



HAL
open science

La qualité des données et métadonnées dans l'implantation d'un système d'archivage électronique

Gora Dieng

► **To cite this version:**

Gora Dieng. La qualité des données et métadonnées dans l'implantation d'un système d'archivage électronique. domain_shs.info.docu. 2020. mem_03172100

HAL Id: mem_03172100

https://memsic.ccsd.cnrs.fr/mem_03172100

Submitted on 17 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

le **cnam**

intd



**Mémoire pour l'obtention du master sciences humaines
et sociales
Mention humanité numérique
Parcours Mégadonnées et analyse sociale
MEDAS**

**La qualité des données et métadonnées dans
l'implantation d'un système d'archivage
électronique**

Soutenu le 09 septembre 2020

Entreprise d'accueil : BRGM (Bureau de Recherches Géologiques et Minières)
Année de formation : 2019/2020
Tuteur Pédagogique : Evelyne BROUDOUX
Maître d'apprentissage : Samuel Le Bideau

Remerciements

En préambule à l'élaboration de ce mémoire, je me permets d'exprimer mes vives et sincères marques de gratitude à l'égard de certaines personnalités dont les différents apports me furent d'un grand atout. D'abord, Je rends grâce à Dieu qui m'a donné l'opportunité de mener cette année d'apprentissage à bout. Je ne saurais enclencher le processus de rédaction du présent rapport sans pour autant exprimer toute ma reconnaissance à l'égard de M Samuel Le Bideau responsable de l'unité gestion des document d'activité (GDA) pour son accompagnement pédagogique, ses orientations, conseils et sa disponibilité durant toute l'année ainsi que toute l'équipe GDA. J'exprimer toute ma reconnaissance à l'égard de Mme Evelyne Broudoux mon tuteur au CNAM, pour l'attention qu'elle porte à mon égard son soutien et ses précieux conseils.

Un immense merci à Mme Beatrice Arruabarrena pour son engagement et ses travaux pour le master MÉDAS. Enfin je remercie tous les intervenants du master MÉDAS pour la qualité de leurs enseignements durant toute cette année universitaire.

Résumé

L'archivage électronique, considéré depuis son invention par les entreprises comme un besoin opérationnel de moindre nécessité est devenu une obligation avec la loi du 13 mars 2000. Au cours des 30 dernières années, le développement rapide de l'informatique et des technologies de l'information a entraîné une révolution dans la création de documents et l'échange d'informations. Autrefois monolithique, les données sont devenues dynamique, augmentant ainsi les possibilités commerciales.

A l'heure actuelle, les organismes publics, les collectivités et les entreprises, ont une obligation de dématérialiser la majeure partie de leurs flux documentaire on peut citer entre autres, les factures pour le secteur public, les réponses aux appels d'offres.

Dès que les données concernées dans le cadre de la dématérialisation présentent une valeur probante, utile pour la gestion à long terme ou ont un intérêt patrimonial, la qualité de ces données doit être optimale pendant son archivage. Seul un archivage électronique peut permettre au juge d'évaluer la valeur juridique des données soumis et la conservation doit répondre aux exigences de la loi ou de la jurisprudence. C'est pour répondre à ces exigences que le BRGM dans son programme de donnée à décider de se doter d'un système d'archivage électronique pour non seulement assurer la qualité de ses données mais aussi fluidifié ses processus métiers et assuré la productivité de ses différents services.

Le rôle central que jouent les données et les métadonnées dans l'exécution des processus de l'organisation est de plus en plus reconnu. Les normes NFZ 42-013 et ISO 15489 place les données au cœur des processus métier et met en œuvre un système de gestion des documents basé sur celui-ci.

Summary (english)

Electronic archiving considered since its invention by companies as an operational need of lesser necessity has become an obligation with the law of March 13, 2000. Over the last 30 years, the fast development of computers and IT technologies. information has brought a revolution in the creation of documents and information exchange. Once monolithic, data has become dynamic, increasing business opportunities.

Nowadays, public establishment, local authority and company have an obligation to dematerialize the major part of their documentary flows, one can cite among others, invoices for the public sector and responses to calls for tenders.

As soon as the data concerned in the context of dematerialization present a probative value, useful for long-term management or have a heritage value, the quality of this data must be optimal during its archiving. Only electronic archiving can allow the judge to assess the legal value of the data submitted and the retention must meet the requirements of law. It is to meet these requirements that the BRGM in its data program has decided to equip itself with an electronic archiving system not only to ensure the quality of its data but also streamline its business processes and ensure the productivity of its different services.

The central role that data and metadata play in the execution of organizational processes is increasingly recognized. NFZ 42-013 and ISO 15489 standards place data at the heart of business processes and implements a document management system based on it.

Sommaire

Remerciements	2
Résumé	4
Summary (english)	5
Liste des abréviations	9
Glossaire	11
Introduction	13
Présentation du BRGM	16
Gestion des documents d'activité (GDA)	18
Première partie : Cadrage du projet SAE au BRGM	20
I. Démarche particulière appliquée au	20
6. Analyse des risques du projet	22
d. Liste des risques	22
d. Matrice des risques	22
d. Gestion des risques et recommandations	23
Deuxième partie : Le système d'archivage électronique définition et mise en place	25
II. État de l'art sur la gestion des archives	25
III. Le système d'archivages électronique et sa mise en place	31
1. Qu'est-ce qu'un SAE	31
2. Texte de référence	31
3. Les exigences fonctionnelles et techniques du SAE	33
4. Modèle de données utilisée	33
5. RÔLES, ACTEURS ET INTERVENANTS DU SAE	46
6. Distinctions SAE, GED CFN	50
III. Mise en place du SAE	51
1. Phases du projet	51

2. Pourquoi la solution Vitam (Valeurs Immatérielles Transférées aux Archives pour Mémoire)	54
3. Livrables attendus	56
4. La politique d'archivage.....	65
Troisième partie : La qualité des données et des métadonnées dans le SAE du BRGM	66
IV. La qualité des données et des métadonnées.....	66
1. Donnée Information Connaissance	66
2. Les données au BRGM	68
3. Les enjeux et le défi scientifique de la DATA pour le BRGM	69
4. Qualité des données	70
a) La Qualité des processus de traitement des données	71
b) Qualité de la valeur des données l'exemple de la BSS	74
5. La qualité des métadonnées.....	77
a) Qu'est-ce que les métadonnées	77
b) Typologie des métadonnées.....	78
c) Quels usages pour quelles métadonnées pour le SAE.....	80
Les bénéfices de la métadonnée	81
Conclusion	87
TABLE DES FIGURES	90
TABLE DES TABLEAUX.....	90
Annexes.....	91

Liste des abréviations

- **AAF** : Association des Archivistes Français
- **ADELE** : ADministration ÉLEctronique
- **AFNOR** : Association Française de Normalisation
- **AIP** : Archival Information Package (Paquet d'Information Archivé)
- **API** : Application Protocol Interface
- **art.** : Article (*législatif*)
- **BnF** : Bibliothèque Nationale de France
- **CFN** : Coffre-Fort Numérique
- **CNAM** : Conservatoire National des Arts et Métiers
- **DIP** : Dissemination Information Package (Paquet d'Information à Diffuser)
- **DTD** : Document Type Definition (Définition de Type de Document)
- **DUA** : Durée d'Utilité Administrative
- **EAC-CPF** : Encoded Archival Context – Corporate Bodies, Persons and Families (Contexte Archivistique Encodé – Collectivités, Personnes et Familles)
- **EAD** : Encoded Archival Description (Description Archivistique Encodée)
- **IHM** : Interface Homme Machine
- **ISAAR (CPF)** : International Standard Archival Authority Record for Corporate Bodies, Persons and Families (Norme Internationale sur les Notices d'Autorité utilisées pour les Archives relatives aux Collectivités, aux Personnes et aux Familles)
- **ISAD (G)** : International Standard Archival Description-General (Norme Générale et Internationale de Description des Archives)
- **ISO** : International Organization for Standardization (Organisation Internationale de Normalisation)
- **METS** : Metadata Encoding and Transmission Standard
- **Moreq** : Model Requirements for the Management of Electronic Records (Exigences Types pour la Maîtrise de l'Archivage Électronique)

- **OAIS** : Open Archival Information System (Système Ouvert d'Archivage d'Information)
- **PDF** : Portable Document Format
- **PRA** : Plan de Reprise d'Activité
- **PREMIS** : PREservation Metadata: Implementation Strategies
- **RM** : Record Management
- **SAE** : Système d'Archivage Électronique (*voir ERMS en anglais*)
- **SEDA** : Standard d'Échange de Données pour l'Archivage
- **SGDA** : Système de Gestion des Documents d'Activité
- **SIAF** : Service Interministériel des Archives de France
- **SIP** : Submission Information Package (Paquet d'Information à Verser)
- **SQL** : Structured Query Language
- **UML** : Unified Modeling Language
- **VITAM** : Valeurs Immatérielles Transférées aux Archives pour Mémoire
- **WORM** : Write Once Read Many
- **XML** : eXtensible Markup Language
- **XSD** : Schéma XML

Glossaire

Collecte : Une des principales missions d'un service d'archive, qui consiste à rechercher et recueillir auprès des producteurs les différents documents à archiver.

Communicabilité : Possibilité de communiquer au public un document archivé en fonction de critères juridiques et réglementaires.

Communication : Mission d'un service d'archive consistant en la mise à disposition de documents archivés communicables au public et aux administrations.

Conservation définitive : Opération qui consiste à conserver les documents qui ne sont pas destinés à être éliminés.

Date de communicabilité : Date à partir de laquelle un document devient communicable, c'est-à-dire accessible au public.

Document : Tout écrit ou enregistrement considéré comme une unité (*extrait norme ISO 15489*).

Dossier : Ensemble de documents constitués de manière organique ou logique.

Élimination : Procédure réglementée qui vise à supprimer ou retirer du système d'archivage un ou plusieurs documents archivés.

Évaluation : Fonction de l'archiviste qui consiste à déterminer l'utilité administrative, l'intérêt historique et le sort final des documents.

Instrument de recherche : Outil papier ou informatique permettant la recherche d'archives.

Jeu de données (ou Dataset) : Liste de données composée d'attributs et de leurs valeurs. Dans un fichier au format CSV par exemple, les colonnes représentent les attributs et les lignes les valeurs. Les données publiques sont très souvent mises à disposition sous forme de jeux de données.

Métadonnées : Données décrivant le contexte, le contenu, la structure des documents ainsi que leur gestion dans le temps.

Pièce : Plus petite unité intellectuelle d'archives, par exemple une lettre, un mémoire, une photo...

Plan de classement : Système fixant l'organisation des archives courantes, intermédiaires et définitives.

Record Management : Gestion des documents d'archives afin d'organiser efficacement et systématiquement la création, la réception, l'utilisation et le sort des documents d'archives. On dit également archivage managérial.

Sort Final : Désigne le traitement final des documents.

Tri : Opération qui consiste à séparer les documents dont le sort final sera la conservation de ceux qui devront être éliminés.

Versement : Opération qui consiste à transmettre des documents à archiver à un service (ou système) d'archivage.

Valeur probante : Qualité d'un document lui permettant de faire valeur de preuve.

Objet Train de bits numériques qui peut être stocké dans un Coffre-fort électronique

Introduction

Les données sont une valeur précieuse et indispensable dans notre société actuelle, elles favorisent le développement des connaissances. L'évolution des textes réglementaires en matière de conservation des données a eu un impact considérable sur les activités de nos entreprises. Si l'entreprise a toujours utilisé le papier pour remplir ses obligations légales de conservation des données liés à ses activités, la loi du 13 mars 2000 [1] a changé la donne en reconnaissant l'écrit sous format électronique comme preuve de l'existence d'un droit ou d'une obligation, au même titre que le papier. Du public au privé en passant par les EPIC, La plupart des communications et des prises de décision deviennent de plus en plus dématérialisées. En effet les factures, les bons de commande, les rapports, les bulletins de salaire et autres données sont automatiquement générés par des logiciels métiers. Cette tendance forte implique un changement dans les manières de penser la conservation à plus ou moins long terme des données, à des fins de sécurisations réglementaires, opérationnelles et patrimoniales.

Contexte

La pratique des acquis théoriques obtenus durant notre formation professionnelle au CNAM est l'une des recommandations auxquelles les responsables du master MÉDAS prête une attention particulière. C'est en ce sens qu'ils proposent en apprentissage à mi-temps la formation avec différentes structures de la place. J'ai choisi d'effectuer mon alternance au bureau de recherche géologique et minières. J'avais comme tâche principale au BRGM le cadrage et l'étude de faisabilité du projet de la mise en place d'un système d'archivage électronique (SAE).

Un SAE est un système informatique qui assure la conservation pérenne et sécurisée de tous les documents numériques qu'il prend en charge garantissant ainsi leurs intégrités leurs traçabilités leurs pérennité et leurs accessibilités. Il permet de conserver ses données tout en maîtrisant ses coûts. Dans un monde de plus en plus interconnecté, la collecte, le traitement, la gestion et l'exploitation des données sont devenu un enjeu majeur pour les administrations. Collectées et exploitées dans divers domaines et pour des finalités variées, les données sont devenues, à l'instar du pétrole et du gaz, la matière première des entreprises dans un marché plus que jamais compétitif. C'est dans ce contexte que le BRGM dans son programme intitulé « collecter, interconnecter et valoriser les données et la connaissance environnementale au travers de services numériques » s'est lancé dans la certification d'un de ses pôles de données.

Notre objectif est l'obtention de la certification "Core TrustSeal" CTS (voir annexe) afin de faire évoluer la culture "data" au BRGM.

Le CoretrustSeal est une certification mise en place conjointement par le DSA (Data Seal of Approval) et le ICSU WDS (the International Council for Science's World Data System) qui vise à décrire les caractéristiques des entrepôts de données utilisables en toute confiance. Cette certification garantit aux déposants que leurs données seront protégées et gérées de manière optimale. Pour l'obtention de cette certification tant convoitée il faut répondre à 16 exigences. Parmi elles, la réponse aux critères d'authenticité, d'intégrité et d'archivages des données.

En France le domaine des archives et de l'archivage est régi par le livre II du code du patrimoine. Il définit les archives comme « Tous les documents produits ou reçus par toute personne physique ou morale et tout service ou organisme public ou privé dans l'exercice de ses activités, quels que soient leur date, leur format et leurs matériaux. » (Art L. 211-1) et les archives publiques comme « Les documents qui procèdent de l'activité, dans le cadre de leur mission de service public, de l'État, des collectivités territoriales, des établissements publics et des autres personnes morales de droit public ou des personnes de droit privé chargées d'une telle mission [...] Les minutes et répertoires des officiers publics ou ministériels » (art L. 211-4).

[2]

« L'intégrité garantit que les modifications apportées aux données et aux métadonnées sont documentées et tracées (raison d'être, auteur de la modification).

L'authenticité couvre le degré de fiabilité des données originales déposées et de leur provenance, y compris la relation entre données originales et données diffusées, et si les relations existant entre les jeux de données et/ou les métadonnées sont effectivement maintenues. »[3]

C'est pour cette mission que l'équipe SG/GDA dont je fais partie est sollicitée pour apporter des réponses aux trois fonctionnalités précitées à savoir la garantie de l'authenticité, l'intégrité et l'archivage des données pris en charge par l'entrepôt de données à certifier. C'est pour répondre aux exigences de cette certification qu'il est nécessaire de mettre en place un système d'archivage électronique (SAE) au BRGM. Ce système d'archivage électronique viendra répondre au-delà de la certification CTS à des besoins fonctionnels du BRGM par exemple s'agissant des marchés publics de l'unité Achat, depuis le 1^{er} octobre 2018 la réglementation impose aux entreprises de répondre sous format électronique via le profil acheteur¹ aux marchés publics d'un montant supérieur à 40 000 € HT. Le BRGM ne disposant

¹ Le profil d'acheteur est une plateforme de dématérialisation permettant de mettre les marchés publics à disposition des opérateurs économiques par voie électronique

pas actuellement de certificat électronique permettant de signer les documents dématérialisés. Il est par conséquent demandé aux entreprises, et surtout à celle qui gagnera le marché, d'une part de répondre électroniquement au marché, la réponse contenant donc un certificat électronique, mais également de matérialiser leurs réponses et d'y apposer une signature manuscrite ou de passer par des prestataires externes pour assurer la signature électronique afin de finaliser le marché. Toutefois se pose la question de la conservation du document signé électroniquement. Outre la commande publique, le flux comptable fait face aux mêmes problèmes notamment le service facturier tend vers une dématérialisation des processus. Au regard de cette tendance il convient de repenser la conservation, la préservation et l'accès à ces données.

L'archivage électronique est une offre de service de conservation à long terme de données à valeur scientifique, légal et patrimonial. La mission d'un SAE est de permettre la conservation des données sur le long terme et de les protéger de toute altération, afin de garantir leur valeur probatoire si besoin en cas de contentieux. Il prend en charge des archives, c'est-à-dire des documents dans leur version finalisée qui doivent être conservés selon les durées de conservation légales.

Pour concrétiser cette offre de service, nous avons porté notre choix technologique sur le logiciel Vitam (Valeurs immatérielles transmises aux archives pour mémoire) développé sous l'égide du service interministériel des archives capables de gérer, préserver et donner un accès à long terme aux archives numériques qui lui sont versées. Nous avons fait le choix de Vitam parce qu'étant développé par les services de l'État pour accompagner les établissements tels que le BRGM tout en étant Open source.

Problématique

La problématique consiste dans un premier temps à faire l'état des lieux sur la mise en place d'un système d'archivage électronique avec Vitam au BRGM, ensuite travailler sur les spécificités de son déploiement, déterminer ce qu'est la qualité des données ? Et qu'est-ce qui confère aux données leurs qualités ? Quel modèle de données et métadonnées on va utiliser pour le SAE ? Enfin comment va-t-on structurer notre modèle de donnée pour qu'elle soit conforme au standard d'échange choisi ?

Le but de cette analyse est de mener les dernières recherches sur les normes et solutions techniques existantes dans le domaine de l'archivage électronique, de définir ces solutions et de déterminer les bons critères de sélection. Ceci nous permettra dans la seconde partie du projet de concevoir, produire et mettre en œuvre un POC de la solution Vitam retenue pour

notre projet, et de créer un connecteur entre le système Teams, Opale et les systèmes informatique métier de sauvegarde et de création de données telle que le SIRH et le SIHA.

Présentation du BRGM

Un peu d'histoire

C'est en 1941, à l'initiative d'Edmond Friedel, ingénieur général des mines, et de Pierre Pruvost, doyen de l'Académie des sciences de Lille, que l'ancêtre du BRGM actuel voit le jour. L'histoire du BRGM mêle deux univers. Celui de la géologie et la géophysique d'abord : services extérieurs du département des mines du ministère de l'Industrie qui est financé que par des fonds publics. Donnant naissance au Bureau de recherches géologiques et géophysiques BRGG. Ensuite celui des mines viendra se rattacher à lui pour devenir Bureau de recherches géologiques, géophysiques et minières (B.R.G.G.M.) de la France métropolitaine. Parallèlement en 1948 et 1949, trois organismes importants sont créés pour opérer hors de la métropole : le BUMIFOM (Bureau Minier de la France d'Outre-Mer), le BMG (Bureau Minier Guyanais) et le BRMA (Bureau de Recherches Minières de l'Algérie)

BRGM actuel

Le BRGM est l'établissement public de référence dans les applications des sciences de la Terre pour gérer les ressources et les risques du sol et du sous-sol.

Le BRGM est créé en 1959 par le décret n° 59-1205 du 23 octobre 1959.⁴ Il est le fruit de la fusion BRGGM, du BUMIFOM, du BMG et du BRMA. Il devient ainsi le service géologique national français, c'est un établissement public à caractère industriel et commercial. Il est sous la tutelle de trois ministères à savoir le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, le ministère de l'Écologie, du Développement durable et de l'Énergie et le ministère de l'Économie, de l'Industrie et du Numérique

Les objectifs

Les objectifs principaux du BRGM sont :

- La mise à disposition d'outils nécessaires à la gestion du sol, du sous-sol et des ressources, à la prévention des risques naturels et des pollutions, aux politiques publiques d'aménagement du territoire.
- La compréhension des phénomènes géologiques.

- Le développement de méthodologies et de techniques nouvelles au service de la recherche scientifique.
- La diffusion de ses connaissances et de son savoir-faire.

Les missions

Les actions du BRGM se situent actuellement dans un contexte socio-économique complexe. Il s'agit en effet d'intégrer conjointement les notions de développement durable, de maîtrise de la globalisation, d'éthique scientifique et également le développement des nouvelles technologies de l'information-communication. Ces actions se déclinent selon quatre missions principales :

- Appui aux politiques publiques : L'activité de Service Public regroupe l'ensemble des actions d'expertise, de surveillance et d'étude menées en soutien des politiques publiques. Ces tâches sont financées soit par une dotation budgétaire soit par des contrats avec des services de l'État, des collectivités, des établissements publics ou des entreprises et acteurs du monde économique.
- Recherche scientifique : La recherche du BRGM se déploie dans trois domaines : la connaissance géologique et la compréhension par l'observation et la modélisation des processus liés au sol et au sous-sol, les développements au service des politiques publiques, les transferts de technologies et d'innovations vers l'industrie.
- Coopération internationale et aide au développement : À l'étranger comme en France, le BRGM apporte son savoir-faire et son expertise principalement dans les domaines d'infrastructure et de cartographie géologiques, des risques naturels, des ressources minérales et des eaux souterraines. Cette mission se traduit notamment par des prestations scientifiques et techniques internationales, la participation aux travaux géologiques internationaux, ainsi que l'appui aux politiques de l'Union Européenne.
- Prévention et surveillance des anciens sites miniers : Outre la surveillance scientifique et technique des anciens sites miniers, cette mission se consacre à la gestion du système d'information après-mine, dont la gestion des archives techniques intermédiaires minières.

Présentation du Secrétariat général

Le Secrétariat général assume toutes les responsabilités liées à la gestion administrative, juridique du BRGM, la gestion des moyens immobiliers et logistique (photocopie et déplacements) mais aussi les ressources archivistiques, la sécurité et à la conformité (loi PRADA, RGPD, Sapin 2), l'éthique et la déontologie et la coordination des actions en matière de responsabilité sociétale des entreprises (RSE), y compris l'ouverture à la société civile.

Le Secrétariat général interagit avec l'ensemble des directions opérationnelles et supports auxquelles il propose son offre de services. Il est en relation permanente avec les autorités de tutelles, les auditeurs externes, les prestataires de services, les entreprises de travaux et les clients.

Les activités et les missions du Secrétariat général sont organisés autour de cinq unités qui sont :

« Patrimoine immobilier » (SG/PI), « Environnement de travail » (SG/ET), « Gestion des documents d'activité » (SG/GDA), « Juridique » (SG/JUR) et « Achats » (SG/HA), chacune étant pilotée par un responsable d'unité.

Gestion des documents d'activité (GDA)

L'unité Gestion des Documents d'Activité (GDA) est issue de la refonte de l'ancienne unité "Mission centrale des archives" (MCA) lui-même issue de l'ancienne unité Service Central des Archives (SCA) qui fut le premier service de gestion documentaire créé au BRGM.

