Received 13 June 2019
Revised 28 July 2019
Accepted 3 September 2019

Peculiarity of the bit rot and link rot phenomena

Karol Król
Department of Land Management and Landscape Architecture, University of Agriculture in Krakow, Kraków, Poland, and

Dariusz Zdonek
Institute of Economics and Informatics, Faculty of Organization and Management, Silesian University of Technology, Gliwice, Poland

Abstract

Purpose – Almost two decades after diagnosing the link rot phenomenon, and in spite of the universal character of the bit rot phenomenon, the need for reminding about these is still valid. This is because it appears that digital device users have become accustomed to them. The bit rot and link rot phenomena describe the inability to access specific contents or the inability to use specific digital resources. This paper aims to answer the question as to whether it is possible to eliminate the link rot and bit rot phenomena from the internet space.

Design/methodology/approach – The link rot and bit rot phenomena were characterised based on a review of various source materials, including the academic literature and internet materials.

Findings – The phenomena known as bit rot and link rot have already been well researched and described. Research has revealed the reasons for their occurrence, and also offered recommendations on how to minimise their occurrence. In the current digital ecosystem, however, it is not possible to completely eliminate the link rot and bit rot phenomena.

Research limitations/implications – One can forecast that the link rot phenomenon will intensify and the internet resources will be even more unstable, because more and more new, highest-level domains which are purchased are emerging.

Originality/value – The study compares the phenomena of bit rot and link rot, even though they are usually considered separately. Moreover, it presents the terms of "digital bumps", which cause "digital tsunami".

Keywords Data decay, Digital bump, Digital ecosystem, Digital tsunami, Link rot, Software rot

Paper type Conceptual paper

Introduction

We are living in the era of lost data referred to as the "digital dark age". These terms are associated with a phenomenon that can be described as "escaping" or "disappearing" ("decay") of formats (files) or the "erasure of files". They refer to the inability to archive all data, which have never so far been becoming obsolete so rapidly; data, which are easily generated in vast volumes these days and, even more easily, deleted. These terms reflect the concern about the inability to restore files deleted on a mass-scale, data lost because of equipment failures, terminated domains and shut down servers. The "information black hole" is enlarging before the eyes of digital device users.

Over time, digital information recorded on various data carriers and in various formats can decay or degrade if not stored properly or subjected to other adverse processes. According to forecasts, within just a few decades data from the beginning of the twenty-first century may become completely unavailable or unreadable (Deegan and Tanner, 2006). Much of what is being done these days is generated in the digital form. This is not

Global Knowledge, Memory and Communication
Vol. 69 No. 1/2, 2020
pp. 20-37

2514-9342
DOI 10.1108/GKMC-06-2019-0067

something that has been digitised or "transferred from an analogue container to a digital one". These are works that have been generated in the digital form and nowadays, more and more frequently, they cease to exist as a digital content while leaving no analogue equivalent (Wernick, 2018).

The bit rot and link rot phenomena describe the inability to access specific contents or the inability to use specific digital resources. As regards bit rot, this phenomenon is mostly because of the "obsolescence" of files, formats, hardware and software, as well as because of a mechanical damage to data carriers or errors during data transmission. Young (2017) described it vividly – because of the obsolescence of technology, the history of the twenty-first century may become an "incomprehensible digital soup". Link rot, on the other hand, describes the lack of specific content (resources) at the indicated internet address.

Because of the popularity of the internet, there is a growing tendency to quote (cite, refer to) resources available online, primarily files and contents, in scientific and popular science articles as well as in internet publications; hyperlinks are, after all, the foundation of the internet. Given the dynamic nature of the Web, however, links may often become inaccessible. The fault does not lie in the links themselves but in the resources, which for a variety of reasons, have ceased to exist at a particular (internet) address. Almost two decades after diagnosing the link rot phenomenon, and in spite of the universal character of the bit rot phenomenon, the need for reminding about these is still valid. These phenomena accompany the users of digital technologies on a daily basis, and are of significance both globally and for each individual user.

Even though the bit rot and link rot phenomena have already been well researched, and the ways to minimise the effects they bring about are known, their intensity is not decreasing; on the contrary, it is on the increase. Moreover, it may appear that the digital media user community has become accustomed to these phenomena, with many users not even being aware of these terms and assuming the data loss to be a common and obvious matter. The paper attempts to answer the question as to whether it is possible to eliminate the bit rot and link rot phenomena from the internet space. To this end, it describes the characteristics of these phenomena while emphasising their particular nature. Next, attention is paid to their contribution to the loss of digital data, and an attempt is made to analyse the relationships occurring between them. Further on, the paper discusses the practical implications of the occurrence of these phenomena, and takes their future significance into consideration.

The peculiarity of the bit rot phenomenon

The information (data) currently generated in the digital form may not be read by devices and software in the future. All these data, our century's digital history, are endangered, as there is no physical documentation for most digital materials (for a vast volume of materials, their digital copies or analogue equivalents are missing). It is feared that selected contents (a certain part of the digital output) will be impossible to recover. This process is in progress. Data stored in an electronic form "decay". Many floppy disks from the early digital era are no longer readable. On the other hand, for fully operational floppy disks, problems with their readout arise because of the lack of equipment capable of reading them (software and workstations). With a bit of luck, copies of various materials will be accessible in public libraries. However, CDs will not last for more than a few decades. This worries archivists and data archaeologists, and is a technological challenge (Wernick, 2018).

The term "bit rot" describes a slow deterioration in the integrity of data stored on digital storage mediums. The phenomenon is also known as (Table I) "bit decay", "data rot", "data decay" (Rouse, 2019) or "digital decay", whereas digital decay is not the wear of a tangible material but decomposition of binary code (digitised information suffers from degradation

Table I.
Selected terms used in the context of the bit rot phenomenon

Term	Source
Bit rot	Hayes (1998)
Bit loss	Hudgins (2011)
Bit decay, data rot, data decay and silent corruption	Baker *et al.* (2005), Bowers (2017), Rouse (2019)
Era of lost data, digital dark age	Kuny (1998), Lyons (2016), Whitt (2016), Wernick (2018)
Born-digital, digital-first	Lor and Britz (2012)
Data degradation	Baker *et al.* (2005)
Digital preservation, digital preservation strategies	Hedstrom (1997), Baker *et al.* (2005), Schlieder (2010), Deljanin (2012), Lor and Britz (2012)
Document persistence	Lor and Britz (2012), Koehler (2004)
Decaying digital artefact	Lyman and Kahle (1998), Tew (2005), Bowers (2017)
Software rot	Odersky and Moors (2009), Wernick (2018)
Software erosion, software entropy, software bloat	Odersky and Moors (2009), Hildenbrand (2017)
Digital heritage, digital archives, digital storage	Leung *et al.* (2001), Baker *et al.* (2005), Knight (2010), Schlieder (2010), Tait *et al.* (2013)
Digital vellum (by Vint Cerf)	Mottl (2015)
Digital obsolescence	Deljanin (2012)

Source: Own elaboration

of binary code) (Haslop *et al.*, 2017). On the other hand, digital decay is a term used to describe any degradation, obsoletion or breakdown of computerised information (Stringfellow, 2018). Bit rot is a phenomenon associated with data loss and the obsolescence of files and formats (Hayes, 1998). Bit rot is the inability of today's generation of computer systems to read a "yesterday's" product (Pitt, 2019).

