or other sensitive data, can be obtained in the Dark Web where they attract buyers who want to open fake accounts, reveal software gaps, steal intellectual property, or commit other scams. Only in 2018 there were 2216 confirmed data breaches, and 76% of them had financial motivation (according to Verizon 2018 Data Breach Investigations Report) [Terbium Labs 2018].

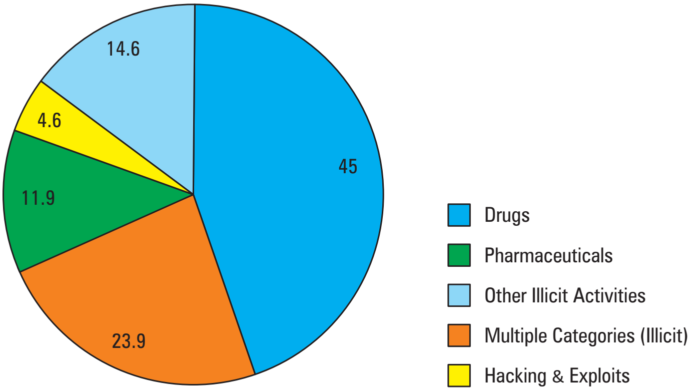
The Dark Internet markets promise undetectable, anonymous transactions. Transactions performed in the Dark Web are based on crypto-currencies that ensure anonymity for buyers and sellers [Maddox et al. 2016]. Bitcoin, a decentralized, international currency based on peer-to-peer technology [Barratt 2012], until now the most-used currency in the Dark Net, is quickly replaced by the Monero crypto-currency [Hurlburt 2017]. The addresses of transactions carried out with the help of Monero are hidden, which makes finding senders and recipients extremely difficult. It is also possible to hide the amount of the transfer. Monero uses ring signatures, passes confidential transactions and hides addresses to obscure the sources, amounts, and destinations of all transactions.

While the Dark Web is often demonized, research has shown that it is a place of primarily legal and even mundane content. Research conducted on the Australian Tor network in a set of 232 792 websites posted on 7651 virtual Tor domains, have led to identifying a broad range of materials therein – from illegal to simply banal [Dalins et al. 2018]. This does not mean, however, that the Dark Web is safe. Other studies have shown that illegal content posted in the Dark Web was dominated by drugs and various illegal services. At the same time, however, legal content accounted for 53.4% of all the domains surveyed (Fig. 2) and 54.5% of all the URLs (Fig. 3) [Gollnick and Wilson 2016].



Source: Author's own study based on Gollnick and Wilson [2016]

Fig. 2. Total content (by URL)



Source: Author's own study based on Gollnick and Wilson [2016]

Fig. 3. Illicit content (by URL)

Avarikioti and his co-authors have shown that contrary to popular opinion, the visible part of the Dark Net is surprisingly well connected through central sites, such as wikis and forums. They conducted a comprehensive categorization of content and showed that about half of the visible "dark content" concerned legal activities. Other content was related to, among other things, the sale of counterfeit goods or drugs [Avarikioti et al. 2018].

The fact that the Tor network also includes legal content is shown in the Dark Web Map. What's apparent from a cursory view of the map, zooming in and out of different areas, is that a significant proportion of mapped sites have been set up for illegal means, but there are also plenty of legal services such as whistleblower and personal pages [Hyperion Gray 2019].

6. Geoinformation versus the Internet's illicit content

All information can be divided into spatial – related to the Earth or to fragments of its surface, i.e. unequivocally located geographically by means of coordinates or spatial reference fields, and non-spatial information about entities and objects that are movable (also called subject and object information). Spatial information is called geographic, geospatial or geoinformation information. Geoinformation concerns objects, processes, phenomena, events and relations between them. There is an opinion that geoinformation absorbs about 60% of all expenditures devoted to applied IT [Ney 2005].

