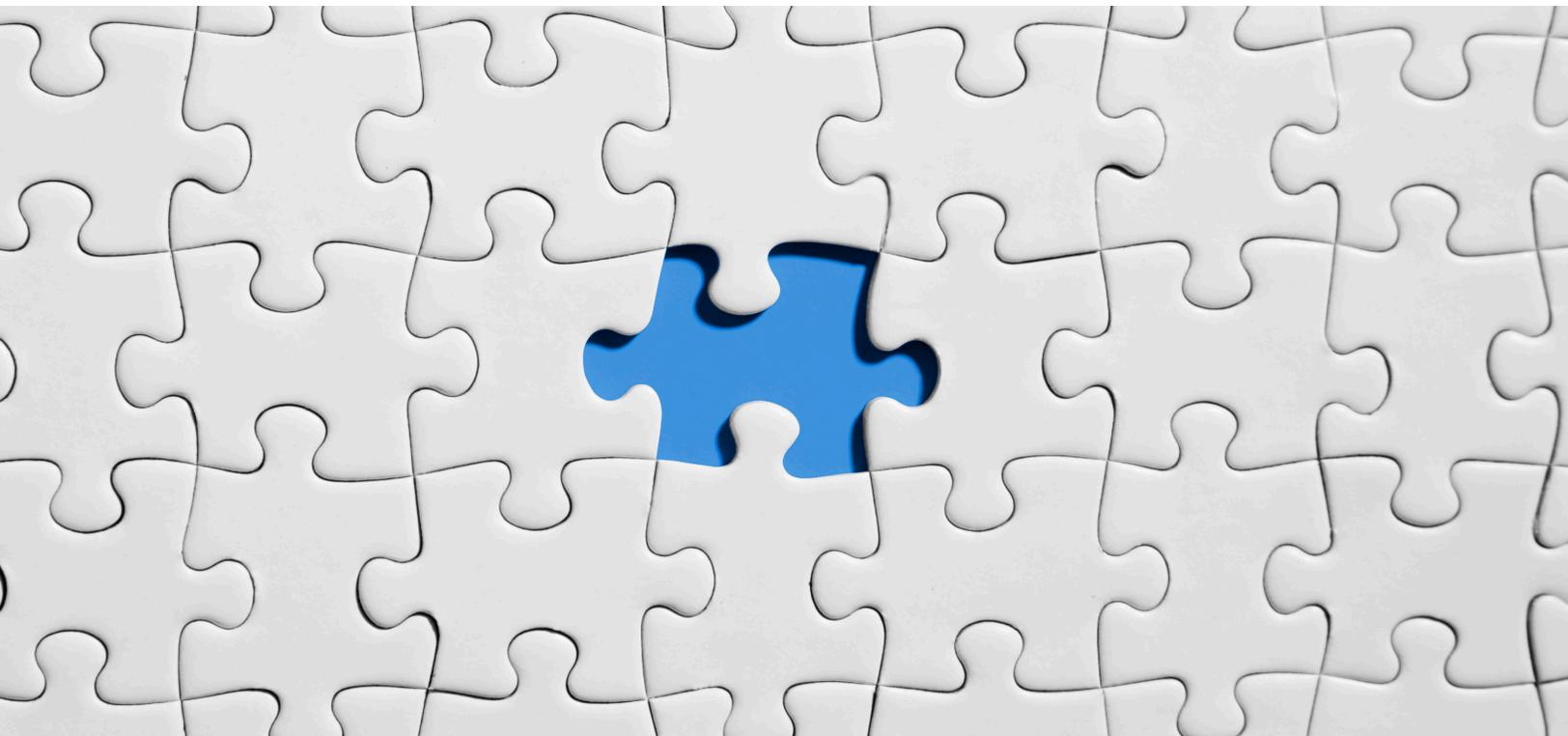


LONG-TERM RETENTION AND ARCHIVING DIGITAL FILES

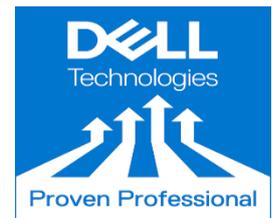


Vinicius Assis

Alliance Lead

Dell Technologies

V_assis@dell.com



The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Learn more at www.dell.com/certification

“Whatever practical men say, this world is, after all, absolutely governed by ideas, and often the most speculative and hypothetical ideas. It is a matter of the utmost importance that our theories about things, even about things that seem very distant from everyday life, must be as true as possible and as far as possible away from error”

T. H. HUXLEY (1893)

Abstract

Companies around the world are becoming increasingly concerned about getting media to protect data from damage caused by unintentional or intentional information loss. Over the years they have created various technologies, including storage devices, automation software and strategies to protect the information that constitute backups. These copies are designed to allow data to be stored on media assumed to be reliable and that can be accessed in case of loss from the original source. Spending by the industry with the use of technologies for data loss prevention reaches billions of dollars annually, and its return is highly sensitive. While acknowledging that the data that generates the financial information is invaluable, it is almost impossible to estimate the rate of financial return for data protection technologies. Since the financial loss only occurs with data loss, calculating Return on Investment for data protection technologies is abstract. Along with this we observe the phenomenon of developing various software for companies around the globe. These software generate files with different structures that are able to be manipulated only by the software companies that created them. Companies that generate files with important data through these private programs create a technological passive over time, forcing keep maintenance, development and licensing costs of software to guard the impossibility of access to your data. The steps directions creating universal files in open formats absolved from proprietary platforms are slow, and more and more companies and users based on their information in proprietary software. The study shows that even with safeguards physical files, access them in the long term may not be possible. International organizations are touting file patterns but technological evolution is fast enough to disregard these standards seeking supply novelties. The proposed work is the creation of a global safeguard standards and methods of file protection in the long term.

Table of Contents

Abstract	4
Tables.....	6
Figures	6
Glossary	7
Introduction.....	8
2 Programs and File Extensions	11
2.1 Types of programs and their functionalities	11
2.2 File extensions	13
3 Software Development Process and Definition Of Use Of Files	18
4 Duplication Of Extensions and Compatibility of File Versions.....	19
4.1 File extension duplication.....	19
4.2 Retro compatibility and free access to reading	25
5 Solution Projects and Initiatives	28
5.1 Open file formats.....	29
5.2 Microsoft and XML for Office.....	31
Opening the PDF file.....	33
Final Considerations	35
References	37

Disclaimer: The views, processes or methodologies published in this article are those of the author. They do not necessarily reflect Dell Technologies' views, processes or methodologies.

Tables

Table 1 – Examples of Files used by an information technology professional	17
Table 2 - List of file types of the Brazilian Federal Revenue program.....	22
Table 3 – Comparison between Microsoft Office file extensions and the Open Document standard.....	32

Figures

Figure 1 - Example of a file using an extension.....	13
Figure 2 - Examples of files used by an information technology professional.....	16
Figure 3 - Some types of files Generated by the Brazilian Federal Revenue program.....	21
Figure 4 - Identification of DEC files by the file-extensions.org website.....	23
Figure 5 - Identification of REC files by the file-extensions.org website.....	23

Glossary

ABNT – Associação BRAZILeira de Normas Técnicas (Brazilian Association of Technical Standards)

CRM – *Customer relationship management*

DEC – Digital Equipment Corporation

ERP – *Enterprise Resource Planing*

USA – United States of America

FSF - Free Software Foundation

IEC - International Electrotechnical Commission

IRPF – Imposto de Renda Pessoa Física (Individual Income Tax)

ISO – International Organization for Standardization

MIT – Massachusetts Institute of Technology

NBR – Normas BRAZILeira (Brazilian Standards)

OASIS – Organization for the Advancement of Structured Information Standards

ODF – Open Document Format for Office Applications

OOXML - Office Open XML

PDF – Portable Document Format

RFB – Receita Federal do BRAZIL (Brazil's Federal Revenue)

IT – Information Technologies

XML - Extensible Markup Language

Introduction

Information technology enables companies and users to create large collections of documents in digital files, without transporting this content to a physical media. Documents are created, disseminated and stored in electronic media, creating silos of information on companies, personal computers or public and private web-based systems. Today, reports, spreadsheets and even books are rarely printed, creating new process needs for saving and retrieving files. Corporate libraries are transitioning toward digitization not only in migration of support of their files from paper to digital, but in the creation of these files already exclusively on electronic media.