The term "bit rot" is related to the term "software rot". Old files, games and other data become useless because there is no longer universally available environment/ecosystem for their readout, i.e. equipment and software that would enable their use (Wernick, 2018). Replacement of either the operating system or, for example, a text editor may lead to a situation where the device or software used so far becomes useless i.e. "incompatible", for example, because the lack of drivers. One day, there may be no Microsoft Word application, photo editors compatible with JPG files, or Excel applications, which render the information stored in one of these currently, used formats useful. In the future, they may be inaccessible to a new generation of programs, and thus to the future generations of users. This changes the perception of owned (created) files, and raises a question concerning the rights of preservation and the rights to read or use files and formats (Mottl, 2015).

Bit rot refers to a gradual loss of data (data decay). Data decay is a silent killer of data (because data is time sensitive). There are two main categories of data decay: mechanical and logical. Mechanical data decay is probably the best-known form of data decay, which may be caused by, for example, a server hang up or damage to the data carrier. Data decay occurs at the mechanical level on a daily basis, even without the users' participation. Every time data are recorded on or read from a data carrier, there is a risk that the data will be lost or the carrier will fail.

Software erosion and software entropy

Bit rot is a constant problem in most long-term IT projects. With the evolution (development) of software, at some point it is preferable to redesign the system from scratch than to update it (however, this is often not the case, and the software systems are "dragged" as the risk of redesigning is regarded to be high) (Odersky and Moors, 2009). Software rot implies that software is "old" and has become slower than before and less useful, or can no longer be used at all. Bit rot is associated with the following terms: software erosion, software entropy and software bloat (Jacobson *et al.*, 1992). Software erosion is a slow but steady deterioration in software capacity, which may ultimately lead to it being replaced by other, more capable and newer software. The phenomena of software erosion and software entropy (a term associated with the increasing software complexity, and thus the increasing number of errors) are sometimes delayed by updates and resetting software to the factory or initial settings.

The term "software bloat" is used to describe software, which provides additional, often superfluous functionalities that (unnecessarily) generate high demand for system resources. Additional modules, components and extensions increase the bit rot phenomenon intensity (Hildenbrand, 2017). Software systems rot not because of rust or material fatigue but because their requirements change. Modifying a software system is comparatively easy, so there's a low threshold to accepting new requirements, and adaptations and extensions are common. However, if not done right, every such change can obscure the original architectural design by introducing a new special case (Odersky and Moors, 2009).

Hardware erosion

The bit rot phenomenon is associated with the wear and ageing of computer hardware. No inexpensive digital carrier is completely fail-safe for a long period of time as it can get degraded (for example, the reflective material used for optical media can start to break down). Disks and magnetic tapes can suffer from bit rot as well. Because bit rot may refer to other (undetected) storage errors, which alter the downloaded content, e.g. errors in the network interface, software buffer over-runs in the operating system and error correction failures in memory (Baker *et al.*, 2005).

The peculiarity of the link rot phenomenon

Two decades after the link rot phenomenon was diagnosed, non-functional links continue to plague the internet, and increasingly so. As early as 20 years ago J. Nielsen (1998) pointed out that it was advisable to regularly run a link validator on every website. As regards small websites, a programmed verification of the "health of links", performed by applications, which can test the website on a monthly basis and generate a list of dead links, may be sufficient. For larger websites, it is more cost-effective to install validation software on the server itself (Nielsen, 1998).

There are two basic situations in which links can be regarded as dysfunctional: (1) link rot (Benbow, 1998) and (2) content drift (Zhou *et al.*, 2015). Both phenomena are most often discussed (and researched) in the context of references to internet sources in scientific publications. For example, Hennessey and Ge (2013) analysed 14,489 unique websites included, in the years 1996-2010, in the Thomson Reuters Web of Science citation index, and determined that the average lifespan for these websites was 9.3 years, and that only 62 per cent of them had been archived. On the other hand, Zittrain *et al.* (2014) demonstrated that over 70 per cent of internet addresses provided in articles published in the Harvard Law Review and other scientific journals, and approx. About 50 per cent of URLs provided in opinions of the Supreme Court of the USA had failed to lead to the originally quoted information. Rumsey (2002) compared the use of links in articles devoted to legal issues to a runaway train. He demonstrated that in scientific publications, only 30 per cent of links, which referred to external sources had been functional after 4 years. According to the APA, a URL address is among the most sensitive (critical) elements of the Web, in particular in the context of compiling a list of quotations; if a link fails to function, readers will not be able to find the quoted material, and the paper's reliability will be adversely affected (APA, 2003).

The link rot phenomenon dates back to the beginnings of the internet. Its nuisance was already noticed more than 20 years ago, when Benbow (1998) equated a click on a non-functional link with a "transfer of the user to cyber-no-man's-land". The phenomenon of non-functional links is referred to using various terms: broken links (Markwell and Brooks, 2002; Kobayashi and Takeda, 2002), linkrot (Denemark, 1996; Taylor and Hudson, 2000), link rot (Fichter, 1999; Markwell and Brooks, 2003) or decay and failure links (Spinellis, 2003) (Table II). Link rot is the process by which website links become irrelevant or broken over time, because the actual websites they link to disappear, change content or move to a new location (Parker, 2007). A dead link doesn't mean the content being accessed doesn't exist anymore but it does make it more difficult to find (Mead, 2013). Somehow link rot is typical of the ever-ephemeral nature of wired Progress. Link rot is the part of the nature of the Web (Merchant, 2014).

Nielsen (1998) presented the link rot phenomenon in a vivid manner: link rot contributes to dissolving the very fabric of the Web: there is a looming danger that the Web will stop being an interconnected universal hypertext and turn into a set of isolated info-islands. In 2004, an IBM research team extended the catalogue of untypical terms associated with the link rot phenomenon. Nielsen's info-islands were defined as "stale neighbourhoods": "Such neighbourhoods are identified only by frustrated searchers, seeking a way out of these stale neighbourhoods, back to more up-to-date sections of the Web" (Bar-Yossef *et al.*, 2004, p. 328). On the other hand, the founder of the World Wide Web (WWW), Tim Berners-Lee noted that:

> It is the duty of a Webmaster to allocate URIs which you will be able to stand by in 2 years, in 20 years, in 200 years. This needs thought, and organisation, and commitment (Berners-Lee, 1998).

