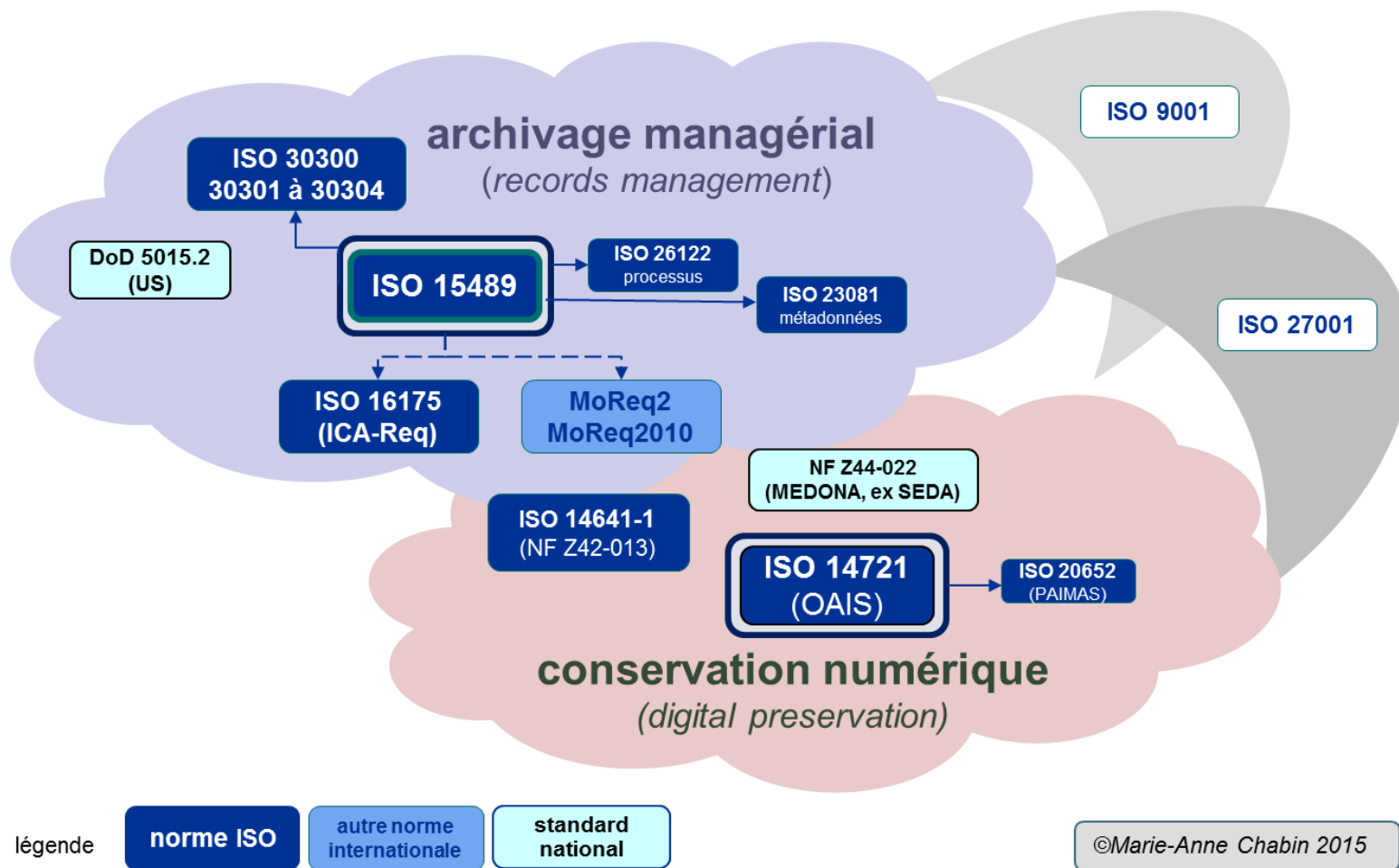
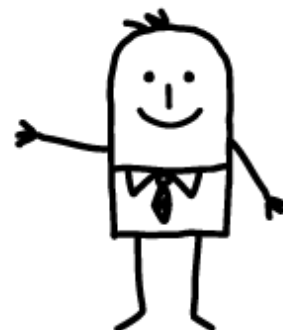


Articulation des normes relatives à l'archivage et à la conservation



Les règles de vie des documents à risque une fois définies, il faut mettre en œuvre leur conservation, sécurisée et pérenne.