Bien que sa mission première soit la collecte, le traitement, la conservation et la communication des documents de l'ensemble de l'établissement. L'impact du numérique au BRGM a créé d'autres missions fondamentale pour le SG/GDA. En effet l'unité GDA est aussi chargé de :

- Apporter son expertise à la définition de la politique d'archivage applicable
- Concevoir et mettre à jour un cadre stratégique de gestion des documents et procédures associées (papier et électronique), de la naissance à la destruction ou au versement réglementaire des documents aux Archives nationales.
- Piloter le processus d'émission, de la validation, de la diffusion et de l'archivage des notes de services pour une perspective de dématérialisation.
- Mettre en œuvre le workflow de validation des notes de service l'archivage des textes originaux
- Accompagner les directions et unités du BRGM dans la mise en œuvre de cette politique par une offre de services permettant d'organiser la gestion des documents papier ou électroniques, de les sécuriser, et de contribuer à la transition digitale de l'Établissement ;
- Enfin, Co-conduire avec la direction en charge des systèmes des infrastructures et des services numériques, le projet de mise en œuvre d'une démarche et des outils pour un

archivage électronique pérenne conforme à l'état de l'art. C'est dans le cadre de cette mission que j'ai intégré l'unité SG/GDA.

Première partie : Cadrage du projet SAE au BRGM

I. Démarche particulière appliquée au

Le cycle de vie d'un projet désigne la décomposition du projet en un ensemble d'étapes nécessaires à la réalisation d'un produit ou service. Ce vaste projet initié depuis 2018 par Samuel Le Bideau, responsable de l'unité SG/GDA doit être Co-conduit avec la direction des infrastructures et services du numérique (DISN). Compte tenu de la dématérialisation croissant des processus et des obligations légales liée à l'archivage le BRGM doit dès maintenant mettre en œuvre les technologies pour la mise en place d'une solution d'archivage électronique. Mon rôle est la conduite de ce projet en phase d'initiation avec l'équipe projet composé de Samuel Le Bideau et la DISN.

1. Planning prévisionnel du projet

1		Archivage électronique - implantation SAE			Affichage de la semaine de lancement			
2		Pilote		SG/GDA	Semaine numéro : 1			
3		Date de début		02/01/2020				
4		Date d'aujourd'hui		22/07/2020				
5		Date de fin prévue		11/03/2021				
6								
7								
8								
3	N°	Tâches	Responsable / Interlocuteur	Commentaires	Début au plus tôt	Charge estimative	Fin au plus tard	charge réalisée
10	1	CERTIFICATION CORE TRUST SEAL PROPOSITION DE DEMARCHE ET APPLICATION			30/01/2019	29	30/09/2020	4,5
16	2	Phase de pré-implantation mis en place du système d'archivage électronique SAE			01/02/2020	110	01/10/2021	35
17	2.1	Délimiter le périmètre du SAE	GDA / DISN	SI BRGM	31/01/2020	6	31/08/20	1
18	2.2	Etude de faisabilité du SAE	GDA / DISN	Etat des lieux / audit préalable	30/01/2020	7	01/09/20	2
19	2.3	Rédaction du cahier de charge	LE BIDEAU S.	Voir EB_SAE sur teams	01/02/2020	4	08/06/2020	2
20	2.4	Choix du logiciel et de l'infrastructure SAE	LE BIDEAU S.	Vitam	01/02/2020	4	08/06/2020	1
21	2.5	Travail de repérage, sélection des documents	LE BIDEAU S.	Format des données typologie document	01/03/2020	15	30/04/21	1
22	2.6	Planification des perennisation	GDA / DISN / équipe logiciel	Assignment des métadonnées de sécurité	16/08/2020	15	01/04/21	0
23	3	Implantation concrète du SAE			01/01/2020	118	01/10/2021	38
24	3.1	Réflexion et traitement de l'existant sur le système d'information	LE BIDEAU S.	A partir des travaux effectués en amont sur place	15/03/2020	26	30/08/2020	3
25	3.2	Représenter le processus/workflow du SI BRGM	LE BIDEAU S.	Voir ce qui existe déjà	01/02/2020	4	08/06/2020	1
26	3.3	Installation d'un système pilote	GDA / DISN / équipe logiciel	Mettre en place un poc en lien avec le travail effectué avec l'unité H.A	01/03/2020	15	30/03/2021	0
27	3.4	Déploiement et gestion du système d'archivage électronique	GDA / DISN		01/10/2020	60	01/10/2021	15
28	3.5	Apporter une réponse aux exigences CTS R7 - R15	BRESSAN H. / LE BIDEAU S.	comment assuré la perennisation et l'intégrité des données	02/02/2020	5	03/09/2020	7
29	3.6	Réunions réflexions collectives équipe logiciel GDA DISN	BRESSAN H. / LE BIDEAU S.		01/01/2020	8	10/09/2021	3
30	3.7	Typologie de versement ou de capture	LE BIDEAU S.	Identification des sources de versement (applications informatiques, répertoires Windows)	01/06/2021	15	01/09/2021	3

Tableau 1 : Diagramme de GANTT - Planning prévisionnel du projet de mémoire

6. Analyse des risques du projet

La transversalité de l'archivage électronique augmente le risque de retard et d'échec du projet. Le but de l'archivage électronique au BRGM est de centraliser tous les archives de l'établissement dans un seul et même système à terme. Ceci implique la participation tous services producteurs d'archives. A cela s'ajoute le fait que le projet est piloté par deux directions différentes et l'équipe qui offre la solution Vitam.

Face aux défis de cette transversalité, il était nécessaire de mener une analyse des risques afin de maîtriser les risques pouvant entraîner la non-réalisation du projet.

d. Liste des risques

RSK1 : Retard de lancement du projet.

RSK1 : Risque financier (budget non disponible)

RSK2 : Non-disponibilité ou Non-garantie de la pérennité de l'équipe projet

RSK3 : Risque de pandémie : Non-disponibilité des acteurs et intervenant.

RSK5 : Espace disque non disponible pour mise en œuvre des POC

RSK6 : Évolution des solutions de SAE.

RSK7 : Évolution de la législation en matière d'archives publiques.

d. Matrice des risques

Cette matrice des risques représente la probabilité d'apparition d'un risque classer dans une échelle de gravité de 1 à 4. 1 correspond à la probabilité d'occurrence la plus faible, et 4 correspond à la probabilité d'occurrence élevée. Le risque est également évalué en fonction de l'impact possible de la survenance du risque. Du plus petit impact au plus grand impact, la plage de cet impact est également de 1 à 4.

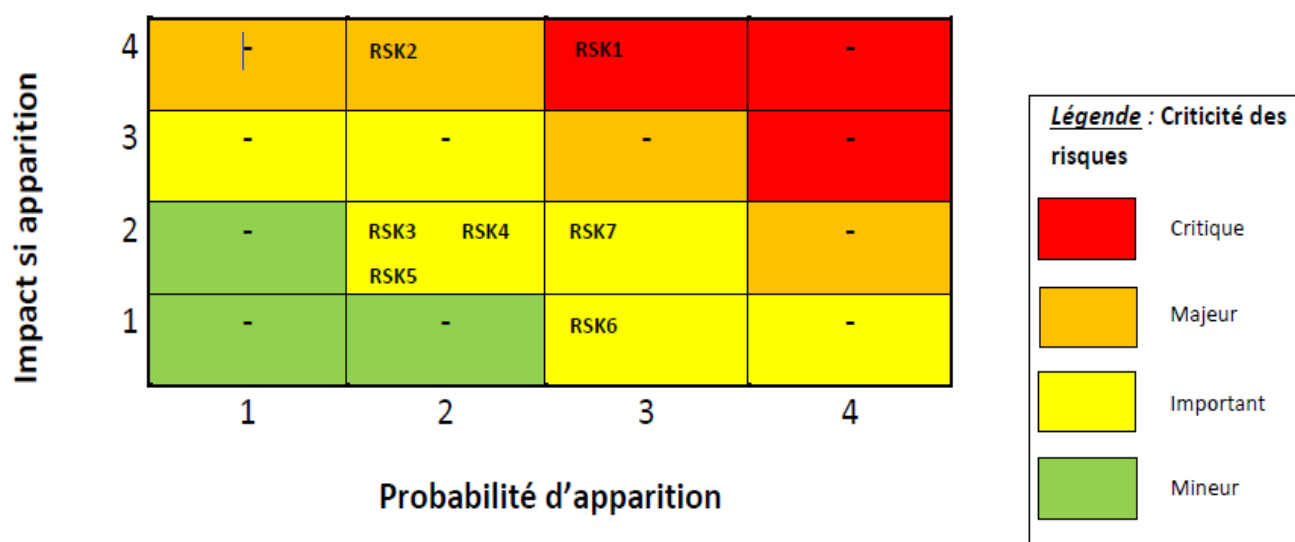


Tableau 2 : Matrice des risques d'échecs du projet

d. Gestion des risques et recommandations

RSK1 : Retard de lancement du projet.

Ce risque peut être causé par une mauvaise entente entre les intervenants dans le projet : risque d'absence de motivation et d'effort commun pour réussir le projet et/ ou d'une mauvaise ou d'une absence de communication entre les intervenants.

RSK1 : Risque financier (budget non disponible)

Ce risque majeur aurait un impact critique sur la tenue du projet d'archivage. Pour réduire à et maîtriser ce risque il faudrait procéder au découpage du coût du projet en phases de développement, définir en amont les risques et coût acceptables ou objectifs globaux en fin de projet et les soumettre aux tutelles avant le lancement du projet.

RSK2 : Non-disponibilité ou Non-garantie de la pérennité de l'équipe projet

L'indisponibilité par les techniciens systèmes ou les autres membres peuvent entraîner des retards dans l'avancement et la livraison du projet. Ceci risque de modifier l'organisation du projet. Ce risque peut être réduit en déplaçant autant que possible la demande de début des travaux mais aussi insister sur la ponctualité des livraisons des différents travaux demandés aux acteurs.

RSK3 : **Risque de pandémie : Non-disponibilité des acteurs des services informatiques et des archives.**

Ce risque peut-être minimiser en prévoyant des alternatifs calendaire lié aux méthodes de travail en privilégiant notamment le travail et les réunions à distance. Surtout dans la phase d'audit documentaire.

RSK6 : Évolution des solutions de SAE.

L'archivage électronique est un domaine en forte évolution. Les infrastructures et technologies permettant sa mise en œuvre sont en constante évolution. Il faut éviter de proposer des architectures ou de s'appuyer sur des technologies obsolètes ou dépassées pour mettre en œuvre notre squelette.

RSK7 : Évolution de la législation en matière d'archives publiques.

Comme pour les risques précédents, au fur et à mesure que le domaine de l'archivage électronique évolue, la réglementation qui le régit peut également évoluer rapidement. Cette évolution peut notamment affecter l'obligation d'archivage définitif dans la sphère publique

Deuxième partie : Le système d'archivage électronique définition et mise en place

II. État de l'art sur la gestion des archives

La gestion des archives consiste à définir, classer ou organiser, conserver, archiver et parfois éliminer les documents dont la durée d'utilité administrative est arrivée à terme (DUA). En entreprise elle doit tenir compte de plusieurs facteurs (juridique, historique, informationnel, mémoriel) et notions (politique d'archivage, records management, management, etc.) qui sont indispensables pour bien comprendre cette problématique. L'archivage correspond au processus défini plus haut. L'association des archivistes français a écrit un ouvrage, *Archiver, guide des procédures existantes, mieux maîtriser l'information en entreprise*⁵ (AAF, 2004), qui répertorie toutes les démarches et les normes nécessaires pour la gestion des archives en entreprise.

La gestion des archives a poussé la réflexion sur ce que l'on appelle les documents d'activités. « Ils sont définis par la norme ISO 15489, comme les documents créés, reçus, et préservés à titre de preuve et d'information par une personne physique ou morale dans l'exercice de ses conduites légales ou la conduite de son activité » (ARNAUD, 2012) [6]. Avec la même norme et la notion de documents d'activités vient le records management. Il s'intéresse à la fois aux supports papiers et électroniques. Ayant la même définition que le document d'activité, le records management « est un ensemble de pratiques visant à contrôler la gestion des documents d'activités » (BAUMANN, 2012).⁷ De manière générale le records management selon Sylvie BAUMANN, « se place du point de vue du producteur, de la preuve et de la « réservation des droits immédiats. Alors que du point de vue de l'archivistique classique on se place du point de vue de l'histoire et de la préservation à long terme » En définitive, elle permet la mise en place d'un système de gestion des documents d'entreprise les plus importants et ayant une portée juridique. Ce qui justifie l'importance de l'intégration du records management et des normes prévues dans la gestion des archives.

Laure MELLIFLUO dans son article *Évaluation des archives : théories et pratiques aux archives communales de la ville de Courage* définit le records management et détermine à quel niveau le critère intervient dans ce processus. « C'est dans la norme 15489 parue en 2001 que l'on trouve un cadre pour mettre en place un système de records management (...). L'on y voit que l'évaluation se trouve à deux niveaux, il s'agit de choisir les documents qui seront intégrés dans le système de records management (...) Ensuite parmi les documents intégrés dans le

système, une deuxième phase d'évaluation est menée afin de définir leur durée de conservation » [8] (MELLIFLUO, 2008).

Cette évaluation, respectant la norme dédiée au records management détermine l'affinité qu'elle a avec ce dernier. Au vue de sa réflexion, la mise en place d'un système de records management en entreprise permet de déterminer les documents à valeur légales à conserver ou à éliminer selon la DUA. Maurice HAMON dans son article sur « Les archives de l'entreprise : Actif matériel et gisement de ressource » explique la ressemblance entre archives et mémoire. « Ces deux notions sont complètement de même nature ». Parce que les archives définitives, une fois classées et la durée de conservation échue (30 ans) sont déversées en archives du patrimoine, représentant ainsi la mémoire de l'entreprise.

Pour définir la portée des archives dans l'établissement de la mémoire de l'entreprise, elle suggère « à partir des documents de base, il faut donc engranger et capitaliser sous une autre forme les connaissances et les savoirs acquis, rapprocher l'expérience humaine des archives produites structurellement, y compris pour reconstituer la part tacite de la mémoire » Par ailleurs, les entreprises sont souvent confrontées à des changements de statuts, à des fusions, des éclatements. Ces éléments amènent ainsi à penser une réorganisation de celles-ci. Cette réorganisation impact également sur les archives étant donné qu'elles représentent l'histoire et le mémoire de l'entreprise. Tous ces changements, doivent être retrouvés, ce qui permettra d'assurer la cohérence de son histoire et que cette réorganisation puisse « influencer sur le travail même de l'archiviste et notamment sur la mise à jour des tableaux de gestion » [(Espace_réservé1)9] (NOUGARET, 2006).

Ainsi Catherine KREBS dans son rapport de recherches les archives d'entreprises, Quelles gestions pour quels enjeux ? Porte sa réflexion sur la nécessité, en identifiant les catégories d'archives (courantes, intermédiaires et définitives), la qualité et les enjeux leur gestion « au regard essentiellement des impératifs spécifiques des entreprises privées dans un secteur concurrentiel » [10] (KREBS, 1989). D'autre part VDALENC rappelle qu'« Il faut en effet souligner que les archives d'entreprises n'intéressent pas seulement l'histoire économique, mais elles apportent des compléments d'éléments indispensables et révélateurs et aussi bien à l'histoire politique qu'à l'histoire sociale»[11] (VIDALENC, 1971).

Une bonne organisation des archives sous-entend l'établissement d'un plan de classement conforme aux besoins d'information de l'entreprise. Le plan de classement varie selon le domaine de l'entreprise dédié, il permet également la normalisation des modalités de nommages et d'enregistrement des documents et facilite l'indexation des documents numériques. Il devient donc le premier instrument de repérage intellectuel des informations,

lequel facilite les démarches de recherches sans s'attarder à des ensembles documentaires non requis. De plus il fournit un cadre référence uniforme utilisé pour tout le personnel. Néanmoins, il est important de souligner que cette activité (classer) nécessite une vision profonde du métier de l'entreprise et ses besoins afin de faire le bon choix de l'outil de gestion. On note ainsi une raison des raisons pour lesquelles un archiviste ou gestionnaire de l'information est important dans une entreprise. Ainsi l'intérêt de la gestion des archives dépasse le cadre réglementaire. Une bonne organisation et classement des archives permettent d'optimiser la recherche comme souligné « L'archivage en entreprise apporte de nombreux bienfaits pour la gérance de l'ensemble des secteurs d'une société. Contenant des informations confidentielles ou publiques, les archives sont souvent le point de départ des actions menées et permettent encore une fois de prouver la bonne tenue de l'entreprise au regard des actions des lois et réglementations qui régissent son secteur. Un archivage tenu à jour permet une optimisation des flux améliorant la traçabilité des documents et favorise de fait même la performance de l'entreprise (...) » [12]

L'étude et l'analyse de cette littérature non exhaustive permet de confirmer la portée du sujet, la valeur et l'impact de la gestion des archives industrielles. De manière générale la mise en place d'une politique d'archivage permet sans contradiction de préserver l'histoire et la mémoire de l'entreprise. Elle facilite et optimise la recherche en respectant la réglementation en vigueur.

Enfin du point de vue stratégique elle facilite la prise de décision, une bonne gestion et conduite de l'information. Les entreprises bien conscientes de l'existence du corps de métier archivistique, ne remarquent pas son utilité d'un coup d'œil. La remise en question et prise de conscience n'intervient qu'au moment de perte de documents à valeur probante, ou un manque d'organisation des documents et dossiers. Or l'organisation des archives ne doit pas seulement être un problème juridique, c'est-à-dire justifier le besoin d'un archiviste afin d'être en accord, de respecter la réglementation en vigueur. En entreprise on retrouve souvent la fonction de manager dans différents domaines, comme manager chargé de la communication. En archivistique, nous également ce que l'on appelle le « management des archives ». Le management est une méthode d'organisation et de gestion d'une activité. Dans le cas des archives cette notion intervient dans tout le processus de mise en place de la politique d'archivage et jusqu'au choix des outils de gestion. Ces outils représentent aujourd'hui des dispositifs de gestion et de conservation des archives dont les systèmes de gestion électronique des documents (GED) et d'archivage électronique (SAE). Avec la multiplication de ces outils à savoir la GED, leur application ne correspond pas forcément aux besoins

d'information de l'entreprise. Cette faille peut être définie du point de vue de la qualité et de la performance de l'outil et de son adaptation au domaine de l'entreprise. Par exemple les outils de GED dans le domaine ferroviaire ne seront pas le même, n'aura pas les mêmes fonctionnalités qu'une GED au BRGM ou d'un autre laboratoire de recherche. Pour faire face aux enjeux de la gestion des archives en entreprise Marie Anne CHABIN dans son ouvrage « Le management de l'archive » présente quelques conseils et solutions pratiques afin de définir une stratégie de mise en place d'une politique d'archivage, les procédures de sélection des outils de gestion et « traitement des documents ». Elle définit également l'utilité de la gestion des archives en entreprise, des démarches et directives à suivre selon les lois et réglementations et de l'évolution ou du changement du métier d'archiviste avec les technologies de l'information. Avec ce nouvel ordre numérique, sachant que toutes les entreprises se bousculent vers les technologies de l'information, rares sont ceux qui utilisent les outils adéquats. En plus du fait qu'elles font face à une multiplication de la production des archives, le domaine archivistique ne peut que s'adapter et évoluer avec cet ordre.

En tenant compte du fait que les entreprises tiennent à la protection et à la confidentialité de leurs données, il serait impossible de traiter les enjeux de la gestion des archives sans citer la réglementation sur la protection des données personnelles, rentrée en vigueur le 28 Mai 2018. Le RGPD n'intervient que dans le cadre de la gestion des archives définitive, il s'applique à la fois aux archives papiers et numériques. Un point qui montre que le domaine archivistique s'adapte aux lois sur la protection des données sachant qu'il « bénéficie d'un statut spécifique, dérogatoire au régime de droits commun, statut justifié par la finalité de la conservation des archives (apporter des preuves, documenter l'histoire) qui implique la conservation des archives intègres (...). Sans oublier que "Le RGPD et les articles 4 (2° et 5°) et 78 de la loi Informatique et Libertés révisée confirment la possibilité de conserver les données au-delà de la durée de conservation dans le traitement initial (durée qui correspond habituellement à la durée d'utilité administrative – DUA) à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques"¹³ (RICARD, 2018). Ainsi la mise en place d'une politique d'archivage assure et garantit la pérennité et l'infaillibilité d'une information cohérente. L'Association des archivistes Français lui attribue « neuf (9) objectifs dont :

- La fonction de preuve : L'archivage permet à l'entreprise de prouver ses engagements et obligations à tout moment.
- La fonction de mémorisation ou de sauvegarde des opérations : L'archivage consiste à s'assurer qu'à partir d'une nouvelle visite, la base de référence pour les nouvelles

opérations ;

- La fonction de compréhension et de positionnement stratégique : L'archivage signifie aider à la prise de décision en examinant l'environnement de l'entreprise à tous les niveaux.
- La fonction patrimoniale : L'archivage est une remise en question constante sur l'identité et l'évolution de l'image de l'entreprise et est-ce qu'elles sont sous contrôle ?»

Enfin « plusieurs facteurs influent directement sur les besoins des entreprises à archiver leurs documents. Il faut savoir que la plupart des entreprises connaissent une croissance exponentielle des documents et des données à gérer, et qu'elles doivent le réaliser le plus efficacement possible à moindre coût. Non seulement une bonne gestion permet de préserver la mémoire de l'entreprise, mais elle facilite grandement son exploitation au quotidien. L'archivage n'est plus considéré comme une contrainte, mais bien comme une nécessité pratique et utile pour accroître son efficacité dans la conduite courante de ses activités » Le traitement et la gestion des archives en entreprise public ou privée, permet une meilleure exploitation de l'information et une préservation de la mémoire de celle-ci.

Par ailleurs, il est primordial voir essentiel de parler d'archivistique sans s'intéresser à la question de la matérialité. La matérialité est vue de manière générale seulement comme l'objet contenant l'information. Mais hormis cette appréhension la matérialité peut être étudiée de différentes manières. C'est d'ailleurs là toute l'étendue de l'article d'Anne KLEIN et Yvon LEMAY sur "Matérialités des archives transmission de l'histoire". Il transcrit l'importance et la portée des archives dans la constitution et la reconstruction de l'histoire, et recentre la place et la définition de la matérialité en archivistique. Pour ce faire ils procèdent à une étude de cas de deux ouvrages : « Archives sur la vie littéraire des français sous l'occupation : à travers le désastre de PAXTON Robert, dont le catalogue augmenté présente une vaste sélection de documents intellectuelles français. Le propos de l'ouvrage est de mieux comprendre les diverses situations auxquelles ont dû faire face la personnalité du monde littéraire français sous l'occupation, leurs réactions et les rôles leurs a assignés ou qu'elles se sont données. » Et le dernier ouvrage : « Paroles de l'ombre : lettres et carnets des français sous l'occupation (1939-1945) Jean Pierre GUENO qui propose une sélection de documents inédits issus des archives de particuliers qui ont accepté de participer au projet consistant à évoquer des faits et des témoignages qui sont essentiellement liés à des années obscures... ». Cette étude faite sur la base de types d'archives différents, présente chacune une façon

particulière d'appréhender la question de la matérialité des archives. Ce qui démontre par ailleurs, toujours selon ces deux cas, que les archives peuvent être l'objet d'une « histoire intellectuelle reposants sur écrits d'historiens » (p.238) d'une part et d'autre part « une histoire populaire visant à permettre de réunir toutes les générations dans l'émotion des documents originaux [...et] réconcilié l'histoire avec la mémoire et la vie quotidienne » (p.238). Ces deux démarches, intellectuelle et émotionnelle rendent compte de deux visions très distinctes des archives. La première soutenant l'idée que les documents nécessitent une interprétation pour être signifiants quand la seconde soutient que les documents archives portent en elles-mêmes une parole qu'il est nécessaire de transmettre mais non commenter ou d'expliquer.