Mead (2014) compared the link rot phenomenon to a specific type of a corrosive force, which consumes contents and contributes to the disintegration and decomposition of the internet

Term	Source
Link rot	Berners-Lee (1998), Nielsen (1998), Benbow (1998)
Link rot	Denemark (1996), Taylor and Hudson (2000)
Link decay	Goh and Ng (2007), Hennessey and Ge (2013)
Broken link, dead link, dangling link	Markwell and Brooks (2002), Kobayashi and Takeda (2002)
Links to nowhere (cyber no man's land)	Benbow (1998)
Decay and failures of Web references	Spinellis (2003)
Cool URI	Berners-Lee (1998)
A link is a promise	Pernice (2014)
Content drift	Burnhill *et al.* (2015), Zhou *et al.* (2015)
Never let any URL die	Nielsen (1998)
Reference rot, web references persistence, stability of Web sources	Taylor and Hudson (2000), Lawrence (2001), Rumsey (2002), Parker (2007), Zittrain *et al.* (2014), Burnhill *et al.* (2015)
Soft-404s, Web's decay	Bar-Yossef *et al.* (2004)
URL half-life	Markwell and Brooks (2003), Koehler (2004)
Ephemeral nature of wired progress	Merchant (2014)
Monument to the fragility of the internet, atrophy of links	Tew (2005)

Source: Own elaboration

Table II.
Selected terms used in the context of the link rot phenomenon

structure. Salmon (2013) noted that the emphasis on resource durability had been declining in recent years –

> we've moved into a world of streams, where flow is more important than stock, and where the half-life of any given piece of content has never been shorter; that's not a world which particularly values preserving that content for perpetuity.

And, of course, it has never been easier to simply delete vast amounts of content at a stroke. In effect, the modern Web has shifted more of the onus of archival onto users.

Content drift

Link rot is becoming not so much a problem of navigation but a threat to truthfulness (reliability) of information on which we rely (Mead, 2014). Even if the location of a quotation on the internet is stable (the source is stable), its content can be modified, so that subsequent readers may not be able to read exactly the same (quoted) material. Content drift means a continuous slow movement of contents from one place to another. Content drift can be obvious where the content of the link (anchor text) suggest a different content, or "deceitful" where the content at a particular address may change dynamically, which can be because of the nature of the source (e.g. a news website), and has actually changed (in most cases without informing users about it) (Burnhill *et al.*, 2015).

Content drift raises the question: "For instance, is a 'webpage' defined by its URI or by its contents?" After all, there is no certainty that the content at a particular URI address will remain (completely or partially) unchanged; the address can remain unchanged, the content not necessarily so (Ashenfelder, 2011). An automated link checker visiting a list of URIs and logging all ultimately successful and failed requests would miss these subtleties.

The content drift phenomenon can be frequently observed in social media where the content is dynamically generated. Various entries are frequently altered, the tone of statements is softened, and significant sentences or important paragraphs are deleted. Content drift has resulted in controversial statements being recorded in image files (screenshots), which is supposed to ensure that they are documented and preserved.

The internet is stitched together by an incalculable number of hyperlinks such as cells in an organism, the sources and destinations have a finite lifespan. Essentially, links can and do die (Routley, 2017). The problem of the link rot phenomenon comes down to the internet entropy. All isolated systems go from a state of order to a state of disorder over time in a process known as entropy (Spacey, 2017). Websites disappear, hosting services are not renewed and URLs are changed during the website update without appropriate redirection, etc. The list of reasons for link breakages seems to have no end, and becomes longer and longer with time (Mead, 2013). A high online entropy rate indicates that older parts of the Web, useful or not, disappear over time in a "swamp of noise and interference".

Link cemetery: a monument to the internet's fragility

The Million Dollar Homepage (Tew, 2005) is a unique initiative that has gone down in the history of the internet. In 2005, Alex Tew put up for sale a million pixels that made up a huge interactive board (digital canvas); it was interactive as each pixel could be a link sending to a specific website indicated by the pixel purchaser. In fact, the space of the "digital canvas" has become a specific link cemetery (which means an accumulation of a large number of non-functional links within a single website). Since 2005, many of the websites to which the graphics (the purchased parts of the board) directed have ceased to exist. Many hyperlinks on the canvas transfer the users to other websites than they used to, and over 20 per cent of them are simply dead. Nowadays, this digital canvas from 2005 is a "decaying digital artefact". Of the 2,816 links that embedded on the canvas (999,400 pixels in total), in July 2017 as many as 547 were completely unreachable. Further 489 websites redirected to a different website (domain) or to a domain resale portal. Moreover, a significant proportion of the remaining links led to websites devoid of content (Bowers, 2017).

At this point, it is worth mentioning that the space of the interactive board was made available in 100-pixel blocks, each of them being the size of 10 × 10 pixels. The graphic was, therefore, divided into 10,000 blocks; the reason for such division was that an area smaller than 10 × 10 pixels would have been too small to present anything of significance. Therefore, to purchase a space of the interactive canvas, an expense of $100, i.e. $1 for a pixel was required. The fee entitled the purchaser to the presentation of an image linked with the indicated website. The website and the digital canvas were intended to be available for at least five years (from the launch date) i.e. at least until 26 August 2010; the website is actually available to the present day.

An analysis conducted by Yanofsky in 2014 showed that 22 per cent of all links from the interactive Million Dollar Homepage board had led nowhere. In reality, however, the scale of the link rot phenomenon is, in this case, even greater, as Yanofsky (2014) failed to include the content drift phenomenon in his analysis. Moreover, many internet addresses from the million dollar homepage currently redirect to other websites. Yanofsky put forward a thesis that the process of link decay on the million-dollar homepage had stabilised, so that the link decay would be reduced in the years to come. The thesis was based on the assumption that websites, which have existed uninterruptedly on the internet for many years posed a lesser risk that they would stop functioning in the future. Is this assumption valid? Practice has shown that any website can disappear from the internet resources at any time, which is obviously not tantamount to its disappearance from digital archives. Research showed, however, that links to selected resources were more stable and more durable than other links (Koehler, 2004).

It is phenomenal that a "crowdfunding" project implemented by a student in 2005 shows the intensity of the link rot phenomenon several years later. "The Million Dollar Homepage" project has become a challenge to archivists, as it demonstrated that the process of archiving a digital object did not end with archiving the object itself. This is because this object can be linked to many other objects, which in this case, were an integral part of the digital canvas. Nevertheless, the canvas mostly remains an intact record of aesthetic and commercial patterns typical of online publications of 2005. The digital canvas is filled with pixel representations of unshapely fonts, ads resembling the "sketched websites" for online gambling and other representations.

Digital dark ages

The term "digital dark ages" is most often mentioned in the context of digital resource preservation (archiving), even though the term has also been used in the context of cyber security. The "digital dark ages" refer to a system where health care, financial – any industry connected to the internet and using data – are always in peril and never fully safe to use (Heikkila, 2018).

We are living in the midst of Digital Dark Ages because enormous amounts of digital information are already lost forever (Kuny, 1998). No access to data is most frequently because of their removal (deletion), creation in one of the obsolete formats or a physical damage to a data carrier. Many data sets are obsolete, which is largely because of technological changes. Many technologies and devices become obsolete when their suppliers provide new product lines, often without backward compatibility or when companies cease their operations. In addition, there are many document formats and numerous types of data carriers, each of which can have its own hardware and software dependencies. If the hardware or software necessary to read particular files is not available, the information recorded in the files will be lost. In addition, because of the increasingly restrictive regulations concerning intellectual property and licensing, many digital resources will never be placed in library collections to be archived. Kuny (1998) indicated three main actions that could support the conservation of digital resources: the preservation of technology (hardware and software); technology emulation; and data migration (the transfer of data to newer, more durable and more capacious data carriers) (Kuny, 1998; Krebs *et al.*, 2012; Deljanin, 2012).

