La gestion des risques pour la conservation des documents numériques

Emmanuelle Bermès
Bibliothèque nationale de France
Département de la Bibliothèque numérique

Réunion Groupe PIN du lundi 23 avril 2007

Sommaire

- Objectifs de la gestion des risques dans le contexte de la mise en place de la préservation numérique
- Présentation de la méthodologie de gestion des risques
- Principaux types de risques et leurs maîtrises
- (pour le groupe PIN) : la gestion des risques et le modèle OAIS
- La gestion des risques comme outil de pilotage

Objectifs et méthodologie

Pourquoi une méthode de gestion de risques ?

- La préservation des documents numériques est une action principalement préventive
  - nécessité d'intervenir avant que les dégradations se produisent
- La préservation des documents numériques est un ensemble de compromis
  - entre les impératifs immédiats des producteurs et les besoins à long terme des utilisateurs
  - entre les besoins et les moyens
- La préservation des documents numériques est un projet comme un autre…
  - toute activité génère des risques
  - la question n'est pas de supprimer les risques, mais de déterminer le niveau de risque acceptable
  - la méthodologie de gestion de risques est ancienne dans le monde industriel, « à la mode » dans le monde du management et des services

Méthodologie de gestion des risques

- 1. définition du contexte – cette étape consiste à fixer les objectifs de la gestion des risques
- 2. liste des risques – identifier les risques et les catégoriser
- 3. évaluation des risques – analyser la probabilité et l'impact de chaque risque, combiner ces deux facteurs pour évaluer le risque
- 4. prise de décision – en fonction de l'évaluation, identifier les risques prioritaires, les moyens de leur traitement et le plan d'action
- 5. maîtrise – mettre en place les actions nécessaires pour diminuer le niveau de risques
- 6. itération - reprendre l'évaluation des risques résiduels ou des nouveaux risques qui surviennent au cours du fonctionnement du système.

1. Objectifs

- **Contexte : mise en place d'un système visant à préserver des documents numériques sur le long terme**
- Une archive (OAIS) s'engage à fournir deux types de services
  - conservation et accès
  - les producteurs de l'information d'une part, et la communauté d'utilisateurs d'autre part
- Notion de transfert de responsabilité
  - formaliser la responsabilité de l'archive et le type d'engagements qu'elle prend en terme de conservation et d'accès.
- La gestion des risques est essentielle
  - elle fournit un référentiel sur lequel peut reposer l'élaboration de ces protocoles
  - elle fait en sorte que l'engagement de l'archive en termes de conservation et d'accès soit proportionné avec les risques qui pèsent sur les objets
  - elle définit les conditions d'une mission continue de veille et de surveillance de risques, définie dans le modèle OAIS comme « planification de la préservation ».

2. Liste des risques : définition du risque

- Le risque est défini par la combinaison de:
  - une vulnérabilité : un état de fait qui rend le risque possible
  - une menace : l'existence d'un potentiel d'exploitation de la vulnérabilité, avec une source et une action
- Ex. : risque d'inondation par crue de la Seine
  - Le risque n'existe que si on est vulnérable (la BnF est construite en zone inondable -*vulnérabilité*)
  - Le risque n'existe que si la menace existe (des causes naturelles -*source*- peuvent provoquer la crue de la Seine -*action*)

2. Liste des risques : identifier et catégoriser

- Il existe différentes méthodes pour identifier les risques :
  - brainstorming, entretiens, réunions
  - expérience, expertise, observation
  - études de cas, tests, analyse de scénarios
  - audit, inspection, retour d'expérience sur incident
  - compréhension globale du contexte : besoins, objectifs, activités particulières (comme la préservation du numérique !)
- Catégories de risques
  - opérationnel, financier, organisationnel, infrastructure, réputation, etc.
- Sources de risques
  - externes – environnement naturel, bâtiment
  - internes – décisions stratégiques, ressources, matériels etc.