Companies worldwide are increasingly concerned with finding ways to protect their digital files containing business information against losses caused by various actions; human error, fraud, problems related to storage equipment and obsolescence of software capable of “reading” these files. Users are also concerned with keeping their personal files protected, which become part of their professional and personal activities, with high added value. To solve this, over the years, the computer industry has created several technologies that include storage devices, automation software and data protection strategies that seek to minimize these losses by offering the possibility to protect files and recover data in the event of an accident.

These technologies aim to make it possible for files to be stored on trusted media, and be accessible in the event of loss of original support. The projects carried out by companies that include the consumption of these technologies turn over billions of dollars annually, and their return is highly sensitive, since the file that stores the data gives access to information and has incalculable value. The *MIT Technology Review* estimates that an average American office worker generates 17.5 GB of data per year ¹.

Here’s what happened. First, the amount of data created each year has grown exponentially; it reached 2.8 zettabytes in 2012, a number that’s as gigantic as it sounds, and will double again by 2015, according to IDC. Of that, about three quarters is generated by individuals as they create and move digital files. A typical American office worker produces 1.8 million megabytes of data each year. That is about 5,000 megabytes a day, including downloaded movies, Word files, e-mail, and the bits generated by computers as that information is moved across mobile networks or the Internet (TUCKER, 2013, p.2)

¹ <<http://www.technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/>>.

This can result in thousands of corporate documents generated with information that needs to be preserved due to business interests, legislation or even preservation of culture and standards.

Consider this: if this amount of data generated by the typical worker was solely from Microsoft Word documents, 10 pages each, completely filled with text, the volume would result in 1,025,144 documents².

A phenomenon observed over the years in the software industry is the great speed in the creation and modification of systems with the launch of numerous versions to add new functionalities, correct errors or adapt the software to the context of the market in which it operates. These modifications can induce a change in the types of files generated or manipulated by this software causing a rupture between the use of new versions of systems and the possibility of manipulating existing data in files from previous versions.

Changes in current behavior also pose a new challenge; data is generated on web-based systems without generating a file, in a model popularly known as “cloud computing”, removing from the user access to the file containing your data and making it available on platforms closed and unknown.

In this context, this Knowledge Sharing article will analyze actions being taken for long-term storage of electronic files, including entities responsible for standardization, hardware and software technologies and storage processes. A preliminary analysis will be carried out between groups of software that generate and do not generate files, in addition to highlighting real cases that occurred in the industry seeking to solve the problems of lack of standardization in extensions and nomenclature. Physical storage technologies are highlighted and an explanation of the view of technology professionals.

²For the calculation, 15.5 gigabytes were considered, resulting in 18350080 kilobytes, with a notation where one kilo is equivalent to 1024. Each Word document with 10 pages of plain text has a total size of 17.9 kilobytes.

The author's experience as a specialist in the information technology project market sometimes makes it difficult to distinguish between what was obtained from an external source or from personal experience. Because it is a developed theme, almost entirely by market agents, and not from academia, very little has been published in IT about its evolution. which On one hand, this makes it difficult to (re) historical and theoretical construction. On the other, it demonstrates the relevance of studying and publishing on this subject.

The research was, therefore, of an exploratory nature, which aims to “gather data, information, patterns, ideas or hypotheses about a research question with little or no previous study (BRAGA, 2007, p.25). Research can also be considered predictive, as it tries to apply to future situations, “analyzes and, based on this analysis, predicts phenomena or predicts the probability of occurrences (...)”(BRAGA, 2007, p. 26).

234/5000

The methodology consists of document analysis of a very specific type; scattered, unstructured documents, of commercial, manual and/or informal origin in collaborative social networks, in addition to historical analysis of technology.

2 Programs and File Extensions

To better understand the long-term technological aspect of the subject of digital storage, some definitions regarding software, hardware and programs that generate digital files will be presented.

2.1 Types of programs and their functionalities

Software or computer programs can be defined as:

Software is a sentence written in a computable language, for which there is a (computable) machine capable of interpreting it. The sentence (the software) consists of a sequence of instructions (commands) and data declarations, which can be stored in digital media. When interpreting the software, the computable machine is directed to perform specifically planned tasks, for which the software was designed (FERNANDES, 2002, p.1).

Hardware can be understood by tool or equipment. According to Silberschatz; Galvin, Cagne (2008), in electronics, mainly in the area of computing, the term refers to the computers' central processing unit (CPU), their storage devices, input and output.

The primary function of any computer program is to perform tasks in an automated way to interact with devices, solve problems, minimize efforts or perform tasks faster than if performed manually. To better identify the types of existing programs, these can be classified into two main categories³.

The first consists of a set of programs that interact with hardware allowing its control, i.e. the operating system and drivers that enable the operating system to control devices. Some authors such as Tanenbaum (1999) conceptualize an operating system as an abstraction of hardware for the user, making all the intermediation between the physical components of the computer and the programs.

The second category involves programs that operate on an operating system and abstract the hardware involved from the user, allowing the performance of specific tasks, called applications. Unlike the first category, programs of the latter type are used to solve business problems, provide entertainment or solve problems in a user's

³ <<http://www.webopedia.com/TERM/S/software.html>>

daily life. They are not specifically developed to relate to computer hardware but, for example, to perform simulation calculations, control accounting spreadsheets, write reports, schedule tasks, play games and perform financial transactions.

On computers, data of any kind – text, audio and images – are stored in files. According to Tanenbaum (2008), a file is “a digital, physical element that occupies space in a storage medium and may not coexist with other files, that is, it is not virtual and they are (sic) an abstraction mechanism” (p. 446). The user's inability or restriction to data access does not characterize that it does not exist, but rather that it is stored in a medium with restricted access.

The programs are capable of manipulating, redirecting or modifying data for users to conduct analysis and make their decisions. To accommodate the large number of specific problems to be solved, the programs can also be categorized as business, entertainment, health, industrial automation, and others. Specifically for this work, two large groups were categorized that do not deal with the identification of features. The categorization below, therefore, does not appear in any academic technical manual or work, but was built within the present work.

Group 01: Such programs aim to process information without exporting data in files for the user. Users of these types of applications send data to be processed and stored, but do not obtain return files for storage in response to this submission. Examples of this set includes web-based applications, i.e. digital social networks, Facebook and twitter, and online stores, i.e. business systems like *Enterprise Resource Planing* (ERP) and cloud-based *Customer relationship management* (CRM), such SalesForce. When entering the data, the user does not receive a return file with the resulting processing from the system. These data are stored in specific systems, structured or unstructured, and are accessed only by the original application. While users know that the data exists and can be accessed through a specific program, they do not have access to the data storage file for any purpose. Some programs perform tasks to export data in files, usually due to legal requirements, but users are unable to manipulate these files, as they do not own or have access to the original program.

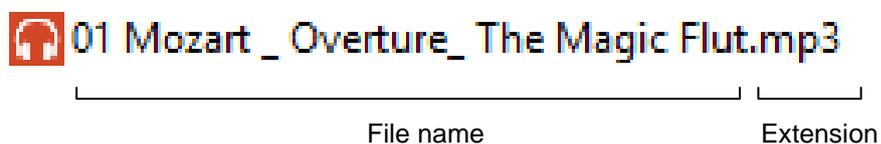
Group 02: This category includes applications that generate a return file according to the users' input data. Office applications such as Microsoft Office, including Excel, Word, Outlook and PowerPoint, are the most common in this category. The information is saved on a specific computer that will serve as storage and, for each new document, a file is generated. For every query to the data, the file must be accessed, as it is a repository for the storage and manipulation of the content. It can be distributed to other users who have programs capable of handling these files.

