

Document Management & Archiving: State of the Art

Egg, December 17, 2021: Normally, a review of the past year follows at this point. For once, this “tradition” is to be deviated from. As part of an invitation, the Managing Director was allowed to write an article for **SwissITMagazine (German)**, which appeared in issue 11/2021. Because the article is a good “navel-gazing” on DMS and archiving in general, it is published here in full length — this in the sense of a review of the year, which for once goes beyond the usual scope.



If you sketched out the basic features of archiving software in 1993, you could not have imagined in your wildest dreams what the state of affairs would be in 2021. Although the Internet and e-mail already existed, the Web played no role in the business environment. Mountains of paper, interfaces to accounting and Word & Co. were the measure of all things. The hardware was demanding. Scanners and hard disks cost a fortune.

Today, almost 30 years later, data carriers are available for a pittance and scanning plays a minor role. In short, no DMS project will fail in 2021 because of the hardware. Nor, incidentally, will it be because it is not possible to work with any end device. All DMS solutions should be web-enabled by now. The challenge is to build systems that are flexible enough to cope with constantly changing conditions.

Stages in the DMS and archiving process

Today, no one even thinks of printing out an invoice in duplicate and filing it in a folder. On the other hand, the processes in a DMS are often structured in a very classic way. The stages in the DMS usually correspond to the processes that are to be mapped in the commercial environment: Purchasing, Sales, Production and Service. Although DMS solutions have sophisticated search capabilities, many use the systems primarily for documentation and less as a digital library (knowledge management), although a high potential can be achieved especially in the service area.

Whereas 20 years ago a great deal of importance was still attached to constant written form (fax, fax and fax again), for some time now contracts have been concluded purely virtually, without a signature. In addition to the telephone, e-mail is the primary means of communication. But messengers are also becoming more common, and even video

chats are being used. The stages themselves remain the same. To put it bluntly, we collect information on the basis of which we act.

A DMS takes over a lot of tedious work by presenting structures and processes in such a way that no one has to worry about how to file a file. In general, it is increasingly no longer a matter of files, but of information chunks. These can be short messages or entire project folders. A DMS product does not replace an ERP solution. While the latter is responsible for the core processes (purchasing, production, sales), the DMS offers the possibility to accompany these processes and to document them in the long term. Regardless of which stages or processes are mapped in the DMS, the end result is legally compliant archiving. A DMS can be used as a pure archiving solution; the reverse case, building a DMS without archiving, makes no sense.

The evolution in the computer industry is gigantic. What seemed unthinkable years ago is now available on every smartphone. What is forgotten is that data volumes are constantly increasing and this increased volume must be kept available in the long term.

Storage concepts: on-premises or cloud

This brings us to the storage concepts that no DMS can do without. An example of how strongly data volumes are growing is given here: One hour of video contains 108'000 individual images. Expressed symbolically in A4 pages or corresponding processes, this results in 54'000 business transactions with two pages per invoice.

