

VITAM, an interministerial archival program

Jean-Séverin Lair

in charge of the Tech.Gouv Program, Direction Interministérielle du Numérique (DINUM)

Abstract:

Archiving is indispensable, even in the digital realm. Public archives are a key to the continuity of the state's activities. Without public digital archives, the state risks losing its memory. Digital archives have to harmonize the nanosecond with the century, flows with stability. VITAM, the interministerial program for digital archives launched in 2015, hinges on developing a scalable, simple and easily interfaced freeware for archiving activities that can be used to securely manage billions of "objects". Initially borne by the ministries of Culture, Foreign Affairs and the Armed Forces, this program has attracted many users, public and private, to become involved in a shared approach. The first phase, 2015-2019, has come up with a product that, supported by a user community, will evolve and be enhanced.

A program for public e-archives

Archives, a basic need... even in the digital realm

For nearly all readers (and probably even more for geeks), the very mention of the word "archives" instantaneously brings to mind images of parchments, dusty shelves and a gray-haired personnel. Let's go modern! With digital technology, we don't need any of that... our memory now has a planetary scale, some would claim. Let me not cast the first stone. After all, I was close to endorsing this caricatured view of the archives (and of archivists!) before getting "educated" (*i.e.*, "led out of" savagery, according to etymologists).¹

Let's come back to the basics of public archives, in the words of the philistine who I am (I beg the pardon of any archivist reading this and ask to be excused for this much too simple account... made with the intent of contributing to popular science):

- The purpose of public archives is to keep the records of the administrative activities and decisions that affect the rights and obligations of physical and legal persons in order to consult these records during their period of effectiveness, invoke them as evidence with probative force in courts of law, make sure they are transparent and capable of being used to hold decision-makers accountable and, finally (for certain archives), to preserve the memory of the past.

¹ Unfortunately, I cannot list all the names of persons on the VITAM work team, but I would like to draw attention to the part played by Edouard Vasseur (current professor at École Nationale des Chartes) and Frederic Brégier in VITAM and by Mélanie Rebours in its promotion and the cohesion of the work team. I also acknowledge the team that has taken up the relay on VITAM, in particular Emmanuel Laborde, who heads it (and was technical director since 2018), and Alice Gripon in charge of diffusion. The series of public servants and service-providers who helped us take up this technological and human challenge can be sincerely proud of the results. My thanks to all. For current events at VITAM, cf. www.programmevitam.fr or Twitter @progvitam.

This article, including quotations from French sources, has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor's approval, completed a few bibliographical references. All websites were consulted in December 2020.

Let me emphasize two points:

- With all due regard for those interested in history and contrary to popular belief, archives are useful, far above all else, not for scientific research but for administrative purposes and to the benefit of organizations and persons. This has, in fact, been so from the very beginning of archives (for preserving titles of nobility, deeds, etc., for royalty). Archives are the key to the continuity of the state's activities and the validity of the decisions made in its name.
- Nowhere in this definition did I mention paper. Obviously, more and more activities and decisions are electronic. Ten years ago, dematerialization was already under way but in stages, some of which still maintained paper documents. Administrative procedures have now reached the point of being fully digital. This means that, in the absence of e-archives, we will be losing essential evidence or that this evidence will lose its effectiveness. What to say about the real estate register with its records of property rights and easements, about court decisions, about naturalization procedures, and so forth? Without e-archives, the state risks "catching Alzheimer's", not to mention the loss for history that would be observable within a few decades.

So, we admit that archives are important. But we know how to store e-documents on hard drives and even in the cloud, which has an enormous storage capacity... so, why do we need e-archives?

Information in archives is stored in safety and in a good enough order so that it can be relocated dozens of years later. For these tasks, proven processes are used; and persons, assigned the duty of warranting that this information has the value of probative, historical evidence. These requirements are very partially satisfied by heaping information onto hard drives, as we have been doing for twenty years now, or dumping it into file repositories (in the clouds) equipped with a search engine. These ordinary procedures work perfectly in the short run, but archives face the long haul and have to provide warranty. Forming an e-archive means that a nanosecond has to be made compatible with a century, flows of data with proof. To transpose age-old professional practices to another medium, new tools have to be used while retaining the requirements related to the long term and the warranty of proof — qualities direly lacked by computers.