Group 2 software, which generates output files from user input, is the central object of this article.

2.2 File extensions

*Digital Equipment Corporation*⁴, commonly known as Digital or DEC, was a North American pioneer in the computing area, with its peak market presence and sales occurring in the 1970s to the 1980s. With the development of operating systems and several file manipulation programs, Digital faced a problem: it had a number of programs that generated different outputs and performed different operations. In order to identify which programs could manipulate certain files, which contained pre-defined or not coding structures identifiable only by certain programs of its own, Digital created the file_name-dot-extension_naming convention or file.extension. This convention ruled that files would have a prefix-based name structure. After the last point of the file name, starting from left to right, the following characters would be the file identifier associating it with a specific program capable of using it. The operating system should implement this convention to associate files with their corresponding programs.

Figure 1 - Example of a file using an extension



SOURCE: Author's collection

⁴ <<http://dec.fileextensionguide.com/>>

Microsoft Corporation in the 1980s had the same problem. With the acquisition of the DOS operating system that used files with extensions based on the Digital model, all Microsoft products started to use this standard, including its adoption in the Office application suite. With the huge success of Microsoft's programs, several other software developers followed this convention, spreading the practice of using extensions and creating a standard in the software market, both for creation and for use.

In addition to programs related to the Microsoft Windows operating system, which fully adopt the file extension convention, other types of operating systems such as UNIX, ignore this convention and do not adopt extension in their files that are not structured. Thus, the programs are fully responsible for the interpretation of the files, leaving the user with the task of choosing which programs will be used to process certain files. This type of operating system operates in a niche market. Its application is based heavily on server application functions, and its use by users for business tasks is not common. As participation of the Microsoft Windows operating system is estimated at more than 90% (FORBES, 2013), the extension-based file convention is prevalent.

With this standardization, the collections of files started to consist of files of different extensions. While physical libraries basically have physical documents, containing all kinds of information such as text, spreadsheet and images in the same medium (paper), digital collections now have the same spreadsheets, images and texts in separate files, supported by different programs with different operating system versions and platforms.

According to the website [file-extensions.org](http://www.file-extensions.org)⁵, which brings together, through the collaboration of users, a list of file extensions used worldwide, on 05/05/2013 there were 26,296 file extensions cataloged in 39 categories! The category with the greatest impact in this analysis is Documents, which comprises text files, spreadsheets, portable document format (PDF), web pages, presentations, graphics, etc. In this category alone, there are 1,476 file extensions.

⁵<http://www.file-extensions.org/filetypes>

Example

The following are some documents that could be on any user's computer, even an IT professional. The files are used by professionals and individuals, some are in common formats, recognized by other users, while there are exceptions, with files in proprietary format and restricted to a group of users.

Figure 2 shows the icons, names and extensions as presented by the Microsoft Windows operating system. Table 1 shows the meanings (file types) of the extensions.

Figure 2 - Examples of files used by an information technology professional

 SFDC 050313.xlsx

 01 Mozart _ Overture_ The Magic Flut.mp3



barco.jpg

 HP_POwerAdvisor_BL_SMC.html

 Configuracao_DataCenter_HP_ProjetoNCC_MAI_2012_ASSIS_v1.sdd

 HP UE603E Datasheet.4AA0-5969PTL.pdf

 platform.ini

 Micro_Tipo4_V2-9_Pub100913_Val180d.doc

SOURCE: author's collection

Table 1 – Examples of Files used by an IT professional

File Name	File Type
SFDC 050313.xlsx	Spreadsheet file in Microsoft Excel OpenXML format. It is a corporate file that must be filed for future reference. Any user with a program compatible with the OpenXML format will be able to manipulate this file.
01 Mozart_Overture_ The Magic Flut.mp3	Audio file in MPEG Layer 3 format controlled by Thomson Consumer Electronics. It is a personal file with no need for storage by companies, but with value for the user. Any user with a program compatible with the MPEG Layer 3 format, usually an audio program, will be able to open this file.
HP_POWerAdvisor_BL_SMC.html	A file in web page format. It is a corporate file that must be filed for future reference. Any user with an HTML-compatible program, such as Web browsers, will be able to manipulate this file.
Configuracao_DataCenter_HP_Proje toNCC_MAI_2012_ASSIS_v1.sdd	File containing project definitions proprietary to Hewlett-Packard. It is a corporate file that must be filed for future reference. Only users with the right to use the HP Sales Builder for Windows software will be able to manipulate this file.
platform.ini	File that contains program configuration settings for the Microsoft Windows operating system. It is a system file that does not need to be archived for future reference. Any user with access to that operating system will be able to manipulate that file.
Micro_Tipo4_V2-9_Pub100913_Val180d.doc	Document file in proprietary Microsoft Word format. It is a corporate file that must be filed for future reference. Unlike the spreadsheet in this table, which uses the open OpenXML format, this file is in a proprietary format and requires the user to use Microsoft Word to manipulate this file.

SOURCE: Author's elaboration

3 Software Development Process and Definition Of Use Of Files

Software that generates files to store information, generally has a model of construction of these files in a closed and proprietary format, aiming at the protection of intellectual and industrial property. This model is controlled and defended by the interests of the companies that developed the format, which can be commercial, strategic, marketing, among others – which impacts the possibility of manipulating these files by other companies' programs. When the application is developed and there is a need to create output files for storing information, the developing company must make the decision on which data structure model the file will contain.

This structure is defined based on several factors, from the storage capacity of the information treated by the software, to the ability of third parties to manipulate these files. It is a decision between using a proprietary format, where the file will be recorded and understood only by programs from that company or from third party programs that hold copyright for that use, or using open formats that allow manipulation of these files by various programs that comprise these formats.

By defining a file in a closed and proprietary format, a company can create a high-cost technological liability for its users. In addition to maintaining software licensing costs, it directly affects the business of organizations: how to ensure access to data recorded in files of proprietary formats after a long period of years, even if they are protected by conventional IT standards dealing with secure storage of files.

Some types of files with a large presence in the computing environment, such as PDF and DOC files currently have their standards open so that third-party programs can perform actions such as execution for reading, printing and in some cases editing. Other files such as AutoDesk's DraWinG (DWG), originally developed in 1982, still remain a closed proprietary format, so only users who have specific software from this manufacturer will be able to run these files.

4 Duplication Of Extensions and Compatibility of File Versions

Problems related to electronic document archiving are diverse, from processes for keeping, searching and indexing to occurrences in the technical environment of generating and maintaining software to access these files. Two problems related to long-term storage will be presented, including practical examples, one related to the technological aspect and the other to the process of keeping these digital documents.

4.1 File extension duplication

Until 2015, there was no regulatory committee or agency that controlled file extensions and structures generated by computer programs. The reasons for this lack of regulation have not been identified, despite the great need for centralized standardization worldwide to create new extensions. The lack of communication between companies that develop programs, together with deregulation and the lack of a public database of available extensions to be used causes several problems, i.e. the possibility of two companies developing different systems that publish files with the same extension or structure. This is common, but not exclusive, in smaller companies that do not have efficient global communication with other companies and have restrictions in the area of research and development.

The software industry creates thousands of programs annually that generate and manipulate an endless number of files. Creating a pattern of structures and extensions is a challenge for all companies involved. A small local company in the interior of Brazil can create a file structure manipulated only by its software. As there is no regulatory committee, this company can choose any file extension for its software. Since extensions with three or four characters are currently used, the possibility of another company, anywhere on the globe, going through the same situation and choosing the same set of characters to represent their files is high. If this happens, one of the companies should give up the right to use the extension, as the software will identify a file from another manufacturer as its own and will not be able to manipulate it, causing problems for the user. Likewise, if the two companies choose to develop programs that generate files with similar structures, problems of intellectual law may arise.