LES NORMES POUR LA CONSERVATION NUMÉRIQUE /*DIGITAL PRESERVATION*

Entre sécurité et pérennité

La conservation des documents archivés est plus complexe avec le numérique qu'avec les documents sur supports papier ou film. À côté des conditions physiques de stockage, il faut à la fois protéger les fichiers numériques contre des accès non autorisés ou des modifications indues, et assurer la possibilité de relire les données dans dix ans, dans cent ans...

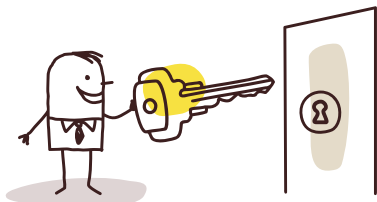
1

La question de la sécurité est omniprésente dans les entreprises.

La sécurité de l'information joue un rôle de plus en plus important dans les projets d'archivage et de conservation numérique.

La norme de référence en matière de sécurité des données est ISO 27001 (2005).

Cette norme aborde plusieurs aspects qui touchent aux exigences de l'archivage.



Entre sécurité et pérennité

2

Au-delà des questions de sécurité de l'information, plusieurs types de normes concourent à une meilleure maîtrise des fichiers numériques dans le temps, avec des approches différentes:

- La norme française **NF Z42-013** met l'accent sur l'intégrité des documents engageants, particulièrement à moyen terme (factures et documents bancaires);
- Le modèle **OAIS (ISO 14721)** organise les données structurées en vue de leur pérennisation sur le long terme ;
- La normalisation porte aussi sur les **formats** d'encodage des données afin que les documents archivés soient produits dans des formats qui anticipent le besoin de conservation;
- Dans tous les cas, on vise à préserver **l'interopérabilité** entre les objets numériques et les outils qui les gèrent.



NF Z42-013: le modèle français

1

La norme Z42-013, promue par l'Association des professionnels de la GED (APROGED) a été diffusée par l'AFNOR dès 1999 et révisée plusieurs fois. Elle décrit « **les mesures techniques et organisationnelles à mettre en œuvre pour l'enregistrement, le stockage et la restitution de documents électroniques afin d'en assurer la conservation et l'intégrité** ».

Nota bene. On parle d'archivage électronique au sujet de cette norme, de même que l'on parle d'archivage électronique pour MoReq et pour ICA-Req. Il y a cependant derrière cette expression unique deux démarches, proches mais distinctes:

1. la conception d'un **système d'information** pour la conservation sécurisée de documents numériques (NF Z42-013);
2. l'organisation des **règles d'archivage** des documents à risque (électroniques) en référence à ISO 15489 (MoReq et pour ICA-Req).

Cible de la norme NF Z42-013: organismes ou entreprises qui souhaitent mettre en œuvre des systèmes informatiques dans lesquels ils pourront conserver des documents numériques de manière à en garantir la fidélité par rapport à l'original ainsi que la lisibilité des données et la traçabilité des opérations sur les documents ; ainsi que les **sociétés de services et éditeurs de logiciels** qui conçoivent des systèmes assurant la fidélité et l'intégrité de documents numériques ; et enfin les **entreprises de tiers archivage**.



NF Z42-013: le modèle français

2

La norme NF Z42-013 énonce **quatre exigences pour l'archivage**, symbolisées par quatre mots dont les initiales forment le sigle PIST :

Pérennité, avec la recommandation d'utiliser des formats ouverts, normalisés ou standardisés, facilitant les nécessaires migrations, **Intégrité**, avec une sécurisation optimale des processus de stockage et de destruction,

Sécurité, avec une duplication des données, un processus d'authentification, le chiffrement et déchiffrement des documents

Traçabilité, touchant l'horodatage, la tenue d'un dossier technique, le journal du cycle de vie des archives et le journal des événements.

La norme insiste sur les aspects de **traçabilité** (au travers de deux journaux systématiques) et met l'accent est mis sur la **sécurisation** du système, notamment avec la pratique de **l'horodatage**.