The availability of digital data is at risk as data are frequently recorded on a transient medium as files of a specific format, which require an appropriate coding scheme to be read (Whitt, 2016). Moreover, the impossibility of archiving all data are largely because of their volume ("data glut"), and the digital mountain of information will just keep growing (Young, 2017). Digital collections facilitate access but do not facilitate preservation. Digital resources are, by their nature, ephemeral. Digital information has been allowed to become a medium for the present, neither a record of the past nor a message to the future (Lyman and Kahle, 1998). Digital places greater emphasis on the here-and-now rather than the long-term, just-in-time information rather than just-in-case (Kuny, 1998).

The situation in which a digital resource cannot be read anymore because of the type of carrier on which it has been stored, because the hardware and software required for the readout are not used anymore, is known as the digital obsolescence phenomenon (Deljanin, 2012). S. Deljanin (2012) listed four major types of digital obsolescence: technical obsolescence and functional obsolescence – new, more efficient and more useful devices and technologies replace the existing ones; systemic obsolescence and postponement obsolescence – planned systemic obsolescence is intentional attempt to make product obsolete by changing its functional system in such a way that it makes its use harder. Postponement obsolescence is a situation in which technological improvement of product is not introduced even if it is possible. Users can decide if they want to buy replacement, to upgrade it or to start using newer software. Rusbridge (2006) listed frequently repeated statements related to the conservation of digital heritage, against which he engaged in polemics: digital preservations is very expensive (because) file formats become obsolete very rapidly (which means that) interventions must occur frequently, ensuring that continuing costs remain high. Digital preservation repositories should have very long timescale aspirations. "Internet-age" expectations are such that the preserved object must be easily and instantly accessible in the format *de jour*, and the preserved object must be faithful to the original in all respects. On the other hand, Lyons (2016) drew attention to the role of the archivists' community in preserving digital data. He noted that "the digital dark age will not happen in the way that the media predict it". We should not be blinded by a fear of the inability to ensure persistence of digital information. The digital dark age will only happen if we, as communities of archives and archivists, do not reimagine appraisal and selection in light of the historical gaps revealed in collections today. He also noted that:

> it is the digital-ness of today's world that may actually allow archives to reach out to and to document the enormity and complexity of society in a way that has never been feasible before (Lyons, 2016).

Preserving information is about maintaining its meaning – or essence – over time. It is achieved through maintaining information's usability, context and content (Howell, 2000). Data preservation (archiving) is associated with the problem pointed out by Howell (2000): it is not possible to archive all data being generated. This leads to the following decision to be taken: what to preserve from the whole gamut of old, new and future resources in a never-ending re-evaluation of their worth? There is more to the preservation of digital heritage than just the technical process of preserving digital resources over a long period of time, as it is also a social and cultural process which concerns, *inter alia*, the selection of contents to be preserved and the form in which they should be preserved (Lavoie and Dempsey, 2004). Without guidance as to what can and should be preserved, it will be difficult.

The conservation of digital heritage faces two major problems: effective data archiving and the selection of data which "should be" archived, as well as ensuring that the archived data can be read (Rosenthal, 2014). While events from bygone eras used to be recorded on paper "carriers" and archived in libraries, currently their records are generated in the digital form and frequently take the form of ephemeral digital messages: tweets, emails, comments, websites with damaged links and files that cannot be read anymore. These digital artefacts may become an obstacle to the reconstruction of historical events by creating a specific "information black hole".

Link rot and bit rot – the common element

Even though the terms bit rot and link rot concern different phenomena, they have one most important feature in common. In both cases, the problem is the inability to access specific contents or the inability to use specific digital resources. The differences that should primarily be indicated include the reasons for non-availability of particular contents or for the inability to use particular resources (Table III). In fact, when it comes to bit rot, it is not possible to read or edit a file content, mostly because of a damage to the file itself (at the software level) or a physical damage to the data carrier. Moreover, the inability to read or edit a file content may be because of the lack of software that might have been replaced by new one, or to the lack of hardware with such software installed. As regards link rot, the situation is slightly different. The unavailability of particular data or the inability to use particular resources is most frequently because of the actual absence of either a file or a

Characteristics	Bit rot	Link rot
Main causes	Data carrier failure; software decay	No target resource at the indicated location; no content or altered content
The main problem	Damages to data carriers, insufficient carrier's capacity or size, insufficient carrier's performance	Improper resource management (deletion of a resource, renaming a file, etc.); content alteration (modification)
Examples of remedies	Back-up copies, emulation, digital vellum, UVC	Digital archives, screenshots, analogue archives (e.g. paper print-outs), link audits, link maintenance
The scale of the phenomenon	Diverse: damages to particular (privately owned) data carriers but also to data server's disks (which maintain data clouds)	Diverse: damages to particular links to resources but also the close-down of long-term hosting services (deletion of hundreds of hosting accounts) and many internet domains
"Achilles' heel"	Data carrier physicalities	Structural changes in the WWW

Source: Own study

Table III.
Summary of selected attributes of the bit rot and link rot phenomena

content (an empty document or a document whose content has been altered). Moreover, problems may arise with the data server, which can prevent (sometimes temporarily) the access to particular resources. What is important, it is more likely that data lost because of bit rot are lost irretrievably. A physical damage to a data carrier is usually irreversible, while data recovery is not always possible and can be costly. In such a case, the lack of a backup copy may result in an irretrievable loss of data. What is more, it will be increasingly difficult to recreate older file formats, operating systems or other software because of the "breaking" of obsolete computer hardware. Original hardware, spare parts or software will be more and more difficult to find. Meanwhile, internet resources that are no longer available at a particular location have most likely been archived in digital archives, or are located at a different location of the Web. There is always a chance of finding them in the immeasurable resources of the internet.

Digital bumps and digital tsunami

In July 2011, social network Google+ (Google Plus) launched as a follow-up to projects such as Google Wave or Google Buzz. In December 2018, Google announced the decision to close Google+ for individual users. The reasons given included decline in the service's popularity and difficulties in its improvement. The service was shut down in April 2019. Thousands of websites had contained links directing users to social network profiles created within Google+. In February 2019, there were approx. 43,000 links to Google+ provided only on Wikipedia (through a query in a web browser window: "plus.google.com site:wikipedia.com"). Whatever happened to other sites?

Hundreds of thousands of photographs and various materials published on Google+ were no longer available. The content of individual users' accounts was deleted. Photos and films from Google+ stored in the "Album Archive" and on Google+ sites were deleted as well. While Google actually made it possible to download the shared content, many users failed to do so, e.g. because of insufficient skills. The service's shutdown led to many other consequences. For example, Google+ interactions with online and mobile applications ceased to be operational in March 2019. The "Logging In with Google+" functionality was completely withdrawn, and shut down in March 2019 as well. Developers were forced to switch to a more universal authentication system i.e. Logging In with Google. The service's shutdown also affected the Blogger service and other services, which had used Google+ to manage comments.