Some companies and users, for different initiatives, created websites with databases containing the extensions currently used and the programs associated with these extensions,

capable of manipulating them. The most prominent case is the website File-Extensions.org which contains a database with tens of thousands of cataloged extensions. According to the website, there is no support team and users are encouraged to send new extensions created or not present in the catalog⁶.

We are trying to provide as much informations as possible for all of the file extensions records in our database and we also encourage site visitors to suggest a new file extension or any other additional information, which can help us improve our database. For such help, we are very thankful.

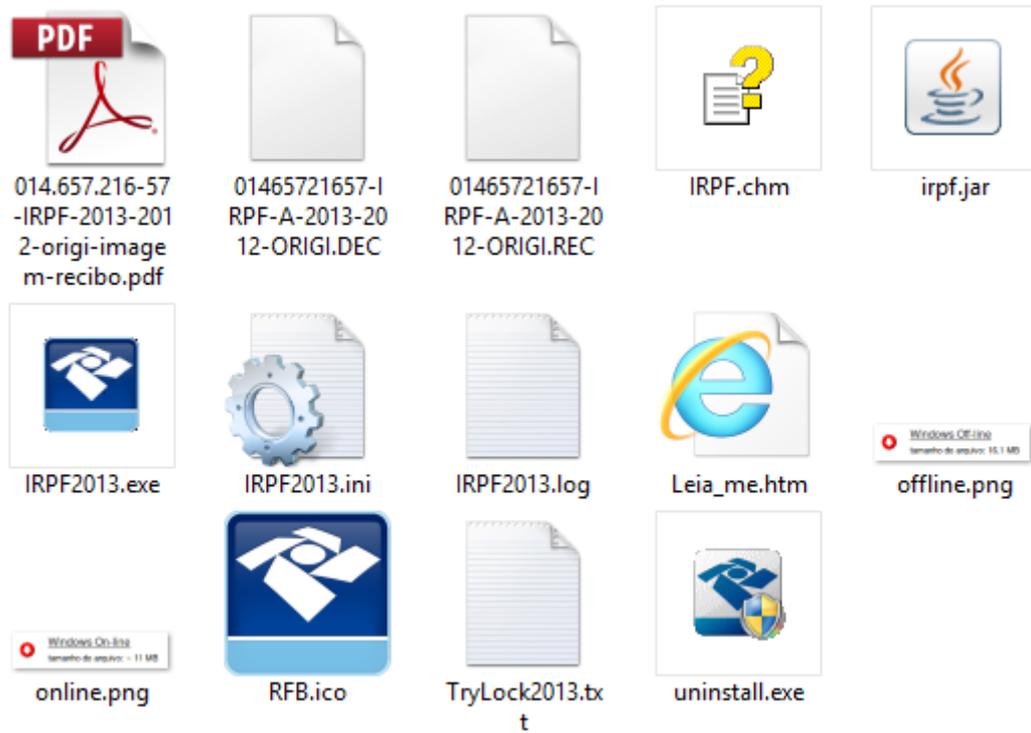
As File-Extensions.org is a resource site only, we regret that we are unable to provide personal support. But we are working very hard to improve the informations available in our database. (<http://www.file-extensions.org/about-us>).

Example

Figure 3 contains examples of program files related to the Brazilian Federal Revenue Income Tax 2013. Table 02 describes these files and their use. This is one of the central points of this study and we highlight here the archives 01465721657-IRPF-A-2013-2012-ORIGI.DEC e 01465721657-IRPF-A-2013-2012-ORIGI.REC. These files have a data structure defined by the Federal Revenue of Brazil and will only be read by programs of that company. When searching the database of the file-extensions.org website, it was identified that the extension .DEC is used by the ArcView program, a geographic information software, in addition to being a support file for UNIX systems. The website suggests that the default program for this type of file is ESRI's ArcView. The same problem occurs for the file with the extension .REC, being informed that it is generally used to store multimedia data, with the suggestion of media programs for its execution. Figures 4 and 5, respectively, show the records of .DEC and .REC extensions.

⁶ <<http://www.file-extensions.org/about-us>>

Figure 3 - Some types of files Generated by the Brazilian Federal Revenue program



SOURCE: author's collection

Table 2 - List of file types of the Brazilian Federal Revenue program

File Name	File Type
014.657.216-57-IRPF-2013-2012-origi- imagem-recibo.pdf	PDF document containing a receipt for delivery of the declaration
01465721657-IRPF-A-2013-2012-ORIGI.DEC	Arquivo com dados da declaração enviada a Receita Federal
01465721657-IRPF-A-2013-2012-ORIGI.REC	File with data from the declaration sent to the Federal Revenue Service
IRPF.chm	File containing instructions for help using the program
irpf.jar	File containing Java computer language programs
IRPF2013.exe	Executable file, main program
IRPF2013.ini	File containing program startup parameters
IRPF2013.log	File containing logging application execution information
Leia_me.htm	File containing a web page for program information
offline.png	Program picture
online.png	Program picture
RFB.ico	Program icon
TryLock2013.txt	Text file with technical information about the program
uninstall.exe	Program responsible for uninstalling the main program

SOURCE: Author's elaboration

Figure 4 - Identification of DEC files by the file-extensions.org website

DEC file extension - ArcView UNIX hyperhelp supporting file

 **Recommended: [Fix DEC file extension related errors now!](#)**

What is .dec file?

File extension is used by ArcView. Unix hyperhelp supporting file.

File type specification:
 [Various data file type](#)

 **Fix .dec file extension errors now! [Run a File Associations Scan](#)**

 **Experiencing system slow-downs? [Run a Free Performance Scan](#)**

 **Fix and update your PC drivers: [Run a Free Drivers Scan](#)**

The default software associated to dec file type:

 **[ArcGIS for Desktop Basic \(ArcView\)](#)**

Company or developer:
 Esri



ArcGIS for Desktop Basic (formerly known as **ArcView**) is GIS (geographic information system) application used to visualizing, managing, creating, and analyzing geographic data. ArcView is a part of the ArcGIS Desktop. ArcView supports creation of interactive maps, spatial analysis, GIS deployment, map viewing and navigation, map printing etc.

SOURCE: File-extensions.org

Figure 5 - Identification of REC files by the file-extensions.org website

REC file extension - Recorded file

 **Recommended: [Fix REC file extension related errors now!](#)**

What is .rec file?

The **REC file extension** is generally used for various recorded files. These can contain any multimedia content, session, macros or more. Such files are pretty common in digital devices that record audio or video. For example Topfield TF5800 uses *.rec files for its recorded videos. Also computer games like Company of Heroes uses *.rec files for its recorded matches - replays.

File type specification:
 [Audio and sound file type](#)
 [Various data file type](#)
 [Digital video and movie file type](#)

 **Fix .rec file extension errors now! [Run a File Associations Scan](#)**

 **Experiencing system slow-downs? [Run a Free Performance Scan](#)**

 **Fix and update your PC drivers: [Run a Free Drivers Scan](#)**

SOURCE: File-extensions.org

As the choice of extension is defined by the company that develops the software, this choice is usually based on an abbreviation of the program name, or something associated with the file type, in order to facilitate the organization of files by users. Due to this logic, the RFB

used the REC extensions as a direct association to the name RECIBO (Receipt in Portuguese), as this file keeps the receipt for sending the Income Tax adjustment statement, and DEC for DECLARACAO (Declaration in Portuguese), which contains the declaration itself. Users who have ArcView software or some other multimedia installed on their computers may experience problems running these files, since an incorrect program will attempt to play these files

This becomes a problem if the file needs to be kept for a long time and needs to be “read” later. A user who is unaware of the programs of the Brazilian Federal Revenue Service, upon encountering this file and trying to run it in the standard programs, may receive an error as a return and assume that the files are in trouble and unusable. You can also assume that, by the extension of the receipt file, REC, it is a recording, an abbreviation of the word Recording. These are the most important files for the Brazilian taxpayer who make adjustments to the income tax return and they must keep it for future reference. Archives are constituted as documents and are required for different purposes of the taxpayer. The Brazilian Federal Revenue Service advises that the file needs to be kept for at least five years⁷.