Geoinformation integrates many fields of science and technology. Geoinformation systems find practical application in almost all areas of life. The science of geoinformation deals with the problems of acquisition, collection, storage, analysis, interpretation, processing, dissemination, transmission, practical application and use of geoinformation [Gajos-Gržetić 2017]. Geoinformation services constitute a powerful tool for information support of decision-making processes. Traditionally, geoinformation systems (GIS) are used for the storage and use of cartographic information. Geoinformation service is a geographic information system's LAN or WAN. It continuously accumulates the map data and provides the access to this data through programming or interactive dialog interface. Google Maps is a typical example of the geoinformation service [Belyakov et al. 2014].

The dynamic development of geoinformation, which took place in the second half of the 1990s, contributed to the mass creation of spatial databases. The stock of digital spatial databases available in Poland includes topographic databases, hydrographic, geological and sozological, geoenvironmental, soil and agricultural maps, and many others (Table 2), as well as numerous regional and private databases. Access to some of the data is limited. Selected spatial data can be obtained via the nationwide Geoportal system [Król et al. 2016]. Thanks to the GEOPORTAL 2 project, implemented by the Main Geodesy and Cartography Authority of Poland (Główny Urząd Geodezji i Kartografii), spatial information infrastructure services are available online, not only to public administration units, but also to natural and legal persons and other organizational units.

The data that is made available to the public via the Geoportal – and yet is located in the Deep Web – includes data provided as part of the INSPIRE service, including: geographical names, administrative units, addresses, cadastral plots, transportation networks, land use, buildings, soil, production and industrial facilities, protected areas, or hydrography. This data is made available via WMS, WMTS, WFS or ATOM.

Table 2. List of selected databases of spatial information

Name	Nominal scale	System of coordinates
Baza Danych Obiektów Ogólnogeograficznych (BDOO) [Database of general geographical objects]	1:250 000	PUWG 1992
Mapa Wektorowa Poziomu Drugiego [Second-order vector map]	1:50 000	WGS84
Baza Danych Obiektów Topograficznych (BDOT) [Database of topographic objects]	1:10 000	PUWG 1992
Mapa Hydrograficzna Polski [Hydrographic map of Poland]	1:50 000	PUWG 1992
Mapa Sozologiczna Polski [Sozological map of Poland]	1:50 000	PUWG 1992
Szczegółowa Mapa Geologiczna Polski [Detailed geological map of Poland]	1:50 000	PUW 1942
Mapa Hydrogeologiczna Polski (główny poziom wodonośny) [Hydro-geological map of Poland]	1:50 000	PUW 1942
Mapa Geośrodowiskowa Polski [Geo-environmental map of Poland]	1:50 000	PUW 1942
Mapa Podziału Hydrograficznego Polski [Hydrological division map of Poland]	1:50 000	PUWG 1992
Mapa Glebowo-Rolnicza [Soil and agricultural map]	1:25 000 (1:5 000)	PUWG 1992
Ewidencja Gruntów i Budynków [Land and building records]	1:5 000	PUWG 2000
Leśna Mapa Numeryczna i System Informatyczny Lasów Państwowych [Forest numerical map and information system of State Forest holding]	1:5 000	PUWG 1992
Numeryczny model terenu z Mapy Wektorowej Poziomu Drugiego [Digital terrain model of the second-order vector map]	1:50 000	WGS84
Baza Danych Pokrycia Terenu Corine Land Cover 2006	1:100 000	PUWG 1992

Source: Author's own study based on Kaczmarek [2011]

Web Map Service (WMS) is an international standard for sharing spatial data on the Internet in the form of a raster. The Geoportal makes available, among other things, the administrative map of Poland, cadastral data, landscape base map of Poland and many more. Web Map Tile Service (WMTS) is an international standard for the provision of spatial data online in the form of raster, predefined portions of the so-called map tiles. With the aid of the WMTS service, digital terrain models, among other things, are made available via the Geoportal. The Web Feature Service (WFS) – which is spatial data retrieval service – makes it possible to download from the PZGiK some or all of the spatial data sets stored therein, according to the set criteria. The implementation of this service requires the use of standards, so that the downloading is interoperable and does not impose on the users the application of specific technological solutions. The spatial data retrieval service in the ATOM profile enables downloading predefined spatial data sets within INSPIRE themes, for instance, BDOT10k for individual regions. Furthermore, Geoportal provides data through the Web Coverage Service (WCS), which is a download service. WCS typically provides the data in the form of a raster. Layers in this case are continuous spatial data, such as aerial and satellite images, as well as terrain and elevation data, whose spatial variability is represented by means of raster covers. However, this does not exhaust the data provided on the website. All of this data is not indexed by regular search engines. These data are specialized, qualitative, and they are contained in the Deep Web. Access to that data requires the use of appropriate software.