Google+ shutdown forcibly illustrated the peculiarity of the link rot and bit rot phenomena. Disabling the service can be equated to the "disruption of the cobweb" made up with links serving as bridges between websites. A lot of time will pass before users repair links on their websites, and many users will not do so at all. A large proportion of the content published on Google+ is irretrievably lost.

In mining, rock bumps are violent discharges of rocks' potential elasticity energy. They may also occur during various underground engineering works (Hirata *et al.*, 2007). In turn, tsunami is a very long-wavelength wave of water that is generated by sudden displacement of the seafloor or disruption of any body of standing water. They are also called "seismic sea waves", although they can be generated by mechanisms other than earthquakes (Helal and Mehanna, 2008). Tsunamis are classified into three categories: distant, regional and local (Fernandez *et al.*, 2000).

Bumps, shocks and tsunami are dangerous and undesirable events, which often have disastrous consequences. A certain analogy can be found in "digital bumps" or "digital shocks", which occur on a different scale and, like in nature, generate a (digital) tsunami with various impact ranges. Major tsunamis are relatively rare and "more discernible". They form a sort of a "gap in digital resources" (usually on servers' disks) and are usually associated with the cessation of provision of (blog, mailing, hosting) services by major service providers. The elimination of thousands of hosting accounts and/or thousands of subdomains (e.g. in the case of free hosting) results in a sudden decay of links ("digital bump", "digital shock"). Because of digital bumps, the intricately woven cobweb of links gets torn apart. A wave of repercussions in the form of a specific "digital tsunami" spreads over the internet resources. On many websites, links become dead and resources unavailable. The greater the digital bump, the further the digital tsunami's wave spreads. Closure of a large website or a failure of a major service provider's servers will have significantly more serious consequences as they will affect more users, a greater volume of data, and more links between these data.

Discussion and practical implications

There is a closed catalogue of the main causes of the decay of links (Lawrence, 2001; Rumsey, 2002; Spinellis, 2003; Sellitto, 2005) and the accompanying technical messages (Koehler, 2004). The main ways of counteracting the link rot (Leighton, 2015) and bit rot phenomena (Deljanin, 2012; Tidelift, 2018) have been developed as well. At this point, it is worth noting that the link rot and bit rot phenomena are actually related to the broadly understood resource management. The link rot phenomenon is, in a way, a consequence of poor management of (Web) resources and (hyper)links between them. The bit rot phenomenon, on the other way, can be prevented through following good practices, back-up copies and organisational procedures. This is a broad issue, which still remains to be addressed by researchers.

Both globally (from the perspective of management of an organisation or a business entity as well as from the perspective of technological development in general) and from the point of view of a particular user of digital technologies, it appears that bit rot is the phenomenon likely to have "more severe" consequences. A loss of digital data because of a data carrier damage is usually much more acute and unexpected than the inability to download a resource because of a damaged link. At this point, it is worth noting the specific "average scale" of incurred losses. There are usually more data on a digital data carrier than in the case of a single link. In addition, copies of files saved on portable disks are made less frequently than copies of internet resources saved on servers (in which case copies are usually made automatically, and the service provider can be asked to restore a specific copy of data).

The link rot phenomenon is, in a way, "embedded" in the bit rot phenomenon. Link decay occurs as part of data decay i.e. in the ecosystem in which the bit rot phenomenon never stops. Both in the first and the second case, specific "digital bumps" or "digital shocks", which cause a "digital tsunami" can be observed.

It is difficult to determine which phenomenon is worse: link rot or content drift. In both cases, the user does not receive what is indicated by the link content. However, in the case of link rot, the particular resource usually does not exist (*404 Not Found*). The message is clear. As regards content drift, however, a resource other than the expected one or even worse, a resource that has been modified from the original one, is requested; the content has changed over time, could have been replaced or modified. Zittrain *et al.* (2014) pointed out that the supplier of a source material may change their views and "update" the original source to reflect the changing views. What is important is that the user is not necessarily aware of the fact that other contents used to be originally available at a particular address. Moreover, many servers, in the absence of a website or other resource, do not return a 404 error code; instead, a substitute page and the "OK" code are presented (200). A substitute page sometimes delivers a written error message, sometimes it redirects to the home page, and sometimes it has nothing to do with the original website. Research showed that such substitutions known as "soft 404's" accounted for over 25 per cent of dead links (Bar-Yossef *et al.*, 2004).

Every link is a promise. Each unfulfilled promise, big or small, may contribute to a loss of trust and reliability. Unfulfilled promises of, for example, downloading a file, viewing a video or reading a specific text can make the user feel irritated, disappointed or even cheated. Broken links may give an impression that the website is neglected and not updated. Moreover, non-functional links may adversely affect the website's ranking in search results. On the other hand, when links fulfil their role, users navigate over the website smoothly and confidently (Pernice, 2014). Regular website maintenance should include a link audit, in particular after major updates/modernisations. The frequency of audit links should be determined according to the website's size and complexity and the frequency of introduced changes (Nielsen, 1998).

The link rot phenomenon is, to some extent, a consequence of the WWW evolution. The beginnings of WWW are referred to as Web 1.0 (the Web of documents, mostly read-only Web). Web 1.0 is a system of interconnected hypertext documents available through the internet. According to Benito-Osorio *et al.* (2013), the Web 1.0 era occurred in the years 1990-2000. From 2000 to 2010, Web 2.0 (the widely read-write Web) developed (Cormode and Krishnamurthy, 2008), while Web 3.0 (the read-write-execute Web) began in 2010 and is supposed to last until 2020 (Fuchs *et al.*, 2010). The future, however, belongs to Web 4.0 (Mobile Web) i.e. the symbiotic web, ubiquitous web, or the web capable of reading users' emotions. The evolution of WWW is based on technological changes that are some kind of "inflection points". It is at such points where changes of, for example, the so-called "website engine" (e.g. the installation of the content management system), the domain or a hosting service provider frequently occur, that there is a major risk of the link rot phenomenon occurrence. Therefore, technological changes require website administrators and editors to focus particularly on maintaining the availability of resources, including link timeliness.

The link rot phenomenon affects all actors, regardless of the scale or "significance" of the activities they are involved in. An example could be a story about a digital artefact regarding the first ever website in the world, which has "disappeared" from the internet resources i.e. ceased to be available at a specific internet address, and its files have "gone astray" somewhere on servers' disks.

The inventor of the WWW is Tim Berners-Lee, a British scientist who developed the concept of WWW in 1989 while working at the CERN (The European Organization for Nuclear Research, Geneva, Switzerland) (Berners-Lee *et al.*, 1994; Hall and Tiropanis, 2012). The first ever website is a text page on a white background, including numerous hyperlinks. It was restored from back-up copies and made available to internet users in 2013 as part of activities undertaken by the CERN. Their aim was to preserve and make available certain digital resources associated with the emergence of WWW (The Associated Press, 2013).