III. Le système d'archivages électronique et sa mise en place

1. Qu'est-ce qu'un SAE

Comme définit dans la partie précédente, un système d'archivage électronique est une offre de service de conservation à long terme de données à valeur scientifique, légal et patrimonial. La mission d'un SAE est de permettre la conservation des données sur le long terme et de les protéger de toute altération, afin de leurs assurés une valeur probatoire si besoin en cas de contentieux. Un SAE doit couvrir diverses fonctionnalités tel que la traçabilité, l'intégrité, la sécurité et la pérennité. Il doit également faciliter la recherche de documents d'archives, leur exploitation et leur communication. Enfin, il permet de conserver ses données tout en maîtrisant ses coûts.

2. Texte de référence

Du point de vue juridique, l'archive est étroitement liée à la preuve parce qu'elle fournit la justification et la défense des droits d'un document ou une donnée jusqu'à son échéance.

Selon le code civil, « *Les documents sous forme électronique peuvent être utilisés comme preuves au même titre que les documents papiers, à condition que la personne qui a émis le document ou les documents puisse être correctement identifiée, et que le document puisse être rédigé et sauvegardé avec la garanti de son Intégrité*» (Code civil, art.1316-1)[14].² L'identité de l'auteur (ou du destinataire) doit passer par un processus d'identification fiable⁶, c'est-à-dire la signature électronique du document. La certification de ce processus est confiée à un tiers de confiance et / ou à un tiers d'horodatage.³

Le livre II du code du patrimoine défini les archives comme « l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité. »[15]

La loi El Khomri texte de référence sur digitalisation entrée en vigueur le 1er janvier 2017 est un véritable accélérateur de la digitalisation des entreprises. Il stipule dans son article 54 : « Sauf opposition du salarié, l'employeur peut procéder à la remise du bulletin de paie sous forme électronique, dans des conditions de nature à garantir l'intégrité, la disponibilité pendant une durée fixée par décret et la confidentialité des données ainsi que leur accessibilité dans le cadre du service associé [...] Un décret en Conseil d'État pris après avis

² Ordonnance n° 2016-131 du 10 février 2016 portant réforme du droit des contrats

³ (D., 30 mars 2001, n°2001-272)

de la Commission nationale de l'informatique et des libertés détermine les modalités de cette accessibilité afin de préserver la confidentialité des données. » [16]

1. Cadre normatif pour l'archivage électronique

De nos jours bien nombre de normes sont devenues de facto des exigences ou des recommandations pour la mise en œuvre de systèmes d'archivage électronique. Ces normes sont principalement dérivées de l'ISO 15489.

Pas moins de trois normes ou standards peuvent apparaître dans le domaine des spécifications techniques d'un SAE. Il s'agit de la norme NF Z042-013 (convertie en ISO 14641-1) et des normes MoReq2010 et ICA-Req (désormais ISO 16175).

On va détailler la norme NF Z42-013 sur cette partie. On traitera la norme MorReq2 dans la partie modèle de données.

En plus de ces textes réglementaires, il existe tout un ensemble de normes couvrant l'archivage.

i. NF Z42-013 (ISO 14641-1)

La norme AFNOR NF Z42-013 est devenue une norme internationale. Elle est publiée sous le titre ISO 14641-1, couvrant les spécifications de conception et de fonctionnement des systèmes informatiques pour le dépôt électronique. Désormais, les entreprises du monde entier peuvent utiliser cette norme. Créée par l'AFNOR, l'NF Z42-013 définit les mesures techniques pour la garantie de l'intégrité, l'horodatage et la conservation des données.

La norme assure entre autres, la traçabilité du processus de dépôt électronique, comme la numérisation ou l'horodatage. Dans sa première version, pour des soucis de préservation il limitait les supports de stockage uniquement au supports non réinscriptibles « WORM » (Write Once Read Many).

Une mise à jour de la norme autorise d'autres supports que le WORM, à savoir le support magnétique, à la condition d'avoir recours à la cryptographie ou la signature électronique afin de garantir l'intégrité du système [17].

La norme stipule que pour garantir la preuve de l'exhaustivité et la fiabilité des données et des documents numériques, les organismes qui mettent en place le SAE doivent périodiquement faire un audit ou état des lieux du système et des procédures définies.

Pour ce qui est des métadonnées la norme est stricte sur leurs assignations sur les documents d'origine numériques en exigeant des métadonnées de description, des métadonnées de provenance et des métadonnées technique. Ceci garantie une « copie fiable a la même force probante que l'original et est présumée fiable jusqu'à preuve du contraire toute copie résultant d'une reproduction à l'identique de la forme et du contenu de l'acte, et dont

l'intégrité est garantie dans le temps par un procédé conforme à des conditions fixées par décret en Conseil d'État. »[18]

Une partie de la norme spécifie également les rythme de sauvegarde, les campagnes de migration à effectuer avant l'obsolescence du matériel ou du format de données. Une autre partie traite la sécurité du système d'information du SAE, plus précisément la sécurité d'accès aux locaux, les règles de sécurité définies pour le personnel, etc. [19]

Enfin, le NF Z42-013 a prévu un recours à des prestataires tels que le tiers-archivage. Elle encadre les services et obligations attendus, tels que les normes, les méthodes de sécurité qu'il a adoptées, ou la réversibilité des archives.

3. Les exigences fonctionnelles et techniques du SAE

Le SAE BRGM devra se baser sur le modèle OAIS. En effet, pour son bon fonctionnement notre SAE doit être conforme à la norme OAIS et s'inspirer donc de son modèle architectural. Elle doit aussi être compatible avec le SEDA (Standard d'Échange de Données pour l'Archivage) pour l'archivage de données et suivre les recommandations de la norme NF Z 42-013.

Nous allons dans ce chapitre présenter les exigences fonctionnelles et techniques auxquelles devra répondre le SAE Vitam.

4. Modèle de données utilisée

a) Le modèle d'information OAIS

Le modèle OAIS se base essentiellement sur le modèle fonctionnel ISO 14721 sur l'idée que l'information est constituée de paquets et que ces paquets diffèrent selon les usages. En effet, la production, la conservation et la diffusion ont chacune des paquets qui leurs sont propre.

On obtient ainsi trois types de paquets

- L'objet numérique sur lequel le producteur travaille avant l'archivage est nommée **SIP** (Submission Information Package ou Paquets d'information à verser) ;
- Une fois archivés, ces SIP sont transformées en **API** (Archival Information Package ou Paquets d'informations à archiver), qui sont des objets internes à l'Archive ;
- Les objets numériques destinés aux utilisateurs sont les **DIP** (Dissemination Information Package ou Paquets d'informations à diffuser), transformés par

l'archive à partir des AIP dans une forme plus facile à communiquer notamment sur le réseau.

Si nous voulons conserver en permanence les informations contenues dans les objets numériques, il est nécessaire de stocker un ensemble d'informations de traçabilité avec l'objet, ce qui garantit la pérennisation et l'intégrité de ces contenus informationnels. Ainsi, à chaque étape de chaque paquet d'information, nous retrouverons le fichier informatique correspondant à l'objet ou au document à conserver, ainsi que des informations sur le document, c'est-à-dire des métadonnées.

Par conséquent, un paquet d'informations est un conteneur conceptuel pour deux types d'informations : le contenu de l'information et les informations de pérennisation.

b) Standard d'échanges de données pour l'archivage

Pour faciliter les échanges de données dans le SAE BRGM nous utiliserons le standard d'échange de données pour l'archivage (SEDA).

Issu de la collaboration entre les Archives de France et l'ancienne direction générale de la modernisation de l'État dans les travaux du programme d'administration électronique ADELE, « Le SEDA vise à promouvoir l'interopérabilité des systèmes d'informations des services d'archives et de ses partenaires (services producteurs, contributeurs, utilisateurs, etc.) dans le cadre de leur échange de données. Le SEDA permet d'échanger des paquets d'information tels qu'ils sont définis dans le modèle OAIS ».[20] Il définit les scénarios à mettre en œuvre pour chaque transaction, sous forme de diagrammes UML⁴, la forme et le contenu des messages sous forme XML⁵ que peuvent ou doivent s'échanger les acteurs.

Le but du standard d'échange est de lui donner une valeur de référence forte et de la protéger des normes internationales incompatibles et inévitables. Le standard est un outil stratégique pour la mise en place de transferts automatisés des données. Les différents types d'échange qu'on peut avoir sont :

Le transfert de données

- La demande de transfert
- L'élimination
- La modification
- La communication

⁴ Unified Modeling Language

⁵ eXtensible Markup Language

- La restitution

Les acteurs

Les acteurs concernés par les échanges sont :

- Le service d'archives
- Le service versant
- Le service producteur
- Le service de contrôle
- Le demandeur d'archives

Les objets échangés et mise en place de profil d'archivage

Pour l'échange des données service d'archives devra mettre en place des profils d'archivage pour chaque transfert. Les différents objets qu'on peut échanger dans le modèle sont :

Archive : qui est constitué de métadonnée c'est-à-dire les informations de représentation et des informations de pérennisations ainsi que d'objets d'archives et de documents.

L'objet d'archives et le Document qui sert à transporter des contenus de données (fichiers) associés à leurs métadonnées.

La configuration d'un profil d'archivage passe par 3 étapes :

1ère étape : Comprendre ses utilisateurs et leurs besoins.

La conservation des archives n'a de sens que s'il est possible de les retrouver : l'archiviste fera un listing des métadonnées descriptives requises pour la recherche archivistique.

2nd étape : établir les prérequis.

Mettre en place les prérequis. En coopération avec le service producteur et avec le tiers de télétransmission, l'archiviste s'assure que les métadonnées sont présentes dans l'outil de production et peuvent être récupérées dans le SAE au moment des échanges SEDA spécifiques aux transferts.

3ème étape : la rédaction du profil d'archivage.

En coopération avec le service informatique, l'archiviste retranscrit les prérequis de transfert des données dans un profil d'archivage afin que ces règles soient compréhensibles par le logiciel de production et le SAE

Outre que les profils simples effectués pour chaque échange, il faut définir un fichier de configuration pour chaque flux série ou échanger des données de prototype. Avec ce fichier de configuration, on peut configurer des informations pré-connues ou décrire les contraintes

à respecter dans l'archive. Il comprend également une balise sur le sort final des archives et la durée de vie administrative.

Le standard d'échange de données s'applique à la gestion et à la description des données qui rentre dans le SAE. Il traite aussi bien la gestion des archives intermédiaires ou des archives définitives. Les données reçues par télétransmission selon le standard sont destinées à être gérées par différents systèmes, autres que le modèle proposé par SEDA voir schéma.

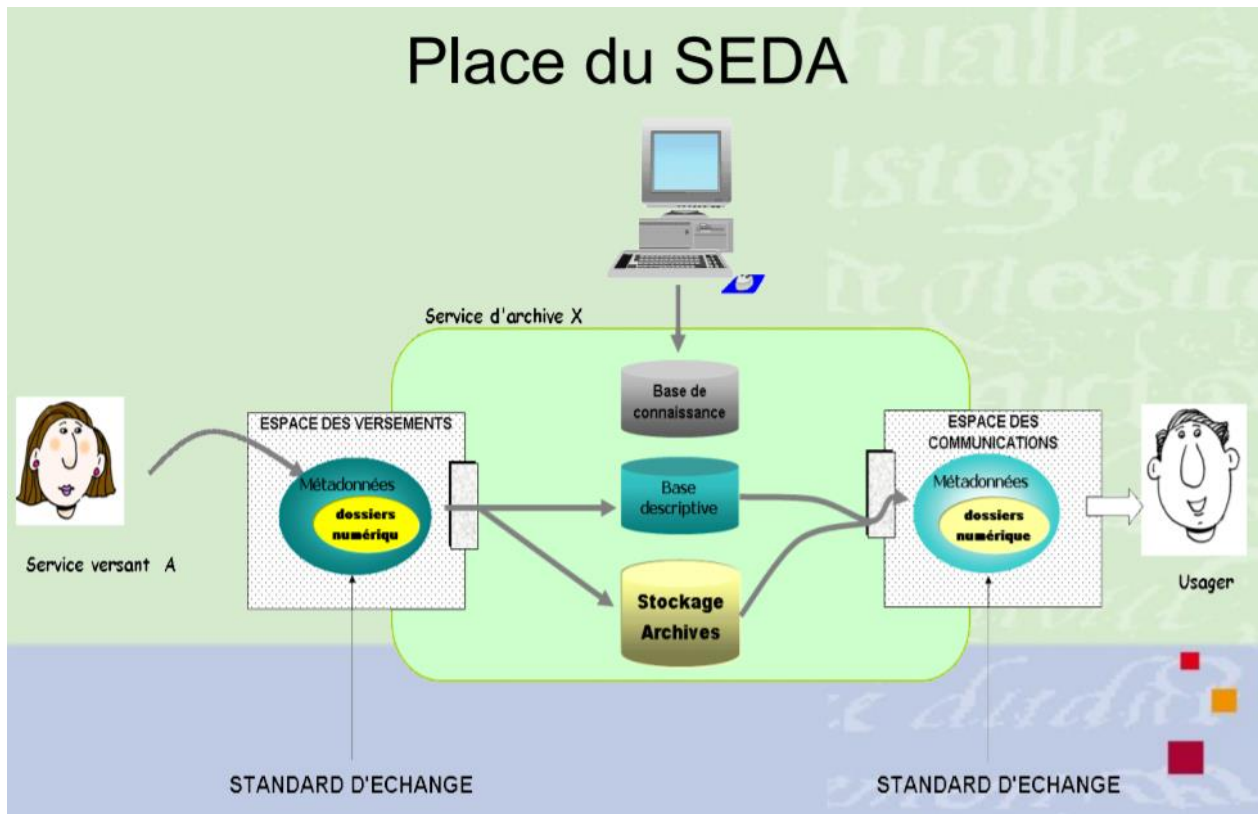


Figure 1 : Modélisation des échanges d'information SEDA

La base descriptive contient des informations sur les archives et fournit des critères de recherche par exemple : la description du contenu du fichier, le nom du service versant, formats des données et leurs modalités de communication, etc.

La base de connaissance comporte des données relatives au cadre d'exécution du processus d'archivage tel que les informations sur les applications métiers faisant l'objet du transfert, les offre de service, la documentation technique et les informations de représentation.

Le système de stockage gère tout ce qui est supports contenant des données.

4. Les entités fonctionnelles du SAE

Le modèle fonctionnel OAIS décrit toutes les fonctions nécessaires dans l'archive de manière très précise pour s'assurer que l'information reste pérenne de la capture à la diffusion.

Ces fonctions sont divisées en six entités fonctionnelles qui sont :

L'Administration, les Entrées, le Stockage, la Gestion des données, l'Accès et la Planification de la Pérennisation.

- **L'Accès** interfacé avec des applications permettant à des utilisateurs d'accéder aux objets archivés à des fins de recherche, consultation, via le front office avec les applications métiers des services producteurs.
- **L'entité Gestion des données** qui permet de gérer le cycle de vie et planifier la pérennisation, via le front office archivistique Atom et les applications métiers des services producteurs.
- Et l'entité **Administration** pour la gouvernance du système appuyé sur des offres de stockage qui assurent la conservation binaire des objets archivés.

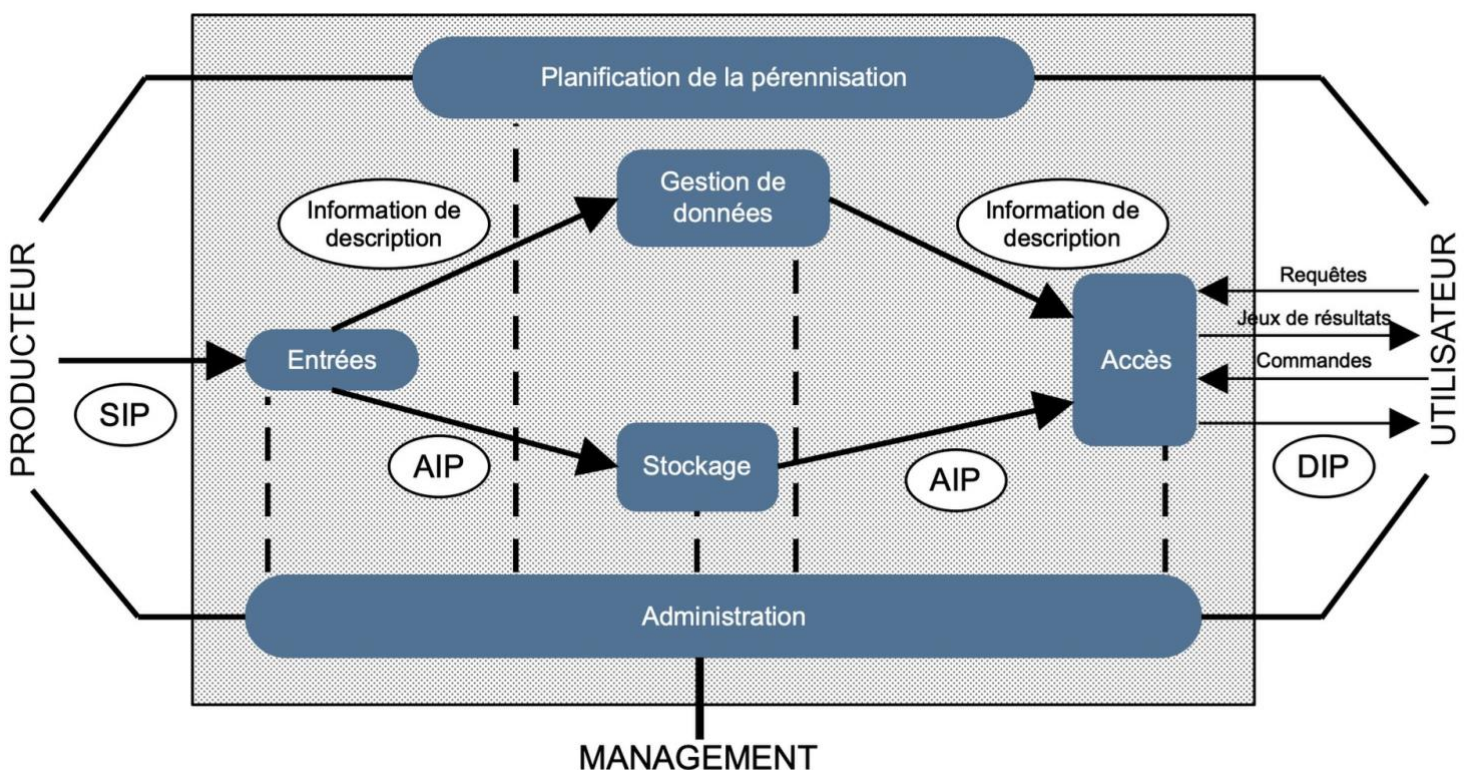


Figure 2 : modèle fonctionnelle OAIS

- Le schéma ci-dessus présente les fonctions du modèle fonctionnelle OAIS comprenant les différentes fonctions que devra offrir la maquette.

Entité Entrée ou Versement

L'objectif principal d'un SAE est la conservation des données archivées. Ainsi la fonction entrée est chargée de Recevoir, contrôler et vérifier les objets à archiver. Le paquet d'information à verser (SIP) reçues sont transférés à l'entité stockage, et les informations nécessaires à leur description et à leur gestion sont transférées à l'entité « gestion des données » au cours du temps.

L'entité versement assure aussi de :

- Prendre en charge les SIP livrés par les producteurs.
- Préparer les AIP en vue de leur stockage.
- Intégrer les AIP et leur Information de description.
- Contrôler la conformité : ce processus vérifie si le dossier d'information fourni est conforme et répond aux conditions définies entre le service soumissionnaire et le service des archives, ainsi que les exigences de construction de toutes les données et de leur structure. Si le format de description est complet et s'il est conforme au format d'encodage de l'objet transmis et de ses composants (conforme à la norme d'échange de données d'archive : SEDA)

Le schéma ci-dessous présente les fonctions de l'entité « Entrée » comprenant différentes les fonctions que devra offrir la maquette.

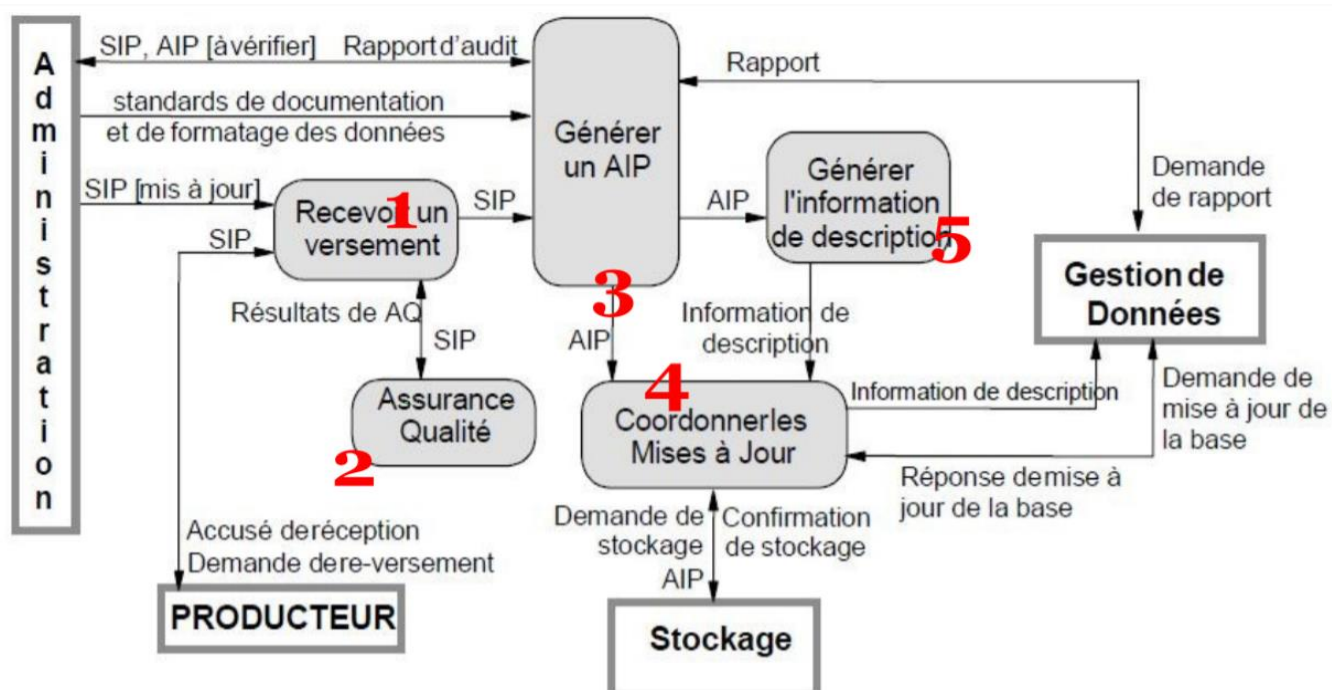


Figure 3 : Fonctions de l'entité « Entrée »

L'entité « Stockage »

L'entité de stockage comme son nom l'indique est chargé de stocker et sauvegarder les paquets d'informations archivés (AIP).