Filing of declaration and receipt

The IRPF 2012 Annual Adjustment Declaration and the delivery receipt must be kept for a period of 5 years from the year of its presentation.

The documents that prove the information provided must be kept at the disposal of the Federal Revenue, until the prescription of the tax credits related to the situations and facts to which they refer occurs. No document should be attached to the declaration. (BRAZIL, 2012)

The example shown suggests that the problem is not restricted to small and medium-sized software developers. The Federal Revenue of Brazil has one of the largest and most modern government computing centers in the world, being recognized for high excellence in various systems.

As there is an informal standardization, but not a limitation, of file extension using up to four characters, being allowed all letters of the alphabet from A to Z and numerals from 0 to 9, using the arrangement formula with repetition we can reach the maximum of files with

⁷Brazilian Federal Revenue page on the 2012 Income Tax declaration and data transmission: <<http://www.receita.fazenda.gov.br/PessoaFisica/IRPF/2012/declaracao/transmissao.htm>>

exclusive extension is 1.847.161. $AR_{n,p} = n^p$ where n is equal to the number of characters available, 27 letters and 10 numerals, and p is equal to the number of positions currently used,

$$AR_{n,p} = 37^4 = 1.874.161. \quad (1)$$

Although almost two million exclusive combinations are possible, an important factor must be considered; the use of extensions with combinations that facilitate their identification and the search that companies have for choosing names with greater commercial appeal. This drastically restricts the extensions that can be used and causes overlap by some companies. In the case analyzed, the choice of the DEC extension for DECLARATION aims to serve the population that does not master computational aspects, as such files are used by contributors from all social classes with varying degrees of education and not only by trained professionals and users. If extensions were regulated and protected, a company could be required to use an available extension that is unrelated to its program.

4.2 Retro compatibility and free access to reading

In public and private organizations, the search for standardization of software used by employees is a challenge for IT managers and financial and cultural barriers hinder this standardization.

Today's IT organizations face a challenge similar to Ford's. While the diversity of hardware and operating systems has grown steadily over the past 15 or 20 years, software applications have rapidly spread. Even worse was the multiplication of the number of add-ins, plug-ins and downloads from the Internet. Few IT companies have the staff and resources to support these components. Keeping them up to date with new application versions can be an even greater challenge for most departments. Since Ford's approach doesn't work in this environment, IT managers need to develop policies and practices to manage the consequences of this growth.⁸ (HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P., 2012).

Based on a Brazilian company⁹ with units present in different parts of the country, employees work with different version programs, generating documents with compatibility problems

⁸ <http://www.hp.com/latam/br/pyme/dicas/padron_soft.html>

⁹The author of this study has been working in the information technology sector since 2003, providing services to companies of different segments and sizes in Brazil. Among them are mining companies,

within the company itself. The value of standardizing the version of the software used, the Microsoft Office suite of applications, was financially unfeasible due to the cost of licenses and labor with updating, training and conversion of the entire base of existing files. As the result of this standardization is not reflected in direct and immediate financial results, projects of this type are placed in lower priorities than other projects directly linked to the company's final product, increasing the long-term technological and strategic liabilities.

Program developers make efforts to maintain backward compatibility with the files generated by their programs between their different versions of previous programs. Microsoft Corporation, for example, includes in its Office suite the possibility of reading files in versions earlier than those used by the user. The problem is more critical when data generated in newer versions needs to be read by older versions of the software, which usually requires installation of additional programs.

This effort is not a guarantee against use of obsolete files from the latest generation programs. The newest and most current version of Microsoft Office allows files from older versions like Office 97 to be opened for reading, but there is a limitation for compatibility with older versions. Excel does not support graphics of Excel 2.0 and 3.0 files from 1987 and 1990 respectively, and PowerPoint does not support opening and editing files from its 1995 and earlier versions¹⁰

Excel 2013 onward does not support the following formats and extensions:

- Excel charts (Excel 2.0, 3.0 and 2.x file formats) (.xlc)
- Lotus 1-2-3 file formats (all versions) (.wk1, .wk2, .wk3, .wk4, .wks)
- Microsoft Works (all versions) (.wks)
- DBASE II file format (.dbf)
- Quattro Pro file format for MS-DOS (.wq1)
- Quattro Pro 5.0 and 7.0 for Windows (.wb1, .wb3)

PowerPoint 2013 onward does not support the following:

- Open or save PowerPoint 95 (or earlier) file formats
- Pack and Go Wizard (.ppz) files.
- Web page files (.htm, .html)
- (Microsoft, 2013)

steelmakers, carriers, public agencies, banks and others. The analyzed company could not have its name mentioned as a source due to corporate governance rules.

¹⁰ File formats supported in Office 2013. <<http://technet.microsoft.com/pt-br/library/dd797428.aspx>>

This is one of the fundamental problems that we pointed out in this article; there is no way to guarantee that the created files will be able to be opened in software versions a few years after their creation. Since documents have value and need to be accessible for decades, in the cases exposed there would be problems in accessing these documents less than 20 years after their creation. Backward compatibility is limited and, for proprietary files, depends exclusively on definitions of the company that owns the software.

This analysis has a greater focus on Microsoft Corporation software, mainly in its Office suite of office applications due to the predominance in the Office program market, but the problem is not restricted to this company. Any other software development company will invariably grapple with the same dilemma about choosing to work with files in a proprietary format and taking care of backward compatibility between versions of their programs or adopting open standards and following the specifications of this standard. In addition, companies are finite and can finalize their operations leaving their customer base without protecting access to their files.

Without going to the unconditional defense of Microsoft, but aiming to study the actions taken by this company, it is valid to analyze their effort of opening previously closed standards. When we look back to the 1980s when Microsoft Office was launched, there were no competitors for this type of application and some other vendors had specific programs like spreadsheets or text editors. The discussion between open and closed formats was not an issue among companies that consume programs and the liability created by that was not a concern. In the 2007 version of Microsoft Office the Open XML format, a mix between the open standard eXtensible Markup Language (XML) and special Microsoft modifications, was adopted.

The Open XML file formats simplify the exchange of data between Office applications and enterprise business systems. Based on open standards, these XML file formats enable the rapid creation of documents from different data sources and speed up document assembly, data mining, and content reuse.¹¹ (MICROSOFT, 2012).

This did not impact previous versions that were still based on closed formats but could work with this new standard through updates. This compatibility was also limited. A company, for

¹¹<<http://technet.microsoft.com/en-us/library/cc179190.aspx>>

example, that held files in Excel 2.0 format should have a compatible program to access those files.

As the program itself is not enough, and there must be compatibility with the operating system and hardware, Microsoft opened its specifications for proprietary formats so that other companies could manipulate these files. With this action, companies such as The Apache Software Foundation, which owns OpenOffice, the most popular open format office application suite on the market, implemented these specifications and this software was able to work with closed Microsoft Office files such as Excel¹². This opening of standards is not a transfer of copyright or industrial rights and the patents involved remain the property of Microsoft. Microsoft Open Specifications¹³ declares that all programs that adhere to these specifications must give credits to Microsoft.

