Journal of Environmental Protection, Safety, Education and Management

2015, Number 5, Volume 3, date of issue 28th Juni 2015 ISSN 1339-5270

LONG TERM ARCHIVING DOCUMENTS IN ENVIRONMENTAL ENGINEERING IN UNSTRUCTURED FORM

Robert HALENAR

Abstract

This article deals with long term archiving documents in environmental engineering, which comes from plenty of sources. Documents should be in text form, or in form of image, it should be a table, or a graph. Many forms need to be stored in many types of document or data files. In this article there is shown how to archive different data sources in one unstructured form, which is suitable for long term archiving. No matter if it is text or image. Data can be stored in heterogeneous data sources in different structures and each other "mixed". Matlab environment addresses the method of data extraction, transformation and alignments records, and then store the data in electronic form in unstructured environment of Photoshop tool. Records are transformed into graphic form and then, using layers, saved in the Photoshop tool - *. psd. The entire process is automated using Matlab environment.

Key words:

Long term archiving, document, data, Matlab, Photoshop

1 Introduction

Data documentation is now common issue for most companies. Especially there is a need for long term data archiving, which consists of text, pictures and schemes. In environmental engineering is lot of problems and solutions, many of them are great ideas, but not realizable in now days. That is why they finish its journey in status of project or analyses in paper form, and they are not developed into full workable solution. In most cases the well analyzed solution is determined for archive. Good idea used in it stills helpful. That is why we need to archive this document. There is a question, how to archive a document which comes from plenty of sources, which are not compatible with each other. The solution should be to archive the document in unstructured form, which is durable in long time period. Data backup must be resistant for data loss due to hardware failure, or in the case of a lay user intervention into the data structure, respectively intervention of foreign persons (criminal act), viruses, natural disasters and the like. The need to back up data today is now widely accepted as correct.

1.1 Backup and archiving an environmental documents

Backup means schedule and safe storage data, and if necessary the company could renew all the systems. The primary objective is therefore recovery backup systems as far as possible in the most current state. Its role is thus not giving insight into the history, but it is in charge of archiving. In most cases, organizations may access the archiving only when the need to retain data, for example, for legislative reasons and then when them already do not work actively, where necessary, free up space for new data on expensive storage systems. When backing up the original data remain in the same place, while the copy is transmitted to another site. It is carried out mainly for the safety reasons in case of failure of primary storage data recovery was possible. In contrast, arise when archiving copies of data, but data is directly transferred to other types of storage with lower operating costs. This is a fundamental difference between backup and archiving. From that then unfolds the target group for both processes. Should back up all types of companies and organizations it depends on their intellectual property. Archive is then worth it when you need to store large amounts of data over a long period of time. [1]

The aim is therefore archiving storage and related ensuring long-term availability of data in digital form. But this will meet the specific requirements. Digital archive must ensure the continuity of data for decades, using control mechanisms must deal with the finality and non - modifiability documents and be able to work with metadata to the archived data was not only a digital chaos. And to make matters worse, it have to be able to use without complex data migration to new types of servers and storage arrays that appear on the market sometime in the future. For those reasons are demands made on filing archiving systems especially high and its complexity exceeds backup solution. Digital archive must be sufficiently

[1] DIVINEC, L.: *Co by měl umět moderní datový archiv ve firmě?,* [online]. [2015-10-06]. Available at: <http://www.ictmanazer.cz/2011/11/co-by-mel-umet-moderni-datovy-archiv-ve-firme/>