The deliberations also raise the question of how "deep" do the link rot and bit rot phenomena reach? The Surface Web used by internet users on a daily basis comprises resources that can be found by search engines and then offered in response to a user's query. However, traditional search engines index only a small portion of data. Other contents are "submerged" in the so-called Deep Web (Bergman, 2001; Devine and Egger-Sider, 2004). Even though the surface web connects a countless number of HTML websites, it was demonstrated that much more information can be found in databases (Bergman, 2001; Gulli and Signorini, 2005; Lewandowski and Mayr, 2006). This type of information is usually not available at a static URL address, as it is generated in real time (He *et al.*, 2007). The bit rot phenomenon poses a danger to databases and can make them inaccessible to users.

A specific panacea for the bit rot phenomena is the so-called digital vellum (the author of the concept is Vint Cerf) – a kind of X-ray snapshot of the content and the application and the operating system together, with a description of the machine that it runs on, and preserve that for long periods of time. All of the above is supposed to preserve software and hardware so that files can be preserved and readable to the generations to come (Mottl, 2015). Digital vellum assumes the preservation of hardware along with software. The hardware, however, wears out or may break down, therefore, this does not exempt anyone from having spare parts and back-up copies. On the other hand, Krebs *et al.* (2012) presented the universal virtual computer (UVC) – a simple yet powerful approach to preserve digital objects on a very long-term scale. Its main attraction is that documents do not have to be processed and transformed during their whole archive lifetime. With UVC, the main effort occurs before ingest time: rendering software for interpreting documents of a given format on UVC must be developed and archived (Krebs *et al.*, 2012, p. 153).

One can forecast that the link rot phenomenon will intensify and the internet resources will be even more unstable as more and more new, highest-level domains, which are purchased are emerging. This, however, is a relatively narrow (selective even) look at link rot. While certain solutions are indeed provided by internet archives, for example, Wayback Machine, they are not "embedded" in the Web structure, which means that many websites slip through the "cracks" (Mead, 2014). Salmon (2013) expressed concern that the entire Web was basically turning into Snapchat (a mobile application used to send videos and images that can be viewed for up to 10 s) in slow motion, where content exists for some unknown period of time before it dies and is lost forever.

Summary

The bit rot phenomenon appears to be natural and is, in a way, a consequence of progress. This progress, however, should rather proceed in an evolutionary, instead a revolutionary, fashion, which would allow hardware and data to be preserved (data transfer, processing, conversion). In the future, however, it may appear to be necessary to maintain the entire infrastructure (both hardware and software) required to read a particular file format.

In the current internet ecosystem, i.e. a computer ecosystem constructed on digital data, it is not possible to completely, or even partially, eliminate the link rot and bit rot phenomena, which are, to a large extent, a consequence of negligence or deliberate efforts. Moreover, they result, to a large extent, from the evolution of software and equipment (including computer hardware) i.e. from technological development. All of this makes them simply natural. As long as the digital ecosystem is only a tool, it will be an ecosystem parallel to the natural ecosystem, and these phenomena will go on.

A thesis can be put forward that bit rot and link rot may disappear completely or be significantly limited by a complete integration of both the digital and natural ecosystem. These phenomena could also be eliminated or minimised (marginalised) either through the complete exclusion of "physical interactions" at the digital data carrier level or through machine intelligence. Intelligent software, artificial intelligence implemented in the digital ecosystem (parallel to the natural one) could limit data decay. Perhaps the future of data storage and archiving lies in atomic storage or data recording in synthetic DNA. However, these are still relationships of the "data carrier – recording device/readout device – software and recording/readout equipment" type. And here the questions arise as to whether data always need a physical carrier, and whether they are going to need such a carrier in the future? A higher level would be the equalisation of the capabilities of a machine with those of the human (artificial intelligence, machine intelligence), and a relationship between these "systems", yet maintained not through further devices, software and cable bundles.

References

APA (2003), "Electronic sources and locator information", *American Psychological Association*, available at: www.apastyle.org/manual/related/electronic-sources.pdf (accessed 24 January 2019).

Ashenfelder, M. (2011), "The average lifespan of a webpage", *Library of Congress*, available at: https://blogs.loc.gov/thesignal/2011/11/the-average-lifespan-of-a-webpage/ (accessed 24 January 2019).

Baker, M., Keeton, K. and Martin, S. (2005), "Why traditional storage systems don't help us save stuff forever", *1st IEEE Workshop on Hot Topics in System Dependability, HPL-2005-120*.

Bar-Yossef, Z., Kumar, R., Broder, A. and Tomkins, A. (2004), "Sic transit gloria telae: towards an understanding of the web's decay", *Proceedings of the 13th international conference on World Wide Web. New York, NY*, available at: https://doi.org/10.1145/988672.988716

Benbow, S.M.P. (1998), "File not found: the problems of changing URLs for the world wide web", *Internet Research*, Vol. 8 No. 3, pp. 247-250, available at: https://doi.org/10.1108/10662249810217867

Benito-Osorio, D., Peris-Ortiz, M., Armengot, C.R. and Colino, A. (2013), "Web 5.0: the future of emotional competences in higher education", *Global Business Perspectives*, Vol. 1 No. 3, pp. 274-287, available at: https://doi.org/10.1007/s40196-013-0016-5

Bergman, M.K. (2001), "White paper: the deep web: surfacing hidden value", *Journal of Electronic Publishing*, Vol. 7 No. 1, available at: http://dx.doi.org/10.3998/3336451.0007.104

Berners-Lee, T. (1998), "Cool URIs don't change", available at: www.w3.org/Provider/Style/URI (accessed 24 January 2019).

Berners-Lee, R., Cailliau, R., Luotonen, A., Nielsen, H.F. and Secret, A. (1994), "The world-wide web", *Communications of the Acm*, Vol. 37 No. 8, pp. 76-82, available at: http://dx.doi.org/10.1145/179606.179671

Bowers, J. (2017), "A million squandered: the million dollar homepage as a decaying digital artifact", *Library Innovation Lab*, available at: https://goo.gl/rppeQV (accessed 24 January 2019).

Burnhill, P., Mewissen, M. and Wincewicz, R. (2015), "Reference rot in scholarly statement: threat and remedy", *Insights the UKSG Journal*, Vol. 28 No. 2, pp. 55-61, available at: https://doi.org/10.1629/uksg.237

Cormode, G. and Krishnamurthy, B. (2008), "Key differences between web 1.0 and web 2.0", *First Monday*, Vol. 13 No. 6, available at: https://doi.org/10.5210/fm.v13i6.2125

Deegan, M., Tanner, S. (Eds) (2006), *Digital Preservation*, Facet Publishing, London.

Deljanin, S. (2012), "Digital obsolescence", *INFOtheca*, Vol. 13 No. 1, pp. 43-53.

Denemark, H. (1996), "The death of law reviews has been predicted: what might be lost when the last law review shuts down?", *Seton Hall Law Review*, Vol. 27 No. 1, pp. 1-32.