Any attempt at editing forces the user to update the file structure and save it in a new and current version. This position is adopted not only by Microsoft, but also by Adobe, Autodesk, Corel and other manufacturers that have heavily file-based programs. The impact of this conversion need is on the preservation of the original file without changes, maintaining its integrity and reliability. When a document is created on a physical medium like paper and is then digitized, its changes in the electronic medium do not compromise the original document that was kept on its primary medium. In the case of an original document created electronically, its original version should be kept unchanged. Such an alteration would be equivalent to altering a paper document compromising its veracity. The custody processes do not change regardless of the medium in which they are inserted. This theme is especially dear to Archivology, in addition to the technological problem shown in this work.

5 Solution Projects and Initiatives

After explaining some problems identified in the long-term storage of files, a few projects and initiatives to solve these problems will be mentioned here. As the cases shown are isolated actions by companies, they do not solve all the problems identified, but they serve as a basis

¹² Microsoft Office Excel 97 - 2007 Binary File Format Specification (*.xls 97-2007 format): <<http://www.microsoft.com/openspecifications/en/us/programs/osp/default.aspx>>.

¹³Open Specification Promise: <<http://www.microsoft.com/openspecifications/en/us/programs/osp/default.aspx>>

for study and implementation by other companies and entities with interests and relationships with this type of subject.

5.1 Open file formats

Concern about file formats and programs with proprietary code is not new in the computer industry. The Free Software Foundation (FSF), founded in 1985, is dedicated to various actions to make and disseminate the use of software free of commercial licenses.

The free software movement is one of the most successful social movements to emerge in the past 25 years, driven by a worldwide community of ethical programmers dedicated to the cause of freedom and sharing.¹⁴ (FREE SOFTWARE FOUNDATION, INC., 2013).

With the growing dependence on files in proprietary format, institutions have shown concern about the possibility of recovering these files in the long term, mainly with the dependence on payment of licenses for programs capable of handling proprietary files. In 2003, the Organization for the Advancement of Structured Information Standards (OASIS)¹⁵ was created to define standards used in information technology. One of its main achievements was the creation of the Open Document Format for Office Applications (ODF) file format, which consists of an open standard for office files including spreadsheets, text documents, presentations and drawings. This standard defines an open, non-proprietary format, which can be used by companies with closed or non-commercial, commercial or free programs. ODF aims that any company that develops office applications software can use this file format to generate their documents. These files can be handled by any other software that supports the standard. The manufacturer could charge licenses for the program for several reasons, such as exclusive features, support for doubts and problems, market presence, sales actions, competitive prices, etc., but the generated files could not be closed and exclusive and the user would have the full condition to use it in any other software compatible with the standard, since technical standards must be observed when generating the file for full compatibility between different programs. A company seeking complete

¹⁴ <<http://www.fsf.org/about/>>.

¹⁵ <<https://www.oasis-open.org/org>>

protection against software dependency could create its own programs that implement the ODF standard.

The standard was so well received in the market that it obtained an ISO approval on May 8, 2006 (ISO/IEC 26300)^{16 17}. This approval strengthens the intentions of using this format as the one prevailing for office files, aiming at its long-term custody. If we define that a long-term period for keeping archives is greater than 20 years, we can consider the risk that the ISO committee does not exist in that period. This could create a technical problem that is easy to get around, since the standard is open and, even if abandoned, software developers can create programs capable of handling these files.

Several countries and their governments have adopted ODF as their preferred office file format. In Brazil, the first law on this recommendation is Law 15742¹⁸ of the state of Paraná 12/18/2007. The Law Project 3070/2008¹⁹, of the Chamber of Deputies and the Brasília Protocol ²⁰, on ODF of August 27, 2008 also recommend using ODF as the preferred format for this type of file. In May 2008, the Brazilian Association of Technical Standards (ABNT) approved the ODF on the standard NBR ISO/IEC 26300:2008²¹

The summary of Law No. 15742 does not make it mandatory to use the ODF standard, but rather recommends that office files be generated preferably in this format, which opens a wide margin of decision on whether or not to use the format by state employees. This recommendation is followed by Project Law 3070/2008, which does not yet implement the requirement and only the Brasília Protocol is willing to adopt ODF as a standard for all public

¹⁶ <<http://www.iso.org/iso/pressrelease.htm?refid=Ref1004>>

¹⁷<http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485>.

¹⁸First Brazilian Law on ODF Law 15742:

<<http://www.legislacao.pr.gov.br/legislacao/pesquisarAto.do?action=exibir&codAto=30&indice=1&totalRegistros=1>>

¹⁹Project Law 3070/2008:

<http://www.camara.gov.br/Sileg/Prop_Detalhe.asp?id=387780>

²⁰Brasília Protocol:

<http://www.softwarepublico.gov.br/4cmbr/arquivos/view/publicacoes/Protocolo_Brasilia_3_3.pdf>

²¹Technical Standard ABNT NBR ISO / IEC 26300: 2008:

<<http://www.abntcatalogo.com.br/norma.aspx?ID=1549>>

entities in the federation. Unlike the aforementioned Law and Bill, the BRAZILia Protocol extends its recommendations with attention to the recommendation of preferential use of free software for manipulating files.

This standardization minimizes, but does not extinguish, the problems generated by proprietary files. While ODF aims to open only files used in the office, other files that are also part of digital collections such as audio files, technical drawings, e-mail, websites and web pages, are not met by this standard.

5.2 Microsoft and XML for Office

Microsoft Corporation is the world's largest software company (Statista, 2020) with total 2020 revenue of \$143 billion (Microsoft Annual Report 2020) with \$46 billion (Microsoft Annual Report 2020) related to Productivity and Business Processes, of which Microsoft Office is part. Microsoft dominates 40.39% (Statista, 2020) of the productivity software market, where office applications are segmented and its suite of office applications, Microsoft Office, which includes programs such as Word text editor, Excel spreadsheet, PowerPoint presentation program and Outlook e-mail. This very strong market share generates a worldwide phenomenon making Microsoft Office applications hegemonic in offices and personal computers. As a consequence, the files of the format of these applications have a massive presence in digital collections, becoming a standard due to the massive use in the daily lives of users. Microsoft invests in the education market, reselling its products at very low prices when compared to the value destined for companies, increasingly implanting the custom and dependence on the use of this software.

Under pressure, for example, by the Commonwealth of Massachusetts, USA, which, in 2005, determined that the Microsoft Office closed format standard should be replaced by the open standard OpenOffice, ODF, due to concerns about the future use of proprietary files, Microsoft, that same year, advanced its project to adopt Open XML format based on open source license for its Office suite. This, in practice, ends with the adoption of files generated in proprietary binary format allowing interested companies to create programs to work with Microsoft files without paying royalties. A file created with Microsoft software could, from now on, be read and edited by programs from other companies. Even with Microsoft's intention to open its file structure, when the new standard was released, a large part of the free software advocacy community expressed concern about the opening of patents by Microsoft.

InfoWorld, in 2008, expressed its opposite position regarding the creation of this new standard, stating: “OOXML was opposed by many on grounds it was unneeded, as software makers could use ODF, a less complicated office software format that was already an international standard” (KIRK, 2008). This was based on the fact that Microsoft created a new standard, not adopting the ODF already widespread and adopted worldwide.

The criticisms and concerns have not stopped, culminating in Richard Stallman's statement, "Microsoft offers a gratis patent license for OOXML on terms which do not allow free implementations." (STALMANN, 2009). The concern was based on the lines of the license published by Microsoft that restricted certain interventions on the standard and that was analyzed by some experts, such as Stallman himself, as the emergence of a new closed standard with nuances of free access.

In 2005, Brian Jones, group program manager for the Solutions Framework Office team, published a series of texts in the Microsoft developer community in response (MICROSOFT, 2013) to community concerns. The most important concerns are: any program that uses Microsoft's XML formats must, in some part of the code, give credit to that company, the second is that companies must meet sublicensing standards. Since licensing is free, there should be no worries about it, but there are no perpetual guarantees that Microsoft will continue with this line of action and, at some point, being a company guided by financial results, modify the terms of the license. Table 3 presents a comparison between the file extensions we refer to.

