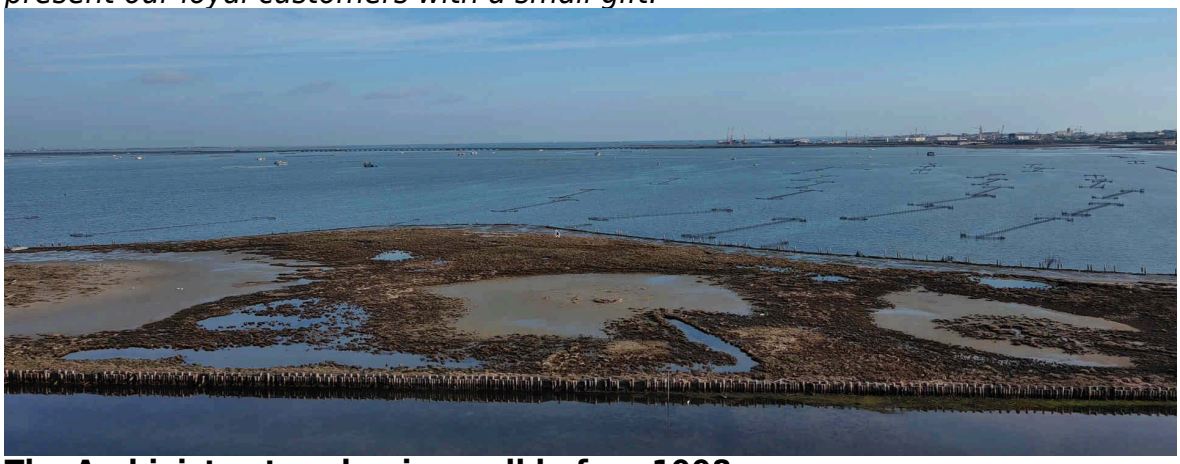
Document management for eternity

Egg, 31 March 2022: *Archivista GmbH is celebrating its 25th anniversary these days. Our solutions have been on the market for a quarter of a century now. In a globalized time, in which many no longer remember yesterday's promises today or tomorrow at the latest, 25 years is a proud proof of continuity. This is not just about promises to keep data cleanly available for decades, but also about the philosophy of maintaining data structures in a way that makes such longevity possible in the first place. The following is a review of the company's history with an outlook into the future. We would also like to present our loyal customers with a small gift.*



The Archivista story begins well before 1998…

In the early 90s of the last century, computers were already widely available in SME companies, but in the end, documents were put on paper and sent by post in such a way that the mountains of paper did not exactly grow to the sky, but still filled archive shelf after archive shelf. The hard disks of the computers ranged around a few GBytes, those who had more than a few MBytes of RAM were in the upper class. If you had a scanner, you could easily live with the fact that capturing an A4 page took a minute or more. OCR (optical character recognition) was in its infancy.

And yet, as early as 1990, there was this dream of having all the information available in paper form digitized on the computer. Ultimately, the project began with programming some macros in Photoshop to make capturing and finding scans easier and faster. In short, the manual work was to be minimized. It was soon realized that even with a few hundred files, the search became not only tedious, but even more time-consuming than sifting through the paper folders of the time by hand.

Years 1993-1998: The search for the ideal solution

Thus, in 1993, the first software was developed to be able to assign structured terms to the scanned documents in a database mask. Initially, it was not planned that a product would emerge from this. Rather, at some point the time had come when the solution created in-house was to be replaced by a product. In Switzerland at that time the Orbit computer fair was the measure of all (new) achievements. Unfortunately, however, all reasonably affordable solutions had the disadvantage that the size of the archive was limited by the size of the hard disk.

There were CD burners for storing 700 MB of data on a non-rewritable CD. Manufacturers waved it off. It was not durable, there were no standards for it, the CD was too slow, it was too complicated etc.… Today it may be added, the CD of that time can still be read. And even if it is now possible to archive 100 GB per disc, the arguments of the manufacturers against outsourcing the data have remained. It is not worth it, redundant hard disks and outsourcing the data to the cloud are much safer. It almost seems as if the arguments remain the same…



