Claire Sibille - de GRIMOÜARD*

THE DATA EXCHANGE STANDARD FOR ARCHIVING: KEY STANDARD FOR THE DIGITAL ARCHIVING POLICY OF FRENCH ARCHIVES

Abstract:
In 2006, the Archives of France Department (Ministry of Culture and Communication) developed a Data Exchange Standard for Archiving (SEDA) in collaboration with central public services charged of the State modernisation. Since then, the SEDA has been incorporated into the governmental recommendation framework for interoperability and public agencies setting up archiving processes applied it as a tool of dialogue with public archives services. Thus, we can say that the SEDA plays a key role for the French public administration. It was standardised by the French standardisation authority, and the national standard NF Z 44-022 has just been released.

Key words:
digital archiving policy, e-government, interoperability, OAIS model, standardisation

INTRODUCTION
Data processing and office automation, already present in public services, in private companies and in households for several decades, continue to develop. However, far from the aim of “zero paper” that the emergence of micro-computers promised us, now we look into the objectives of productivity improvement or services of better efficiency and better quality that data processing could bring.

* Claire Sibille - de Grimoüard, Interdepartmental Service of French Archives, Ministry of Culture and Communication, 56 rue des Francs-Bourgeois, 75 141 Paris Cedex 03, France, contact: claire.sibille@culture.gouv.fr.
The public services take an active part in this evolution of our society. A typical example is the take-off of e-administration, which is constantly accelerating nowadays with different governmental programs. Electronic exchange plays an important role within a context where mainly the organisation efficiency and the exchange facilitation are looked for. The Data Exchange Standard for Archiving or SEDA is placed at this strategic level of the information system. It aims at facilitating the interoperability between the information system of an archival agency and the information systems of its partners and it provides a framework for the various specific transactions which may occur: transfer, delivery, destruction, etc.

Section 1 describes the legal framework of French e-administration. Sections 2 and 3 present the scope and design concepts of the Data Exchange Standard for Archiving as well as the process of its standardisation.

1 E-GOVERNMENT ISSUES

Since the late 1990s France has been involved in the development of e-government, like its European neighbours. Of course, the dematerialization of administrative documents is not a new phenomenon. However, e-government added another dimension to data exchange between public services and actors interacting with them (citizens, private companies) and to automated administrative workflow performed by information systems. These requirements assume that seamless data interoperability is ensured.

French law was the vehicle for the dematerialization process defining the need for data archiving as well as authenticity and integrity requirements. In the Act No 2000-230 of 13 March 2000 on legal evidence,\(^1\) it is stated that a digital document has a legal value if its author is identified and if its storage has ensured its integrity. In addition, the Order of 8 December 2005 identified two important requirements for the successful implementation of this legislative framework: information systems security ensuring data protection and data interoperability.\(^2\)

Since 2006 the take-off of e-government has been accompanied by the Directorate-General for Modernisation of the State (DGME), which became in 2012 the Interdepartmental Directorate for Public Service Modernisation (DIMAP). In order to apply the recommendations of the Order of 2005, the DGME developed with different partners two recommendation frameworks: the Recommendation Framework for Interoperability (released in November 2009) and the Recommendation Framework for Security (released in February 2010).\(^3\) Both sets of recommendations which were highly-anticipated play a structuring and unifying role as an implementing decree obliges public services to apply them.

Emphasis is here on the philosophy of these recommendations frameworks, which are to be considered as a set of best practices based as far as possible on standards and open tools. For example, the Interoperability Framework recommends

\(^1\) Act No. 2000-230 of March 13, 2000 adapting the law of evidence to information technology and relating to electronic signature (may be consulted on the Legifrance website: http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT0000003999095).


\(^3\) The reference documentation for French e-administration is available at: http://references.modernisation.gouv.fr [accessed 23.10.2013].
to use standardised data formats (ODT and OOXML, PDF/A), as files formats are crucial for long-term preservation and readability.