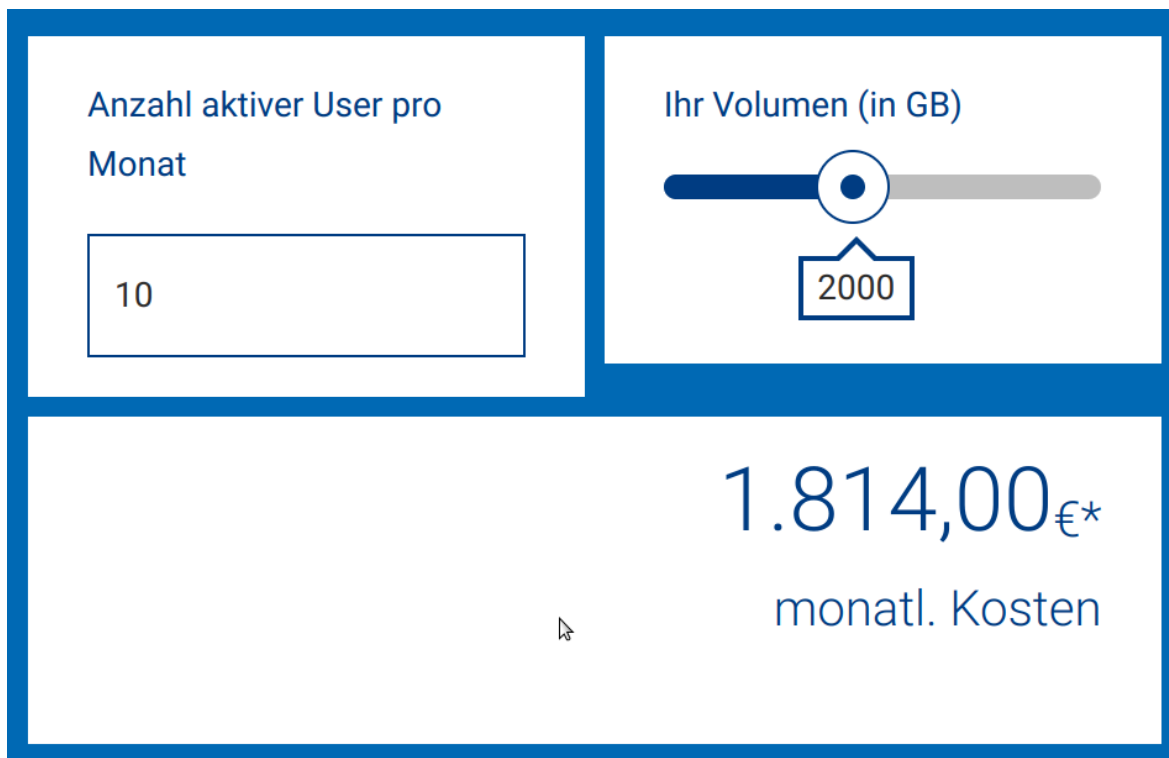
The question for companies is whether the data should be stored locally (on-site = on-premises) or remotely with an external service provider in a cloud. From a purely data perspective, it seems foolish not to store it locally. After all, hardware prices couldn't be lower. However, DMS solutions (like any software) rely on many components that all need to fit together. Systems need to be maintained and secured.

This is precisely what causes difficulties for many SMEs. Even companies with several hundred employees find it difficult to recruit suitable specialists or consider the risk of being dependent on one IT person to be too high, so they prefer to leave IT in external hands (and thus in the cloud). To be fair, it must be added that the DMS industry is not entirely innocent of this situation. Particularly as a representative of open source systems, one experiences again and again that it is not possible, or not possible at all, to keep solutions so simple that they can also be maintained by non-experts.





Cloud solutions offer an alternative here. A complete package is available for fixed monthly costs. The provider takes over the care, security and maintenance of the solution, while the customer can concentrate on his core competencies. Anyone looking for DMS offerings in the cloud will quickly discover that the prices are billed in gigabytes. Representative of many is the offer from a provider with server location in Switzerland: With ten users and two terabytes of storage, the solution costs around 5000 francs per month.

<p>Anzahl aktiver User pro Monat</p> <input type="text" value="10"/>	<p>VPN-Verbindung?**) <input type="text" value="Ja"/></p>
<p>Ihr monatl. Traffic (in GB)</p> <p><input type="range" value="50"/></p>	<p>Ihr Volumen (in GB)</p> <p><input type="range" value="2000"/></p>
<p>Ihr Serverstandort</p> <input type="text" value="Schweiz"/>	<p>4.746,50 €* monatl. Kosten</p>

The same solution, running on its own server, results in costs of approximately 2000 francs per month.



The price difference is therefore 3000 francs per month. Comparable hardware (with redundant hard disks) costs around 1200 francs; admittedly with 10 instead of 2 terabytes.

	WD Gold (10 TB, 3.5")	
–	2	+
		542.–
	HPE ProLiant MicroServer Gen10 Plus (Intel Xeon E-2224, 16 GB, Micro Server)	
–	1	+
		679.–
	Zusatzleistungen	
Total		1221.–

It can be deduced from this that cloud solutions are significantly more expensive than local servers, purely from a financial point of view.

Special case of mail archiving

Somewhat provocatively, it can be said that what used to be received in paper form is now received digitally as e-mail messages. It goes without saying that these have to be integrated into the DMS. The question is rather how the mails should be archived. Mail messages are in IMAP format. The structure seems archaic

Received: from XXX.XXX.XXX.XXXX (MailZZZ authenticated user mailyyyy)

Message-ID: <baa56acb35454f59b1821d2a840346cf@mail.archivista.ch>

In-Reply-To: <7B79@AS8PR08MB5942.xxxx.yyy.com>

Date: Wed, 13 Oct 2021 10:07:11 +0200 (CEST)

Subject: Re: WG: Anfrage Fachbeitrag "Swiss IT Magazine"