1998-2000: Company foundation with Archivista version 4.x
At the end of March 1998, the time had come. The first Archivista solution was available for purchase. **Our web museum (only in German) recalls the beginnings of our solution.** The beginning was not exactly easy. Even then, a Swiss standard software solution, how does this work? The country manager of a Japanese company, which still sells document scanners today, summoned me to his office and tried to make me understand that it was crazy to launch a product on the market for a few hundred francs.
Kodak's national representative acted better. With a lot of tact, he sounded out whether the company Archivista GmbH could be integrated into the group. A partnership was even formed. Archivista GmbH bought the text recognition from Kodak Switzerland and we sold the Archivista solution to Kodak so that they could offer an SME product. Half a year later, he called to say that the invoice still had to be sent today, and that he would arrange this tomorrow, and that the Swiss branch would be "closed" in a few days. Ten years later, Kodak would no longer exist… As much as the takeover by the giant at that time had been tempting, the Archivista solutions would probably have disappeared from the market sooner or later.
Another small anecdote should be added here: An employee of a Swiss bank wanted to order an Archivista solution. Apparently he could not really get his way with his superiors, which led him to order a system for several thousand francs on 25 May 2000, using the bank's copied logo. The bank paid without complaint, but the employee had to vacate his post — and the project petered out. At that time, the disappointment at Archivista GmbH was quite big, today it is known that the said bank is currently disappearing from the scene. Even if large corporations have their appeal as customers, it is probably more in line with the philosophy of our products if they are developed and maintained by an SME for SMEs.
And something else should be reported from this period. A large magazine tested the Archivista solution at the time and found that the data was mothballed for eternity. At the time, this did not cause us to go into raptures. In retrospect, it can be said that keeping data available for decades with the right "care product" (i.e. ArchivistaBox) is no small feat and, after 25 years of company history, now provides the right context.



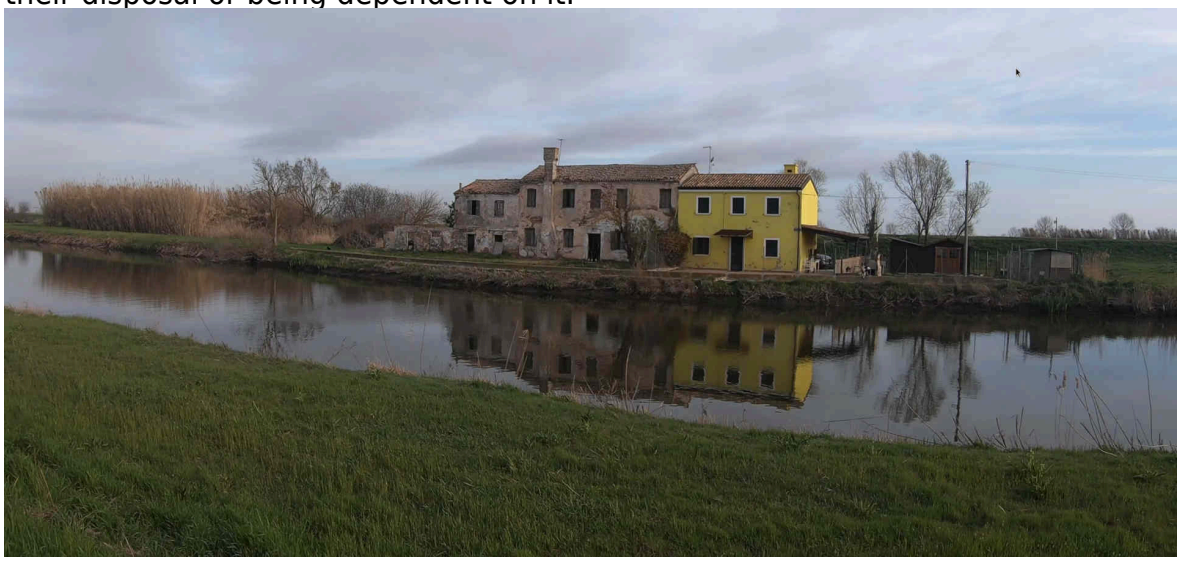
2001-2010: From Archivista to ArchivistaBox

The Archivista solutions were originally developed for Windows. Over the years, unfortunately, we experienced that Windows became more and more difficult to maintain and service for our products. It also became apparent on the horizon that the Internet would sooner or later begin its triumphant march.

For this reason, we first positioned the database in 2001 and the entire solution in the years 2004 to 2007 as platform-independent(er). This led to the fact that today we work almost exclusively from and with open source products.