Table 3 – Comparison between Microsoft Office file extensions and the Open Document standard

Document Type	Microsoft Office extension in binary format	Microsoft Office extension in Open XML format	Extension to the OpenDocument standard
Text	doc	docx	odt
Spreadsheet	xls	xlsx	ods
Presentation	ppt	pptx	dop

SOURCE: from the author, based on MICROSOFT, 2013

Aiming to close the subject, ensuring the opening of the standard and no future vision collections, Microsoft posted on its license terms:

Microsoft irrevocably promises not to assert any Microsoft Necessary Claims against you for making, using, selling, offering for sale, importing or distributing any implementation to the extent it conforms to a Covered Specification [...] (Open Specification Promise, 2006)

It is important to note that the new formats used by Microsoft Office, despite being open, do not obey the ODF standard, becoming a strong competitor of this format. Office can process ODF files, but its file standard is XML. As users are strongly accustomed to Office extensions, such as DOC for Word documents, XLS for Excel workbooks and PPT for PowerPoint presentations, continuing to use files with this extension is easier, as it is part of everyday life, without having to change the work routine, than using the ODF standard file extensions.

Opening the PDF file

One of the most striking episodes of long-term digital file storage efforts is the opening of the PDF standard by Adobe Systems, seen as a victory by the software industry and library professionals.

In 2003 Adobe Systems Software created the PDF, a file format independent of hardware and operating system. This standard was intended to give the user the ability to manipulate documents on any computing platform. Its success was enormous in the industry and documents that were previously associated with certain software, such as those in the Microsoft Office suite, can now be read on any computer. As it was a closed format, Adobe created a market where file changes could be made only with its proprietary software, but users could view the files for free. Document creators could also apply restrictions such as printing or sharing files, making files more secure.

From this creation, a document that was previously manipulated only on desktops could now be seen on cell phones, web pages and several other platforms. Due to the file editing restriction made only by Adobe software, the market quickly associated PDF files as a safe way to generate documents and maintain their immutability. Aiming at a market to be explored, software manufacturers and programmers started to create programs capable of manipulating PDF files, removing from the user the need to purchase Adobe software

licenses. This was an illegal practice, since the software was protected by intellectual and industrial property laws.

With the need to make digital documents available and accessible in the long term and with the wide use of PDF files worldwide, in May 2005 ISO approved the PDF/A standard through the ISO 19005-1: 2005 standard; the first standard created with the intention of preserving archives in the long term. When held by Adobe, the PDF standard contained a number of technical aspects that could prevent manipulation of files created in the long run. The PDF/A standard consists of removing these aspects, allowing a created file to be read decades after its creation. To enable the use of these files in an open format and standardized by ISO, Adobe assigned its copyright on this standard from that moment. This allowed several software manufacturers to develop programs capable of reading, creating and editing PDF/A files without the need to pay licenses for Adobe, which made the format even more widespread. This widespread use has made it so ubiquitous that the Federal Government of Brazil, Science and Technology, Open Standards, 2010, recommended that the final version of documents for processing between agency or filing be saved in the PDF/A standard.

An important point about the ISO 19005-1: 2005 standard is that it does not describe archiving processes for electronic documents; only the file format. Storage techniques, such as disposal, signature, and periodicity, must be addressed by the interested companies, with the standard focusing on technical aspects of generating and maintaining the file format.

Final Considerations

In this study, we sought to expose models that can be improved to solve the problem explored in this article, with the main points regarding long-term access to files and normalization of extensions for programs that generate output files.

There is a latent need for an international entity to consolidate the file extensions created by companies that develop computer programs. Several entities, such as Ecma International, standardize interoperability and communication models between systems, but have restrictions on operations in regions of the globe. The International Organization for Standardization (ISO) does not act as a regulatory agency, does not issue patents and has no legal powers, but is recognized as the main international body for issuing standards in several areas. Its adoption as a centralizer of extensions in use can be easily accepted by the industry, which already adopts technology standards indicated by that organization. However, ISO is recognized for setting standards and issuing opinions after study times, which could result in an inefficiency in the speed of publishing extensions.

We believe that the presence and reliability of the ISO organization could lead to adoption of more careful procedures, even if they impact the agility of publications by the companies that develop the files. At the same time, ISO could issue standards for archiving and retrieving the generated digital documents. The collections of initial extensions are easy to obtain, since there are several entities with published sites, as mentioned in the study, with an existing collection.

Regarding backwards compatibility, in addition to the technical requirement, archiving processes should be considered. Adoption suggestions for implementing long-term file access process could be listed as:

- Every software development company, with or without proprietary file, should provide a free application to read the generated files, as soon as their production is finished and / or updated.
- Reading programs could preferably be distributed with source code for maintenance by the interested community for compatibility with future systems.
- Reading programs should preferably be developed in a layer independent of hardware and operating system, providing the greatest possible compatibility with future devices.

- Digital collections should store, in addition to files with documents, reading programs made available by companies and compatible with the files contained in these collections.

These simple actions can guarantee that the programs responsible for reading will accompany the files, preserving the integrity of the documents and their future access. Since digital collections, as well as libraries and archives, do not have permission to change documents, editing permission on these programs would not be required.

This solution would be a step in the process of archiving digital documents – inevitable to be implemented – in companies with digital collections of added value. Some program developers have already spontaneously adopted this practice. This could cause a future disruption if a company today adopts an extension that has a published reading program and, in the future, for commercial reasons, these programs undergo modifications in the distribution model or even have their production closed.

I therefore suggests creation of:

- a long-term storage model for digital files, with a focus on systems normalization
- a collective and collaborative archiving process
- an international standardizing organization

To implement these suggestions, an alignment of interests and research between the areas of Librarianship, Archivology, Information Science and Information Systems will be vital.

References

ARELLANO, Miguel Angel. Preservation of digital documents. **Information Sciences**, Brasília, v. 33, n.2, p. 15-27, 2004.

BRAZILIAN ASSOCIATION OF TECHNICAL STANDARDS. **NBR ISO/IEC 26300:2008**: <<http://www.abntcatalogo.com.br/norma.aspx?ID=1549>>

BODÊ, Ernesto Carlos. **Preservation of digital documents**: the role of file formats. 2008. 153 f. Dissertation (Master) - Information Science Course, University of Brasília, Brasília, 2008.

BRAGA, Kátia Soares. Relevant aspects for the selection of appropriate methodology for social research in Information Science. In: MUELLER, Suzana Pinheiro Machado. *Research Methods in Information Science*. Brasília: Thesaurus, 2007. p.17-38.

BRAZIL. Chamber of Deputies. **Project Law 3070**, presented on March 25, 2007 - author: Paulo Teixeira - PT/SP: <http://www.camara.gov.br/Sileg/Prop_Detalhe.asp?id=387780>

BRAZIL. Planning Ministry. **E-Ping**: <<http://eping.governoeletronico.gov.br/>>

BRAZIL. Brazil Federal Revenue Service. Ministry of Finance. **Information on the 2012 Income Tax declaration and data transmission**: <<http://www.receita.fazenda.gov.br/PessoaFisica/IRPF/2012/declaracao/transmissao.htm>>

BRAY, Hiawatha. State may drop Office software. Report urges adopting new file format to ease sharing of information, 2005: <http://www.boston.com/business/globe/articles/2005/09/02/state_may_drop_office_software/>

DIGITAL Equipment Corporation: <<http://dec.fileextensionguide.com/>>

ESPINIER, Tom. **Legal worries led Massachusetts to open standards**. 2005. <<http://www.zdnet.com/legal-worries-led-massachusetts-to-open-standards-3039216391/>>

FEDERAL UNIVERSITY OF PARANÁ. **Files – Formats**: <<http://www.ufpa.br/dicas/progra/arq-exte.htm>>

FEDERAL UNIVERSITY OF PARANÁ. **Programs – Functions and Types**: <<http://www.ufpa.br/dicas/progra/protipos.htm>>