L'entité « stockage » assure la conservation physique des objets archivés. L'entité entrée met les données archivées à la disposition de l'entité « accès ». Comme définit par l'entité « administration », il prend en charge la réalisation des copies multiples et le renouvellement des supports anciens. Il gère aussi les supports de stockage tout en veillant à leurs obsolescences par cette mission il assume la migration et le renouvellement des supports. Enfin il transmet les AIP à l'entité **Accès** en réponse aux commandes des utilisateurs

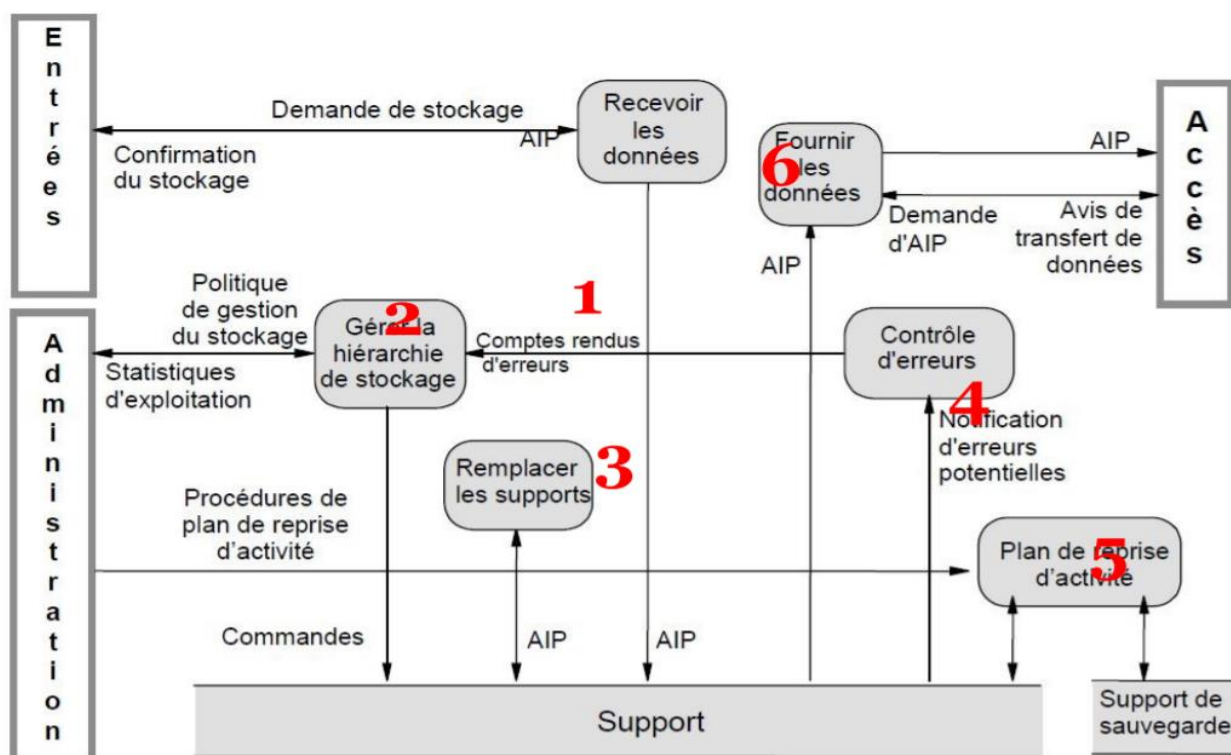


Figure 4 : Fonctions de L'entité « Stockage »

L'entité « gestion des données »

L'entité « gestion des données » se charge de la mise à jour des bases de données pour le fonctionnement du SAE. Elle fournit aux autres entités du système les informations de données descriptives nécessaires pour générer des lots de résultats.

L'entité « Administration »

L'entité « administration » est la fonction la plus complexe du SAE. Gérer par le service d'archives il est en charge du pilotage du système. Elle en établit les règles internes. Elle veille à la qualité du service du SAE pour apporter des améliorations à d'éventuelles défaillances. Elle rend compte au management. Elle garantit la sécurité d'accès au système, via l'authentification des différents intervenants et l'assignation des droits d'accès. En cas de redondance des informations, elle gèrera également la réplication d'un site vers l'autre. L'entité administration devra également respecter le plus possible le plan de reprise d'activité PRA et le Référentiel Général d'Interopérabilité (RGI), afin de rendre interopérable le système, permettant alors sa réversibilité. Ces missions principales sont :

- La proposition et la négociation des protocoles de versement avec les producteurs.
- L'allocation des moyens techniques pour contrôler et améliorer l'exploitation du système et pour inventorier et migrer les contenus.

L'établissement et le maintien des normes et règles applicables.

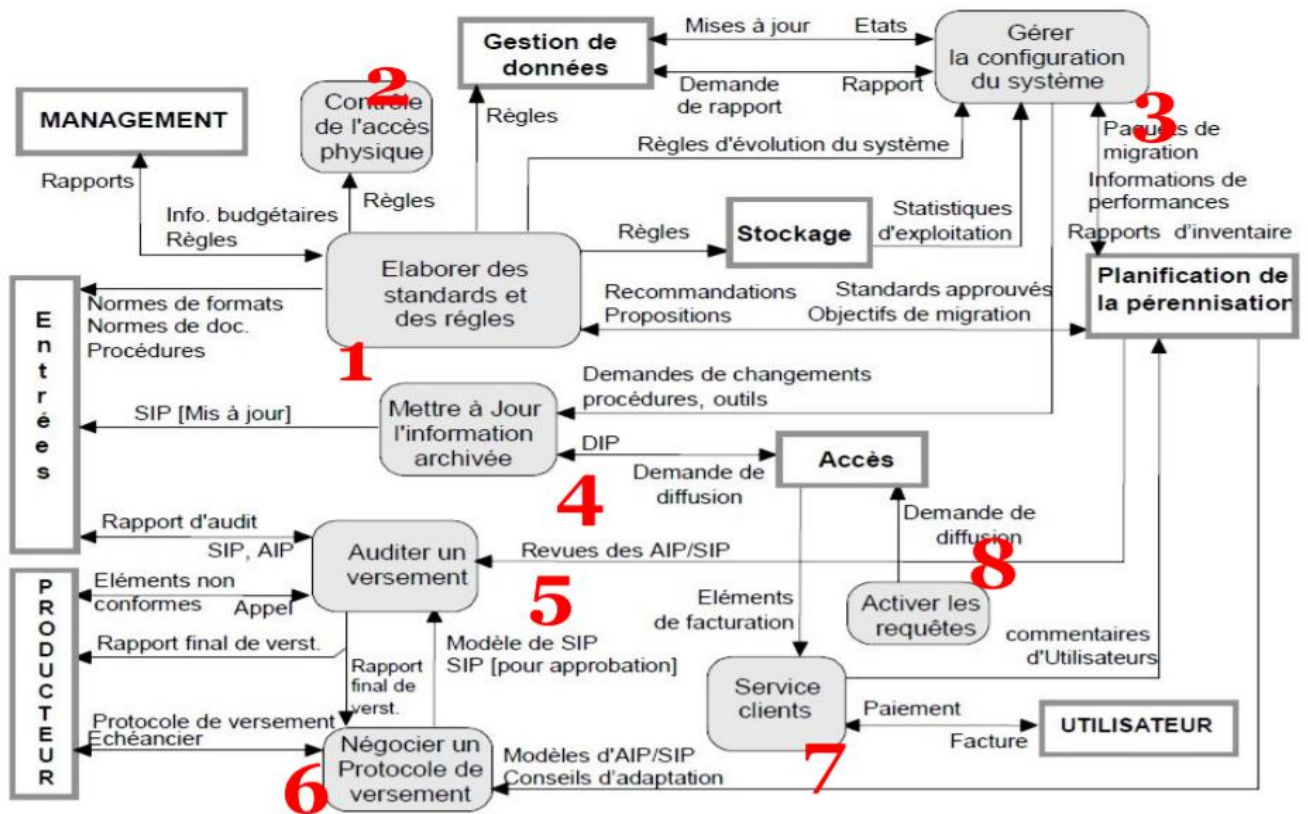


Figure 5 : Fonctions de L'entité « Administration »

L'entité « planification de la pérennisation »

L'entité « planification de la pérennisation » est la cellule de veille et de planification du système. Elle veille sur l'environnement extérieur et émet des recommandations en vue de procéder aux évolutions nécessaires, notamment aux évolutions de type technologiques. Elle assure une veille technologique et propose les évolutions et les stratégies pour prévenir l'obsolescence. Elle est ainsi responsable du suivi des changements qui peuvent s'opérer dans la « communauté d'utilisateurs et des archives en vue de garantir que le service d'accès reste conforme aux attentes nouvelles des usagers.

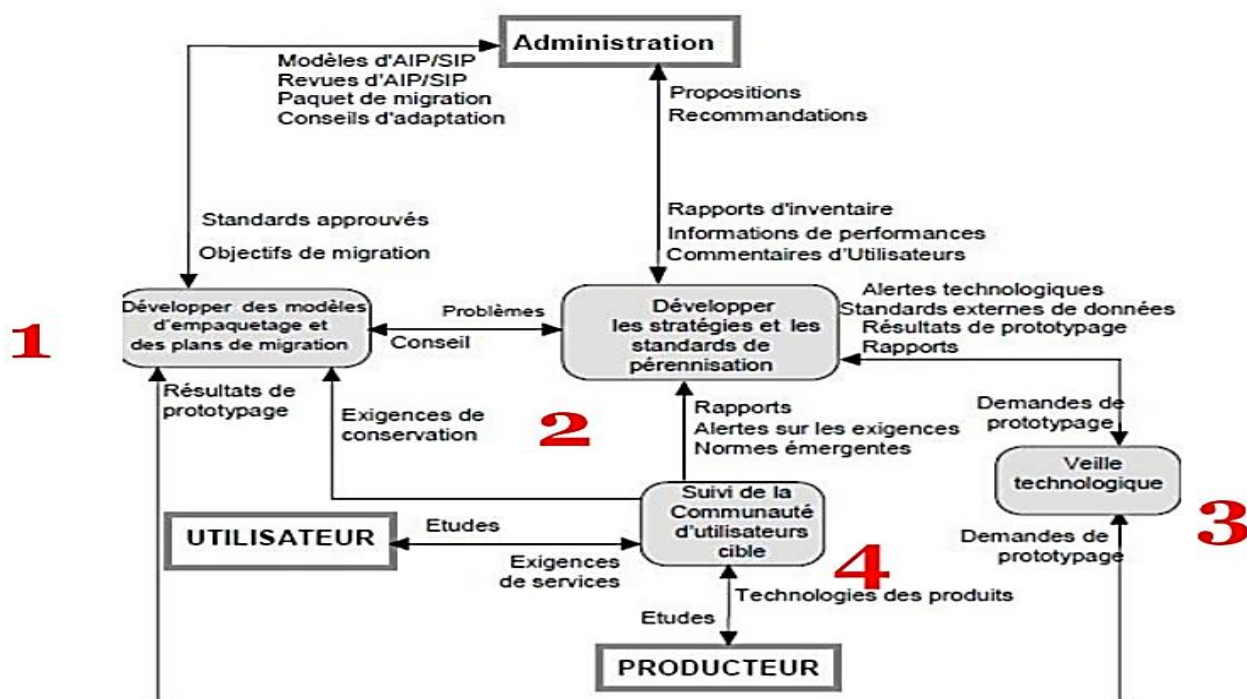


Figure 6 : Fonctions de L'entité « planification de la pérennisation »

L'entité « Accès »

Situé de l'autre bout de la chaîne du système d'archivage électronique L'entité « accès » regroupe tous les services qui sont en interface directe avec les utilisateurs. Il reçoit la demande des utilisateurs et veille au contrôle des droits d'accès. Il traite les demandes et permet aux utilisateurs de disposer des Paquets d'Information à Diffuser (DIP) dans le « catalogue » des objets archivés.

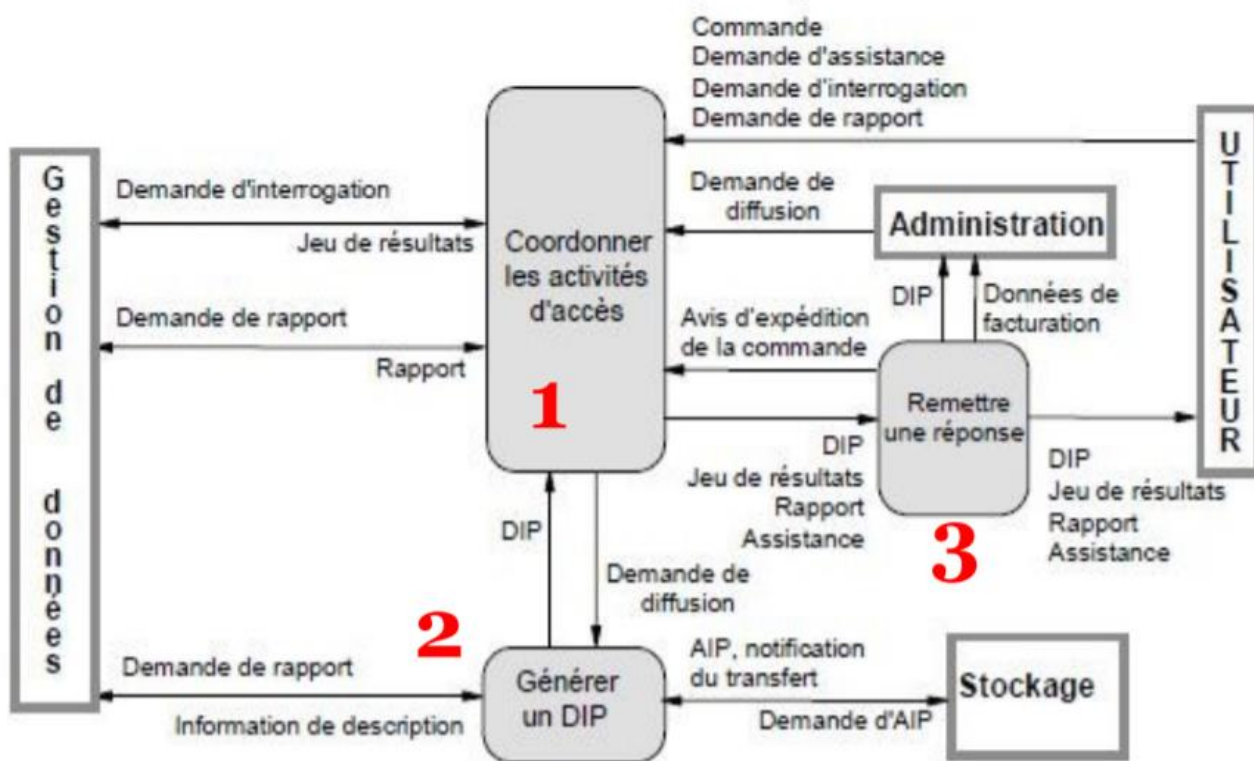


Figure 7 : Fonctions de L'entité « accès »

Nous avons à présent fini de montrer les exigences fonctionnelles du modèle OAIS. Ces exigences couvrent les différents aspects nécessaires pour garantir la confiance dans le SAE et la pérennité des données à savoir les aspects couvrant :

- l'intégrité
- la confidentialité
- la traçabilité
- la sécurité
- la compréhensibilité
- la lisibilité
- la réversibilité

5. RÔLES, ACTEURS ET INTERVENANTS DU SAE

Cette sous partie détermine les acteurs du SAE leurs rôles et responsabilités. Les acteurs pouvant échanger avec le système et ses intervenant donc assumer les rôles définis conformément au Standard d'échange de données pour l'archivage. Il convient de bien distinguer les rôles des acteurs. Selon le cas d'utilisation, un acteur peut ainsi jouer des rôles différents lors de ses interactions avec le système. L'acteur ne doit pas être confondu avec l'intervenant (ou personne physique) qui peut assurer différents rôles suivant la transaction utilisée.

1. RÔLES ET RESPONSABILITÉS ASSOCIÉES

Le service d'archivage

L'Autorité d'archivage est responsable de l'ensemble des services fournis par le Système d'archivage électronique. Le rôle d'Autorité d'archivage est assuré par le Service d'archives dans notre cas le SG/GDA met en œuvre le SAE. L'Autorité d'archivage garantit la pérennité et l'intégrité, dans le temps, des archives qui lui sont confiées et pour lesquelles elle a délivré une attestation de prise en charge. Elle est responsable de l'ensemble des engagements de services définis dans la Politique d'archivage et dans les Contrats de services et de versement. Enfin, l'Autorité d'archivage est chargée de rédiger et de faire évoluer la Politique d'archivage en fonction des évolutions législatives, réglementaires, technologiques ainsi que des besoins fonctionnels exprimés par les acteurs.

Equipe Vitam

L'équipe Vitam est l'équipe logicielle, il fournit de manière opérationnelle les services du SAE et maintient le système en conditions opérationnelles dans le cadre contractuel qui le lie à l'Autorité d'archivage.

L'autorité de contrôle

L'Autorité de contrôle certifie les modalités de conservation des archives et préserve la conformité des services rendus par le SAE avec la réglementation et les engagements de services décrits dans la présente Politique d'archivage.

1) ACTEURS ET INTERVENANTS

Le Service producteur

Le Service producteur est l'entité qui a produit les données/documents et leurs métadonnées destinées à être archivées dans le SAE, c'est-à-dire qui les a créées ou reçues dans le cadre de son activité.

Dans le cadre de notre SAE les Services producteurs sont aussi les services versant c'est-à-dire les services dont les archives pourront être versées sont tous les directions ainsi que les partenaires dont le BRGM prend en charges les données qu'ils produisent dans le cadre de leurs activités.

Le Service d'archives

Le Service d'archives est l'entité destinataire du transfert des archives et assurant la gestion des informations transférées par les Services producteurs et versants et destinées à être communiquées aux Demandeurs dans le respect des conditions légales, réglementaires et/ou contractuelles.

Dans le cadre de notre, Le SG/GDA est le Service d'archives. Il assume le rôle d'Autorité d'archivage et dispose à ce titre la responsabilité des services fournis dans le cadre du SAE et décrits dans la Politique d'archivage accompagnant le projet.

Le demandeur

Le Demandeur désigne toute personne physique ou morale qui souhaite consulter les informations conservées par le Service d'archives dans le respect des conditions légales, réglementaires ou contractuelles en vigueur.

Dans le cadre du SAE BRGM, les Demandeurs d'archives sont les services producteurs de documents.

Le contrôle scientifique et technique

L'autorité de contrôle scientifique et technique de l'État pour les archives publiques examine et contrôle le respect des services fournis par la SAE. Elle autorise également, le cas échéant, la communication et l'élimination des archives.

Dans le cadre du SAE BRGM, l'autorité chargée du contrôle scientifique et technique sur les archives publiques, est le Service Interministériel des Archives de France (SIAF).

Voici le tableau récapitulatif des acteurs/intervenants et leurs rôles sur le SAE

Tâches	Acteurs/Intervenants
Déploiement SAE	GDA-DISN-Equipe Vitam
Gestion de l'archivage	GDA
Administrateur Fonctionnel	GDA-Equipe Vitam
Administrateur technique	DISN-Équipe Vitam
Versement de paquet d'information	Service versant : BRGM HA, RH, DAT...
Demande de communication	Service versant, Utilisateur citoyen

Tableau 3 : Rôle des acteurs et intervenants dans le SAE

6. Distinctions SAE, GED CFN

La Gestion électronique de document

Quand on effectue un capture d'écran avec notre téléphone portable on fait volontairement ou involontairement de la gestion électronique de document. La GED est l'ensemble des techniques informatique utiliser dans un but de collecte, d'acquisition, de classement, ou de stockage de document. La Gestion Électronique des Documents regroupe toutes les techniques permettant de gérer les flux d'information qui circulent au sein de l'entreprise.

Le coffre-fort numérique

Un CFN est un service qui a pour objet :

La réception, la conservation, la suppression et la transmission de données ou documents électroniques qu'il reçoit dans des conditions permettant de justifier de leur intégrité, leur traçabilité et de l'exactitude de leur origine.

Les distinctions

Dans les applications de Gestion Électronique de Document telle que Teams et Opale au BRGM l'intégrité des documents générés ou reçus n'est pas assurée parce que l'auteur ou les intervenants peuvent modifier dupliquer ou supprimer des données. A la différence d'une GED, le SAE assure l'intégrité des données qu'il reçoit selon les spécifications de la **norme NF Z42-013**. Un Chaînage des Journaux (Date et heure de vérification, algorithme de chiffrement, Date et heure de déchiffrement) sont associés à chaque document. Ces informations ainsi que chaque opération réalisée sont tracées (dépôt, migration, élimination, ...). Ces objets ou traces capturées constituent des **journaux**, eux-mêmes déposés dans le SAE et chaînés entre eux. Tandis qu'un coffre-fort numérique intervient sur le périmètre de la conservation la pérennité et l'intégrité des objets qu'il conserve. Il ne tient pas compte de spécificités liées au Code du Patrimoine ou aux processus métiers liés à l'archivage comme le ferait un SAE.

Voici le tableau récapitulatif de la comparaison d'un système d'archivage électronique, d'un coffre-fort numérique et des différents outils de GED dont dispose le BRGM

Fonctionnalités	SAE	GED	CFN
Modifications du document	Non	Oui	Oui
Différentes versions du même document	Non	Oui	Oui
Suppression du document par le créateur	Non	Oui	Oui
Traçabilité des opérations	Oui	Non	Oui
Workflow pour suppression	Oui	Non	Non
Contrôle des accès aux données	Oui	Non	Oui
Authentification forte	Oui	Oui	Oui
Gestion de l'intégrité	Oui	Non	Oui
Gestion des volumes de stockage	Oui	Oui	Non
Migration des formats	Oui	Non	Parfois
Migration des supports de stockage	Oui	Oui	Non
Modification des métadonnées	Oui	Oui	Parfois

Tableau 4 : Fonctionnalités comparées SAE / GED / CFN

III. Mise en place du SAE

1. Phases du projet

La présente partie décrit les différentes phases du projet d'archivage électronique du BRGM. L'objectif est de montrer les étapes à suivre pour la mise en place d'une offre de service d'archivage électronique. La mise en place de l'offre de service d'archivage électronique nécessite une vraie collaboration GDA/DISN en lien avec les services producteurs et utilisateurs des données.

Pour concrétiser cette offre de service, nous avons porté notre choix technologique sur le logiciel du service interministériel des archives Vitam.

Le projet d'archivage électronique devra respecter les six phases suivant :

- Faisabilité et périmètre d'application du SAE
- Audit documentaire
- Analyse des données collectées et analyse des risques du rapport d'audit et élaboration de scénarios
- Rédaction du rapport d'audit, des scénarios et des livrables
- Rédaction du cahier des charges
- Le choix technologique

Faisabilité et périmètre d'application du SAE

Il s'agit de vérifier l'alignement entre les objectifs du BRGM et le projet pour sa validation.