Geoinformation in the Dark Web

Tor is a circuit-based low-latency anonymous communication service. Its main design goals are to prevent attackers from linking communication partners, or from linking multiple communications to or from a single user. Tor relies on a distributed overlay network and onion routing to render anonymous TCP-based applications like web browsing, secure shell, or peer-to-peer communications. When a client wants to communicate with a server via Tor, he selects n nodes of the Tor system and builds a circuit using those selected nodes. Messages are then encrypted times. As a result of this onion routing, each intermediate node only knows its predecessor and successor, but no other nodes of the circuit. In addition, the onion encryption ensures that only the last node is able to recover the original message. A Tor client typically uses multiple simultaneous circuits. As a result, all the streams of a user are multiplexed over these circuits [Chaabane et al. 2010]. All this makes geolocation in the Tor network very difficult or impossible.

Dark Web Map

Web cartography has been happening for many years now. For a long time, it has focused on the surface web, which forms a tiny proportion of the entire Internet. Dark Web can be associated with "Dark Maps", a kind of graphical style of map presentation, for example: Styled Maps – Night Mode for Google Maps Platform. Dark Web Map is also available online. Hyperion Gray's Dark Web Map was the product of a 2015 experiment conducted under DARPA's Memex project. The objective of Memex was to construct search engines to be used by the legal facet of dark web. The map is mostly a collection of dark web homepage images that a user may view to gain an understanding of the hidden sites existing on the Tor network. The Dark Web Map provides visual insight into the hidden web. The Dark Web Map (a visualization of 3.7k Tor onion services) is a visualization of the structure of Tor's onion services. The map consists of 3,747 dark web sites crawled during March 2019. Each site is represented in the map as a screenshot, and sites with structural similarity are connected with a line [Hyperion Gray 2019].

The mapping process reveals that a general viewpoint of the Dark Web is rather shallow, and quite deceiving. The first-ever comprehensive Dark Web mapping study highlighted the attributes of hidden websites. Concerning this, it was found that the Dark Web possessed characteristics that mostly set it apart from conventional websites. Scientists discovered approximately 7 000 websites that were connected to each other via 25 000 links. The research findings, however, proved that more than 85 percent of these Dark Web sites lacked linkages to other pages. Essentially, this means that the Dark Web is comprised of sets of isolated "dark silos" [Dark Web 2018].

7. Conclusions

Most industry data, high quality data, i.e. data collected according to a verified, recognized methodology, often also peer-reviewed (verified), can be found within the Deep Web. These mostly include the data that finds some kind of specific, specialized application. Some data extracted from the Deep Web is also "less distinguished" data, everyday use information, such as meteorological data.

Geoinformation does occur in Surface Web, however, only a small fragment thereof is available in the search results. Most of the main websites of geoinformation portals, online maps and databases are made available through search indexes. The majority of geoinformation is available on the Deep Web, and it requires the use of specialized search engines or exploration of thematic maps.

Geoinformation denies the idea of Dark Web. The Tor network, which is the foundation of the Dark Web, was created in order to provide anonymity, and to hide the identity and location in space. Dark Web users want to be anonymous. Entities appearing in the Dark Web in most cases do not give their location, nor do they use Internet maps.

References

Avarikioti G., Brunner R., Kiayias A., Wattenhofer R., Zindros D. 2018. Structure and Content of the Visible Darknet. arXiv preprint arXiv:1811.01348.