Devine, J. and Egger-Sider, F. (2004), "Beyond google: the invisible web in the academic library", *The Journal of Academic Librarianship*, Vol. 30 No. 4, pp. 265-269, available at: https://doi.org/10.1016/j.acalib.2004.04.010

Fernandez, M., Molina, E., Havskov, J. and Atakan, K. (2000), "Tsunamis and tsunami hazards in Central America", *Natural Hazards*, Vol. 22 No. 2, pp. 91-116, available at: https://doi.org/10.1023/A:1008102600622

Fichter, F. (1999), "Do I look like a maid? Strategies for preventing link rot", *Online*, Vol. 23, pp. 77-79.

Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M. and Bichler, R. (2010), "Theoretical foundations of the web: cognition, communication, and co-operation. Towards an understanding of web 1.0, 2.0, 3.0", *Future Internet*, Vol. 2 No. 1, pp. 41-59, available at: https://doi.org/10.3390/fi2010041

Goh, D.H.L. and Ng, P.K. (2007), "Link decay in leading information science journals", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 1, pp. 15-24, available at: https://doi.org/10.1002/asi.20513

Gulli, A. and Signorini, A. (2005), "The indexable web is more than 11.5 billion pages", *Proceeding WWW'05 Special interest tracks and posters of the 14th international conference on World Wide Web*, pp. 902-903, available at: https://doi.org/10.1145/1062745.1062789

Hall, W. and Tiropanis, T. (2012), "Web evolution and web science", *Computer Networks*, Vol. 56 No. 18, pp. 3859-3865, available at: http://dx.doi.org/10.1016/j.comnet.2012.10.004

Haslop, B., Schnabel, M.A. and Aydin, S. (2017), "Digital decay", *International Journal of Parallel Emergent and Distributed Systems*, Vol. 31 No. 6, pp. 149-156, available at: https://doi.org/10.1080/17445760.2017.1390092

Hayes, B. (1998), "Bit rot", *American Scientist*, Vol. 86 No. 5, pp. 410-415.

He, B., Patel, M., Zhang, Z. and Chang, K.C.C. (2007), "Accessing the deep web: a survey", *Communications of the Acm*, Vol. 50 No. 5, pp. 94-101, available at: https://doi.org/10.1145/1230819.1241670

Hedstrom, M. (1997), "Digital preservation: a time bomb for digital libraries", *Computers and the Humanities*, Vol. 31 No. 3, p. 189, available at: https://doi.org/10.1023/A:1000676723815

Heikkila, A. (2018), "How to prevent the digital dark ages", *Innovation Excellence*, available at: https://www.innovationexcellence.com/blog/2018/03/01/how-to-prevent-the-digital-dark-ages/ (accessed 24 January 2019).

Helal, M.A. and Mehanna, M.S. (2008), "Tsunamis from nature to physics", *Chaos, Solitons and Fractals*, Vol. 36 No. 4, pp. 787-796, available at: https://doi.org/10.1016/j.chaos.2007.08.044

Hennessey, J. and Ge, S.X. (2013), "A cross disciplinary study of link decay and the effectiveness of mitigation techniques", *BMC Bioinformatics*, Vol. 14 No. 5, available at: https://doi.org/10.1186/1471-2105-14-S14-S5

Hildenbrand, J. (2017), "Bit rot explained: why your phone is slower than when it was newAndroidcentral", *Mobile Nations*, available at: https://goo.gl/rdeXYq (accessed 24 January 2019).

Hirata, A., Kameoka, Y. and Hirano, T. (2007), "Safety management based on detection of possible rock bursts by AE monitoring during tunnel excavation", *Rock Mechanics and Rock Engineering*, Vol. 40 No. 6, pp. 563-576, available at: https://doi.org/10.1007/s00603-006-0122-7

Howell, A. (2000), "Perfect one day – digital the next: challenges in preserving digital information", *Australian Academic and Research Libraries*, Vol. 31 No. 4, pp. 121-141, available at: https://doi.org/10.1080/00048623.2000.10755130

Hudgins, A.M. (2011), "Preservation of the video game", *Provenance, Journal of the Society of Georgia Archivists*, Vol. 29 No. 1, pp. 32-48, available at: https://digitalcommons.kennesaw.edu/provenance/vol29/iss1/4 (accessed 24 January 2019).

Jacobson, I., Christerson, M., Jonsson, P. and Övergaard, G. (1992), *Object-Oriented software Engineering: a Use Case Driven Approach*, ACM Press, Addison-Wesley.

Knight, S. (2010), "Early learnings from the national library of New Zealand's national digital heritage archive project", *Program*, Vol. 44 No. 2, pp. 85-97, available at: https://doi.org/10.1108/00330331011039454

Kobayashi, M. and Takeda, K. (2002), "Information retrieval on the web", *ACM Computing Surveys*, Vol. 32 No. 2, pp. 144-173, available at: https://doi.org/10.1145/358923.358934

Koehler, W. (2004), "A longitudinal study of web pages continued: a consideration of document persistence", *Information Research*, Vol. 9 No. 2, p. 24 paper 174, available at: www.informationr.net/ir/9-2/paper174.html (accessed January 2019).

Krebs, N., Schmitz, L. and Borghoff, U.M. (2012), "Implementing the universal virtual computer", in Moreno-Díaz R., Pichler F., Quesada-Arencibia A. (Eds), *Computer Aided Systems Theory – EUROCAST 2011. EUROCAST 2011. Lecture Notes in Computer Science*, Springer, Berlin, Vol. 6927, available at: https://doi.org/10.1007/978-3-642-27549-4_20

Kuny, T. (1998), "The digital dark ages? Challenges in the preservation of electronic information", *International Preservation News*, No. 17, pp. 8-13.

Lavoie, B., Dempsey, L. (2004), "Thirteen ways of looking at... digital preservation", *D-LIB*, Vol. 10 Nos 7/8, p. 24, available at: www.dlib.org/dlib/july04/lavoie/07lavoie.html (accessed January 2019).

Lawrence, S. (2001), "Free online availability substantially increases a paper's impact", *Nature*, Vol. 411 No. 6837, p. 521.

Leighton, W.K. (2015), "The growing problem of internet 'link rot' and best practices for media and online publishers", Journalist's Resource, available at: https://goo.gl/n1T3aP (accessed 24 January 2019).

Leung, S., Perl, S., Stata, R. and Wiener, J.L. (2001), "Towards web-scale web archaeology", *SCR Research Report*, Vol. 174, pp. 1-5.

Lewandowski, D. and Mayr, P. (2006), "Exploring the academic invisible web", *Library Hi Tech*, Vol. 24 No. 4, pp. 529-539, available at: https://doi.org/10.1108/07378830610715392

Lor, P.J. and Britz, J.J. (2012), "An ethical perspective on political-economic issues in the long-term preservation of digital heritage", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 11, pp. 2153-2164, available at: https://doi.org/10.1002/asi.22725

Lyman, P. and Kahle, B. (1998), "Archiving digital cultural artifacts: organizing an agenda for action", *D-Lib Magazine*, available at: www.dlib.org/dlib/july98/07lyman.html (accessed 24 January 2019).