Dans cette partie il convient de :

- Annoncer le contexte et l'objectif du projet (Centralisation des archives et conformité de la certification CTS)
- Définir le périmètre du projet (système d'information BRGM)
- Identifier les acteurs et les attendus



Audit documentaire

Il s'agit ici de collecter la matière nécessaire pour alimenter la phase de cadrage du projet. Il conviendra ici

- Identifier les éléments nécessaires à collecter
- Identifier les sources et interlocuteurs clé (DISN, HA...)
- Les pratique outil utilisé
- Identifier ou créer les outils de collecte
- Suivi de la collecte et documentation des données (grille de collecte)
- Analyse des données collectées

Analyse des données collectées et analyse des risques du rapport d'audit et élaboration de scénarios

Il s'agit d'identifier les risques encourus en matière d'archivage

- A partir des éléments collectés identifier les risques à plusieurs niveaux

- *Respect de la réglementation (RGPD)*
- *Risque financier*
- *Impact sur le SI*
- [Analyse des risques du rapport d'audit](#)

Rédaction du rapport d'audit, des scénarios et des livrables

Dans cette partie il faut proposer des scénarios répondant aux besoins exprimés mais avec des caractéristiques différentes [file:///Users/w/Desktop/HA livrable/Evaluation SI existant.pdf](file:///Users/w/Desktop/HA%20livrable/Evaluation%20SI%20existant.pdf)(voir diagnostique des SI dernière partie de ce mémoire)

- Le rapport est un document qui donne une vision globale des flux, des typologies, des fonds, de leurs emplacements et de l'espace qu'ils prennent...)
- A partir de cette analyse, seront notifiés des constats avec des propositions de scénarios (tenant compte de l'analyse des risques

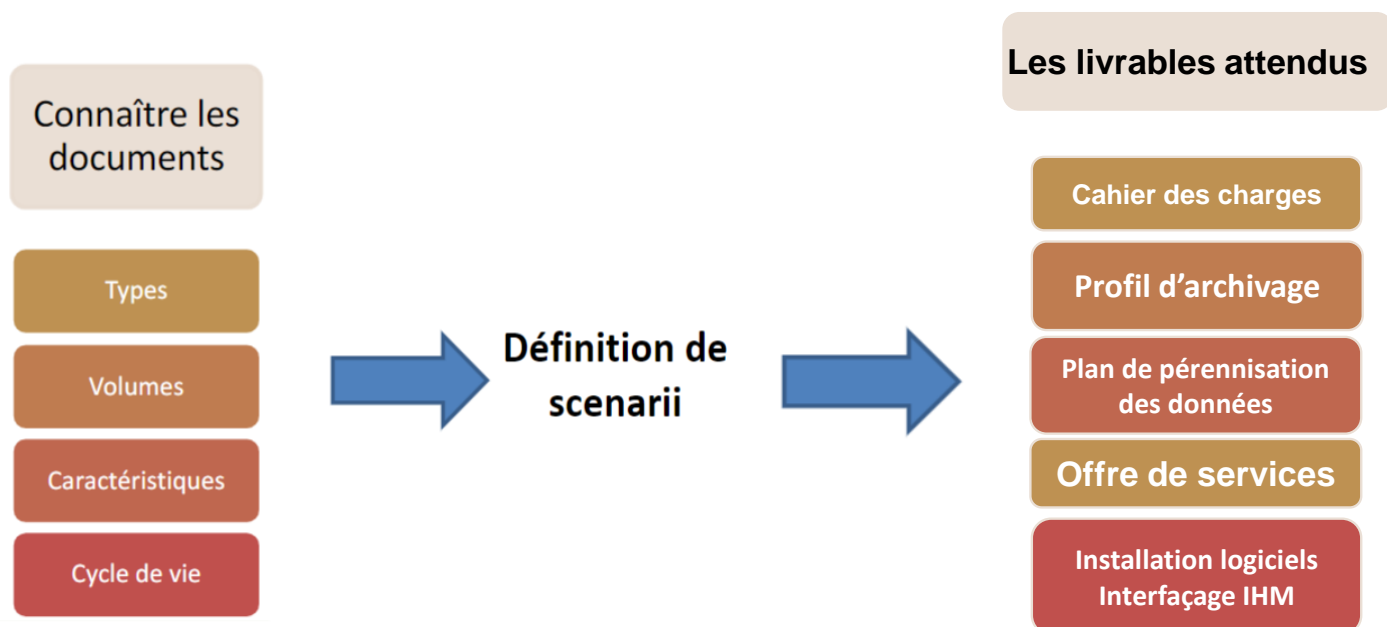


Figure 8 : Définition des scenarii et des livrables attendus

Rédaction du cahier des charges

Il s'agit ici de collecter la matière nécessaire pour alimenter la phase de cadrage du projet en trois étapes :

Modélisation du processus exemple de l'unité HA

Une fois modélisés ces processus sont traduits en langage exprimant un besoin, ceci aidera à faire les choix organisationnels et technologiques

Le choix technologique

C'est la jonction entre la MOA et la MOE dans notre cas GDA DISN.

La mise en place de la fonction archivage électronique nécessite une vraie collaboration archivistes/informaticiens. Chacun doit faire sa part du chemin pour comprendre les besoins et contraintes fonctionnels et techniques,

Le choix technologique est prépondérant pour la suite avec le choix du scénario pour le projet ainsi que l'arbitrage sur les risques.

Pour le projet de SAE du BRGM le choix logiciel est porté sur la solution VITAM qui s'appuie sur le modèle OAIS

2. Pourquoi la solution Vitam (Valeurs Immatérielles Transférées aux Archives pour Mémoire)

Étant développé sous l'égide du service interministériel des archives C'est un logiciel libre d'archivage électronique capable de gérer, préserver et donner un accès à long terme aux archives numériques qui lui sont versées (back office). Pour être utilisé, il doit être interfacé avec :

Des applications métiers et les logiciels des services d'archives, largement adaptables aux besoins et contraintes organisationnelles des entités utilisatrices.

Des infrastructures de stockage (file system ou virtualisées).

Notre objectif principal est la construction d'une plate-forme d'archivage électronique, à interfacé avec une instance VITAM de back-office et avec les systèmes d'information métier existant au BRGM tels qu'Opale SIRH ou TEAMS en FrontOffice. Positionnée comme une brique d'infrastructure, Vitam prend en charge toutes les opérations nécessaires pour garantir le maintien de leur valeur probante et pour assurer la pérennisation des documents numériques versés ci voici un schéma du positionnement du brique architectural de VITAM au sein du BRGM relié avec les applications de GED existant avec le plan de reprise d'activité prévu en cas d'indisponibilité d'un des serveurs ou système de stockage avec une réplication systématique des données conservé sur deux sites.

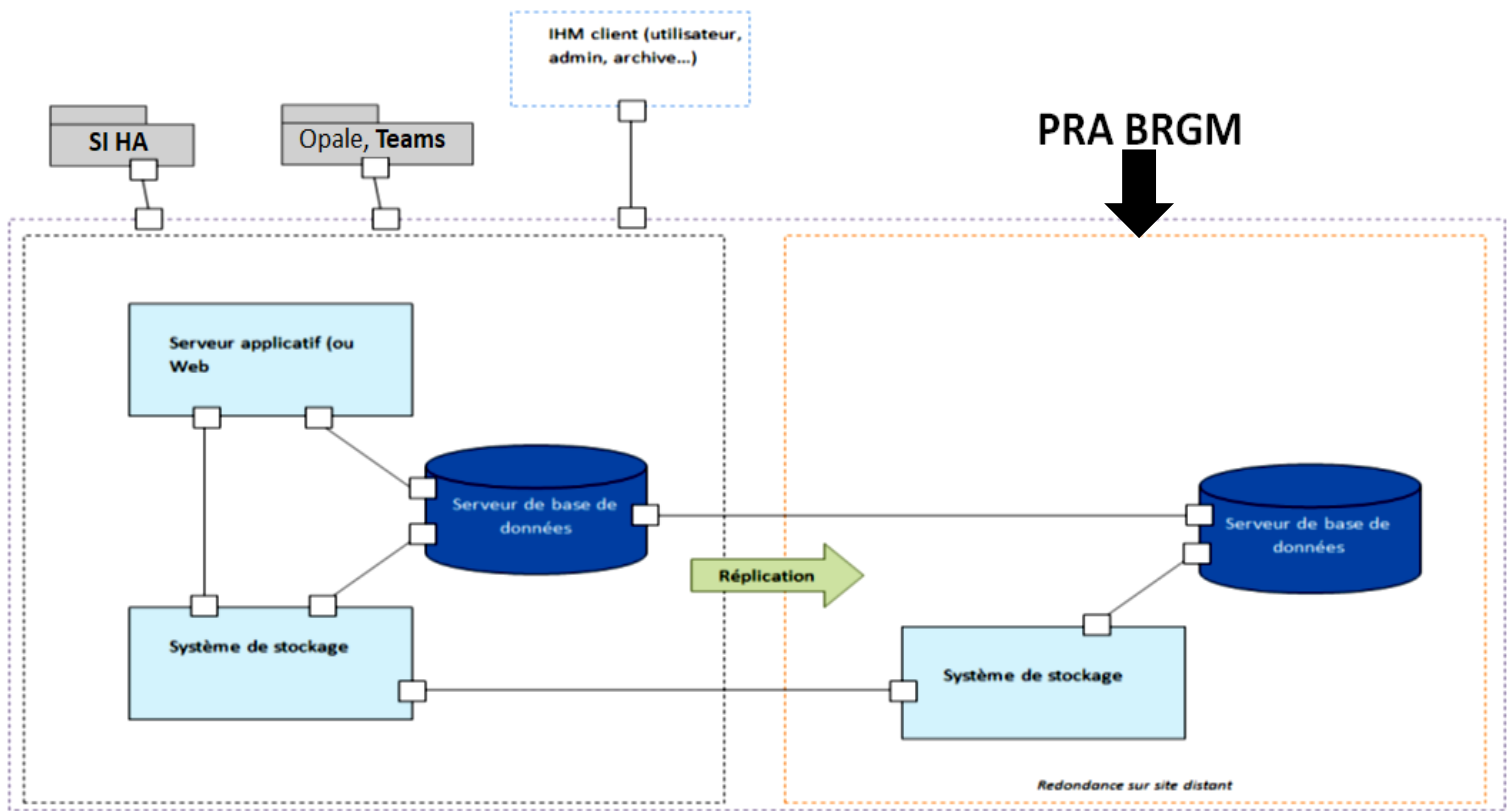


Figure 9 : Positionnement de Vitam dans le SI BRGM

3. Livrables attendus

a. Modélisation du processus et étude des besoins des Achats

i. Modélisation du processus achats

Afin de mieux comprendre les procédures de marchés publics et de mieux évaluer l'applicabilité des plateformes d'Achat public actuellement disponibles en France, nous avons dans cette partie modéliser les procédures de marchés publics afin de les rendre facilement utilisable, pratiques et compréhensible. Cette modélisation aidera à formuler des spécifications de système d'information adaptées aux marchés publics

Selon le site Alaide un processus métier est « l'ensemble des activités internes d'un métier dont l'objectif est de fournir un résultat observable et mesurable pour un utilisateur interne du métier ». Au BRGM, un achat est déclenché à l'initiation d'un agent quelconque par une demande d'achat saisie sur l'application OPALE. Cette dernière enclenche automatiquement la procédure en avisant la personne autorisée a approuvé la transaction à l'unité Achat. Via Opale l'unité Achat à tous les informations nécessaires concernant l'achat et l'Acheteur. En effet Les salariés créent leurs demandes d'achats dans Opale qui suit le "Workflow" de validation jusqu'à l'unité Achat. La demande d'achat est ensuite traitée par le personnel des achats qui donne renseignements sur le prix conclu avec les fournisseurs. Une fois validée, elle est acheminée vers le responsable du service auquel le salarié demandeur est affecté pour une potentielle approbation. Celui à deux options soit de valider la commande qui repart vers l'unité achat ou l' invalide ce qui annulerait la procédure. Vous trouverais ci-dessous les différentes modélisations effectué pour le processus des achats, les données générées par ce processus et les outils utilisés.

Reproduction du processus HA



Figure 10 : reproduction du processus des Achats

Les outils utilisés

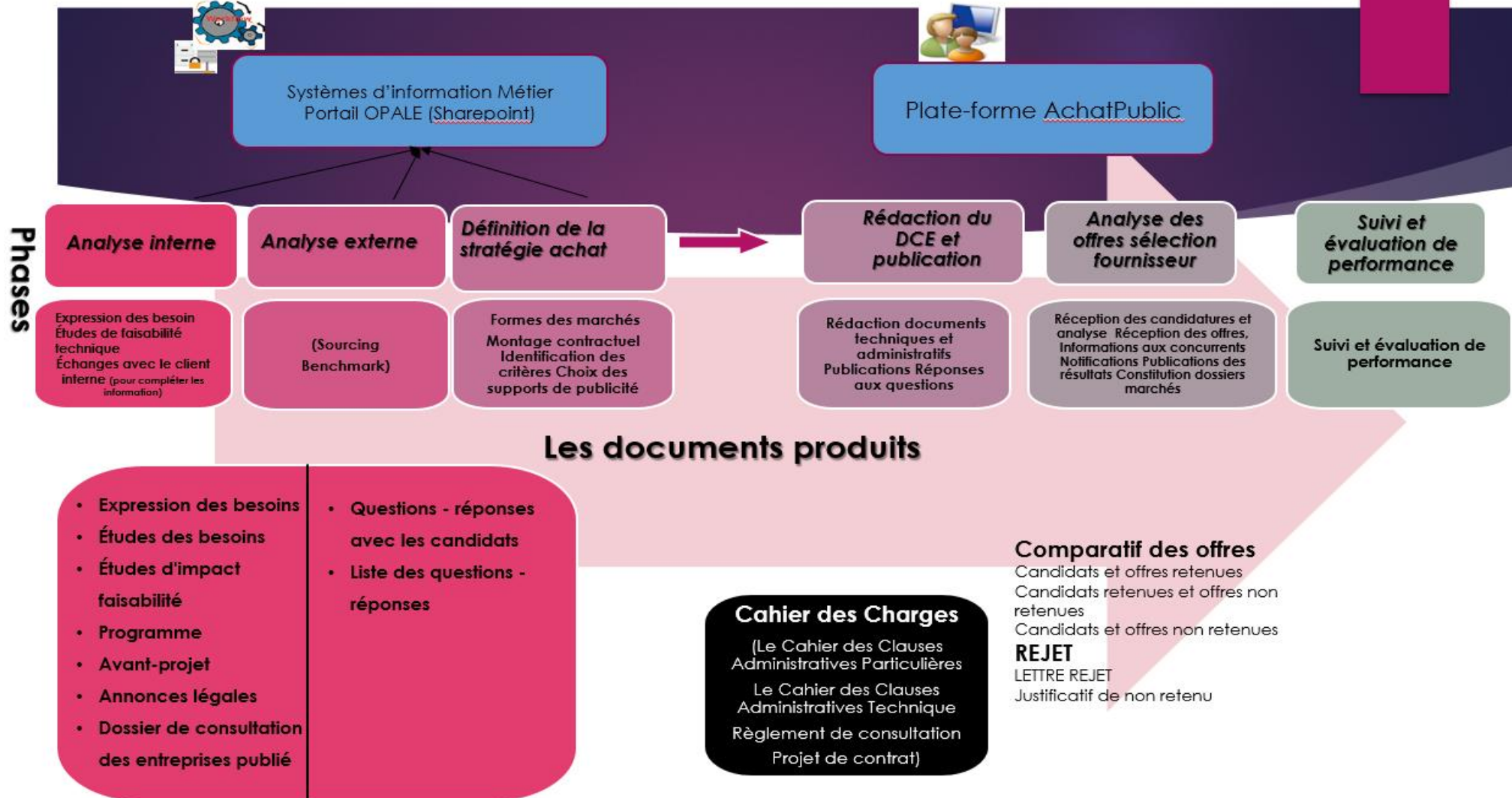


Figure 11 : liste des documents produits dans le processus Achat BRGM

Diagnostic du système informatique existant pour les Achats

Cette partie présente le diagnostic des système d'information actuel du BRGM. C'est une cartographie applicative existante qui présente des différents composants applicatifs et matériels du SI du BRGM servant à la création et la sauvegarde des données on a concentré notre diagnostic à l'unité Achat des Marchés publics parce qu'ayant des priorités plus élevé pour la conservation de leurs données à valeur probant. Voici une présentation de la liste des applications utilisées dans le cadre des activités liées au processus achats :

- Opale édite la proposition d'achat et saisi le visa technique (FE019, FE039)
- Oracle Enclenche le workflow de validation de la commande et valide l'engagement juridique (**approbateur**), certifie le service fait ; ordonnance la dépense (**ordonnateur**) et fait le control de cohérence (**service facturier**)
- BRGMachat.fr permet d'effectuer une proposition d'achat en cas de bug ou de non-fonctionnement d'OPALE.
- Chorus portail Pro pour éditer la commande d'achat lancer l'appel d'offre et analyser les offres.

ii. Description de l'architecture logicielle existante

Il s'agit de décrire les applications utilisées dans le processus achat de BRGM.

Fiche descriptive d'OPALE – Portail utilisateur

Description métier	
Processus concerné	Achat
Entités organisationnelles utilisatrices	Toutes les directions du BRGM
Acteurs	Tous les agents utilisateurs d'OPALE
Objectifs	Permettre l'édition de la demande d'achat par le personnel qui l'envoi à l'unité achat et leurs approbateurs respectif pour validation.

Classification (Criticité, importance, utilité)	Application importante dans le cadre du suivi de l'évolution des besoins des usagers. Les formulaires doivent être actualisés en permanence. Par contre son arrêt n'est pas critique
Entrées	Besoins du salarié
Sorties	Demande d'achat
Fonctions	« Éditer la demande d'achat » utilisée par les salariés qui saisissent les biens à commander dans le formulaire ; « Traiter la demande d'achat », utilisée par le personnel du service achats pour compléter les prix des biens fournis par les fournisseurs ; « Valider la demande », utilisée par les approbateur pour valider la demande d'achat saisie dans le formulaire ;
Classes concepts générées	Création de la demande, Modification, Suppression, Visualisation Approbateur : Modification /Visualisation Personnel service Achats : Modification /Visualisation



Description applicative et technique	
Type d'application	GED, background ORACLE
Année de développement	Version 2015
Volumes traités	20 demandes par jour
Disponibilité	5 jours sur 7, de 8h30 à 19h
Période critique ou de pointe	10h à 14h30
Perception de l'application par les Utilisateurs du point de vue fiabilité	????
Lieux de stockage	Serveur BRGM
Premiers éléments de diagnostic	
Points forts de l'application	Application disponible pour les utilisateurs avec des droits d'accès bien précisés. Répartition hiérarchique des rôles (approbateur, ordonnateur, et acheteur).
Points faibles de l'application, problèmes	Impossibilité de sauvegarder une proposition d'achat initié sur Opale réel problème pour la constitution du dossier à archiver. Les agents de l'unité Achat ne savent pas où parte les propositions d'achat lorsqu'ils sont validés Processus d'alerte et de Workflow pour la validation des demandes. Processus d'archivage électronique non défini
Nouveaux besoins	Besoin d'un dispositif d'archivage électronique à valeur Probante pour l'archivage central des commandes au BRGM

Fiche descriptive Achat public

Description métier	
Processus concerné	Commande publique
Entités Organisationnelles utilisatrices	Unité Achat
Acteurs	Personnel de l'unité achats
Objectifs	achatpublic.com propose une offre complète de solutions couvrant toute la chaîne de l'achat public : du sourcing à l'archivage en passant par la rédaction des pièces de marché, le profil acheteur, la gestion des procédures et le suivi des actes.
Classification (Criticité, importance, utilité)	Site important dans le cadre du pilotage des activités liés au marché public des administrations publiques, les EPIC et collectivité territoriale
Entrées	Demande d'achat validée par le chef de service
Sorties	Commande
Fonctions	« Éditer la commande d'achat » utilisée par le personnel du service achats ; « Modifier la commande d'achat » utilisée par le personnel du service achats ; « Supprimer la commande d'achat » utilisée par le personnel du service achats ;
Classes concepts générées	Personnel du service achats : Création de la commande / Modification, Suppression, Visualisation
Description applicative et technique	
Type d'application	Site internet
Année de développement	Version 2017

Perception de l'application par les utilisateurs du point de vue Fiabilité	Site internet perçue comme sûre et fiable avec un niveau de performance acceptable
Type de sites concernés	Site de l'État
Premiers éléments de diagnostic	
Points forts de l'application	<p>Application disponible pour les utilisateurs du service achats avec leur profil acheteur.</p> <p>Tous les logiciels de marchés publics sont accessibles via le site internet sans aucune installation à réaliser sur le poste de l'utilisateur.</p> <p>Processus d'alerte et de Workflow pour informer le salarié de l'envoi de la commande.</p>
Points faibles de l'application, problèmes	<p>Risque d'indisponibilité des données</p> <p>Impossibilité de garantir l'intégrité, l'authenticité et la pérennité des données</p> <p>Externalisation des données à valeurs probante pour archivage</p>
Nouveaux besoins	Besoin d'un dispositif d'archivage électronique à valeur probante pour l'archivage central des commandes au BRGM

Description métier	
Processus concerné	Commande publique
Entités Organisationnelles utilisatrices	Unité Achat et service facturier
Acteurs	Personnel de l'unité achats
Objectifs	Permet de recevoir toutes les factures des fournisseurs puis de les transmettre à l'Administration destinataire correspondante afin qu'elles soient traitées ; et offre toutes les fonctionnalités propres à un portail web gratuit pour l'émission, le téléchargement et la consultation des différents statuts du cycle de vie des factures dématérialisées.
Classification (Criticité, importance, utilité)	CHORUS est le nouveau système d'information budgétaire et comptable de l'État. Site important dans le cadre du pilotage des activités liés au marché public des administrations publiques, les EPIC et collectivité territoriale
Entrées	Demande ou saisie de facture XY
Sorties	Consultation ou réception de facture XY
Fonctions	« Factures Émises » utilisée pour le dépôt ou la saisie de facture ; « Facture à valider » utilisée par le personnel du SFACT après rapprochement ; « Suppression ou archivage des facture » utilisée par le personnel du service achats ;
Classes concepts générées	Personnel du service achats : Création de la commande / Modification, Suppression, Visualisation

Points forts de l'application	Site à disposition permanent pour les unités achats et service facturier. Création de facture fiable et intègre avec les métadonnées nécessaires pour la pérennisation des données.
Points faibles de l'application, problèmes	3. Saturation et panne des serveurs

Fiche descriptive CHORUS Portail pro

Résultat du diagnostic des besoins

La cartographie applicative du BRGM montre que l'architecture applicative existante repose essentiellement sur les applications standards qui sert à piloter l'ensemble des activités des directions.

Les documents issus de l'application Opale sont souvent stockés dans un dossier bureautique sur un serveur de fichiers. On constate dès lors que le système de GED actuelle ne permet pas de garantir la valeur probante des documents stockés.

Voici à présent :

Les points forts du système applicatif existant :

- Architecture applicative bien maîtrisée par l'ensemble des utilisateurs
- Architecture applicative du SI existant en partie urbanisée
- Bonne sécurité des données malgré les difficultés de retrouver les informations

Les points faibles du système applicatif existant :

- Altération des documents et données dans le temps
- Risque d'indisponibilité des données
- Impossibilité de garantir l'intégrité, l'authenticité et la pérennité des données
- Externalisation des données à valeurs probantes pour archivage

Ces avantages et inconvénients identifié lors de l'analyse doivent lors de la conception du SAE être au cœur des enjeux. L'objectif est de garder les points fort afin de continuer à maintenir leurs avantages et essayé de couvrir tout ce qui n'est pas couvert par ces applications de GED du processus Achats par le SAE.

4. La politique d'archivage

Une politique d'archivage est une déclaration cosignée par l'archiviste d'une part et la direction générale de l'entreprise d'autre part englobant les critères juridiques, fonctionnelles, techniques et sécuritaire auxquelles un système d'archivage électronique doit répondre pour être conforme aux exigences réglementaires de gouvernance et répondre au besoin l'entreprise.

Elle est rédigée à l'attention de l'ensemble des services du BRGM. C'est un outil de référence pour l'archivage du versement, à la, communication, la politique d'archivage définit les rôles et responsabilité des acteurs et intervenants dans la mise en œuvre du SAE, tout en rappelant l'environnement réglementaire qui la sous-tend. C'est un engagement entre les producteurs et le SG/GDA, pour une méthodologie de gestion des données rationaliser sur le long terme.

Avantage d'une politique d'archivage

- Un soutien officiel de la direction générale du BRGM Cosignataire de la politique d'archivage
- Un contenu structuré par les règle établies et approuvé par les autorités compétentes
- Un outil de construction et d'évolutions de la fonction archive

Un exemple de politique d'archivage destiné à l'unité Achat est rédigé et peut être édité pour une réutilisation pour les autres services

Troisième partie : La qualité des données et des métadonnées dans le SAE du BRGM

IV. La qualité des données et des métadonnées

1. Donnée Information Connaissance

A la rentrée de septembre 2015 alors que je découvrais l'université en première année de licence pro en science de l'information documentaire. Dès la première semaine de cours, tous les intervenants avaient fait le tour de la question de la différence entre donnée information et connaissance. Ces trois concepts, bases de la théorie de l'information se concorde et les spécialistes du domaine ne s'accordent pas tous sur leurs définitions [21].