In 2011, the French state realized it was time to act

The Ministry of Culture has the duty of keeping interministerial archives. It was alone in 1982 when it adopted a procedure ("Constance") for e-archiving documents. Though mainly based on manual procedures and other means that would now be deemed amateurish (magnetic bands, databases, etc.), it did archive some major documents, namely: big surveys conducted by the French National Institute of Statistics and Economic Studies (INSEE) (e.g., the 1962 census) and statistics from various ministries. Meanwhile, the volume of information to be stored was growing exponentially. In 2011, the National Archives reported that the contents to be stored by various ministries were of a larger volume than all the archives collected since 1982. Action had to be taken!

In France, three ministries manage most of the state's historical archives: the Armed Forces and Foreign Affairs for their own needs, and Culture for all other ministries (except Finance). In 2011, all three were mulling over the possibility of establishing e-archives. For want of sufficient means, they had no choice but to cooperate. However (and this is an important point) it was not just the departments that managed the existing archives in these three ministries that would work together and draft what, at the start, was a mere project. In addition, the departments of information systems in these three ministries also participated.

It took nearly four years to raise the means needed. The key person behind the launching of this project was Jacques Marzin, the interministerial director of Systems of Information and Communication. He was able to muster backing from the Future Investments Program under the CGI (Commissariat Général à l'Investissement) to complete funding from the ministries. At the start of 2015, the program Valeurs Immatérielles Transmises aux Archives pour Mémoire (VITAM) was born for “immaterial values conveyed to the archives for memory”.

The VITAM program

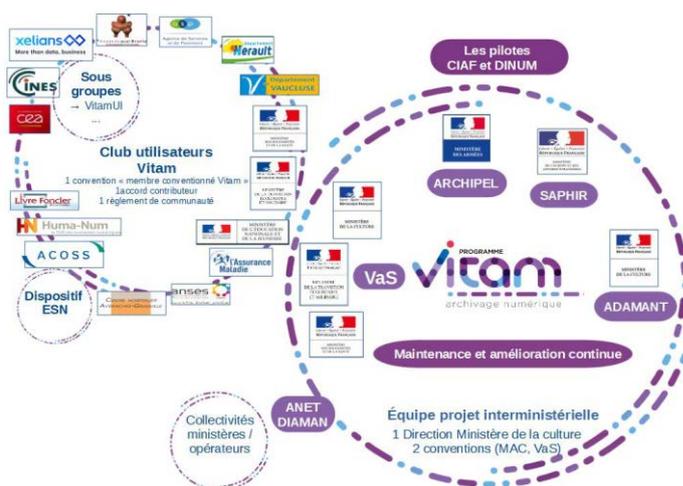
Zooming the basics

The program’s objectives are:

- To invent a free e-archiving software, VITAM, that will manage the conservation of very large volumes of e-documents, which can even be classified, and oversee the security of consultations of them.
- To set up archival platforms with the help of the aforementioned software in each of the three ministries (via ministerial programs: Saphir for Foreign Affairs, Adamant for Culture and Archipel for the Armed Forces).
- To diffuse and reuse as much as possible VITAM for historical and intermediate archives.

From Vitam’s “manifesto”, a term that summarizes the main points of a project using the standards of “agile software development”: *“For the services of the administration, in order to satisfy the need to have access over time to their e-documents, Vitam proposes a free software solution for archiving activities that is simple, adaptable and easy to interface, that serves for the secure management of single objects or of billions of objects, and that is intended for adoption by as many public services as possible.”*

Figure 1: The actors involved (at the start of 2020)



Vitam was introduced at the Ministry of Culture (Adamant) in 2018 and the Ministry of Foreign Affairs (Saphir and Xelians-Locarchives) in 2019. Software development:

- Agile Rhythm
 - Sprint (3 weeks), release on the third of the month; major version every year.
 - V1 early 2018, V2 early 2019, V3 early 2020
 - Up to four “scrum teams” in parallel (more than 10000 commits)
- Architecture:
 - Microservices full REST
 - Stubs “BigData and schemaless”: MongoDB, ElasticSearch
 - Storage “agnostic and extensible”: File system, Objet (SWIFT, S3), Bande (LTO)
 - Proofs: the chaining of log files, Merkle trees, time stamps on batches of operations (close to the concepts used for blockchains)
 - Multisite, multiholder, 3-75 VM, even more.
 - Automated by Ansible
 - Languages and OS: Java 8 and 11, Debian and CentOS
- Performance
 - Designed for billions of documents (but single units can be manipulated) in petabytes.
 - Regularly tested on more than 400 million objects
 - Test example: 160,000 office documents archived (with full guarantee of legal conservation) in 18 minutes.
- Community
 - A principal stub and satellites (libraries, human-computer interactions) free (CECILL) on GitHub
 - More than fifteen partners, public services are users (with their own or pooled implementation procedures), including two active contributors (CEA, CINES).
 - More than ten (among the biggest) companies that provide digital services have been trained on VITAM.
 - Xelians, the leader of private archives in France, has built its offer around VITAM and is contributing to the code.