From: "Archivista GmbH" <master@archivista.ch>

To: "Simon Wegmueller" <wegmueller@swissitmedia.ch>

Content-Transfer-Encoding: 8bit

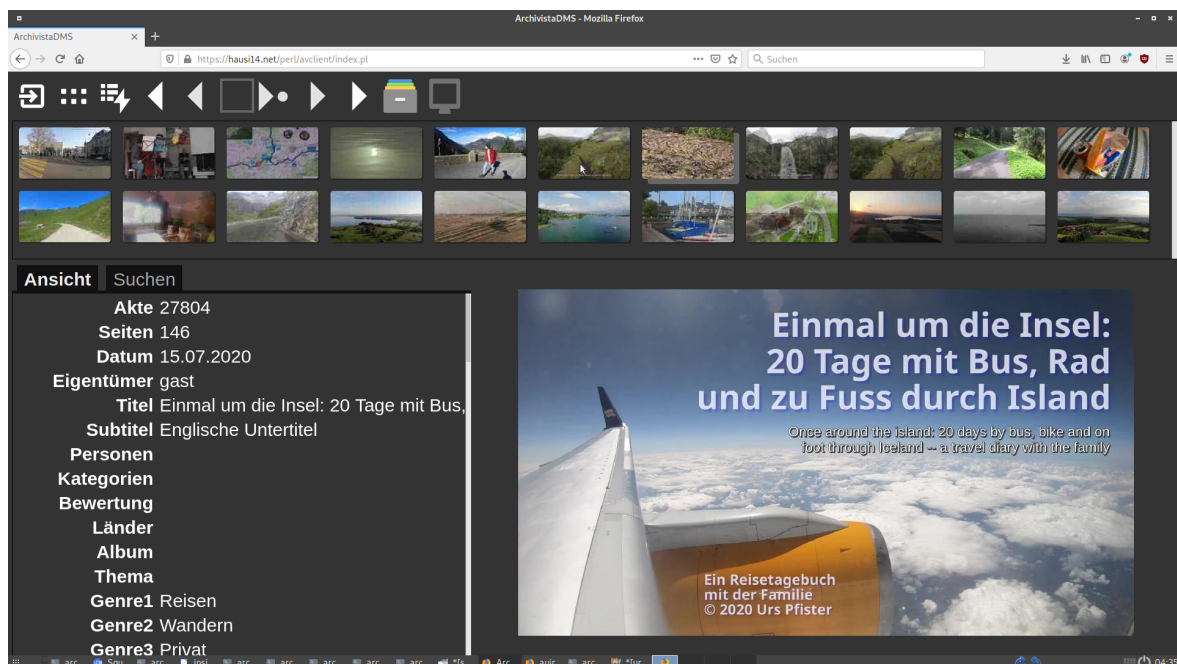
In fact, mail messages are still sent with 7-bit as text files. The problem now is that each mail program (sender as well as recipient) interprets the messages differently. To simply keep the text messages seems risky, because years later the mail program X

can only display the messages of the original sender more badly than right. For this reason, it is recommended to create additional PDF or image files from the text files.

Attachments can prove to be treacherous. On the one hand, compressed data is inefficiently converted into lines of text, and on the other hand (precisely because it is inefficient), only the links are often sent in the case of larger attachments. When such mails are opened, the file is downloaded on-the-fly with the link from the sender's external server. This works well as long as the files are available for download. Since the half-life (the period of time after which a web link no longer works) is hardly more than six months, a timely transfer to the DMS (while retrieving the links) is highly recommended.

Stumbling blocks had to be observed recently with Microsoft cloud solutions. Even the old Exchange servers had the peculiarity that, in addition to the IMAP format, a proprietary copy (Msg format) was always kept for viewing in Outlook. If such a mail message is deleted in Outlook, this does not mean that the IMAP copy is also deleted. In the opposite case, i.e. the IMAP copy is archived and deleted, the message still remains in Outlook.

This leads to the fact that when a certain size (usually 50 GB) is reached, no more mails are received or sent at all. This means that the entire mail traffic comes to a standstill. These are always the moments when the hotline is running hot, even though the problem itself is not one of archiving. Even if this sounds old-fashioned, the mails are usually deleted manually. Whoever thinks that 10'000 mails are quickly "removed" will find themselves reminded of old Explorer times in the Microsoft Cloud. No one could ever explain to me why the data could be deleted in seconds via command.com with "del *.*", whereas the file explorer in Windows sometimes took hours for the same process.