3. Évaluation des risques

- **Probabilité**
  - **improbable (1) :** la vulnérabilité est limitée, et le risque ne s'est jamais produit par le passé
  - **rare (2) :** la vulnérabilité est limitée, le risque s'est déjà produit par le passé mais il a été considéré comme un événement exceptionnel et non susceptible de se répéter
  - **occasionnel (3) :** la vulnérabilité est importante, mais le risque ne s'est jamais produit par le passé
  - **probable (4) :** la vulnérabilité est importante, le risque s'est déjà produit par le passé et est susceptible de se répéter
  - **fréquent (5) :** la vulnérabilité est importante, et le risque se produit régulièrement de façon répétée

3. Évaluation des risques

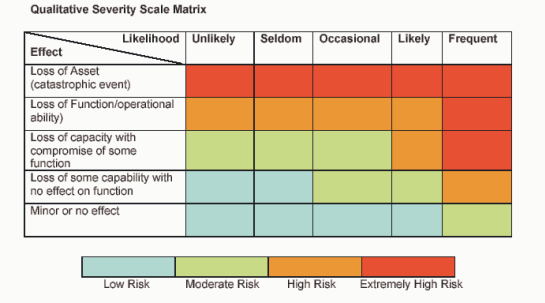
- **Impact**
  - l'impact est évalué uniquement concernant les collections numériques, en terme de conservation et d'accès.
    - **insignifiant (1)** : sans effet sur l'objet numérique
    - **mineur (2)** : l'objet numérique est altéré, mais cela ne cause pas de perte de fonctionnalités d'accès
    - **modéré (3)** : l'objet numérique est altéré, et cela cause des pertes partielles de fonctionnalités d'accès
    - **majeur (4)** : perte complète des fonctionnalités d'accès
    - **catastrophique (5)** : perte complète de l'objet lui-même.

3. Évaluation des risques

- **Temps**
  - Ce critère de mesure a pour objectif de placer les risques dans l'évolution temporelle. Ici on considère pour un risque qui doit se produire à un instant T, dans combien de temps cet instant T va arriver.
    - **lointain (1)** : dans plus de 10 ans
    - **moyen (2)** : d'ici 5 à 10 ans
    - **proche (3)** : d'ici 1 à 5 ans
    - **imminent (4)**: dans les 12 prochains mois
    - **immédiat (5)** : le risque existe dès aujourd'hui

4. Prise de décision

- Méthode matricielle



- Les risques sont-ils acceptables ?
  - Objectifs et activités de l'établissement
  - Producteurs et communautés d'utilisateurs
  - Coûts et bénéfices des opérations de maîtrise
  - Obligation légale

5. Maîtrise

- Les maîtrises sont les mesures à prendre pour endiguer le risque.
  - Les mesures de maîtrise peuvent être déjà existantes ou être recommandées par la gestion des risques
- Il y a trois niveaux de triplets impact / probabilité / temps :
  - risque brut : avant toute action de maîtrise
  - risque actuel : triplet pondéré en fonction des actions déjà prévues et opérationnelles
  - risque résiduel : triplet pondéré en fonction des actions recommandées par le groupe

5. Maîtrise

- Maîtrises qui diminuent la probabilité d'un risque
  - Ex. choisir un format ouvert, normalisé et répandu pour faire baisser la probabilité d'obsolescence du format
- Maîtrises qui diminuent l'impact d'un risque
  - Ex. mettre en place une sauvegarde complète hors site pour pouvoir restaurer le système en cas de catastrophe
- Éviter le risque
  - On peut éviter le risque en renonçant à l'activité qui l'engendre (ex. refuser d'archiver certains types d'objets considérés comme trop risqués)
- Partager le risque
  - Ex. souscrire une assurance
- Tolérance au risque
  - En dessous d'un certain niveau de risque, on tolère ou accepte le risque.

6. Itération

- L'itération est prévue périodiquement sur l'ensemble des risques
- Risques résiduels
  - Une fois un risque pris en compte, on vérifie le niveau de risque résiduel et on s'assure que la maîtrise mise en œuvre n'a pas créé de nouveaux risques
- Nouveaux risques
  - À chaque fois que le système est modifié, on vérifie si la modification entraîne l'apparition de nouveaux risques ou la modification de l'évaluation de risques existants
  - En cas de problème non prévu par la gestion des risques, il faut mettre à jour la liste des risques
- Risques évolutifs
  - Certains risques peuvent évoluer naturellement avec le temps (s'aggraver ou disparaître), sans qu'aucune évolution du système ou action de maîtrise n'ait été effectuée : certains risques peuvent avoir une périodicité de révision

Principaux risques et leur maîtrise

Exemple de la BnF

Risques naturels

- Sont classés dans cette catégorie les risques dont l'origine est liée à l'environnement et à un événement naturel.

Liste des risques :
- **inondation**
- **terrorisme, guerre**
- **autres catastrophes naturelles**
- **épidémies**

Maîtrise :
- **Plan d'urgence**
- **Duplication hors site**

Risques liés à la sécurité

- Cette catégorie concerne les risques liés à une intrusion malveillante dans le système.
- Cette intrusion peut être physique ou virtuelle. Elle peut avoir différentes conséquence sur les collections suivant les intentions de l'intrus.