Lyons, B. (2016), "There will be no digital dark age", *Issues and Advocacy*, available at: https://issuesandadvocacy.wordpress.com/2016/05/11/there-will-be-no-digital-dark-age/ (accessed 24 January 2019).

Markwell, J. and Brooks, D.W. (2002), "Broken links: the ephemeral nature of educational WWW hyperlinks", *Journal of Science Education and Technology*, Vol. 11 No. 2, pp. 105-108, available at: https://doi.org/10.1023/A:1014627511641

Markwell, J. and Brooks, D.W. (2003), "Link rot' limits the usefulness of web-based educational materials in biochemistry and molecular biology", *Biochemistry and Molecular Biology Education*, Vol. 31 No. 1, pp. 69-72, available at: https://doi.org/10.1002/bmb.2003.494031010165

Mead, D. (2013), "The supreme court has a serious problem with link rot", Motherboard Blog, available at: https://goo.gl/R6EjGT (accessed 24 January 2019).

Mead, D. (2014), "When will the internet defeat link rot?", Motherboard Blog, available at: https://goo.gl/JQiGL9 (accessed 24 January 2019).

Merchant, B. (2014), "What it was like to surf the web in 1989", Motherboard Blog, available at: https://goo.gl/RSt2Pf (accessed 24 January 2019).

Mottl, J. (2015), "Internet pioneer Vint Cerf warns of bit rot and digital dark age but don't panic yet", Tech Times, available at: https://goo.gl/Wmo4nu (accessed 24 January 2019).

Nielsen, J. (1998), "Fighting linkrot", Nielsen Norman Group, available at: www.nngroup.com/articles/fighting-linkrot/ (accessed 24 January 2019).

Odersky, M. and Moors, A. (2009), "Fighting bit rot with types (experience report: scala collections)", *Leibniz International Proceedings in Informatics*, Vol. 4, pp. 427-451, available at: https://doi. org/10.4230/LIPIcs.FSTTCS.2009.2338

Parker, A. (2007), "Link rot: how the inaccessibility of electronic citations affects the quality of New Zealand scholarly literature", *New Zealand Library and Information Management Journal*, Vol. 50 No. 2, pp. 172-192.

Pernice, K. (2014), "A link is a promise", Nielsen Norman Group, available at: www.nngroup.com/articles/link-promise/ (accessed 24 January 2019).

Pitt, R. (2019), "From moldy books to bit rot", available at: https://goo.gl/fZZ1eu (accessed 24 January 2019).

Rosenthal, D. (2014), "The half-empty archive", DSHR'S BLOG, available at: http://bit.do/HalfEmptyArchive (accessed 24 January 2019).

Rouse, M. (2019), "Definition bit rot", TechTarget, available at: https://searchstorage.techtarget.com/definition/bit-rot (accessed 24 January 2019).

Routley, N. (2017), "Error 404: a look at digital decay", Visual Capitalist, available at: www.visualcapitalist.com/digital-decay/ (accessed 24 January 2019).

Rumsey, M. (2002), "Runaway train: problems of permanence, accessibility, and stability in the use of web sources in law review citations", *Law Library Journal*, Vol. 94 No. 1, pp. 27-39.

Rusbridge, C. (2006), "Excuse me... some digital preservation fallacies?", *Ariadne (Web Magazine for Information Professionals)*, No. 46, available at: www.ariadne.ac.uk/issue/46/rusbridge/ (accessed 24 January 2019).

Salmon, F. (2013), "The spread of link rot", Reuters Blog, available at: https://goo.gl/VKrsWR (accessed 24 January 2019).

Schlieder, C. (2010), "Digital heritage: semantic challenges of long-term preservation", *Semantic Web*, Vol. 1 Nos 1/2, pp. 143-147, available at: https://doi.org/10.3233/SW-2010-0013

Sellitto, C. (2005), "The impact of impermanent web-located citations: a study of 123 scholarly conference publications", *Journal of the American Society for Information Science and Technology*, Vol. 56 No. 7, pp. 695-703, available at: https://doi.org/10.1002/asi.20159

Spacey, J. (2017), "6 types of bit rot", Simpicable, available at: https://simplicable.com/new/bit-rot (accessed 24 January 2019).

Spinellis, D. (2003), "The decay and failures of web references", *Communications of the Acm*, Vol. 46 No. 1, pp. 71-77, available at: https://doi.org/10.1145/602421.602422

Stringfellow, A. (2018), "Digital decay: understanding digital decay, its impacts on modern business, and best practices for preserving digital assets and data", *MerlinOne*, available at: https://merlinone.com/digital-decay/ (accessed 24 January 2019).

Tait, E., MacLeod, M., Beel, D., Wallace, C., Mellish, C. and Taylor, S. (2013), "Linking to the past: an analysis of community digital heritage initiatives", *Aslib Proceedings: New Information Perspectives*, Vol. 65 No. 6, pp. 564-580, available at: https://doi.org/10.1108/AP-05-2013-0039

Taylor, M.K. and Hudson, D. (2000), "Linkrot and the usefulness of web site bibliographies", *Reference and User Services Quarterly*, Vol. 39 No. 3, pp. 273-277.

Tew, A. (2005), "The million dollar homepage – own a piece of internet history!", available at: www.milliondollarhomepage.com/ (accessed 24 January 2019).

The Associated Press (2013), "CERN searches for 1st web page", CBC/Radio-Canada, available at: http://bit.ly/2wuGzGa (accessed 26 July 2019).

Tidelift (2018), "Bit rot: the silent killer", Tidelift, Inc., available at: https://blog.tidelift.com/bit-rot-the-silent-killer (accessed 24 January 2019).

Wernick, A. (2018), "Scientists warn we're entering a digital dark age", *The Week*, available at: https://goo.gl/URWTqY (accessed 24 January 2019).

Whitt, R.S. (2016), "Through a glass, darkly: technical, policy, and financial actions to avert the coming digital dark ages", *Santa Clara High Technology Law Journal*, Vol. 33 No. 2, pp. 117-229.

Yanofsky, D. (2014), "The million dollar homepage still exists, but 22% of it has rotted awal", Quartz, available at: https://goo.gl/rycFbh (accessed 24 January 2019).

Young, L.J. (2017), "Data reawakening. The 'file not found' series: part 3 of 3", Science Friday, available at: https://apps.sciencefriday.com/data/reawakening.html (accessed 24 January 2019).

Zhou, K., Grover, C., Klein, M. and Tobin, R. (2015), "No more 404s: predicting referenced link rot in scholarly articles for pro-active archiving", in *Proceedings of the 15th ACM/IEEE-CE on Joint Conference on Digital Libraries. JCDL'15, ACM, New York, NY*, pp. 233-236, available at: https://doi.org/10.1145/2756406.2756940

Zittrain, J., Albert, K. and Lessig, L. (2014), "Perma: scoping and addressing the problem of link and reference rot in legal citations", *Legal Information Management*, Vol. 14 No. 2, pp. 88-99, available at: https://doi.org/10.1017/S1472669614000255

Corresponding author
Karol Król can be contacted at: k.krol@onet.com.pl