A program deliberately conducted in an unusual way...

Having drawn lessons from the National Pay Office and wanting to use VITAM to demonstrate “another way” for “successful” big projects, the DISIC (forerunner of the Interministerial Direction of Digital Technology: DINUM) laid down the following requirements:

- The project has to use agile software development procedures.
- Several service-providers have to be involved so as not to fall under the control of a single firm that dominates the market.
- Public servants, experts, operatives and technicians have to control the choices made.
- Plans for implementation have to be part of the project of software development so that each stakeholder is accountable to the others.

Besides the DISIC, the CGI has also set requirements since it has provided funding under a program for stimulating economic development. It has asked for an approach that develops uses and, if possible, economic activities around the solution to be retained.

Given the force of these two sets of what were, at the time, atypical requirements for project management, the VITAM program has become an agile program, with choices controlled by the administration and with the priority of responding to the needs of users, who will promote for its wide usage. This strongly cooperative approach has been established both with archivists from all horizons (ministries, local authorities, etc.) in order to better detect their real needs (and not just their wishes) and, too, with computer scientists who vouchsafe that the proposed solutions are technically realizable in the environment.

So, the VITAM program is no longer just an information system project. It is an operation for building a product for use by a committed professional community. By gaining credibility, it has attracted interest in the business world.

Collaborating on change...

VITAM has been able to exist because it responded to a need strongly felt by the whole professional community, which was being forced to evolve as it faced the digital revolution. Since the medium of information was changing, new tools were needed to handle this information and make it durable. But what about archiving practices?

Most of the century-old ideas used to manage paper archives had a reason that reaches far beyond the form or medium of the information. So, they still have meaning in the digital era. Regardless of the medium, the objective of archiving is still to recuperate information that is useful to keep under conditions that warrant for its validity over time. Furthermore, it should be possible to relocate the information when needed, even years, decades or centuries later. As work on this project advanced, digital technology led us to ask questions about some of these central ideas.

For example, archiving natively electronic documents has opened onto the possibility of processing a document or even an elementary piece of information. This could not be imagined for paper archives, which were managed manually. The purpose of archivists and computer scientists working together is to find the right way to put digital technology to work. Should we rely fully on searches “in bulk” like those done by the big search engines? A more granular view of a hierarchy with metadata (selected at all points in the hierarchy down to the final piece of information) was adopted. Search engines thus put their full power to use on a huge volume of quality data.

Other ideas, like the reception and destruction of documents, the change of depositary, or the life-cycle management of documents have come under review. In each case, practices could be improved thanks to digital technology, instead of just being criticized or eliminated.

Some new concepts have appeared that need time to grow. The most complicated one has to do with the perpetuity of information. An old piece of paper, if properly kept can always be read (even if this means taking a course in paleography). But an old electronic file? Who has never had the experience of coming upon a document, photo, or memo on a hard drive that no one is able to open or that, if opened with current software, turns out to be illegible?

The work done with the community assembled around the VITAM program has helped many archivists to adapt their professional practices to digital technology. It has also made the computer scientists who are working on the software to understand both the reasons for existing practices and the real needs of archivists, to understand requirements and to lay out paths for the future opened by digital technology.

A conclusion compels recognition: archivists are still qualified for a probative governance of information in the long run, maybe even more so in the volatile world of virtual reality.

And the program goes on...

VITAM was initially foreseen up to V3, a version that, released at the start of 2020, has all essential features for digital archiving activities. However the program has not come to an end. For one thing, the decision was made to see to the software's maintenance and ongoing improvement by maintaining the work team on the job, using the means provided by the three ministries at the origin (Culture, Foreign Affairs, Armed Forces) and, too, facilitating the activities of the community of users and contributors. For another thing, an interministerial project for pooling archives in a service mode (VAS: VITAM As a Service) has rallied the ministries of Culture (for its management of our cultural heritage), of the Environment and of Social Affairs. This project completes the development of VITAM's back-office thanks to the strides made in human-computer interactions. The genuine user community that has formed around VITAM supports and sustains the product, which will thus be able to evolve and keep up with the digital transformation of the jobs and practices of archivists.

I was lucky to have been at the helm of this adventure from mid-2015 till the end of 2019. I thus discovered the world of archives with its extraordinary professionals who have an underestimated but essential spirit of public service.