La norme NF Z 42-013 s'est diffusée en même temps que la profession de tiers-archivage et a permis de normaliser cette nouvelle activité.



NF Z42-013: le modèle français

3

Un des points forts de la norme NF Z42-013 est **la sécurité**.

Les recommandations concernent l'administration et l'organisation de la sécurité, la sécurité physique, la sécurité des matériels et logiciels, ainsi que les aspects de maintenance et d'évolution des systèmes et de migration des supports.

Concernant **les supports de conservation**, la norme distingue :

- **les supports non réinscriptibles**, dit WORM (*Write Once Read Many*) ; les données sont enregistrées sur le disque une bonne fois (inscription irréversible), et on pourra lire le disque X fois ; les disques WORM se subdivisent en deux groupes :
 - le **WORM logique**, techniquement non modifiable ;
 - le **WORM physique** qui est théoriquement modifiable mais cette modification est contrôlée par un dispositif logiciel ;
- **les supports réinscriptibles**, ce qui peut surprendre pour l'archivage ; le principe est que si le support est modifiable et est modifié, l'intégrité est garantie par une empreinte des fichiers, cette empreinte étant conservée avec les fichiers pour attester de la non-modification des données archivées.



NF Z42-013: le modèle français porté à l'ISO

4

La norme française NF Z42-013 a été portée au niveau international en 2012 sous le nom : **ISO 14641-1** - *Electronic archiving - Part 1: Specifications concerning the design and the operation of an information system for electronic information preservation.*



Une certification fondée sur un référentiel basé sur la norme NF Z 42-013 et sur la norme ISO 14641-1 est opérationnelle depuis début 2013. Il s'agit de **la marque NF 461** qui permet de certifier l'application de la norme dans un contexte donné.



Cette certification est accessible aux tiers archiveurs et entreprises qui souhaitent mettre en œuvre un système d'archivage électronique conforme aux exigences de la norme.

Voir AFNOR Certification, organisme de certification indépendant et accrédité par le Comité français d'accréditation (COFRAC).



Le modèle OAIS pour la pérennisation des données



1

Le *Reference Model for an Open Archival Information System* (OAIS) a été publié en l'an 2000 par le CCSDS (Consultative Committee for Space Data Systems) puis sous le titre **ISO 14721** en 2003. Le CNES (Centre national d'Études spatiales) a contribué pour la France à l'élaboration de cette norme.

Le sujet de la norme est « l'archivage à long terme des informations existant sous une forme numérique ».

Dans le sigle OAIS, le **I**, « **Information** », désigne toute connaissance qui peut être échangée, indépendamment de sa forme (physique ou numérique) – par opposition à « **donnée** », définie comme une forme de représentation de l'information.

Nota bene. *La norme utilise le terme **Archive** (au singulier) pour désigner l'organisation (personnes et systèmes) chargée de conserver des informations et de les rendre accessibles à une Communauté d'utilisateurs cible, sans oublier le rôle de veille technologique de ces services.*



Le modèle OAIS pour la pérennisation des données



2

Un des préalables de l'OAIS (qui rejoint complètement la démarche d'archivage managérial) est qu'il faut « **bien identifier ce que l'on veut conserver** ».

Au centre du dispositif se trouvent les « **paquets d'informations** ». Ils sont constitués par des agencements logiques et signifiants des informations à gérer. **Il existe différents types de paquets, appropriés à chaque opération:**


Pour les versements dans l'Archive: le **paquet d'informations à verser** (*Submission Information Package* - **SIP**) livré par le producteur à l'OAIS pour l'élaboration d'un ou plusieurs paquets d'informations archivés (AIP) ;

Pour la conservation: : le **paquet d'informations archivé** (*Archival Information Package* - **AIP**), constitué d'un contenu d'information et de l'information de pérennisation associée ;

Pour la communication) : le **paquet d'informations diffusé** (*Dissemination Information Package* - **DIP**), reçu par l'Utilisateur en réponse à sa requête à l'OAIS et formé à partir d'un ou de plusieurs paquets d'informations archivés (AIP).