File name EXTENSION. Wikipedia: <http://pt.wikipedia.org/wiki/Lista_de_extens%C3%B5es_de_ficheiros#Lista_de_extens.C3.B5es>

FERNANDES, Jorge. **What is a program (software)?:**
<<http://www.cic.unb.br/~jhcf/MyBooks/iess/Software/oqueehsoftware.html>>

FILE-EXTENSIONS ORG. **FAQs:** <<http://www.file-extensions.org/articles/faq>>

FILE-EXTENSIONS.ORG: <<http://www.file-extensions.org/filetypes>>

FILE format. Wikipédia:<http://pt.wikipedia.org/wiki/Formato_de_arquivo>

FREE SOFTWARE FOUNDATION. **Worldwide mission to advance software freedom:**
<<http://www.fsf.org/about/>>

HEWLETT-Packard Development Company (BRAZIL). **Standardization of software installations:** <http://www.hp.com/latam/br/pyme/dicas/padron_soft.html>

TECHNOLOGICAL INSTITUTE OF AERONAUTICS – ITA. **Open Document Format (ODF) / BOffice. The standard adopted by the Government to replace the Microsoft Office®:**
<http://www.ita.br/sinf/faqs/odf.html>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (E.U.A). ISO 32000-1:2008 **Document management - Portable document format - Part 1: PDF 1.7:**
<http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (E.U.A). **ISO and IEC approve OpenDocument OASIS standard for data interoperability of office applications:** <<http://www.iso.org/iso/news.htm?refid=Ref1004>>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 26300:2006:**
<http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485>

JONES, Brian: **Office Solutions:**
<http://blogs.msdn.com/b/brian_jones/archive/2005/09/22/472826.aspx>

KIRK, Jeremy. ISO publishes Office Open XML specification. **Inforword**, 2008:
<<http://www.infoworld.com/t/applications/iso-publishes-office-open-xml-specification-918>>

LIST of PDF software. Wikipédia:
<http://en.wikipedia.org/wiki/List_of_PDF_software>

TEACHING MATHEMATICS. Combinatorial analysis:
<<http://www.matematicadidatica.com.br/CombinacaoSimples.aspx>>

TEACHING MATHEMATICS. Simple arrangement and combination calculator:
<<http://www.matematicadidatica.com.br/CalculadoraArranjoCombinacaoSimples.aspx>>

MICROSOFT (BRAZIL). **Supported file formats in Office 2013:**
<<http://technet.microsoft.com/pt-br/library/dd797428.aspx>>

MICROSOFT (BRAZIL). **Open XML file name formats and extensions:**
<<http://office.microsoft.com/pt-br/word-help/formatos-e-extensoes-de-nomes-de-arquivo-open-xml-HA010354237.aspx>>

MICROSOFT (BRAZIL). **Introduction to new file name extensions and open XML formats:** <<http://office.microsoft.com/pt-br/word-help/introducao-a-novas-extensoes-de-nome-de-arquivo-e-formatos-xml-abertos-HA010006935.aspx>>

MICROSOFT (E.U.A). **Microsoft Office Excel 97 - 2007 Binary File Format Specification (*.xls 97-2007 format):**
<<http://www.microsoft.com/openspecifications/en/us/programs/osp/default.aspx>>

MICROSOFT (E.U.A). **Open Specification Promise:**
<<http://www.microsoft.com/openspecifications/en/us/programs/osp/default.aspx>>

MICROSOFT (E.U.A). **Overview of the XML file formats in Office 2010:**
<<http://technet.microsoft.com/en-us/library/cc179190.aspx>>

MICROSOFT (E.U.A). **XML Paper Specification Copyright License:**
<<http://msdn.microsoft.com/en-us/library/windows/hardware/gg463420.aspx>>

MICROSOFT (E.U.A). **XML Paper Specification Licensing Overview:**
<<http://msdn.microsoft.com/en-us/library/windows/hardware/gg463418.aspx>>

MICROSOFT (E.U.A). **XML Paper Specification Patent License:**
<<http://msdn.microsoft.com/en-us/library/windows/hardware/gg463420.aspx>>

MORTARA, Bruno. The new part of the standard ISO 19005 or, simply, PDF/A-2. **Graphic Technology, Sao Paulo**, n. 76, 23 mar. 2011:
<http://www.revistatecnologiagrafica.com.br/index.php?option=com_content&view=article&id=1762:a-nova-parte-da-norma-iso-19005-ou-simplesmente-pdf-a-2&catid=60:normalizacao&Itemid=185>

MOUTEVELIS, Alexa. **CAGW Criticizes OpenDocument Policy in Massachusetts.**
<<http://cagw.org/media/press-releases/cagw-criticizes-opendocument-policy-massachusetts>>

MS-DOS. Wikipédia: <<http://pt.wikipedia.org/wiki/MS-DOS>>.

ORGANIZATION for the advancement of structures information standarts - OASIS (E.U.A). **Advancing open standarts for the information society:** <<https://www.oasis-open.org/org>>

PARANÁ. **Law n. 15742, 2007:**
<<http://http://www.legislacao.pr.gov.br/legislacao/pesquisarAto.do?action=exibir&codAto=30&indice=1&totalRegistros=1>>

PORTABLE Document Format. Wikipédia:
<https://pt.wikipedia.org/wiki/Portable_document_format>

PROTOCOL of Brasilia. Public Protocol of Intent for Adoption of Open Document Formats:
<www.softwarelivre.gov.br/protocolo-BRAZILia-1/protocolo%20BRAZILia.pdf>

PWC'S TECHNOLOGY INSTITUTE. **PwC Global 100 Software Leaders**, 2013:
<http://www.pwc.com/en_US/us/technology/publications/assets/pwc-global-software-100.pdf>

RUDMAN, Robbins Geller; DOWD LLP. Files Class Action Suit against Microsoft Corporation.
BusinessWire: <<http://www.businesswire.com/news/home/20130812006136/en/Robbins-Geller-Rudman-Dowd-LLP-Files-Class>>

SILBERSCHATZ, Abraham; GALVIN, Peter Bauer; CAGNE, Greg. **Operating Systems with Java**. 7 ed. Rio de Janeiro: Campus, 2008.

STALLMAN, Richard. **We Can Put an End to Word Attachments**:
<<http://www.gnu.org/philosophy/no-word-attachments.en.html>>

TANENBAUM. Andrews S. **Computer Networks**. 4 ed. Translation by Vandenberg D. de Souza. Rio de Janeiro: Campus Elsevier, 2003.

TANENBAUM, Andrew. **Operating Systems: Project and Implementation**. Porto Alegre: Bookman, 2008.

TEAM, Trefis. **An Overview Why Microsoft's Worth \$42**:
<<http://www.forbes.com/sites/greatspeculations/2013/01/09/an-overview-why-microsofts-worth-42/>>

TRUCKER, Peter. **Has Big Data Made Anonymity Impossible?**:
<<http://www.technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/>>

VBMANIA. **Numerical possibilities**:
<<http://www.vbmania.com.br/pages/?varModulo=Forum&varMethod=abrir&varID=241905>>

WATERS, Richard. Massachusetts set to switch off Microsoft. **Financial Times**, 2005:
<<http://www.ft.com/cms/s/2/80033a76-1a71-11d1-b7f5-00000e2511c8.html#axzz2fZDOU2VL>>.

WEBOPEDIA. **Software**: <<http://www.webopedia.com/TERM/S/software.html>>.

XIPPH.ORG. **About**: <<http://www.xiph.org/about/>>

YAHOO. **Questions and answers**:
<http://br.answers.yahoo.com/question/index?qid=20070209102652AAF65hr>

Dell Technologies believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." DELL TECHNOLOGIES MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying and distribution of any Dell Technologies software described in this publication requires an applicable software license.

Copyright © 2020 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.