One of the lessons that can be learned from this overview is to adopt, as far as possible, standards and tools to overcome the constraints of the IT market: digital data are sensitive and critical, but also inherently very fragile and its long-term readability is not secured. However, you must be able at any time to transfer back documents for legal reasons. In this context, the choice of standards is strategic and suitable as standards are shared and open and as the revision of an existing standard does not deny its previous recommendations.

2 A TOOL FOR DIGITAL ARCHIVING DATA: A NATIONAL DATA EXCHANGE STANDARD

The Archives of France Department of the Ministry of Culture and communication is charged of the national digital archiving policy, in collaboration with central public services charged of the State modernisation. One of the results of this collaboration was the development of a national Data Exchange Standard for Archiving (SEDA), which was first released in 2006. Since then, the SEDA was revised twice. The current version 1.0 was released in September 2012. The Data Exchange Standard is maintained by a steering committee made up of representatives of the Archives of France Department, software vendors, storage service providers and public archival institutions.

The Data Exchange Standard is recommended by the Interoperability Framework: “to set up an archiving process, it is recommended that public archives services and their partners use the Data Exchange Standard.” A public agency that implements archiving and document retention is advised to apply the Data Exchange Standard which is the tool for dialogue with the network of public archives services.

Information on the Data Exchange Standard is available on the website of the Archives of France. This site provides access to specifications of the different versions and the corresponding XML schemas. It also provides some additional information (technical documentation, training material, composition of the steering committee), tools (stylesheets to support the version changes or conversion to other formats, codes examples for implementing XML validation technologies, AGAPE tool for editing XML profiles), and use cases (archiving profiles).


8 The source code of the AGAPE software (Application de Génération Automatisée de Profils Électroniques) can be downloaded at: http://agape.adullact.net/.
2.1 Principles

The Data Exchange Standard is also based on interoperability issue, but in a specific context: data exchanges between public originating agencies and public archival agencies involved in transactions. It uses the methodology defined by UN/CEFACT (United Nation/Centre for Trade Facilitation and Electronic Business). Moreover, it includes the UN/CEFACT components library to describe transactions, actors and exchanged objects using metadata sets.

The Data Exchange Standard also relies on the OAIS model (Open Archive Information System). In addition to the Archival which is an Electronic Records Management System (ERMS) managed by a public archival agency, the followings are defined:

- **Actors** that represent coherent sets of roles. The actor should not be confused with the user (or individual) that can play different roles depending of the used transaction. For instance, an agent that transfers information generated by his agency will serve as a Transferring agency during the transfer, as an Originating agency during the restitution and as an Access Requester during his access request:
  - The **Archival agency** is the actor receiving the transfer and is responsible for the management of the Archives transferred by the Transferring agencies and their delivery back to Access Requestors in respect of the legislation applicable in the field of archive delivery;
  - The **Transferring agency** refers to the actor that transfers an Archive to an Archival agency;
  - The **Originating agency** refers to the entity that produced the archives, i.e. that created them or received them within the context of its activities;
  - The **Control agency** refers to the entity that authorises or not delivery and destruction;
  - The **Access requester** refers to any person or legal entity wishing to consult the Archives kept by the Archival agency in respect of the legislation applicable in the field of archive delivery;

- **Transactions** that occur between these actors around the Electronic Records Management System:
  - the **Archive Transfer** by a Transferring agency to an Archival agency in order to ensure an Archive's preservation. The transfer can be preceded by an Archive Transfer Request;
  - the **Archive Destruction** refers to exchanges linked to the archive destruction, either at the request of the Archival agency with the authorization of the Originating agency and of the competent Control agency, or at the request of the Originating agency, with the authorization of the competent Control agency;

---

• the **Archive Restitution** refers to the Transfer of an Archive from an Archival agency back to the Originating agency;

• the **Archive Modification** refers to information sent by an Archival agency to an Originating agency to inform it that Archives coming from it have been modified by the latter in order to ensure its correct preservation (for instance format conversion, adding, correcting, or updating metadata);

• the **Archive Delivery** refers to the delivery of an Archive by an Archival agency to an Access requester, with the authorization, if necessary, of the Originating agency and of the competent Control agency;

• **Information packages** that are received and managed in these transactions, which are to be understood in the sense given by the OAIS model: Submission Information Package (SIP), Archival Information Package (AIP) and Dissemination Information Package (DIP).