robust to allow authentication and storing unchanging data using a digital signature and at the same time the possibility of verifying that the data from the time of deposit remained unchanged.[2]
Data Archiving Information System works exclusively with data that has somehow arisen box. This is a structured data. Data structure maintains the information system. Logic data archiving information system is built on a simple principle - the value of data changes over time. No data have the highest value at the time of its creation. Then their value decreases. Nevertheless, the organization shall, on the basis of a legislative obligation to retain for a long time (about 5 to 10 years depending on the type of data). After that time-is not cost-effective to retain transaction is closed, the old data in the database information system. Data Archiving solves the transfer of these data from the information system database to another type of storage. An important principle which must be respected, says the data must still be viewable from the information system, as it owns them, and still maintain their structure. [3]
Sometimes there is the need to handle archive data as big data. Specific solution is described in the literature. [4]
In the case of archiving data stored in the information system is a structured data archiving. Their reconstruction is necessary not only compliance software compatibility (thus buying the next version of the program from the same supplier), but also hardware compatibility, which is virtually impossible conditions for decades. Thus the old structured data will have to be either from this long-term archiving excluded completely, or at a high cost transformed into a new form (structure).
One approach to make archiving process not only to be more efficient, but over the decades allow it, is the change of structured data to unstructured. This usually means transforming them into a commonly used file formats (export) documents such as Office applications. At the cost of losing data structures (and thereby reduce the ability of searching and matching) we ensure easy and inexpensive availability of the data. In this process, we should also choose file formats that are time-tested and have proven their ability to maintain the compatibility over decades. One of these formats is mainly image file formats. The advantage is also a number of free programs for their display (i.e. free licenses).

2 The tool for editing graphical data – Photoshop

In my work I focused on the creation of electronic archive, which stores records intended for long-term archiving in unstructured form. Thus, in contrast to a database or data warehouse data are organized. For long-term storage of data in electronic form, it is important to choose such a format that is not only widely supported (such as * .doc), but also a well-known (in terms of its internal structure). This condition meets formats * .jpg, * .gif, * .bmp, and so on. Their disadvantage is the way of data representation in the form of 2D images. They thus have the opportunity (capability) to represent complex structures. Data are ordered chronologically and therefore would need to keep it on several such files, which represent the chronology, i.e. number sequence in the file name. (obr001.jpg, obr002.jpg obr063.jpg ... ... etc.) It was necessary to find such a format which would respect the conditions of the broad support known internal structure and representation with the ability chronology. A compromise which satisfies this purpose is a graphic format * .psd, it is a Photoshop file.
Photoshop (often referred to as PS) is a raster program produced by Adobe designed for creating and editing 2D graphics. From the title it is clear that was originally created as a tool to adjust the then scanned photographs, but his options and advanced tools made from it quickly universal tool for work with 2D graphics for professionals. Photoshop is by far the most advanced and most professional tool for working with raster graphics. Photoshop features are almost unlimited. [5]

Adobe Photoshop represents not only the tip of photo editing, but also in the field of 2D graphics in general. It is a professional tool which, not suitable only for the work of the photographer (as might whisper the name) but also, for example, architects, 3D designers, scientists and doctors. Layers are one of the characteristic features of Photoshop. We understand under layers of photos, texts, filters and the like. [6]
There is a clear parallel with the card of patient health records, when the individual layers may contain both text information and the graphic records (outputs EEG or ECG, X-ray images, namely records MRI or CT) and even video (automatic mode imaging sequences, MRI, ultrasound video recordings, etc.).

[2] DIVINEC, L.: *Co by měl umět moderní datový archiv ve firmě?,* [online]. [2015-10-06]. Available at: <http://www.ictmanazer.cz/2011/11/co-by-mel-umet-moderni-datovy-archiv-ve-firme/>

[3] MRÁZ, J.: IW: *Archivácia dát informačného systému*, [online]. [2015-10-06]. Available at: <http://www.itnews.sk/tituly/infoware/2010-08-23/c135372-iw-archivacia-dat-informacneho-systemu>

[4] TRNKA. A.: *Big data analysis*, In: European Journal of Science and Theology, Vol. 10, suppl. 1 2014, ISSN 1841- 0464. s. 143-148.

[5] MAK, J.: *Photoshop – úvod*, [online]. [2015-10-06]. Available at: <http://www.photoshopbook.net/photoshop-manual/photoshop-start.html>

[6] ONDROUŠEK, L.: Adobe Photoshop, [online] [2015-10-06], Available at: <http://www.ephoto.sk/fototechnika/recenzie/softver/adobe-photoshop/>

Thus, Photoshop can save several types of information in a single file, and offers the possibility of a chronological representation and at the same time we know (at least partly) the internal structure of stored information. Last property from the said (the interior structure of stored information) is crucial in the automated transformation of various types of data from different types of storage. Using MATLAB, this process can capture and transform data to automate (depending on specific conditions partially or fully), and then stored in * .psd.