When we chose Linux as the platform for the ArchivistaBox in 2004, it was a courageous decision. At that time, nobody assumed that one day the majority of all computer devices would run on Linux. And likewise, no one assumed that all tech companies or their business models would not function without Linux.

However, unlike these giants, our solutions do not serve to collect as much data as possible from our customers in order to “exploit” it commercially. The purpose of our products is to enable our customers to manage their data independently and to have control over it at all times. In the short term, this may yield little(er) profit. In the medium to long term, we will see which companies have more substance. Those that have a high degree of autonomy over their data, or those that either live from having other companies’ data at their disposal or being dependent on it.



2011-2018: The ArchivistaBox comes of age

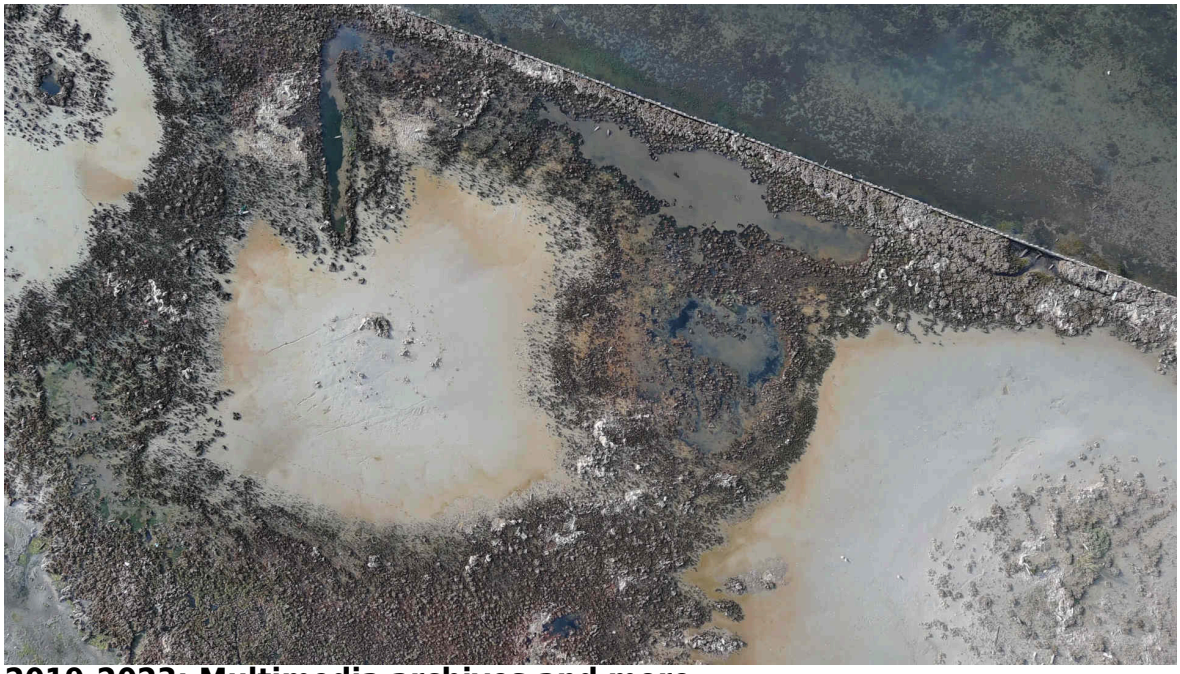
While Archivista products were originally pure software solutions, this changed with the widespread use of ArchivistaBoxes. Suddenly it was not only a matter of making sure that Archivista ran smoothly as document management software, but also the substructure, in short, the entire ArchivistaBox had to be looked after and “cherished”.

This led in the years from 2010 onwards to an in-depth examination of the interaction between the operating system and the software based on it. This resulted in our own Linux distribution running completely in main memory. Only this made it possible for us to maintain the ArchivistaBox with minimal effort. The fact that the operating system is built up completely automatically at every start means that there is no need for installation. Start the USB stick or ISO file from the hard disk, and within a few seconds every ArchivistaBox is ready to work. There are no more updates as a result. Simply use a new USB stick or ISO file to start, and that’s it.

This automation is still unique on the market, and to this day it is hardly noticed. Users don’t notice the “magic”, because they don’t even recognize what’s different about starting ArchivistaBox, and the experts seem to prefer to shovel thousands of packages back and forth every day. Be it because it is a habit, be it because this complexity means that no one wants to operate local servers anymore, making it all the easier to propagate cloud solutions.