Le modèle OAIS pour la pérennisation des données



3

La norme OAIS - sans jamais utiliser le mot « métadonnées » - détaille les catégories d'**information** qui précisent le « contenu d'information » mis en paquet : information d'identification, information de contexte, information de provenance, information de description, information de structure, information d'empaquetage, information d'intégrité, information de pérennisation ou **PDI** (*Preservation Description Information*).

À noter (dans la version 2011 de la norme), la notion d'"**information d'intégrité**" (*fixity Information*) ; il s'agit de la description rigoureuse des modifications inévitables des fichiers numériques en cas de migration technologique.

Les relations entre l'Archive OAIS et les producteurs (en amont) ou les utilisateurs (en aval) sont régies par des **protocoles**. La norme **ISO 20652** (2006) dite PAIMAS pour "Producer-Archive Interface Methodology Abstract Standard" précise ces règles.



Le modèle OAIS pour la pérennisation des données



4

Un des points forts de la norme est la description des processus de migrations des paquets d'informations.

Quatre principaux types de migration numérique sont présentés par ordre croissant de risque de perte d'information :

le **rafraîchissement de support** consistant à remplacer le support initial par un support du même type par copie bit à bit ;

la **duplication** qui ne modifie ni l'information d'emballage, ni le contenu d'information, ni le PDI ;

le **ré-emballage** qui modifie l'information d'emballage ;

la **transformation** qui conserve l'intégralité des contenus d'information mais produit quelques changements dans les trains de bits.

La nouvelle version d'OAIS (2011) intègre les nécessaires ajouts liés à l'évolution des technologies et des problématiques rencontrées.





Les formats de données

Les deux principaux formats de données pour la conservation des données dans le temps sont :



PDF pour les **données non structurées ou semi-structurées**, autrement dit les documents textes tels que notes, factures, rapports, contrats ;

XML pour les **données structurées** : extractions de bases de données ou documents de type formulaire ou qui correspondent à des contenus très structurés (un dictionnaire par exemple).

Le format **PDF/A** a été normalisé en 2005 par l'ISO (ISO 19005) avec des développements en 2010 et 2012 qui permettent d'insérer des images, tandis que **XML** est proposé depuis une quinzaine d'années par le consortium W3C et permet d'encapsuler des documents textes.

Il faut également mentionner le format texte **TXT**, très simple, limité aux mots sans aucune mise en forme. L'avantage est qu'il est accessible par tous les postes informatiques. Par exemple, pour ce MOOC, les transcriptions des vidéos sont proposées en PDF avec la charte graphique du MOOC, et en TXT où le texte est brut.

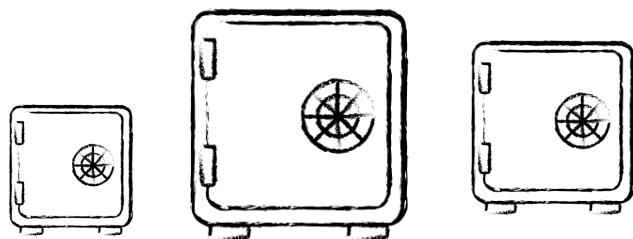
Pour les **images fixes**, le principal format normalisé est **JPEG**. Pour les **images animées**, la référence est **MPEG4**. Pour les images issues d'une opération de **numérisation**, le format **TIFF** (image bitmap) fait l'objet également d'une norme ISO.

À mentionner encore des **formats non standardisés** mais qui dont l'usage est très répandu, comme le **DWG** pour les plans.

Les Référentiels de l'Administration française

L'administration française a diffusé en 2005 deux référentiels en matière de sécurité et d'interopérabilité des systèmes d'information.

- 1 Le référentiel général de sécurité (RGS)** est un recueil de règles et de bonnes pratiques en matière de sécurité de systèmes d'information à destination des services administratifs qui proposent des services en ligne aux usagers.
L'idée est de promouvoir une approche globale de la protection des données et une analyse systématique du risque.



- 2 Le référentiel général d'interopérabilité (RGI)** énonce diverses recommandations pour favoriser l'interopérabilité au sein des systèmes d'information lors d'échanges électroniques entre les usagers et les autorités administratives, et entre les autorités administratives.
Une nouvelle version de ce référentiel est annoncée pour 2016.