3 The tool for automatize archiving – Matlab environment

In my work I focused on the transfer of structured data to unstructured text format and then to the image data format.
Since the transfer of data from the structure is relatively complex process, it is necessary to use sophisticated tools upgradeable options and a high algorithmic variability. Often times the data are stored in dozens of tables and "mixed" with the other. Thus, export (migration) data in unstructured form must respect all recorded dependencies, and use them to obtain the required information from the entire data structure (database) and assign them just one record.
Such a tool is Matlab. Matlab simulation programming language is the 4th generation in integrated environment for technical calculations, modeling, simulation and analysis. It allows interactive work, but also to create an application. It provides users relatively powerful graphics and computing tools, as well as an extensive library of functions that are useful in scope in virtually all areas of human activity. Thanks to its architecture Matlab is also designed for those who need to solve computationally intensive tasks without detailed examination of the mathematical nature of the problem. Matlab own language is much easier than Delphi or C and a high potential productivity and creativity. A major strength of Matlab is fast computational core with optimized algorithms and powerful mathematical base. Matlab implementations are key platforms - Windows and Linux and followers. [7]

4 Extraction, transformation and loading data in unstructured form

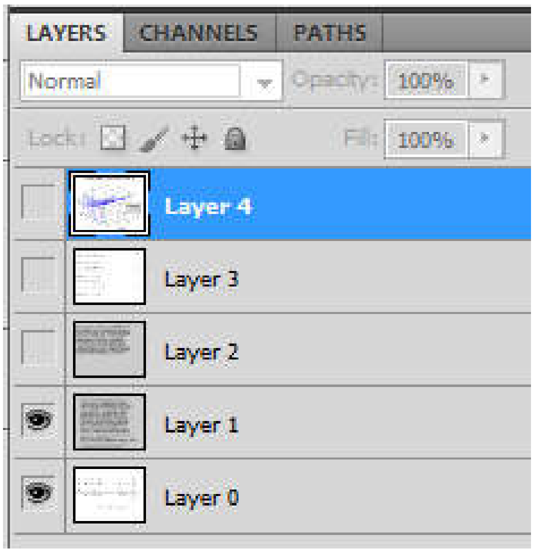
We have focused mainly on the text informations and the images, because many programs used for creating for example schemas are able to store its data in form of images. In result we have several images (pictures, schemas, tables…) and some descriptive text, which is bounded into one report. We have a choice – to create some the document of the common file extensions (office document, portable document format known as PDF and so on…) or store it as a package of the pictures. First choice is handmade (needs some editorial work) and second choice is automated via Matlab and stored as a package of the pictures (layers in *.psd file). The following example is a sample processing of one image, and some of the records using MATLAB into * .psd. First, it is necessary to accurately identify input sources:

- Location – network storage, IP address, database, and "connect string" …
- File format – format internal DBMS (database management system) Oracle or another format, and file extension, the need to transfer into electronic form ...
- The export options into the standard formats - must be considered or benefits of the export