Dans ce mémoire, on retiendra les définitions suivantes :

- Données : « les données sont des faits et des statistiques qui peuvent être quantifiées, mesurées, comptées, et stockées. » (Dr Donald Hawkins, technologie de l'information).

Selon Wikipédia « une donnée est une description élémentaire d'une réalité. C'est par exemple une observation ou une mesure ». Une donnée sert de point de départ pour une recherche, un élément sur lequel on peut fonder un raisonnement
Une donnée est un résultat d'observation ou d'expériences (Srce : Larousse)

- Information et connaissance : « une information est un ensemble de données organisées selon une ontologie qui définit les relations entre certains sujets. L'information peut être communiquée. La connaissance est un ensemble de structures conceptuelles présentes dans le cerveau humain et imparfaitement représentées par des informations pouvant être communiquées.»[22]

Cette dernière définition est la parfaite illustration de l'acheminement d'une donnée vers la connaissance. C'est le cycle de vie de l'information. L'Homme reçoit et manipule des données à l'aide des progiciels. En effet, Une fois collectés ou capturés, les acteurs traiteront les données via plusieurs systèmes d'information métier (SIRH OPALE ou SIHA). Cela implique l'ajout de métadonnées et d'autre données. La finalité de ce traitement est l'obtention d'une donnée contextualisée : Information.

Cette information va être confronté à d'autre processus pour son interprétation via les systèmes d'information pour devenir une connaissance. Ainsi les acteurs en charge de ces données avec cette nouvelle connaissance font le lien avec d'autres Connaissances déjà acquises pour obtenir ainsi des Compétences qu'ils pourront mettre à disposition de l'entreprise de manière répétée dans le temps.

Le processus logique unissant nos trois notions peut être représentée sous la forme de ce schéma ci-dessous.

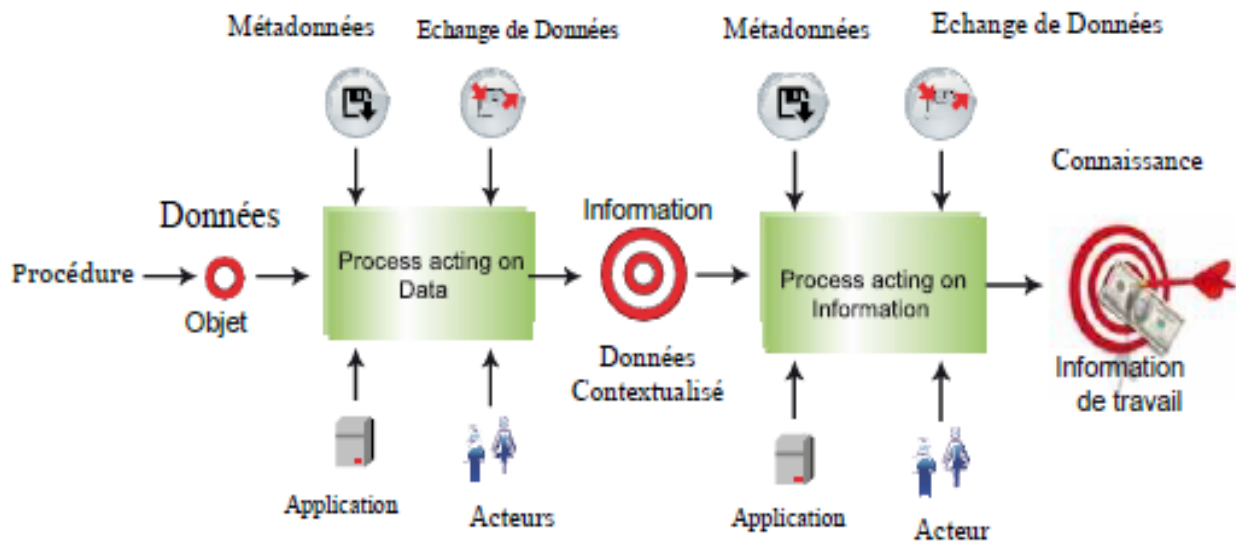


Figure 12 : Processus d'acheminement d'une donnée vers la connaissance

2. Les données au BRGM

a) Différents types de données et différents usages

Au BRGM les données scientifiques sont créées ou reçues soit par les chercheurs, les équipes de laboratoire, géophysiciens, etc. Ces données sont issues de divers environnements et de systèmes d'information. Chaque type de données diffère dans la manière dont il est créé, stocké et analysé. Pendant notre audit documentaire les producteurs questionnés sur les données et leurs qualités dans leurs services ont fait ressortir trois typologies de données à savoir les données structurées/semi-structurées et non-structurées. Mais aussi les Données de référence et données opérationnelles et enfin les données observées et données calculées

b) Distinction des typologies de données

Données structurées, semi-structurées et non-structurées

Les données sont structurées si on peut les stocker dans des champs de bases de données. Il concerne toutes les données pouvant être stockées dans une base SQL ou une table avec des lignes et des colonnes. Ils ont une clé relationnelle et constituent aujourd'hui l'un des moyens les plus simples de gérer les informations.

Les données semi-structurées sont les données ne résidant pas dans une base de données relationnelle, mais qui possèdent des propriétés organisationnelles facilitant leur analyse. Exemple les données XML.

Les données non structurées, enfin, sont définies par Bill Inmon comme « Tout document, fichier, image, rapport, formulaire, etc. qui n'a pas de structure standard définie qui permettrait de le stocker facilement dans un dispositif de traitement automatisé. Il ne peut pas être défini en termes de lignes et de colonnes ou d'enregistrements. (...) Les données non structurées sont les e-mails, les feuilles de calcul, les documents, etc. Certaines des informations les plus précieuses de l'entreprise résident dans ses données non structurées. » [23] Elles sont de plus en plus présentes dans les systèmes informatiques et sont utilisées par les services dans diverses applications. Exemple : **Word PDF**

Données de référence et données opérationnelles

« Les données de référence, aussi appelées données maîtres ou en anglais master data, sont des données partagées par l'ensemble des processus qui soutiennent l'activité courante d'une entreprise (ou plus généralement d'une organisation) et ses prises de décision »²⁴ Au BRGM

comme données de référence on peut citer entre autres les référentiels contrats et de partenariats, référentiels clients.

Les données de ces référentiels sont relativement stables : le nombre de clients, les contrats de partenariat sont des données assez stables sur une période de temps donnée.

Contrairement aux données de référence, les données opérationnelles sont liées aux activités journalières de l'entreprise et sont plus éphémère. Les données opérationnelles évoluent vite, et s'apparente à des événements ponctuels par exemple un marché d'achat de fourniture passé un jour tel par tel agent. Au BRGM il peut s'agir des données de prospection prélevée sur une zone donnée, passation de commande aux Achat, etc.

Précisons que les données de référence peuvent-être des données structurées, semi-structurées ou non-structurées.

3. Les enjeux et le défi scientifique de la DATA pour le BRGM

Comme défini dans son programme collecter, interconnecter et valoriser les données et la connaissance environnementale au travers de services Numériques le BRGM place entrevoie le développement de sa culture DATA et place les données au cœur de son activité.

- **Description de l'enjeu sociétal**

- L'accès aux données géoscientifiques et environnementales est de plus en plus critique pour répondre aux enjeux de l'environnement, du changement climatique, des ressources, des risques
- La demande de « casser les silos » pour traiter les problèmes complexes
- Le secteur est en plein bouleversement : open data, open science, production exponentielle de données, montée en puissance de « nouveaux » acteurs (académiques, entreprises, citoyens)

- **L'enjeu pour le BRGM**

- Dans sa mission d'établissement de recherche et de service géologique, il occupe une position centrale. Il s'agit de :
 - Garantir l'accès, la fiabilité et la pérennisation des données

- Asseoir sa position sur son expertise des données

- **Le défi scientifique pour le BRGM**

- Les données produites par la science : Développer des référentiels de données et connaissances des géosciences pour répondre à la diversité des enjeux et des acteurs

4. Qualité des données

Dans un contexte où les défis des organisations sont divers et croissants avec une contrainte de collecte de plus en plus importante, la possession de données de qualité devient une nécessité incontournable. Ces données sont un patrimoine de la génération numérique.

L'explosion des volumes de données dans les administrations, a suscité une grande attention pour la conservation des données et de leurs qualités.

Bien que la sauvegarde apporte une sécurité actuelle des données, elle reste insuffisante pour la garantie de l'intégrité, l'authenticité et la traçabilité et donc garanti pas la qualité des données dans le temps. Garantir la qualité des données dans le temps exige une démarche cohérente et ordonnée afin de prévenir les risques lié ce défi.

Qu'est-ce qui confère aux données leur qualité

Tout d'abord, il faut savoir qu'une donnée ne dispose pas une qualité intrinsèque c'est-à-dire prendre une caractéristique bien identifiée et le définir comme mesure de qualité universelle pour toutes les données. Parce que la qualité d'une donnée peut être évaluée que quand on sait ce que l'on veut en faire : Quel est l'objectif final de la collecte de cette donnée ? Quel traitement lui est réservé ? Pour quelle finalité ? Autrement dit, c'est l'usage qu'on a prévu qui détermine la qualité des données. Ceci dit, au BRGM l'usage des données sont diverse et variées selon les typologies de données. On peut distinguer deux grandes familles de données au BRGM : Les données Scientifique et les données administratives.

Les données administratives sont tous les données relevant du fonctionnement de l'administration du BRGM c'est essentiellement les données produit par les fonctions supports alors que les données scientifique relève de l'activité scientifique du BRGM à savoir les activités géophysiques, les activités de prospections, la météorologie, etc.

Comme nous l'avons défini au début de ce chapitre la qualité des données fait référence à la conformité des données aux usages prévus. Elle désigne l'aptitude de l'ensemble des critères ou caractéristiques des données à savoir la fraîcheur, la disponibilité, l'intégrité, la fiabilité, la

traçabilité, la sécurité, et l'exhaustivité. Le respect de ces critères permet une utilisation des données en toutes confiances.

On peut répartir ces différents critères en fonction des usages prévues que sont l'usage légal qui concerne souvent les données administratives et l'usage scientifique pour les données de la science. Pour l'usage légal on va s'intéresser à la qualité des processus de traitement des données alors que pour l'usage scientifique on parlera de la qualité de la valeur des données

a) La Qualité des processus de traitement des données

L'information produite, traitée, sauvegardée, partagée et diffusée par l'administration du BRGM est souvent une information sous forme numérique. La question de sa préservation se pose naturellement pour assurer les obligations légales de l'administration. Parmi les critères de qualité des processus de traitement des données on peut citer :

- La pérennité
- L'intégrité ;
- L'authenticité ;
- La traçabilité ;
- La confidentialité ;
- La disponibilité

La pérennité

Garantir la pérennité de données exige une démarche rigoureuse afin de prévenir les risques d'obsolescence. Pour assurer la pérennité, les données doivent être conservées dans un format standard du marché pour pouvoir en restituer le contenu informationnel. Il s'agit de favoriser la conservation des données dans des formats bien maîtrisés, dont les spécifications sont publiées (on peut prendre l'exemple de PNG - ISO 15948 pour les images) et normalisées si possible (comme le PDF/A normé avec l'ISO 32000-1), mais aussi et surtout les formats fortement utilisés par les producteurs d'information.

Pour la lisibilité et la convertibilité dans le temps, les fichiers doivent respecter les caractérisations de leur format. Dans le domaine de l'archivage électronique il existe des outils libres comme JHOVE et DROID, ainsi que des bases de données bien définies consacrées aux

formats on peut citer dans ce sens la base de données PRONOM qui pour la pérennité permet l'identification, l'approbation et la détermination des fichiers assurant ainsi leur éligibilité à une conservation à long terme. Dans le contexte d'un SAE la pérennité est assurée par la migration des formats défini par une politique de migration des formats. La migration peut se faire en amont vers un format pérenne avant le transfert des données dans le système. Ceci doit être spécifié dans la politique d'archivage en déterminant les formats acceptés ou préféré. Mais le cas de figure de pérennisation le plus fréquent est la migration des formats de conservation au cours du temps.

L'intégrité

Notion qui assure que les données n'ont pas été modifié ou altéré depuis leur création. Cette notion englobe à la fois la stabilité du contenu dans le temps et la traçabilité des tâches réalisées sur les données (édition, élimination). L'intégrité garantit que les modifications apportées aux données et aux métadonnées sont documentées et tracées (, auteur de la modification, qualité, motif, etc.).

Le maintien de l'intégrité des données nécessite de présenter la preuve que les données n'ont pas été endommagé intentionnellement ou non. Ceci passe par la non-modification du contenu et de la forme initiale. En cas de modification il faut pouvoir garantir la vérification de l'identité de l'auteur mais aussi de toute personnes ayant contribué à la création du document. En d'autres termes on s'assure que les données ne soient pas altérées.

L'authenticité

L'authenticité couvre le degré de fiabilité des données originales déposées et de leur provenance, y compris la relation entre les données originales et les données diffusées, et si les relations existant entre les jeux de données et/ou les métadonnées sont effectivement maintenues. Il s'agit d'assurer la non-modification du contenu et de la forme d'origine des données. La mise en place d'une Signature électronique lors de la création des données et leur signature en amont du versement et la signature électronique après le versement facilitera l'authentification.

La traçabilité

Les enjeux liés à la traçabilité sont divers et variées surtout dans le domaine de la gouvernance de l'information. La traçabilité renseigne le détail des évènements qu'a connus les données dans un journal. Elle fournit la description de toutes les opérations effectuées (consultation, migration, droit, etc.). Pendant toute la durée de conservation des données, les opérations effectuées sur elle sont enregistrées dans le journal des événements celui-ci est horodaté

signé et conservé avec les données dans le système. Les tentatives ou demandes d'accès aux données conservées doivent aussi être identifiées et datées.

Cette traçabilité permet aux détenteurs de données d'éviter les mauvaises surprises en cas de contentieux légaux, mais également assouvir les demandes des usagers. En effet, la traçabilité assure le traçage des opérations effectuées sur les données de sa capture sa diffusion en passant par le traitement. Elle répond à une obligation légale et ou réglementaire, pour des raisons liées à la sécurité des données ou à la protection de l'entreprise. La traçabilité peut prendre quatre formes :

- ❑ Une traçabilité fonctionnelle
- ❑ Une traçabilité technique
- ❑ Une traçabilité normative
- ❑ Une traçabilité applicative

La confidentialité

La confidentialité des données est une nécessité présente pendant tout le cycle de leur exploitation mais aussi de leurs archivages. Il s'agit d'abord d'assurer la conservation des données mais aussi de s'assurer que seules les personnes habilitées ou autorisées puissent avoir accès aux données. C'est la garantie d'une diffusion suivant la législation en vigueur. Elles doivent être protégées de toute intrusion. Elle recouvre la notion d'habilitation : autorisation d'accéder aux données et est assurée par une bonne gouvernance des données. Cette fonction est primordiale pour le bon fonctionnement de l'administration.

La disponibilité

C'est la garantie du fonctionnement normal du système d'information. Cela signifie que le personnel autorisé peut accéder aux données dans le créneau horaire prévu avec un temps de réponse approprié.

En effet, la disponibilité est la caractéristique à laquelle un usager est autorisé à utiliser des données au moment et à l'endroit indiqués. Cette disponibilité dépend souvent de certaines conditions, telles que : L'horaire de disponibilité (pendant les heures de bureau ou uniquement) ou le lieu de consultation (consultation sur place ou dans un environnement sûr). C'est ces mêmes conditions qui déterminent le niveau d'indisponibilité qui peuvent être tolérables ou pas tolérables selon le type de données. Ainsi, plus les informations sont « essentielles » pour la prise de décision ou l'action, plus les conditions fournies sont précises et contrôlables.

b) Qualité de la valeur des données l'exemple de la BSS

« Les données scientifiques sont un trésor à partager ».

Selon l'OCDE, les données scientifiques sont « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche. Ce terme ne s'applique pas aux éléments suivants : carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris).»²⁵

Au BRGM les données sont collectées et stockées dans la base de données BSS banque du sous-sol. La BSS est une base des bases de données du BRGM. La BSS est la base nationale qui conserve toutes les données sur les ouvrages souterrains du territoire Français.

La mission du référentiel à l'égard de sa communauté est de conserver toutes les données sur les ouvrages souterrains (forages, sondages, puits et sources) du territoire sont collectées pour être conservées dans la base de données, la BSS, organisée et gérée par le BRGM.

Elle contient

Plus de 2 millions de documents numérisés

Public plus de 800 000 descriptions d'ouvrages

Aujourd'hui, la qualité des données scientifiques au BRGM présente un défi socio-économique non négligeable tant sur la productivité que sur l'innovation. Face au défi de la conservation des données scientifique les enjeux dont le SAE BRGM doit apporter une réponse seront :

- L'identification des besoins pour la communauté scientifique afin de proposer un service utile.
- Proposer un poste de travail et un outil mieux structuré et facilement utilisable pour attirer les utilisateurs.
- Garantir une conservation à moyen et long terme

La banque du sous-sol est composée de deux catégories de données d'une part les données archivées qui ont été finalisés classer et d'autre part les données hyperactives qui bouge quotidiennement. Cette dernière pose un certain nombre de problèmes sur la capture et le transfert de ces données pour l'archivage. En effet, ces données étant changeante parce que produit et éditer quotidiennement, à cela s'ajoute l'important volume de ces données (deux

millions de documents numérisés) et une documentation incomplète des éléments de la base de données parce que produit par des géoscientifiques pas assez sensibilisés sur la question de la qualité des données et des métadonnées. Ces manquements sont fortement ressentis au niveau de la description des données importé dans la base de données. Ces errements sur les métadonnées sont la cause de la difficulté à retrouver ces données mal décrit qui peut aller jusqu'à la perte définitive de ces données. Pour la mise à disposition des données conservées dans la Banque du sous-sol on distingue deux portails pour la communication :

- Intraterre pour la communication en interne des données de la BSS dédié aux agents BRGM et aux entreprises partenaires
- Infoterre pour la communication en externe qui mets à disposition les données de la BSS via le site internet institutionnel du BRGM de manière à ce que l'ensemble des données soient accessibles au public.

Pour pouvoir élargir le périmètre du SAE sur cette partie hyperactif de la BSS, la banque de données doit être en mesure de répondre à certaines exigences parmi elle, les questions liées à la

- Documentation des données versées
- Volumétrie de données important
- Présence d'objets non figée
- Base de données hyperactif
- Dissociation des portails infoterre intraterre

Le schéma ci-dessous montre les différents scénarios élaborés pour l'élargissement du SAE au sein de la banque du sous-sol. Celui-ci pose deux hypothèses. La première est la création d'un entrepôt de données au sein de la BSS et appliquer le SAE à cet entrepôt de données qui prendra en charge que des données fixe communicable et très bien documentées. Cette hypothèse nous éviterait les difficultés liées à la présence de données non figées et donc qui peuvent changer quotidiennement. Et la seconde hypothèse est celle où l'on applique le SAE sur l'ensemble de la BSS. Celle-ci permettrait de prendre en charge tous les données présentes dans la banque de données mais avec des contraintes non négligeable de coût mais aussi surtout de gouvernance de la qualité des données prises en charges.

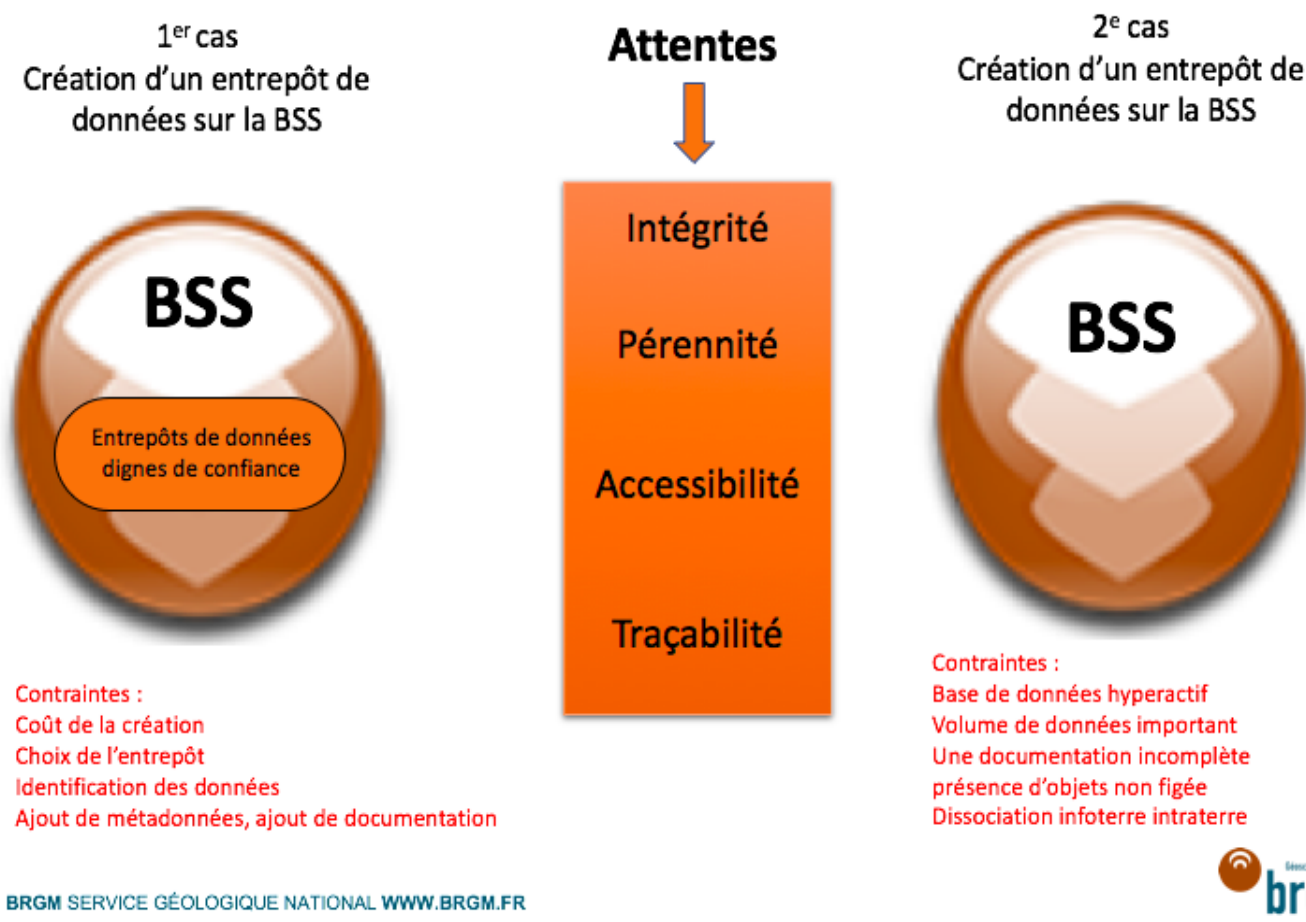


Figure 13 : Périmètre d'application du SAE dans la banque du sous-sol

Les exigences sur l'aspect qualitatif de la valeur des données scientifiques citées plus haut doivent être précédé par le respect des critères intrinsèque lié à la qualité du contenu des données. Dans cette partie nous allons aussi lister et expliquer chacun de ces critères.

Fraîcheur

La fraîcheur est l'ensemble des facteurs déterminant le caractère actuel et récent d'une donnée entre l'instant de sa création, sa collecte ou sa capture et le moment où elle est présentée à l'utilisateur. Ce critère est capital pour les données météorologiques pour leur caractère temporel éphémère.