Barratt M.J. 2012. Silk Road: eBay for drugs. Addiction, 107(3), 683–683. https://doi.org/10.1111/j.1360-0443.2011.03709.x

BBC 2014. Russia Offers $110,000 to Crack Tor anonymous Network. BBC News. https://www.bbc.com/news/technology-28526021 [accessed: 09.05.2019].

Belyakov S.L., Bozhenyuk A.V., Belykova M.L., Rozenberg I.N. 2014. Model of Intellectual Visualization of Geoinformation Service. ECMS, 326–332.

Bergman M.K. 2001. White paper: The deep web: surfacing hidden value. Journal of Electronic Publishing, 7(1). http://dx.doi.org/10.3998/3336451.0007.104

Bradbury D. 2014. Unveiling the dark web. Network Security, 4, 14–17. https://doi.org/10.1016/S1353-4858(14)70042-X

Cardullo P. 2015. 'Hacking multitude' and Big Data: Some insights from the Turkish 'digital coup'. Big Data & Society, 2(1), 1–14. https://doi.org/10.1177/2053951715580599

Chaabane A., Manils P., Kaafar M.A. 2010. Digging into anonymous traffic: A deep analysis of the tor anonymizing network. In 2010 Fourth International Conference on Network and System Security, IEEE, 167–174.

Chertoff M., Simon T. 2015. The impact of the dark web on internet governance and cyber security. Global Commission on Internet Governance Paper Series, 6.

Dalins J., Wilson C., Carman M. 2018. Criminal motivation on the dark web: A categorisation model for law enforcement. Digital Investigation, 24, 62–71. https://doi.org/10.1016/j.diin.2017.12.003

Dark Web 2018. This Map Gives You a Peek into the Dark Web. Dark Web News, https://darkwebnews.com/dark-web/darknet-map/ [accessed: 09.05.2019].

Devine J., Egger-Sider F. 2004. Beyond Google: The invisible web in the academic library. The Journal of Academic Librarianship, 30(4), 265–269. https://doi.org/10.1016/j.acalib.2004.04.010

Dingledine R., Mathewson N., Syverson P. 2004. Tor: The second-generation onion router. Naval Research Lab. Washington DC.

Ehney R., Shorter J.D. 2016. Deep web, dark web, invisible web and the post ISIS world. Issues in Information Systems, 17(4), 36–41.

Ford N., Mansourian Y. 2006. The invisible web: An empirical study of "cognitive invisibility". Journal of Documentation, 62(5), 584–596. https://doi.org/10.1108/00220410610688732

Gajos-Gržetić M. 2017. Reprezentacja nauki o geoinformacji w wybranych językach informacyjno-wyszukiwawczych. Wydawnictwo Uniwersytetu Śląskiego, Katowice.

Gehl R.W. 2016. Power/freedom on the dark web: A digital ethnography of the Dark Web Social Network. New Media & Society, 18(7), 1219–1235. https://doi.org/10.1177/1461444814554900

Gollnick C., Wilson E. 2016. Separating Fact from Fiction: The Truth about the Dark Web. Terbium Labs.

Gulli A., Signorini A. 2005. The indexable web is more than 11.5 billion pages. Proceedings of the 14th International Conference on World Wide Web (WWW) – Special interest tracks and posters, 902–903.

He B., Patel M., Zhang Z., Chang K.C.C. 2007. Accessing the deep web: A survey. Communications of the ACM, 50(5), 94–101.

Hurlburt G.F. 2017. Shining Light on the Dark Web. IEEE Computer, 50(4), 100–105. https://doi.org/10.1109/MC.2017.110

Hyperion Gray 2019. Dark Web Map. A visualization of 3.7k Tor onion services, http://bit.ly/2PPNDpQ [accessed: 09.05.2019].

Jardine E. 2015. The Dark Web dilemma: Tor, anonymity and online policing. Global Commission on Internet Governance Paper Series, 21.