However, this outsourcing of data and services ultimately means that autonomous work is no longer possible. The future will show whether cloud solutions will enable data stocks to remain securely “controllable” over several decades. A certain skepticism would be appropriate here (viewed very pragmatically). For example, it is relatively easy to transfer daily data volumes to any point in the world. It is more difficult or even impossible to copy only a few dozen GBytes, let alone TBytes, back to the local instance within a reasonable time (e.g. a few hours).

Currently, about 1 TByte per hour can be copied from hard disk to hard disk of conventional design, with SSD disks it is about 2 TByte per hour and with NVME (admittedly only via the right interfaces) up to more than 10 TByte per hour can be both backed up (backup) and restored (restore). Such a data throughput is still unimaginable in the network. Of course, it may be mentioned here that only a few SMEs have data stocks with several TBytes. However, this is only the case as long as only “normal” data (documents, images) is available; with multimedia data, on the other hand, several TBytes are quickly reached.



2019-2023: Multimedia archives and more

This marks the last 5 years of the company's history. In **2018, to celebrate the 20th anniversary of the company, we released the multimedia Linux distribution AVMultimedia.** At that time, it was not planned that ArchivistaBox would ever become multimedia. However, in the last few years, it has become impressively clear how strongly multimedia content pushes conventional information to the side, if not displacing it altogether.

For example, press releases used to be published everywhere. Today, there are almost only live streams. A current example: the end of CreditSuisse is streamed directly from the Federal Palace to the entire world at prime time on Sunday evening. If you look for a press release to go with it, you will find links to the streams.

This transformation from conventional document management (DMS) to a multimedia management system (MMS) is demanding, as multimedia data means several times larger data volumes. For our ArchivistaBoxes, this means that the current Box systems have significantly more capacity than was the case just a few years ago.

In addition to the multimedia capabilities of our solutions, automation has been another focus of the last five years. Thirty years ago, it was often the case that the majority of documents had to be scanned and indexed. Today, between 60 and 95 percent of documents are already delivered digitally and processed automatically. In addition to the classic 1D barcode, many files today contain 2D barcodes (QR codes).

The challenge here is that digitally delivered documents occur in batches, i.e., in the past, a few tens of thousands of documents might have been "shipped" to the DMS per day at peak times, whereas today a multiple of files can occur per day. Likewise, files occur that are much larger. Until 2019, the maximum size of a file was 512 MByte; currently it is 64 GByte per file.



What does the future hold?