Integration of audio and video files

A few decades ago, just storing audio files caused significant problems. An uncompressed audio file requires about 10 megabytes per minute. For example, in a call center where ten employees make 8 hours of phone calls per 250 days, this results in about 11 terabytes of data. If the data is saved in MP3 or OGG format, the amount of data is reduced to one tenth.

The situation is even more accentuated for video files. Here, too, the volume can be illustrated with an example. The analog VHS cassette could record one feature film at a time. At that time, the local video store had about 3000 films on offer. This required a store space of about 50 square meters. Current streaming providers also offer several thousand films. If we consequently calculate with 3000 feature films of 100 minutes each, this results in 5000 hours.

To store this amount of uncompressed data, $25 \text{ MB} * 25 \text{ frames/second} * 3600 \text{ seconds per hour} * 5000 \text{ hours}$ would be required. This would require 536 hard disks of 20 terabytes each. In MP4 format, the data can be extremely compressed by saving only the differences in the colors between the images (and this also somewhat inaccurately). This reduces the 5000 hours to about two to four terabytes.

In short, without compression, we would all still be staring at our analog TV sets in 2021, or struggling with vast collections of VHS tapes. It is only through efficient encoding that we can move into dimensions in which multimedia data can be handled efficiently. In order to be able to archive these collections in the long term, outsourcing to M-Disk can be highly recommended – especially in the SME environment.

The millennium data carriers (meaning the manufacturer's promise of a service life of 1000 years; realistically, a few decades are likely) offer 100 gigabyte capacity and are compatible with Blu-ray discs. Drives cost a bit over 100 francs and a good 25 francs per data carrier. Thus, swapping out two terabytes generates costs of 600 francs (including the drive). Compared to the hard disk, the price seems high. If the DMS cloud offer is used, where the monthly subscription costs are higher by factors, the financial outlay is modest. Optimally compressed, about 200 hours of video can be archived on an M-Disk.

Data saved in this way has the advantage that subsequent changes are impossible, since the media can only be written to once. Loss is therefore only imminent if the data media are stolen or damaged. One hundred percent of the data can be prevented from being hacked with encryption. Some DMS systems offer interfaces for outsourcing to M-Disk. Otherwise, every SME can be advised to manually create unchangeable copies of sensitive data every year.

DMS solutions that have an internal media player can play audio and video files directly in the web browser. This saves time and bandwidth, since only the part that is played is transferred to the web browser. It should also be added that listening to audio and video files to find a specific part is time-consuming. DMS systems that have speech

recognition offer the advantage that searches can also be text-based. For example, if calls are recorded in a call center, knowledge archives can be built automatically. In the meantime, there are several open source products that ensure that the data does not fall into the wrong hands.

Current market situation in Switzerland

The number of providers and the way in which DMS solutions are implemented could hardly be more different in 2021. The customer is therefore spoiled for choice. In recent years, there has been a certain amount of consolidation among providers. Examples include the acquisition of Docuware by Ricoh and that of Habel and Proxess by Beta-Systems in 2019. In 2020, the acquisition of Easy Software by Deltus 36 was announced and completed. On the other hand, there are always new players, the market remains not quite easy for interested parties in the overview.

On the technological side, there has been a strong emergence in recent years: Solutions that are not Internet-capable will have a more than difficult time. What's more, a DMS today must not only be operable via a web interface, but access via an API (Application Programming Interface) is a must-have. This is the only way to combine or automate solutions independently of devices and locations.

Whether the solutions run on-premises or in the cloud is not even that central. Depending on what a customer wants to do himself or delegate, he will opt for on-premises or the cloud. Prices in the cloud are rather dignified high. After an initial euphoria about the cloud (with low purchase prices), sobriety is likely to have set in. The all-round carefree package with a Swiss location has its price.

From this perspective, the question is not primarily whether a DMS runs locally or in the cloud, but whether sufficient resources and expertise can be built up for security at the local site and whether this could be obtained more cheaply in the cloud. Because one thing is clear: In times when the home office is booming, location-independent access is so central that no SME can or wants to do without it. Regardless of this, it is simply a pleasure to be able to work from anywhere on the road with any device.

Company vacations 20.12.2021 to 4.1.2022

The time between Christmas and New Year is used to recharge our batteries. For this reason, our store will be closed during this time. Of course, it is guaranteed that customers will receive support at any time during this period. We wish everyone a relaxing holiday season and a happy new year.

P.S.: A security vulnerability is currently circulating in systems that use the log4j library. This Java library is not used by the ArchivistaBox or AVMultimedia. Nevertheless, an updated version 2021/XII is available for customers. This fixes a problem with user accounts for certain special characters.