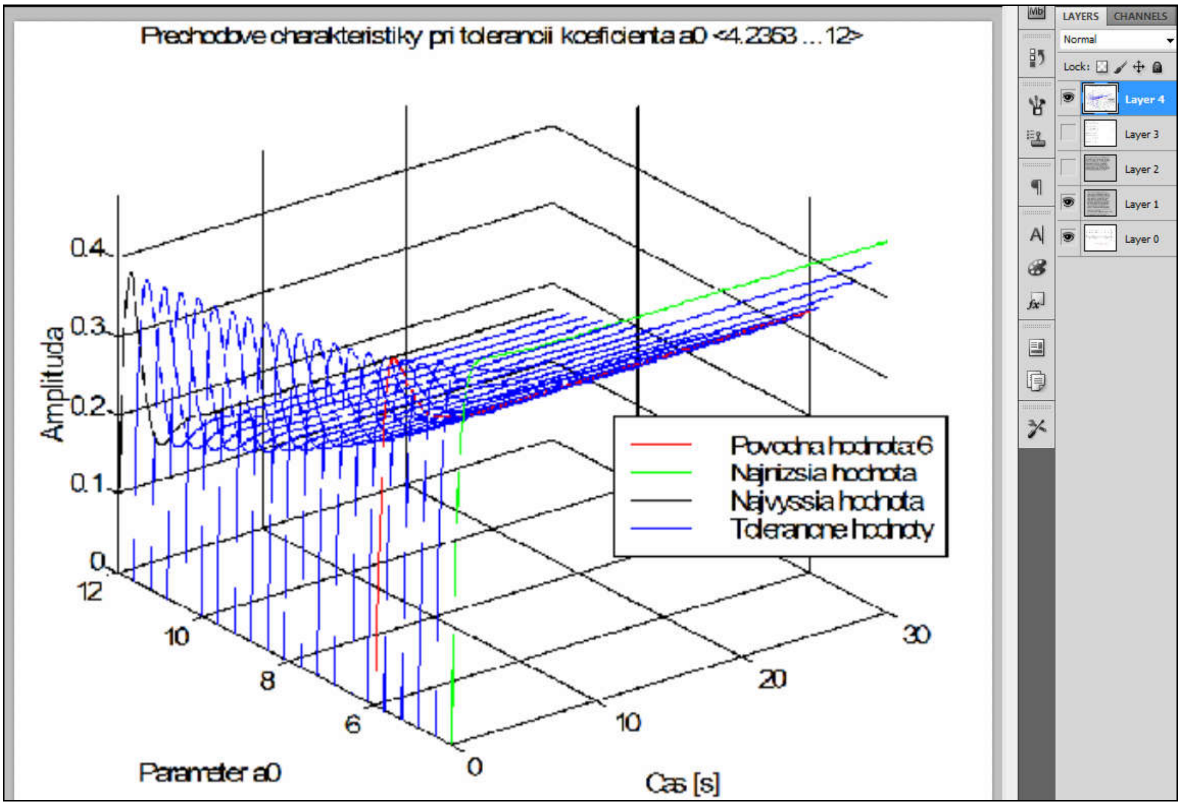
In our illustrative example we have an image (scanned and stored in * .jpg) and records about the picure (transcribed into * .txt). Textual records were further modified by inserting a special character "%" indicating the end of the line, because the line width is to be adjusted according to the size and font type and resolution used by the next file in Photoshop - u * .psd. Subsequently the text records were stored in * .xlsx for further processing (parceled into individual lines of text ready for publication and further automated processing in Matlab).
In automatic data processing from specific sources directly in Matlab can easily solve data extraction followed by pretreatment (preprocessing) text records into the required format, respectively, add a special characters to the text.

Matlab has subsequently launched a transformation script that images in * .png and text entries in the * .xlsx processed and stored as layers in Photoshop, which is shown in Fig. 1-5.