2.2 **Transaction modeling**

The Data Exchange Standard makes a distinction between:

• the exchange processes (transactions) and the exchanged objects (information packages), for which a model is proposed;

• the effective implementation of an Electronic Records Management System (ERMS).

The formalism used to express the model is UML (Unified Modeling Language).

10 Three types of diagrams are used:

The **use case diagrams** provide an overview of the system by representing the actors and their actions on the system. For instance, during the Archival transfer (see figure 1), the Transferring agency transfers to the Archival agency a set of information concerning the transfer itself (identification of the Transferring agency and of the Archival agency, the type of convention (or service contract) agreed between these two parties, level of description...) in addition to the information on the objects to be preserved. If the objects are digital, the objects themselves can be attached to the transfer.

![Figure 1: Transfer use case](image-url)
The **sequence diagrams** show every application case and provide a temporal representation of the progress of each action. These diagrams show the scenarios of actors. They usually linked together four messages: a request, a request acknowledgement, a request reply and a reply acknowledgement. For instance (see figure 2), the transfer request (optional) allows the Transferring agency to check with the Archival agency that the planned transfer is acceptable by sending just the metadata for agreement. It is immediately acknowledged, then, in asynchronous context, by a notification from the archival agency that the transfer proposal is accepted or rejected.

![Figure 2: Transfer exchange sequence](image)

Lastly, the **class diagrams** are used to provide a set of properties to all items handled when scenarios are implemented by actors. For instance, the Message class defines the properties used for each message (date, identifier, signature) and the child classes that specialise the messages (reply, acknowledgement receipt, etc.). In the diagram below (figure 3), the Organization class enables us to describe the actors that interact in the exchanges (transferring agency, archival agency, originating agency, etc.). They must be identified (Identifier). In addition, technical identifiers (xml:id) can be used to provide a reliable mechanism for managing references within the same XML document.

![Figure 3: Organization class](image)
2.3 Formal expression of exchange messages

Furthermore, exchange messages are formalised in XML schemas to describe:

- the type of transfer;
- the content of the exchanged information packages (metadata and data objects);
- the archival hierarchy of information packages.

Information packages include:

- data objects (data files produced by the Originating agency);
- a description of the structure and of the contents of the package (XML file), including metadata produced by the Originating agency or by the Archival agency.

For example, regarding a business application whose data will be transferred in an Electronic Records Management System, the principle is to:

- export data from an Originating agency (files);
- export metadata from an Originating agency and add and integrate archival metadata into an XML document related to data;
- ingest data and metadata in an Electronic Records Management System.

So, both an Originating agency and an Archival agency should implement the SEDA.

2.4 Standardisation references

Another aspect which should be emphasised here is that the Data Exchange Standard relies on standards like governmental Recommendations Frameworks. The development methodology was inspired by five standards and models related to data exchange:

- UN/CEFACT Modelling Methodology (CEFACT/TMG/N090R10, November 2001);
- UN/CEFACT ebXML Core Components Technical Specifications Version 2.01 - ISO 1500-5;
- UN/CEFACT NamingAndDesignRules_1.1a 16022005;
- UML-XML Guide v2.0;
- Common Data Model (CDM) for the French Administration (available at: http://references.modernisation.gouv.fr/volet-semantique).