Kaczmarek L. 2011. Potencjał informacyjny krajowych baz danych przestrzennych w kartograficznych badaniach środowiska przyrodniczego. Rozprawa doktorska pod kierunkiem prof. UAM dr hab. Beaty Medynskiej-Gulij, Zakład Kartografii i Geomatyki UAM w Poznaniu.

Khare R., An Y., Song I.Y. 2010. Understanding deep web search interfaces: A survey. ACM SIGMOD Record, 39(1), 33–40.

Kim W., Jeong O.R., Kim C., So J. 2011. The dark side of the Internet: Attacks, costs and responses. Information Systems, 36(3), 675–705. https://doi.org/10.1016/j.is.2010.11.003

Król K., Prus B., Salata T. 2016. Geoportal 2: nationwide network node of spatial information – description of its characteristics and an attempt at evaluation of selected functionalities.

Geomatics, Landmanagement and Landscape (GLL), 1, 47–63. https://doi.org/10.15576/GLL/2016.1.47

Lewandowski D., Mayr P. 2006. Exploring the academic invisible web. Library Hi Tech, 24(4), 529–539. https://doi.org/10.1108/07378830610715392

Lin K.I., Chen H. 2002. Automatic information discovery from the invisible Web. Proceedings of the International Conference on Information Technology: Coding and Computing, IEEE, 332–337.

Lu J. 2008. Efficient estimation of the size of text deep web data source. Proceedings of the 17th ACM Conference on Information and Knowledge Management CIKM'08. ACM Press, New York, NY.

Maddox A., Barratt M.J., Allen M., Lenton S. 2016. Constructive activism in the dark web: Cryptomarkets and illicit drugs in the digital 'demimonde'. Information, Communication & Society, 19(1), 111–126. https://doi.org/10.1080/1369118X.2015.1093531

Minkle W. 2002. The invisible web. School Library Journal, 49(12), 29.

Moore D., Rid T. 2016. Cryptopolitik and the Darknet. Survival, 58(1), 7–38. https://doi.org/10.1080/00396338.2016.1142085

Ney B. 2005. Geoinformacja w społeczeństwie informacyjnym. Roczniki Geomatyki – Annals of Geomatics, 3(3), 11–18.

Owen G., Savage N. 2015. The Tor dark net. Global Commission on Internet Governance Paper Series, 20.

Pederson S. 2013. Understanding the Deep Web in 10 minutes. BrightPlanet. Whitepaper.

Pedley P. 2002. Why you can't afford to ignore the invisible web. Business Information Review, 19(1), 23–31.

Raghavan S., Garcia-Molina H. 2000. Crawling the hidden web. Technical Report. Stanford.

Sherman C., Price G. 2003. The invisible web: Uncovering sources search engines can't see. Library Trends, 52(2), 282–298.

Sixgill 2018. Forging Documents in the Deep and Dark Web. Sixgill Report.

Stats 2019. Total number of Websites. Internet Live Stats, http://www.internetlivestats.com/total-number-of-websites/ [accessed: 09.05.2019].

Symanovich S. 2019. How to safely access the deep and dark webs. Symantec. https://nr.tn/2Vb8ycG [accessed: 09.05.2019].

Terbium Labs 2018. A Buyer's Guide to Dark Web Monitoring. Six Questions to Ask When Developing a Dark Web Intelligence Strategy. Terbium Labs.

UC Berkeley 2010. Invisible or Deep Web: What it is, How to find it, and Its inherent ambiguity. University of California, Berkeley. Teaching Library Internet Workshops. http://bit.ly/2DDaoZ0 [accessed: 09.05.2019].

Weimann G. 2016. Going dark: Terrorism on the dark web. Studies in Conflict & Terrorism, 39(3), 195–206. https://doi.org/10.1080/1057610X.2015.1119546

Dr inż. Karol Król
Uniwersytet Rolniczy w Krakowie
Katedra Gospodarki Przestrzennej i Architektury Krajobrazu
30-059 Kraków, al. Mickiewicza 24/28
e-mail: k.krol@onet.com.pl
website: http://homeproject.pl
ORCID: https://orcid.org/0000-0003-0534-8471