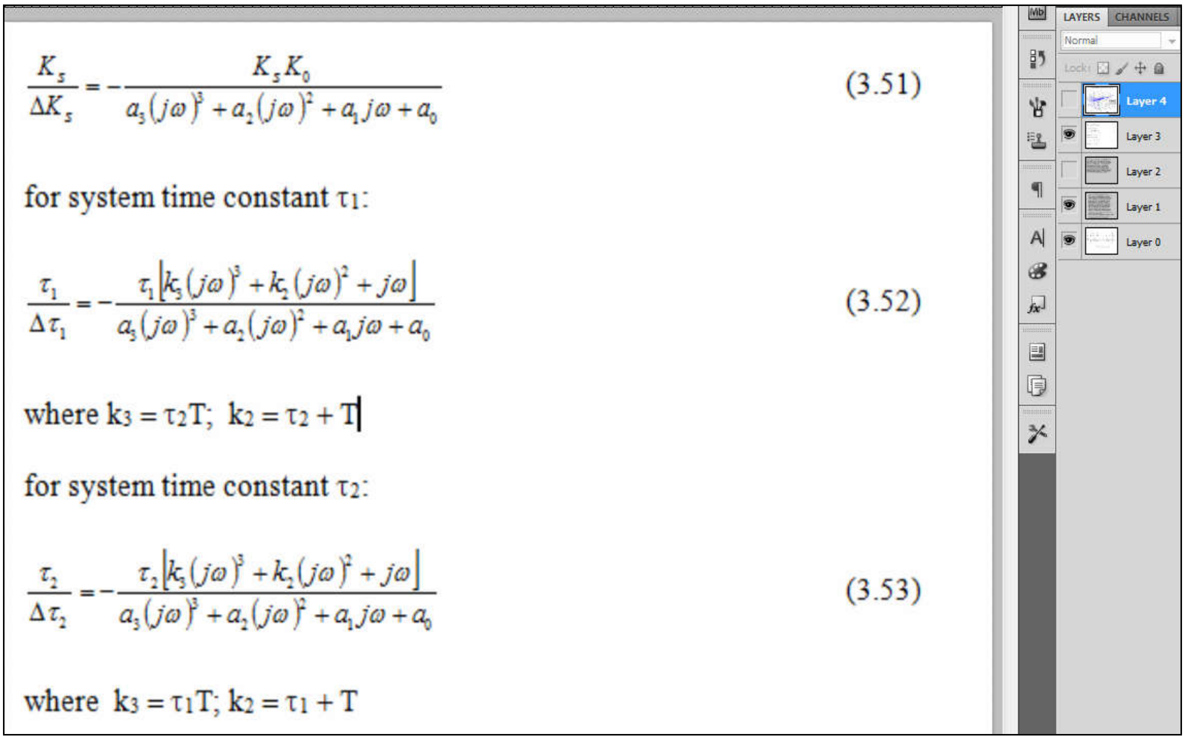
[7] HALENAR, R..: MATLAB - *Properties of dynamical systems : research of dynamic system properties*, Saarbrücken : LAP LAMBERT Academic Publishing, 2012, ISBN 978-3-659-18082-8, s. 62.



Picture 1: *Layers in Photoshop, created by Matlab*
Source: *Own creation*



Picture 2: *Photoshop with active and visible layer of grapg*
Source: Own creation



$$\frac{K_s}{\Delta K_s} = -\frac{K_s K_0}{a_3(j\omega)^3 + a_2(j\omega)^2 + a_1 j\omega + a_0} \quad (3.51)$$

for system time constant τ1:

$$\frac{\tau_1}{\Delta\tau_1} = -\frac{\tau_1\left[k_3(j\omega)^3 + k_2(j\omega)^2 + j\omega\right]}{a_3(j\omega)^3 + a_2(j\omega)^2 + a_1 j\omega + a_0} \quad (3.52)$$

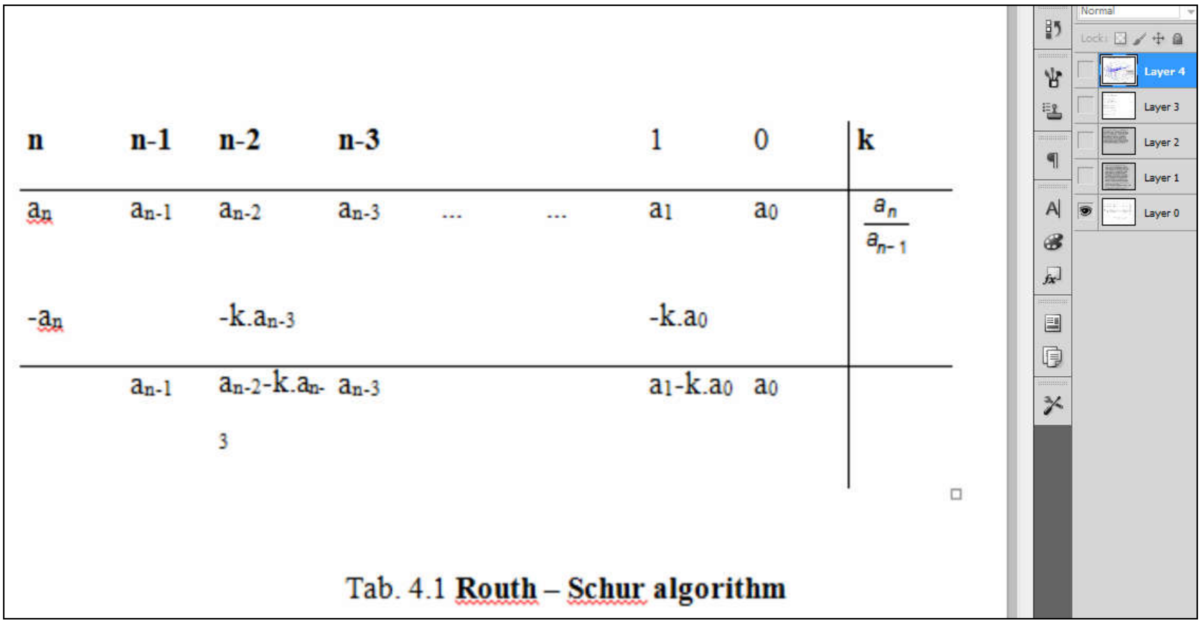
where $k_3 = \tau_2 T$; $k_2 = \tau_2 + T$