Other standards were also considered:

- the standard ISO 14721:2003 -- Space data and information transfer systems - Open archival information system - Reference model, better known as the OAIS (Open Archival Information System), was used to define actors, exchanges and exchanged information objects;
- ISO 20652:2006 -- Space data and information transfer systems - Producer-
Archive interface - Methodology abstract standard (PAIMAS): in the SEDA the interactions which take place between an information producer and a deposit archive regarding the first stages of the ingest process were inspired by PAIMAS;

- MIME types served to describe formats of the exchanged data files;
- identifiers of data formats were taken from the PRONOM database;
- language codes were taken from the ISO 639-3 standard serve to indicate the language of the exchanged documents and of the description;
- the International Standard on Archival description (General) ISAD(G) and the DTD (Definition Type of Description) EAD (Encoded Archival description) 2002 served to define elements which are necessary for describing exchanged data;
- the SEDA is positioned before the Model Requirements for the management of electronic records or MoReq2 as data transfer from a Transferring agency to an Archival agency;
- PREMIS (Preservation Metadata: Implementation Strategies) version 2.0 (http://www.loc.gov/standards/premis/) was used to check that no important long-term preservation information to be provided by the transferring agency had been overlooked.

Lastly, code tables were created to express the regulations concerning access and the disposal of records (retention or destruction).

2.5 Use cases of implementation

Since its first release in 2006, SEDA has been used in two types of context:

- for data transfers from Originating agencies to Archival agencies: SEDA was implemented to enable data exports;
- in projects setting up Electronic Records Management Systems: SEDA was implemented to enable data import and access to data.

In the first use case, data producers worked with public archival institutions and the Data Exchange Standard was adapted to specific administrative workflow, including for instance: dematerialization of public procurement, data produced in databases of social welfare services, or departmental applications (database of children going to school in France, applications for the management of dematerialised legal texts). Public administration implemented in its applications features to export data and to describe them according to SEDA rules (platform for public e-procurement, CDC-Fast, S²LOW, and Demat-Aube data transmission systems).

While SEDA was used to allow public services to export their data, it was also used to develop tools that can import data which was exported into SEDA format and run according to the scenario of the transactions it defined:

- project of free open source platform As@lae being developed by the ADULLACT association (French association of Open Source software developers and users in local and central government) regarding territorial governments;
- platforms of territorial government such as M@rine for the territorial
government of Aube;

• proprietary electronic records management systems Mnesys, Maarch and Cimail used by territorial governments;

• storage service providers accredited by the ministry of culture and communication to carry on activities of outsourced storage of public records and that implemented SEDA (Navaho, Atos-WorldLine, Locarchives, Bull companies...).

Thus, we can say that the SEDA plays a key role in public administration.  

3 THE STANDARDISATION PROCESS

3.1 The approach

Since 2013, the Archives of France initiated a process of standardisation of the SEDA within AFNOR (French Association of Standardisation: the French standardisation authority). This process was launched to address the need for stability, generic nature and independence expressed by current users of the SEDA (software editors, storage service providers, archival agencies) as well as by other communities (banking sector, vendors of secure storage solutions) and legal authorities (interdepartmental delegate of French archives for the public archives, interdepartmental directorate for information and communication systems for IT public services).

The standardisation work is being carried out within the Commission “Technical interoperability” of the National Commission “Information and documentation” (AFNOR/CN46-4). The standardisation process started with the creation of an experts group composed of volunteers from the steering committee or from other bodies, in liaison with other AFNOR standardisation committees: the French mirror commission of the international ISO sub-committee on Records management (TC46/SC 11), and the National Commission “Applications for archiving and managing the lifecycle of the document” (CN171). During meetings, members of the CN46-4 drafted specifications of the future standards based on the version 1.0 of the SEDA, and they tried to clarify the terminology and to position the future standard “Exchange Model Data for Archiving” or MEDONA from the OAIS model. Then, a draft of the future standard NF Z 44-022 “Exchange Model Data for Archiving” or MEDONA was submitted to a public review. All the comments received were positive and were analysed by CN46-4. The national standard has just been released.