La fraîcheur est importante dans deux cas :

- Le temps qui se passe entre la capture ou la collecte d'une donnée et son analyse,
- Le temps qui se passe entre le reporting sur une donnée et l'amélioration et le perfectionnement qui en découle.

La cohérence

La cohérence vise à assurer que les valeurs données à un ensemble ou jeux de données sont compatibles avec les valeurs comprises dans un autre ensemble de données. La cohérence peut être définie entre un ensemble de valeurs d'attribut et un autre ensemble d'attributs dans le même enregistrement (la cohérence au niveau des enregistrements), entre un ensemble de valeurs d'attributs et un autre attribut figurant dans les registres différents (cohérence entre enregistrements), ou entre un ensemble de valeurs d'attribut et l'attribut du même ensemble pour un même enregistrement à différents points dans le temps (cohérence temporelle). [26]

Exactitude

L'exactitude est liée à la cohérence des informations contenues ou dérivées des données par rapport à la réalité qu'il doit mesurer. C'est une mesure capitale de qualité pour une donnée parce que son non-respect a tendance à fausser toute analyse prévue des jeux de données en question. Son observation sur les données brutes géologiques doit permettre de faire la concordance entre ces données brutes et une source de donnée de référence. On pourra ainsi attester ou pas que les données sont porteuses de la bonne information.

La complétude

La complétude des données est l'attribut qui indique les données ont toujours une valeur comprise dans un certain ensemble de données. Autrement dit si les données sont suffisamment volumineuses. Selon Thomas Redman pour la complétude d'un ensemble de données « *Chaque item de données nécessaires doit y être présent* » [27] l'absence d'une ou d'un ensemble de pièce ou d'item pourrait compromettre la complétude des données.

Le volume d'informations scientifiques ne cesse de croître et en l'absence d'actions spécifiques, une part conséquente risque d'être définitivement perdue. Le respect de ces critères qui assure la fiabilité du contenu des données reste primordial.

Ceci ouvre des perspectives pour la publication et la consultation des données scientifique d'où cette forte tendance de la communauté scientifique pour l'Open Data.

5. La qualité des métadonnées

« *Celui qui contrôle les métadonnées contrôle le web* » [28]

a) Qu'est-ce que les métadonnées

En partant de son préfixe "méta" qui vient étymologiquement du grec "méta" indiquant l'autoréférence ou ce qui englobe, le mot signifie donc proprement « donnée de ou à propos

de donnée ») On peut littéralement en déduire qu'une métadonnée est une donnée sur une donnée. Elles décrivent le contexte, le contenu, la structure des documents ainsi que leur gestion dans le temps permettant ainsi de préserver des informations qui décrivent des objets numériques : il s'agit d'informations de pérennisation (description du corps de l'information, de sa source, et son historique) et d'informations de représentation (aspect technique de la donnée, de l'architecture, la forme, et les droits d'accès des données).

Au regard de tous ces éléments on peut dire que les métadonnées sont un ensemble structuré d'informations techniques, de gestion et de description attachées à un document servant à décrire les caractéristiques de ce document en vue de faciliter son repérage, sa gestion, son usage ou sa préservation.

L'objectif principal est de décrire les données de manière à ce qu'elles soient utilisables par l'ensemble de ses usagers en créant une cartographie des données précise et compréhensible. On peut faciliter cette description en optant pour le choix de standards et de norme de métadonnée de métadonnées : en effet, les métadonnées standardisées sont décrites dans un référentiel, qui peut facilement être utilisé comme base de vérification pour garantir la qualité.

b) Typologie des métadonnées

Les métadonnées ne doivent pas être vu uniquement que de par leurs aspects descriptifs. En effet à côté des métadonnées descriptives existe d'autre typologies de métadonnées comme les métadonnées de gestion. Parmi elle on peut citer les métadonnées administratives (cycle de vie d'un document dans son contexte de production, et de conservation), les métadonnées juridiques et les métadonnées de sécurité. Outres que ces métadonnées on peut citer Ils décrivent l'origine des données et précisent comment elles sont traitées.

Les métadonnées de provenance

Les métadonnées de provenance sont nécessaires pour la compréhension, et l'usage des données. Elles décrivent l'origine des données et précisent comment elles sont traitées en donnent les informations telles que

- D'où viennent les données ?
- Pourquoi la collecte ou la capture données ?
- Qui les a collectés ?
- Quels outils ont été utilisés pour la capture des données ?
- Comment traiter ces données ?

Les métadonnées droits et accès

Les métadonnées des droits et accès indiquent les droits d'accès, les licences et Règles d'utilisation des ressources.

- Comment accède-t-on aux données ?
- Ceux qui ont le droit d'afficher ou de modifier des données ou des métadonnées, et sous quelles conditions ?
- Qui a le droit de gérer les données ?
- Ya - t-il des frais associés à la consultation ?

Les métadonnées de citation

Elles sont nécessaires pour une personne de citer correctement les données elle est composée des champs suivants :

- Le Titre
- L'auteur ;
- Éditeur ;
- L'année de publication ;
- Identifiant, etc.

Type de métadonnées	Exemples de propriétés	Usages primordiaux
Métadonnées descriptives	Titre Auteur Sujet Genre Date de Publication	Découverte Affichage Interopérabilité
Métadonnées techniques	Type de fichier Taille du fichier Date de création Schéma de compression	Interopérabilité Gestion d'objets digitaux Conservation
Métadonnées de conservation	Somme de contrôle (<i>checksum</i>) Évènement relatif à la conservation	Interopérabilité Gestion d'objets digitaux Conservation
Métadonnées juridiques	Droits d'auteur Termes du contrat de licence Détenion des droits	Interopérabilité Gestion d'objets digitaux
Métadonnées structurelles	Séquence Place dans la hiérarchie	Navigation
Langages de balisage	Paragraphe En-tête Liste Nom Date	Navigation Interopérabilité

Tableau 5 : des typologies de métadonnées et leurs finalités

c) Quels usages pour quelles métadonnées pour le SAE

Les métadonnées descriptives se réfèrent au standard Dublin Core ; les autres métadonnées, qualifiées de métadonnées de gestion, servent à décrire : le contexte administratif souvent prédéfini par les producteurs de documents. (Intitulé, Objet du document, direction, service, etc.), les métadonnées juridiques pour les données propriétaire ou pour les droits d'auteur, et les modalités de conservation (format accepté par le SAE, le standard d'encodage, es caractères...). Le tout est embrumé en XSD conformément à la norme METS qui permet de placer des métadonnées à des fins diverses (telles que EAD, Dublin core et même les métadonnées PREMIS) dans différentes catégories. Ces métadonnées, spécifiques et obéissant à des normes et standards informatiques répandus, seront utiles à la conservation et la diffusion des données qu'il s'agisse d'une diffusion au sein du BRGM ou pour les citoyens.

Les bénéfices de la métadonnée

La partie définition nous apprend que grâce aux métadonnées, la finesse cartographique des données des de replacer les informations dans leur contexte et offre de nombreux avantages telle que le gain de temps, la gestion de la conformité, et la qualité.

Les métadonnées présentent de nombreux avantages, notamment :

- L'accès à l'information pour des utilisateurs sans connaissances techniques. Les données créent de la valeur lorsqu'elles sont utilisées professionnellement à des fins commerciales. Grâce au contexte détaillé, chacun peut utiliser un moteur de recherche simple pour indexer la structure, le contenu, la qualité et la nature de chaque donnée, accédant ainsi facilement aux informations requises.
- La qualité des données. Il est plus facile d'évaluer la qualité des données. Une fois qualifiés, il n'est plus nécessaire de filtrer les données pour voir si elles sont utiles, opportunes et pertinentes. Les métadonnées font le travail
- Gain de temps. En fournissant un fichier de configuration complet et détaillé pour chaque donnée, l'utilisateur consacre son temps à l'exploitation plutôt qu'à l'évaluation des informations.
- Les métadonnées permettent de protéger les données sensibles et confidentielles. Le RGPD pose les base d'un nouveau cadre pour l'utilisation des données personnelles. La cartographie des données personnelles et des données sensibles assure cette protection avec opérations de cryptage et d'habilitation et de gestion des accès.
- Une exploitation et une collaboration facilitées. La traçabilité fournie par les métadonnées apporte de la transparence au traitement effectué. Cette connaissance rend les utilisateurs plus confiant et rassurés lors de l'exploitation des données. En effet la traçabilité des opérations conservé dans le journal des évènements garantie la maîtrise des impacts des actions effectuées sur les données.

En plus des informations de base, les métadonnées s'ils sont bien renseignées doivent permettre à une entreprise d'avoir une cartographie complète pour comprendre ses données et les méthodes de traitement associées. Plus précisément, les métadonnées doivent pouvoir répondre à la méthode du questionnement QOQCCP en français, à savoir :

- Qui a créé la donnée ? Qui l'utilise ? À qui appartient-elle ? Qui en assure le traitement et la maintenance ?
- Quelle en est la définition métier ? Quelles sont les règles de gestion ? Quel degré de sécurité ? Quelles en sont les dénominations standards au sein des bases de données ?

- Où est conservée la donnée ? D'où vient-la donnée sa provenance ? Où est- utilisée et partagée la donnée ? À quelle standard et norme réglementaire répond-elle ?
- Pourquoi avons décidé de conserver cette donnée ? Quel est l'usage prévu ?
- Quand a été créé la donnée ? Quand arrive-t-elle à son terme ?
- Comment est structurée la donnée ? Combien de fois est-elle répliquée ? Combien de fois est-elle présente ?

Les réponses à cet ensemble de questions apportent une valeur ajoutée significative aux données. Ils deviennent précis, compréhensibles pour tous les domaines de l'entreprise, accessibles et faciles à partager.

d) Modèles de métadonnées

Maintenant que nous avons posé les bases de la signification du concept de métadonnées c'est-à-dire la signification du vocabulaire en utilisant la logique théorique, nous présenterons les modèles de métadonnées. Il existe des modèles de métadonnées, proposés par des standards ou d'autres données élaborés par des institutions. C'est des listes de rubriques ou champs à renseigner pour documenter ou gérer les données destinées à l'archivage. L'un des modèles de métadonnées d'archives les plus complets est la norme européenne MoReq2 publiée en 2008. MoReq2 identifier 158 métadonnées mais tous ne sont pas obligatoires et certains sont exclusifs. En se limitant à l'essentiel et en automatisant autant que possible la capture de métadonnées, parmi les 158 métadonnées possibles, l'utilisateur n'a plus que 4 à 5 données à saisir.

Petit aperçu des types de métadonnées de MoReq2 :

- *Métadonnées de description : auteur, mot-clé*

- *Métadonnées de capture : date et heure de capture dans le système d'archivage, identifiant système, code de classement, type de document*


- *Métadonnées de dossier : identifiant dossier, date ouverture, date clôture, nombre de documents*

- *Métadonnées de classement : entité de rattachement, existence de copie, existence d'extrait*

- *Métadonnées de signature électronique : date et heure de vérification, algorithme de chiffrement, date et heure de déchiffrement*

e) Cadre normatif et standards couvrant les métadonnées

Les métadonnées sont organisées selon l'ordre d'importance des sujets traités, il est donc important de se référer sur une norme ou standard pour un gain de temps considérable et de qualité. Le but d'un standard ou d'une norme est de fournir un ensemble d'éléments caractéristiques pour la description de la production scientifique. Ainsi, la recherche peut être facilitée en se concentrant les typologies de métadonnée sur les critères définis. La description de l'élément lui-même peut être spécifiée en utilisant un vocabulaire dédié. La norme est

82 
La qualité des données et métadonnées dans l'implantation d'un système d'archives électronique

choisie en fonction de la finalité des données, conservation, diffusion, élimination, etc. Il se distingue par son vocabulaire, et peut être spécialisé par discipline c'est-à-dire une norme pour un métier ou une communauté spécifique, par type de données, etc.

La panoplie des jeux de métadonnées est très diversifiée, du plus générique pour décrire les ressources numériques (ex. Dublin Core - ISO 15836) aux plus spécifiques à un domaine précis comme les données géographiques et géologique (ISO 19115), en passant par les métadonnées plus tourné vers la pérennisation et les techniques de préservation (PREMIS, METS) ou encore les métadonnées administratives pour la propriété intellectuelle et les droits d'auteur (MPEG-21). On va à présent détailler les différentes normes qu'on vient de citer la partie suivante afin de connaître leur utilité et leur contexte d'utilisation.

i. Normes de métadonnées archivistiques

Pour la description des données versé dans notre SAE on se basera sur le langage de programmation XML pour les descriptions des objets conservés. Le XML est un langage de programmation informatique qui permet un échange électronique standardisé des descriptions archivistiques au moyen des DTD EAD (pour les descriptions de ressources archivistiques) et EAC (pour les descriptions de producteurs)

Parmi les standards de métadonnées axés sur le domaine des archives, on peut citer EAD DTD (*Encoded Archival Description ou Description Archivistique Encodée*), EAC-CPF⁶ qui permet de modéliser une notice ISAAR(CPF) et même Dublin Core. Tous les trois basés sur le Schéma XML.

ISAD-G (International Standard Archival Description-General)

Indépendamment du format ou du support du matériel d'archives, l'ISAD-(G)⁷ pose les bases de la description archivistique. Le modèle définit vingt-et-six éléments de description répartis dans sept zones d'informations on peut citer parmi elle l'identification, le contexte, le contenu ou les conditions d'accès.

Son objectif est d'autoriser la recherche et l'échange de données entre les services mais aussi les chercheurs sur le contenu et les intérêts des fonds d'archives afin de garantir la rédaction de descriptions compatibles, pertinentes et claires, en permettant l'utilisation pour rendre possible l'intégration de descriptions de différents lieux de stockage.

La première rubrique ou zone, qui est la plus importante, comprend l'identifiant, le titre qui fait un bref résumé du contenu, les dates extrêmes, le niveau de description choisi et des

⁶ (*Encoded Archival Context - Corporate Bodies, Persons and Families ou Contexte Archivistique Encodé - Collectivités, Personnes et Familles*)

⁷ Norme générale et internationale de description archivistique (International Standard Archival Description-General)

indications sur le volume de données ISAD(G) ne traite pas de la gestion matérielle, de l'indexation et des règles d'écriture et du vocabulaire. La norme ne définit pas les formats d'éditions.

EAD DTD

L'EAD est un standard de description basé sur le langage XML qui permet de structurer des descriptions de manuscrits ou de documents d'archives.

L'EAD DTD⁸ permet de structurer les descriptions des documents d'archives, et donc de standardiser les instruments de recherche des archives. Il fonctionne en corrélation avec la norme l'**ISAD(G)** en permettant de structurer les descriptions définies par cette dernière. C'est devenu un format courant de catalogage en France, principalement dans les services d'archives, les musées ou les bibliothèques. [29]

Dans L'EAD on parle d'[Attribut] et de [Valeur]. Les attributs sont associés à la plupart des éléments contenus dans l'EAD. Chaque attribut a un nom correspondant à l'attribut d'élément. Sa valeur dépend du contexte dans lequel il apparaît. Afin d'attacher un ou plusieurs attributs à un élément, il faut inclure le nom de chaque attribut dans la balise d'ouverture et la valeur qu'il doit prendre.

C'est-à-dire,

```
<[balise] [attribut]="[valeur]">
```

Ou

```
<[balise] [attribut1] = "[valeur1]" [attribut2] = "[valeur2]">
```

Dublin Core

Créé en 1995 à Dublin dans l'Ohio par OCLC (Online Computer Library Center) et le NCSA (National Center for Supercomputing Applications), la norme Dublin Core (transposé dans la norme ISO 15836) est une norme largement utilisée par les bibliothèques et les archives. L'objectif du Dublin Core est de fournir une base commune pour les éléments descriptifs afin d'améliorer les balisages et la recherche de diverses ressources au-delà de la communauté et les nombreux formats descriptifs spécifiques à chaque profession, tout en maintenant une structure adéquate.

Il permet ainsi la description par domaine ou thème et dispose également de balises permettant la gestion des droits d'auteur. Il est généralement implémenté dans les outils de GED et de SAE.

⁸ EAD - Encoded Archival Description

Le Dublin Core est composé de 15 éléments réparti en trois champs qui sont :

La description du contenu :

Title ;

Subject ;

Description ;

Source ;

Language ;

Relation ;

Coverage ;

La description de la propriété intellectuelle :

Creator ;

Contributor ;

Publisher ;

Rights ;

La description de l'instanciation :

Date ;

Type ;

Format ;

Identifier ;

Le Dublin Core a vocation à permettre une interopérabilité minimale entre des systèmes conçus indépendamment les uns des autres. Les règles internationales et nationales applicables à l'information bibliographique doivent être respectées. [30]

ii. Normes généralistes

Pour assurer la pérennisation des données, il est obligatoire de décrire la donnée par des métadonnées et de formats maîtrisés normalisé ou standardisés. Dans la sphère du BRGM et l'univers scientifique, il n'y a pas beaucoup de standards de métadonnées et de formats publiés, et largement utilisés.

METS (Metadata Encoding and Transmission Standard)

METS (Metadata Encoding and Transmission Standard) est un standard basé sur un XSD⁹ hébergé et maintenu par la bibliothèque du congrès aux USA,¹⁰ tout comme PREMIS. Il définit un format d'emballage, c'est-à-dire un format qui permet de structurer un ensemble d'informations liées de manière explicite. Le document XML généré permet d'organiser clairement diverses informations décrivant les données archivées.

Ce standard définit un schéma XML qui est utilisé dans des systèmes d'archivage électronique pour donner une forme aux paquets d'informations, en particulier pour l'AIP (le paquet qui est conservé au sein du système). Le schéma permet d'encapsuler et de placer dans des catégories distinctes des métadonnées d'usages divers telles que des métadonnées EAD, Dublin core ou encore PREMIS. [31]

Voir Annexe 6 : Exemple METS.

PREMIS (PREservation Metadata : Implementation Strategies)

PREMIS est un standard pour la description des métadonnées de préservation. Il permet de gérer une partie des données de préservation (ou de pérennisation au sens OAIS). Par exemple, pour qu'une donnée conserve sa qualité ou sa valeur probatoire dans le temps, il devra retracer tous les événements décrivant les dates de modifications des autres métadonnées, ou encore bien décrire la ou les signatures électroniques de l'archive. Ce modèle concerne à la fois ceux qui conservent et ceux qui transmettent de l'information à conserver. Il définit un modèle de données utile pour décrire les métadonnées de préservation. On peut être conforme au standard PREMIS si on utilise le vocabulaire pour exprimer les métadonnées même si on n'utilise pas le schéma XML qui en est proposé. Il est possible d'embarquer PREMIS dans le standard METS ce que font déjà quelques éditeurs de systèmes d'archivage électronique. Tout comme METS, ce standard est hébergé et maintenu par la Bibliothèque du Congrès.

Voir Annexe 5 : Exemple PREMIS.

⁹ Schéma XML

¹⁰ Library of Congress

Conclusion

L'arrivée de l'ordinateur a naturellement pour conséquence le boom informationnel que connaît le monde depuis deux décennies. Avec cette explosion des données il est tout à fait légitime d'aborder la question de la qualité des données. Mais si on étudie le sujet en profondeur on s'aperçoit que le concept en lui-même est problématique. En effet, si l'importance des données se mesure sur ce qui est de qualité ou pas, il reste à savoir qui sera habilité à définir la notion de qualité d'une donnée ? Le producteur ? L'archiviste ? Ou l'utilisateur ? Parce que tous les trois sont étroitement liés à la production de l'information. A cela s'ajoute la production de plus en plus variée allant des outils bureautiques passant par les bases de données et applications métiers aux vidéos et image 3D, les modalités de collecte et de capture sont profondément modifiées.

Dans le cadre du projet d'implantation d'un système d'archivage électronique, il m'a été demandé de réaliser le cadrage et la faisabilité du projet avec les livrables associés pour chaque partie afin de préparer le terrain pour son lancement en 2021. De mon point de vue personnel, ce projet sera à coup sûr bénéfique pour le BRGM parce que les agents bénéficieront de tous les avantages qu'offre un SAE respectant la réglementation spécifique de leur cœur de métier.

L'archivage électronique apparaît comme un domaine émergent qui deviendra une nécessité pour les entreprises. Le secteur public n'étant pas épargné par cette règle il devra se doter d'un SAE pour préserver les données produites ou reçues afin d'assurer leur mission de service public. La largesse et la complexité du domaine de l'archivage électronique mais les obligations et les restrictions liées à sa mise en place a tendance à démotiver ses aspirants. Bien qu'elle offre toujours plus de fonctionnalités à moindre coût, l'information numérique est forcément très fragile dans le temps. En effet, sans mesures préventives spécifiques, elle sera confrontée à trois risques principaux :

- La dégradation et le vieillissement des supports de stockage qui peut aboutir à la méconnaissance du contenu des objets numériques et l'incapacité à lire les formats de fichiers ;
- La disparition dans le temps des matériels ou logiciels de lecture. Pour cela, les méthodes éprouvées à mettre en œuvre sont : l'utilisation de métadonnées et d'identificateurs persistants, choisir des formats de fichiers persistants standardisé tel que le PDF/A, mettre en place une gestion proactive des supports de stockage vieillissants avec des migrations dans le temps ;

- Le milieu de l'archivage électronique étant récent et changeante, la législation qui l'encadre peut évoluer prochainement. Pour prévenir ce risque on peut effectuer une veille technologique constante pour anticiper des changements.

Si ces différents risques sont maîtrisés, les données conservées seront de qualité et pourront ouvrir la voie à d'autres usages tel que le web sémantique ou le web de données lié.

Pour faire le bilan sur cette année d'apprentissage, il est important de distinguer les points de satisfaction et d'amélioration.

Points faible

Au début de mon stage j'ai été confronté à des contraintes d'ordre logistique pour accéder à mon lieu de travail. En effet le site du BRGM se situant à Orléans à plus de 130 kilomètres de mon lieu de résidence conjugué cela aux diverses périples de la fin de l'année 2019 (grève des transports) et le début de l'année 2020 (la crise du Covid-19 le confinement généralisé,) plusieurs jours de travail ont été perdu. Ceci a créé un décalage sur le calendrier initial de plusieurs entretiens notamment ceux qui nécessitaient un déplacement. Seul les entretiens pouvant être assuré à distance ont été maintenus.

Sur le plan de la qualité et du traitement des données j'aurai aimé aller jusqu'au bout de mon travail de versement c'est-à-dire aller jusqu'au lancement du projet mettre en place les fonctionnalités détaillées depuis le début de ce mémoire et assister à la mise en marche du système et l'obtention du CoreTrustSeal. A cela s'ajoute la durée du contrat d'apprentissage qui est de onze mois contrairement à mes condisciples qui ont effectué deux ans de contrat ou plus. En effet, ayant intégré la formation MÉDAS en deuxième année de master, j'avais qu'une seule année d'étude à faire et devais faire l'alternance sur une année.

Points fort

L'avantage majeur tout au long de mon apprentissage au BRGM fut la progression l'autonomie dont j'ai pu disposer pour mener à bien mon travail. L'oreille attentive et les conseils précieux de mon tuteur M. Samuel Le Bideau m'ont beaucoup aidé. Cette autonomie s'est matérialisée par les moyens importants mobilisés pour la réussite de ma mission. En effet, lors de cette année d'apprentissage le BRGM m'a placé dans un bureau au sein du secrétariat général à côté de l'unité Achat pour faciliter les entretiens et faire l'état des lieux.

Mon travail de saisi des profils d'archivage et de recherche se faisait essentiellement en local avec Microsoft Teams ce qui était très rentable en en gain de temps car permettant à mon

tuteur M. Samuel Le Bideau d'y avoir accès depuis son bureau à temps réel sans avoir à se déplacer pour faire le point.