for system time constant τ2:

$$\frac{\tau_2}{\Delta\tau_2} = -\frac{\tau_2\left[k_3(j\omega)^3 + k_2(j\omega)^2 + j\omega\right]}{a_3(j\omega)^3 + a_2(j\omega)^2 + a_1 j\omega + a_0} \quad (3.53)$$

where $k_3 = \tau_1 T$; $k_2 = \tau_1 + T$

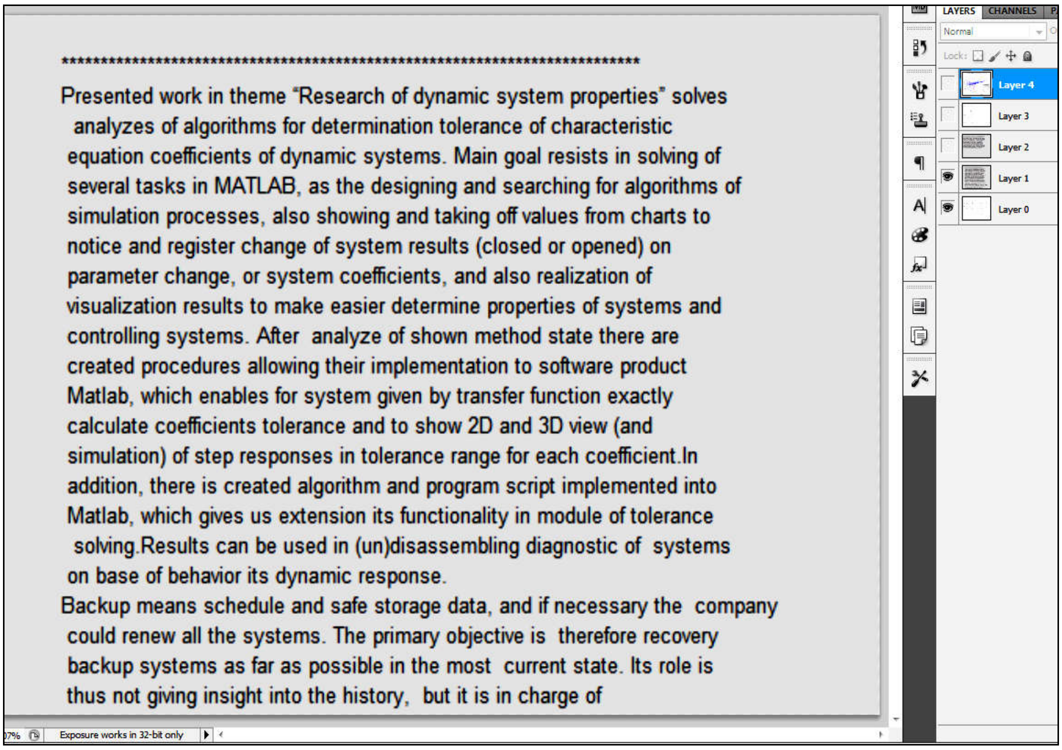
Picture 3: *Photoshop with active and visible layer of expression*
Source: Own creation