Normative work done within AFNOR should then lead to a proposal for an international ISO standard. The project was already presented during the meeting of TC46 “Information and documentation” in Paris in June 2013. A resolution was adopted by ISO/TC46/ SC 4: AFNOR should provide the documentation describing the existing national draft standards for archival information interchange and ISO Technical Committee 46 resolved to ask volunteers in subcommittees 4 and 11 to review the AFNOR draft standard for consideration as possible new work item. One of the issues raised during the TC46/SC 4 meeting was the role of MEDONA in comparison of other exchange standards such as PAIMAS (ISO 20652: Space Data and

Information Transfer Systems - Producer-Archive Interface - Methodology Abstract Standard) which only cover the first stages of the ingest process defined by OAIS.

3.2 Differences with SEDA

The MEDONA standard is not a mere repetition of the SEDA specifications but a fully rewritten deliverable introducing some changes in the model as well as a redefinition of the scope: the data exchange for archiving and not just the exchanges about public records and archives. Another major change is that the MEDONA standard no longer refers to archival metadata element sets such as EAD whereas SEDA provides a model which is compliant with the International Standard on Archival Description (General) or ISAD (G). In MEDONA, the model should be specified by users.

Thus, standards and best practices are no longer used for file formats (although it is strongly recommended to use the PRONOM database), data encoding, integrity algorithms, types of relations between files, signature status code specifying the presence or absence of a (certified) signature and reply codes related to types of anomalies. The use of other standards should be negotiated by the actors in the Archival agreement.

In addition, only a minimal set of technical information should be provided for each data object: size, format, integrity, encoding, place. Other information about data objects of the exchanged package should be compliant with a model to be defined before by the actors in the Archival agreement. It is the same for organizations. Whereas only the identification of organizations is specified, other information should be compliant with a model to be defined before by the actors in the Archival agreement. Lastly, the MEDONA standard does not oblige actors to apply the legislation regarding access to public records. So, information on access conditions should also be compliant with a model to be defined before by the actors in the Archival agreement.

3.3 Perspectives

So, after the release of MEDONA, the SEDA should be updated to be compliant with the AFNOR standard. The most part of the work will be to specify how public archival agencies should use MEDONA in the specific administrative context. In particular, it will be necessary to determine what metadata elements sets public services should use for intellectual descriptions of data objects, and for information on actors and access conditions, and how they should take into account the different versions of these metadata sets. So, the protocol should refer to:

- the Encoded Archival Description (EAD) developed by the Society of American archivists with regard to intellectual description (the revised XML schema should be released late 2013);
- the XML schema EAC-CPF (Encoded Archival Context - Corporate bodies, Persons, Families), released by the Society of American Archivists in March 2010; http://eac.staatsbibliothek-berlin.de/) as regards description of actors (Originating agencies, Transferring agencies, Archival agencies and Access requesters);
• PREMIS (Preservation Metadata: Implementation Strategies) version 2.0, to give more precise preservation information, for instance for events related to formats identification, control and conversion;

• METS (Metadata Encoding and Transmission Standard) version 1.9 (http://www.loc.gov/standards/mets/) for the packaging format if there is a need to integrate different schemas.


It will be also necessary to develop mappings between the current types of information included in the Data Exchange Standard and the target standards in order to help archival institutions to switch between SEDA 1.0 and SEDA 2.0. Stylesheets based on these mappings will be proposed.

A format should be defined to express the choices of target descriptive models and standards. This format should be compliant with the standard so that each instance valid with the format would be validated by the schema of the standard. As far as possible, this format will be defined as a XML schema favouring the reuse of eternal components (other chosen XML schemas: EAD, EAC-CPF, etc.). This schema will eventually be completed by a schematron in order to complete the toolkit necessary to validate exchanges.

As far as regards specific business applications or types of data such as public procurement or the administrative review of the legality of acts, profiles have been developed to specify how to use the Data Exchange Standard in these contexts. These profiles enable you to define in advance how a number of properties will be coded (management rules, hierarchical structure of archival description, indexing, etc.). The tool which currently enables to edit these profiles (http://agape.adullact.net/) expresses these properties in the form of constraints in addition to those already defined by the schema of the Data Exchange Standard. Recommendations should be given in order to guide the development of this tool, which should of course reflect the new release 2.0 of the SEDA.