Sur le plan du traitement et de la gouvernance des données, j'ai trouvé très avantageux d'avoir travaillé sur ce genre de projet parce qu'il me permet de mettre en pratique les enseignements reçus au CNAM. D'emblée, j'ai effectué beaucoup de recherche sur la qualité des données et la conservation à moyen et long terme archives intermédiaire car ce domaine m'intéressait fortement. A cela s'ajoute la masse et la variété des documents que l'on retrouve au BRGM. Le travail effectué m'a aussi permis de mieux connaître le processus de création et de conservation des données dans les administrative et entreprise publique notamment l'évolution des format et logiciel métier mais aussi les conditions d'utilisation. A travers ma mission dans le service Gestion des Documents d'Activités j'ai pu appréhender la diversité des tâches de l'archiviste ou dirais-je du gestionnaire des données mais aussi la haute responsabilité qui lui incombe pour le bon fonctionnement de l'administration à travers la gestion des flux d'information. Contrairement à l'image archaïque que lui donne la société, l'archiviste n'est nullement cette personne isolée dans sa cave en train de s'occuper de la paperasse. Il est l'interlocuteur qui relie les agents, les lecteurs et les chercheurs aux données. Ceci s'est matérialisé par la présence de mon bureau au sein de l'unité Achat.

TABLE DES FIGURES

N° Page	Nom de la figure
34	Figure 1 : Modélisation des échanges d'information SEDA
35	Figure 2 : modèle fonctionnelle OAIS
36	Figure 3 : Fonctions de l'entité « Entrée »
37	Figure 4 : Fonctions de L'entité « Stockage »
39	Figure 5 : Fonctions de L'entité « Administration »
40	Figure 6 : Fonctions de L'entité « planification de la pérennisation »
41	Figure 7 : Fonctions de L'entité « accès »
49	Figure 9 : Positionnement de Vitam dans le SI BRGM
51	Figure 8 : Définition des scenarii et des livrable attendu
53	Figure 10 : reproduction du processus des Achats
54	Figure 11 : liste des documents produits dans le processus Achat BRGM
63	Figure 12 : Processus d'acheminement d'une donnée vers la connaissance
72	Figure 13 : Périmètre d'application du SAE dans la banque du sous-sol

TABLE DES TABLEAUX

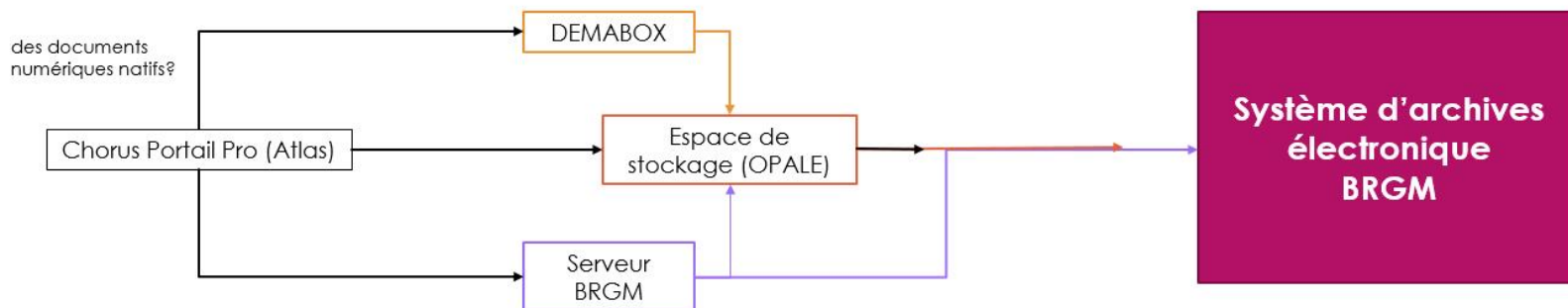
N° Page	Nom du tableau
12	Tableau 1 : Diagramme de GANTT - Planning prévisionnel du projet de mémoire
14	Tableau 2 : Matrice des risques d'échecs du projet
38	Tableau 3 : Rôle des acteurs et intervenants dans le SAE
40	Tableau 4 : Fonctionnalités comparées SAE / GED / CFN
61	Tableau 5 : des typologies de métadonnées et leurs finalités

Annexes

Annexe 1 Différents lieux de capture des données

Espace de collecte des données au BRGM

des documents numérisés ?



Annexe 4 Différents type de métadonnées et services concernés

Métadonnées

H.A
SFACT

Métadonnées descriptives				
Titre document	Entité responsable de la création du document/ identifiant HA	Objet du document	Dates	Formats
Métadonnées de gestion				
Exemple de support/sous dossier	DUA	Sort final	Communicabilité 4 niveaux	

SG/GDA

Métadonnées de sécurité					
intégrité	perennité	Accessibilité	Formats	Traçabilité	Disponibilité
Chainage des Journaux (Date et heure de vérification, algorithme de chiffrement)	Date et heure de capture dans le système d'archivage, identifiant système, code de classement, type de document	entité de rattachement, sous dossier et existence de copie	Dublin core XML/EAD	Journal des événements, Journal du cycle Chainage des traces Horodatage...[choix selon exigence légal (niveau 3)]	Tolérable mais doit être limité

Annexe 5 : Exemple PREMIS

Cette exemple PREMIS repris intégralement de la traduction française de la documentation de PREMIS. [32]

L'exemple ci-dessous est la simplification d'un exemple conçu par la Bibliothèque du Congrès pour un tutoriel PREMIS. Il montre les unités sémantiques PREMIS et les valeurs utilisées pour décrire une image TIFF. La LoC utilise des identifiants handle (identifiants créés et gérés par une application appelée le système Handle) pour les fichiers stockés dans l'entrepôt (cf. 1.1 objectIdentifier et 1.10.3 relatedObjectIdentification). L'objet décrit est reconnu dans l'entrepôt par le biais de son identifiant unique (1.1 objectIdentifier) mais son nom d'origine avant le versement était « 001h.tif » (1.6 originalName). Nous savons qu'il s'agit d'un fichier grâce à la catégorie de l'objet (1.2 objectCategory). Le fichier possède une empreinte MD5 calculée par le système de gestion de contenus de la LoC (1.5.2 fixity). Son format est décrit de deux manières différentes, par un type MIME et un numéro de version (1.5.4.1 formatDesignation) et en pointant vers le répertoire de formats PRONOM qui détient une copie de la spécification TIFF. La LoC considère qu'il y a eu deux applications d'origine pour ce fichier. Un élément non déterminé (probablement un fichier RAW sans nom) a été créé sur un scanner par l'application ScandAll et a été transformé en TIFF en utilisant Adobe Photoshop (1.5.5 creatingApplication). Il n'y a pas de cryptage ou d'autre type de restriction que la LoC ait jugé pertinent de consigner explicitement (1.5.6 inhibitors). Le fichier est stocké sur disque à un emplacement identifié par son chemin dans l'arborescence (1.7.1 contentLocation) ; le type de support correspondant est décrit très précisément, sans doute afin que les gestionnaires de l'entrepôt puissent déterminer le moment où les données nécessiteront une migration vers un type de support plus récent (1.7.2 storageMedium). La LoC a pris la peine de consigner un environnement recommandé pour visualiser le fichier : utiliser Adobe Acrobat version 5.0 sur un processeur Intel x86 fonctionnant sous Windows XP. La vitesse du processeur doit être au moins de 60 MHz et il doit y avoir au moins 32 Mo de mémoire, et de préférence 64 Mo (1.8 environment). Le fichier TIFF fait partie d'une représentation dont l'identifiant interne est R200802948. Cette information est donnée par le troisième bloc d'informations sur les relations (1.10 relationship). Il y a deux autres fichiers dans cette représentation, identifiés par des identifiants Handle. Nous savons qu'il doit s'agir de fichiers car le type de relation consigné est « frère ». Il nous faut consulter la description des objets [loc.music/gottlieb.09602](#) et [loc.music/gottlieb.mets](#) pour plus d'informations sur ces derniers, par exemple pour savoir de quel type de fichiers il s'agit. Nous n'avons pas d'information descriptive sur l'entité intellectuelle correspondant à cette

représentation. Toutefois, nous avons un lien vers son identifiant, un LCCN (1.12 linkingIntellectualEntityIdentifier). En toute logique, si l'on recherchait cet LCCN, nous trouverions une description bibliographique de l'entité intellectuelle en question (pas dans cet exemple toutefois car le LCCN a été forgé). »

Unité sémantique	Valeur
1.1 objectIdentifier (identifiant de l'objet)	
1.1.1 objectIdentifierType (type d'identifiant de l'objet)	hd1
1.1.2 objectIdentifierValue (valeur de l'identifiant de l'objet)	loc.music/gottlieb.09601
1.2 objectCategory (catégorie de l'objet)	file
1.3 preservationLevel (niveau de préservation)	
1.3.1 preservationLevelValue (valeur du niveau de préservation)	full
1.5 objectCharacteristics (caractéristiques de l'objet)	
1.5.1 compositionLevel (niveau de composition)	0
1.5.2 fixity (intégrité)	
1.5.2.1 messageDigestAlgorithm (algorithme de l'empreinte numérique)	MD5
1.5.2.2 messageDigest (empreinte numérique)	36b03197ad066cd719906c55eb68ab8d
1.5.2.3 messageDigestOriginator (origine de l'empreinte numérique)	localDCMS
1.5.3 size (taille)	20800896
1.5.4 format	
1.5.4.1 formatDesignation (dénomination du format)	
1.5.4.1.1 formatName (nom du format)	image/tiff
1.5.4.1.2 formatVersion (version du format)	6.0
1.5.4.2 formatRegistry (répertoire de formats)	
1.5.4.2.1 formatRegistryName (nom répertoire de formats)	PRONOM
1.5.4.2.2 formatRegistryKey (identifiant du répertoire de formats)	fmt/10
1.5.4.2.3 formatRegistryRole (fonction du répertoire de formats)	specification
1.5.5 creatingApplication (application d'origine)	
1.5.5.1 creatingApplicationName (nom de l'application d'origine)	ScandAll21
1.5.5.2 creatingApplication Version (version de l'application d'origine)	4.1.4

1.5.5.3 dateCreatedByApplication (date de création par l'application d'origine)	1998-10-30
1.5.5 creatingApplication (application d'origine)	
1.5.5.1 creatingApplicationName (nom de l'application d'origine)	Adobe Photoshop
1.5.5.2 creatingApplication Version (version de l'application d'origine)	CS2
1.5.5.3 dateCreatedByApplication (date de création par l'application d'origine)	1998-10-30T08:29:02
1.5.6 inhibitors (restrictions)	
1.5.6.1 inhibitorType (type de restriction)	none
1.6 originalName (nom d'origine)	001h.tif
1.7 storage (stockage)	
1.7.1 contentLocation (emplacement du contenu)	
1.7.1.1 contentLocationType (type d'emplacement d'origine)	filepath
1.7.1.2 contentLocationValue (valeur de l'emplacement d'origine)	amserver/
1.8 environment	
1.8.1 environmentCharacteristic (caractéristique de l'environnement)	recommended
1.8.2 environmentPurpose (fonction de l'environnement)	render
1.8.2 environmentPurpose (fonction de l'environnement)	edit
1.8.5 software (logiciel)	
1.8.5.1 swName (nom du logiciel)	Adobe Acrobat
1.8.5.2 swVersion (version du logiciel)	5.0
1.8.5.3 swType (type de logiciel)	renderer
1.8.5 software (logiciel)	
1.8.5.1 swName (nom du logiciel)	Windows
1.8.5.2 swVersion (version du logiciel)	XP
1.8.5.3 swType (type de logiciel)	Operating System
1.8.6 hardware (matériel)	
1.8.6.1 hwName (nom du matériel)	Intel x86
1.8.6.2 hwType (type de matériel)	processor
1.8.6.3 hwOtherInformation (autres informations sur le matériel)	60 Mhz minimum
1.8.6 hardware (matériel)	
1.8.6.1 hwName (nom du matériel)	64 MB RAM
1.8.6.2 hwType (type de matériel)	memory
1.8.6.3 hwOtherInformation (autres informations sur le matériel)	32 MB minimum
1.10 relationship (relation)	

1.10.1 relationshipType (type de relation)	structural
1.10.2 relationshipSubType (sous-type de relation)	has sibling

Annexe 2 : Exemple METS

Ci-dessous un exemple de fichier METS pour ce document créé sous Word 2010, une fois transformé en PDF-A. Des métadonnées Dublin Core et PREMIS y sont ajoutées.

```
<?xml version="1.0" encoding="UTF-8"?>
<mets:mets xmlns:mets="http://www.loc.gov/METS/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:premis="info:lc/xmlns/premis-v2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/METS/
    http://www.loc.gov/standards/mets/mets.xsd
    http://www.w3.org/1999/xlink
    http://www.loc.gov/standards/mets/xlink.xsd
    http://purl.org/dc/elements/1.1/
    http://dublincore.org/schemas/xmls/qdc/2003/04/02/dc.xsd
    http://www.loc.gov/standards/premis/v1
    http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd">
  <!-- Renseignement sur le fichier METS -->
  <mets:metsHdr CREATEDATE="2020-08-08T04:04:00" RECORDSTATUS="Complete">
    <mets:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
      <mets:name>Dieng Mouhamed</mets:name>
    </mets:agent>
  </mets:metsHdr>
  <!-- Metadonnées descriptives internes, -->
  <mets:dmdSec ID="memoire_cnam_sae_qualité_données">
    <!-- type Dublin Core (DC) -->
    <mets:mdWrap MDTYPE="DC">
      <mets:xmlData>
        <dc:creator>Gora Dieng</dc:creator>
        <dc:title>La qualité des données et métadonnées dans
          l'implantation d'un système d'archivages
          électronique</dc:title>
        <dc:format>application/pdf</dc:format>
        <dc:date>2020-08-08</dc:date>
        <dc:type>Mémoire de master MEDAS</dc:type>
        <dc:coverage>France</dc:coverage>
        <dc:coverage>BRGM Orléans</dc:coverage>
        <dc:coverage>Conservatoire National des Arts et
          Métiers</dc:coverage>
        <dc:coverage>Archivage Électronique</dc:coverage>
        <dc:subject>Archivage Électronique</dc:subject>
        <dc:subject>qualité des données</dc:subject>
        <dc:subject>SAE</dc:subject>
        <dc:subject>Métadonnée</dc:subject>
        <dc:language>fr</dc:language>
        <dc:identifiant>mémoire_GD.pdf</dc:identifiant>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
  <!-- Métadonnées Administratives -->
  <mets:amdSec>
    <mets:techMD ID="premis_pdf">
```

```

    <mets:mdWrap MDTYPE="PREMIS">
      <mets:xmlData>
        <premis:object>
          <premis:objectIdentifier>
<premis:objectIdentifierType>Mémoire GD PI f99d23a2eba4</premis:objectIdentifierType>
          <premis:objectIdentifierValue>uuid:0b83a11e-cdee-47e7-8968-
</premis:objectIdentifierValue>
          </premis:objectIdentifier>
          <objectCharacteristics>
            <fixity>
              <messageDigestAlgorithm>SHA256</messageDigestAlgorithm>
<messageDigest>b6411d3ddb804077d721277452f6312e905fadb1be8973fc85b3f207fac7
2bd7d</messageDigest>
              <messageDigestOriginator>BRGM
Orléans</messageDigestOriginator>
            </fixity>
            <size>18437</size>
            <format>
              <formatDesignation>
                <formatName>application/pdf</formatName>
                <formatVersion>PDF-A</formatVersion>
              </formatDesignation>
            </format>
          </objectCharacteristics>
          <environment>
            <software>
              <swName>Micorsoft Word</swName>
              <swVersion>2020</swVersion>
            </software>
            <software>
              <swName>Microsoft Windows</swName>
              <swVersion>10</swVersion>
              <swType>Operating System</swType>
            </software>
          </environment>
        </premis:object>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:techMD>
</mets:amdSec>
<!-- Description de l'ensemble des fichiers utilisés -->
<mets:fileSec>
  <!-- Fichiers rassemblés par groupes de fichiers -->
  <mets:fileGrp ADMID="premis_pdf">
    <mets:file ID="memoire_v7.pdf" MIMETYPE="application/pdf">
      <mets:FLocat LOCTYPE="URL" xlink:href="memoire_v7.pdf"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
<!-- Carte de structure -->
<mets:structMap>
  <mets:div>
    </mets:div>
  </mets:structMap>
</mets:met

```

Bibliographie

- [1] Site de Légifrance, loi du 13 mars 2000. Consulté le 09 mai 2020 à l'adresse <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000005629200>
- [2] Site de Légifrance. Livre II code du patrimoine. Consulté le 20 mai 2020, à l'adresse : <https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000006845559&idSectionTA=LEGISCTA000006159940&cidTexte=LEGITEXT000006074236&dateTexte=20080716>
- [3] Site internet de RDA France. The coretrustseal. Consulté le 23 Juillet 2020, à l'adresse <https://www.rd-alliance.org/coretrustseal-criteres-de-conformite>
- [4] Site de Légifrance. Décret de création BRGM. Consulté le 15 Juin 2020, à l'adresse : https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=B3F6EEF2F2667AC58F66EF1A5029674A.tpdjo08v_3?cidTexte=JORFTEXT000000304238&categorieLien=cid&dateTexte=
- [5] Association des archivistes français, Archiver, guide des procédures existantes, mieux maîtriser l'information dans l'entreprise. 2004. Consulté le 19/07/2020, à l'adresse : <https://www.archivistes.org/Archiver-le-guidedes-procedures>
- [6] Jules Arnaud. Gestion des documents d'activité : définition, principes et concepts. La Gazette des archives, n°228, 2012-4. Normalisation et gestion des documents d'activité (records management) : enjeux et nouvelles pratiques pour notre profession. pp. 103-117. Consulté le 12/07/2020, à l'adresse : www.persee.fr/doc/gazar_0016-5522_2012_num_228_4_4988
- [7] Dessolin Baumann Sylvie. Des chartriers aux bases de données. Les enjeux de la gestion des documents d'activité. Consulté le 19/07/2020, à l'adresse : [https://www.persee.fr/doc/gazar_0016-5522_2012_num_228_4_4987](http://www.persee.fr/doc/gazar_0016-5522_2012_num_228_4_4987)
- [8] MELLIFLUO Laure. Evaluation des archives : en théorie et en pratique aux Archives communales de la ville de Carouge, p.25. Mémoire de bachelor : haute école de gestion de Genève, 2008. Consulté le 19/07/2020, à l'adresse : https://doc.rero.ch/record/11281/files/Travail_de_bachelor.pdf

[9] Nougaret Roger, Zuber Henri. Les archives d'entreprises en France. La Gazette des archives, n°204, 20064. Les archives en France. pp. 171-187. Consulté le 19/07/2020, à l'adresse : www.persee.fr/doc/gazar_00165522_2006_num_204_4_3834

[10] KREBS Catherine. Les archives d'entreprise : Quelle gestion pour quels enjeux. Projet de recherche, Ecole nationale supérieure de bibliothécaire, 1989. Consulté le 19/07/2020, à l'adresse : <https://www.enssib.fr/bibliothequenumerique/documents/62252-les-archives-d-entreprises-quellegestion-pour-quels-enjeux-l-experience-ducentre-rhenan-d-archives-et-de-rechercheseconomiques.pdf>

[11] Vidalenc Jean. Les archives d'entreprise et leur intérêt pour l'histoire. La Gazette des archives, n°73-74, 1971. pp. 83-91. Consulté le 19/07/2020, à l'adresse : www.persee.fr/doc/gazar_0016-5522_1971_num_73_1_2198

[12] Blog Scriba. Les archives d'entreprise, un enjeu important. Consulté le 19/08/20, à l'adresse : <https://www.scriba.fr/archivesentreprises-enjeuimportant/>

[13] RICARD Bruno. Le RGPD et les archives. Consulté 01/06/2020, à l'adresse : <https://siafdroit.hypotheses.org/792>

[14] Site de Légifrance. Code civil. Consulté le 12/07/2020, à l'adresse : <https://www.legifrance.gouv.fr/affichTexteArticle.do?idArticle=JORFARTI000032005161&cidTexte=JORFTEXT000032004939&categorieLien=id#:~:text=soit%20leur%20support.-%C2%AB%20Art.%C3%A0%20en%20garantir%20l'int%C3%A9grit%C3%A9.>

[15] Site de Légifrance. Livre II Code du patrimoine. Consulté 01/06/2020 à l'adresse : <https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000006845559&idSectionTA=LEGISCTA000006159940&cidTexte=LEGITEXT000006074236&dateTexte=20080716>

[16] Site de Légifrance. Loi 2016 elkomri. Consulté le 13/07/2020 à l'adresse <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000032983213&categorieLien=id>

[17] International Organization for Standardization, (ISO). Normes NFZ42-013 Consulté le 08/07/2020, à l'adresse : <http://www.iso.org/iso/fr/home/standards.htm>.

[18] Blog de Numen opérateur de confiance. Norme internationale ISO 14 641-1. Consulté le 13/07/2020, à l'adresse : <https://numen.blog/iso-14641-1-nf-z-42-013-nf-z-42-026-psdc-quesaco/>

[19] F. Banat-Berger et C. (. Huc, « Section 5 : Modèle de référence OAIS et normes associées, » novembre 2011. Consulté le 08/07/2020, à l'adresse : http://www.piaf-archives.org/espace-formation/file.php/9/section05_web/co/module_section05.html.

[20] Sibille Claire, Nichele Baptiste. Le Standard d'échange de données pour l'archivage (SEDA), un outil structurant pour l'archivage. La Gazette des archives, n°240, 2015-4. Voyages extraordinairement numériques : 10 ans d'archivage électronique, et demain ? sous la direction de Frédérique Fleisch et Anita Frieh. pp. 153-164. Consulté le 12/08/2020

[21] Le site Knowledge Mapping Research recense en 2006 130 définitions différentes formulées par des experts. Consulté le 15/08/2020 à l'adresse http://www.success.co.il/is/zins_definitions_dik.pdf

[22] Dr H.M. Gladney, HMG Consulting livre Blanc la qualité des données en entreprise Consulté le 12/08/2020

[23] Site internet de informatica. Consulté le 12/07/2020, à l'adresse : <http://www.information-management.com/glossary/u.html>

[24] Philippe Tassin, Systèmes d'information et développement durable, Hermès Lavoisier, p. 224-228. Consulté le 20/07/2020

[25] Site de l'OCDE. Les données scientifiques. Consulté le 21/08/2020, à l'adresse : <http://www.oecd.org/fr/science/inno/38500823.pdf>

[26] The Data Management Association, The DAMA Guide to The Data Management Body of Knowledge, 1 re éd, DAMA International, New Jersey, États-Unis, 2009, 406p. Consulté le 12/08/2020

[27] Redman, T.C., Data Quality for the Information Age, 1re éd., Artech House, Boston, MA, 1996, 261 p. Consulté le 12/08/2020

[28] Fabien, gandon, @fabien_gandon. Qualité des métadonnées. Consulté le 12/07/2020, à l'adresse : <http://fabien.info>

[29] Direction des Archives de France, (DAF), « ICA-Req, ». Consulté le 04/08/2020, à l'adresse : <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/standard/ica-req/>

[30] Bibliothèque National de France. dublin core standards. Consulté le 19/08/2020, à l'adresse : <https://www.bnf.fr/fr/dublin-core#bnf-ressources>

[31] Direction des Archives de France (DAF), « Métadonnées, » Consulté le 04/08/2020, à l'adresse : <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/standard/metadonnees/>

[32] Library of Congress (Loc), « METS - Metadata Encoding & Transmission Standard, » Consulté le 25/08/2020, à l'adresse : <http://www.loc.gov/standards/mets/>