n	n-1	n-2	n-3			1	0	k
a_n	a_{n-1}	a_{n-2}	a_{n-3}	…	…	a_1	a_0	$\frac{a_n}{a_{n-1}}$
$-a_n$		$-k.a_{n-3}$				$-k.a_0$		
	a_{n-1}	$a_{n-2}-k.a_{n-3}$	a_{n-3}			$a_1-k.a_0$	a_0	

Tab. 4.1 **Routh – Schur algorithm**

Picture 4: *Photoshop with active and visible layer of table*
Source: Own creation



****

Presented work in theme "Research of dynamic system properties" solves analyzes of algorithms for determination tolerance of characteristic equation coefficients of dynamic systems. Main goal resists in solving of several tasks in MATLAB, as the designing and searching for algorithms of simulation processes, also showing and taking off values from charts to notice and register change of system results (closed or opened) on parameter change, or system coefficients, and also realization of visualization results to make easier determine properties of systems and controlling systems. After analyze of shown method state there are created procedures allowing their implementation to software product Matlab, which enables for system given by transfer function exactly calculate coefficients tolerance and to show 2D and 3D view (and simulation) of step responses in tolerance range for each coefficient.In addition, there is created algorithm and program script implemented into Matlab, which gives us extension its functionality in module of tolerance solving.Results can be used in (un)disassembling diagnostic of systems on base of behavior its dynamic response.

Backup means schedule and safe storage data, and if necessary the company could renew all the systems. The primary objective is therefore recovery backup systems as far as possible in the most current state. Its role is thus not giving insight into the history, but it is in charge of

Picture 5: *Photoshop with active and visible layer of text data*
Source: Own creation

5 Conclusion

We can see that any data stored as graphic object in form of layers are fully readable and recognizable. Advantages of this method are several:

- No need to keep more data object, just one can handle it
- No need to keep structuralized data, data are stored in unstructured form
- view records, it is possible to use some commercially available browsers * .psd files (eg. Free PSD viewer, IrfanView, MyViewPad, Picasa 3.6, etc.)
- virus resistible – psd data format cannot be virus infected, because is not executable
- method is fully automated by Matlab environment

Analyzes and project of environmental engineers contains many graphics and text objects where data sources can be different. This method can join many sources, which needs to be store in structured form to unstructured and that means, archiving would be more simple and robust.

References:

DIVINEC, L.: *Co by měl umět moderní datový archiv ve firmě?.* [online]. [2015-10-06]. Available at: <http://www.ictmanazer.cz/2011/11/co-by-mel-umet-moderni-datovy-archiv-ve-firme/>

HALENAR, R..: *MATLAB - Properties of dynamical systems : research of dynamic system properties*, Saarbrücken : LAP LAMBERT Academic Publishing, 2012, Pp. 62, ISBN 978-3-659-18082-8

MAK, J.: *Photoshop – úvod*. [online]. [2015-10-06]. Available at: <http://www.photoshopbook.net/photoshop-manual/photoshop-start.html>

MRÁZ, J.: IW: *Archivácia dát informačného systému.* [online]. [2015-10-06]. Available at: <http://www.itnews.sk/tituly/infoware/2010-08-23/c135372-iw-archivacia-dat-informacneho-systemu>

ONDROUŠEK, L.: Adobe Photoshop, [online] [2015-10-06], Available at: <http://www.ephoto.sk/fototechnika/recenzie/softver/adobe-photoshop/>

TRNKA. A.: *Big data analysis*, In: European Journal of Science and Theology, Vol. 10, suppl. 1 2014, ISSN 1841-0464. s. 143-148.

CONTACT ADRESS

Author: Robert Halenar, Ing. PhD.
Workplace: University of Ss. Cyril and Methodius in Trnava, Faculty of massmedia communication
Address: Nam. J. Herdu 2917 00 Trnava, Slovak republic
E-mail: robert.halenar@ucm.sk