CONCLUSION

In conclusion, we would like to insist on the fact that the Data Exchange Standard for Archiving does not stand in isolation, but fit within a broader archival framework which consists of international standards (e.g. OAIS) or codes of practice. From the very beginning it was decided to apply standards that would bring to our work a longevity which would allow it to be shared with foreign partners. So, SEDA is compatible with ISO 15489, OAIS, and MoReq2.

In addition, it can be used as a common exchange language between records systems. Without such a standard, any transfer of digital records between records systems requires specific custom arrangements between each pair of systems, and such custom arrangements have a number of issues: high risk of loss or compromised records and high cost. So, the use of SEDA means that less custom software is
required to transfer records, and this software can be re-used; its use also means less intervention is required during the transfer of records. In other words, SEDA is part of risk minimization and therefore ultimately results in cost reduction.

Lastly, the targeted audience is very large: those responsible for records within organisations (e.g. records managers, archivists, IT specialists, and corresponding senior managers); those responsible for carrying out a transfer of records between systems (e.g. producers, archivists, and information technology staff); those responsible for designing and implementing software for exporting and importing records from/to records systems (e.g. vendors, software developers); and those responsible for testing systems to ensure that systems accurately implement the transfer process and/or submission information package. And as the producer may be an archival institution or a secondary storage supplier and as the archive may be an agency (for example when records are transferred back from the archival institution or the secondary storage supplier to the agency), many cases are covered: agency to archive, archive to archive, archive back to agency, archive to secondary storage, and so on.

SOURCES AND LITERATURE

POVZETEK

Claire Sibille – de GRIMOÜARD∗

STANDARD ZA IZMENJAVO PODATKOV ZA ARHIVIRANJE: KLJUČNI STANDARD ZA POLITIKO DIGITALNEGA ARHIVIRANJA FRANCOSKIH ARHIVOV

1. E-vlada
Od konca devetdesetih let prejšnjega stoletja se je Francija ukvarjala z razvojem e-vlade, kot tudi njene evropske sosedje. Francoski zakon je bil pogon za proces dematerializacije, ki je definiral potrebo po arhiviranju podatkov kot tudi zahteve po avtentičnosti in celovitosti. Službe, odgovorne za modernizacijo države, so razvile priporočila, ki imajo v e-upravi strukturno vlogo. Med njimi je tudi standard za izmenjavo podatkov za arhiviranje, t. i. SEDA.

2. Orodje za digitalno arhiviranje podatkov: narodni standard za izmenjavo podatkov

3. Proces standardizacije
Leta 2013 je Arhiv Francije začel proces standardizacije SEDA znotraj ANFOR (Francoskega združenja za standardizacijo), ki je standard NF Z 44-022 MEDONA tudi standardizirala. Delo, opravljeno znotraj združenja ANFOR, vodi kasneje do vloge za mednarodni standard ISO. Standard MEDONA ni le ponovitev specifikacij SEDA, ampak je popolnoma prenovljen in prinaša nekatere spremembe v modelu kot tudi ponovno definicijo obsega: izmenjavo podatkov za arhiviranje, ne samo izmenjavo javnih dokumentov in arhivskega gradiva. Druga velika sprememba je, da se MEDONA ne nanaša več na sklope arhivskih metapodatkovnih elementov. Torej, po objavi MEDONE bo SEDA moral biti prenovljen, da bo primerljiv s standardom ANFOR. Največ dela bo usmerjenega v specifikacijo uporabe MEDONE s strani javnih arhivskih ustanov.

∗ Claire Sibille – de Grimoüard, Medoddela služba francoskih arhivov, Ministrstvo za kulturo in komunikacijo, 56 rue des Francs-Bourgeois, 75 141 Paris Cedex 03, Francija, kontakt: claire.sibille@culture.gouv.fr.