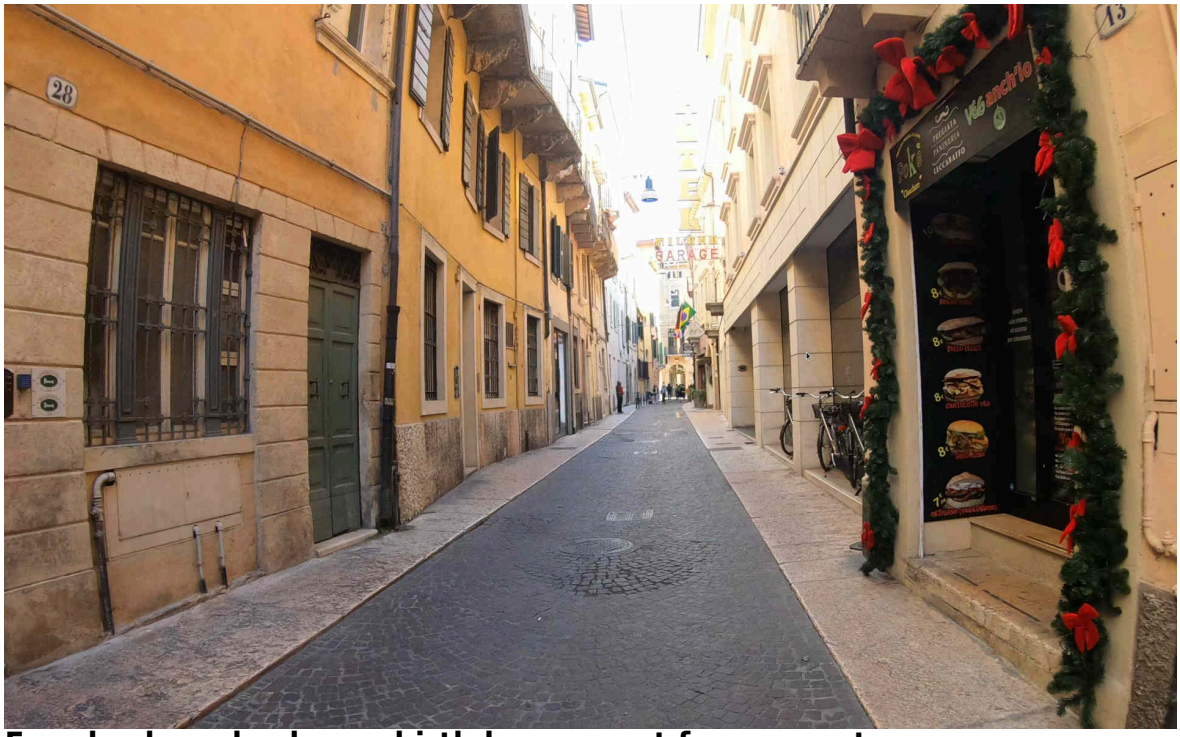
First of all, a rolling stone gathers no moss. With this in mind, the ArchivistaBox is constantly being further developed. What the future will bring remains to be seen. It is foreseeable that the complexity of the systems will increasingly represent a greater challenge. Here is an example: The Internet would not have come into being without standardized protocols. In addition to the HTML standard, mail messages have played a central role up to now. In itself, mail messages are sent via relatively simple protocols; 7-bit coding (normal characters, without umlauts) will still be used in 2023. In addition to SMTP for sending messages, IMAP has become established for fetching them (originally there was still POP). However, after the mail servers were increasingly running in the cloud, the largest provider decided to discontinue both sending via SMTP and fetching via IMAP.

The argument that new protocols lead to greater security is open to question. Primarily, standards are being destroyed that were never implemented 100% standard-compliant by the same companies before. Unfortunately, it is quite possible that in a few years we will find a web that is dictated even more or all-encompassing by the tech giants.

It must also be mentioned here that the largest search giant is currently trying to ensure that HTTPS certificates are no longer valid for approx. 400 days, but that the time span is shortened to 90 days. Apart from the fact that the majority of the information would not have to be transmitted in encrypted form (especially since the manufacturer in question is well known for copying all the information in its own browser), this means that the services are not primarily made more secure, but that significantly more technology is required to create the impression of trustworthiness, which does not exist.

Digitally prepared information that is only displayed on screens is not secure in principle. On the one hand, all data can be changed at will in fractions of a second and in a very user-specific manner, and on the other hand, it is never possible to say with certainty what is true and what is false with digitally created content. As much as the continuous capture of information directly on the computer is (more) cost-efficient than if paper and pen were used for this purpose, it is impossible to determine whether the data was captured by humans or machines.

Artificial intelligence (AI) is currently being praised to the skies as if it were almost a matter of establishing a new world religion. This lurid choice of words is forgiven, but in the end it must simply be added that this AI does nothing other than to come from a more or less large pile of data via algorithms to results that are not completely unknown to our memory. However, the algorithms are not transparent, so that it will be all the more difficult in the future to distinguish the correct from the incorrect information.



Free backup check as a birthday present for our customers
Finally, what would a company anniversary be without a gift to our customers? Archivista GmbH is known for investing resources in its products instead of throwing parties and giving presents.
Nevertheless, after 25 years there is a gift for our customers with very practical benefits. The best data on the best ArchivistaBox is only useful as long as backups are made of it. And of course it is part of the duty of our products to have good backup concepts. Ultimately, the best backup is only as good as it is regularly executed and checked.
Customers can check backups at any time. However, just because a log file says the data is backed up does not necessarily mean the data is correct. It should be noted that for the past 25 years, none of our log files have falsely led our customers to believe that the backed up data did not exist according to the log file. However, there were and still are cases where the control of the backups was neglected.
In order to establish more awareness, short and well more security here, we offer a free backup check to all customers on the occasion of our 25th anniversary. Just call +41 44 350 05 60 and we will have a look together if/how the backup concept is implemented, if there were or are problems and what could be done better in the future. **The offer is valid for all customers with a valid maintenance contract until May 31, 2023.**