Liste des risques :
- **intrusion dans le système**
  - **sécurité logique**
  - **sécurité physique**

Maîtrise :
- **Plan de sécurité des systèmes d'information**

Risques organisationnels

- Risques liés aux acteurs et aux personnels
  - **Cette catégorie concerne les risques liés à l'absence de personnel compétent.**
  - **Le personnel compétent peut être absent pour des raisons environnementales ou ponctuelles (maladie).**

Liste des risques :

- **carence de ressources humaines compétentes**
- **impossibilité de faire adhérer l'ensemble de la bibliothèque au projet**

Maîtrise :

- **Visibilité du système**
- **Formation, accompagnement au changement**

Risques technologiques

- Cette catégorie regroupe les risques liés à l'environnement matériel et logiciel concernant la lecture d'un document.
- Cette catégorie recense aussi les risques liés aux migrations en fonction de la maîtrise que l'on a sur le format, et tout ce qui concerne les plateformes d'émulation et leur utilisation.

Liste des risques :

- **obsolescence des formats et de leur environnement**
- **obsolescence de la plateforme matérielle requise**
- **perte des compétences usagers**

Maîtrise :

- **Trajectoires de migration, émulation**
- **contrôle, alertes, veille**
- **collecte d'informations (métadonnées, logiciels, plateformes)**

Risques concernant l'accessibilité sémantique

- Cette catégorie prend en compte les risques liés aux données et aux outils qui donnent accès aux documents.
- L'accessibilité sémantique correspond à la compréhension par la communauté des utilisateurs de l'objet auquel on donne accès.

Liste des risques :

- **absence des métadonnées descriptives appropriées**
- **absence de référentiels sémantiques pour l'interprétation des documents**
- **perte du contexte affectant la signification ou la complétude du document**

Maîtrise :

- **collecte d'informations (métadonnées, documentation associée)**

Risques concernant l'accessibilité technique

- Cette catégorie prend en compte les risques liés aux données et aux outils qui donnent accès aux documents.
- L'accessibilité technique correspond aux informations nécessaires pour la restitution (« rendering »). Il peut y avoir des entraves à l'accessibilité technique (dispositifs anticopie).

Liste des risques :

- **absence des métadonnées techniques appropriées**
- **absence des métadonnées de structure appropriées**
- **copie ou consultation de la copie empêchée par un système de protection**

Maîtrise :

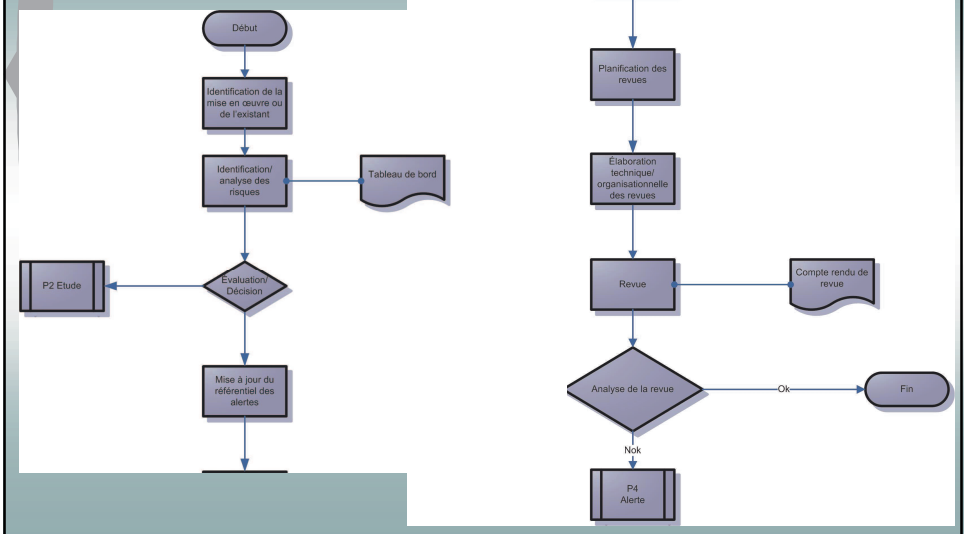
- **collecte d'informations (métadonnées, mots de passe)**

Gestion des risques et Planification de la Préservation

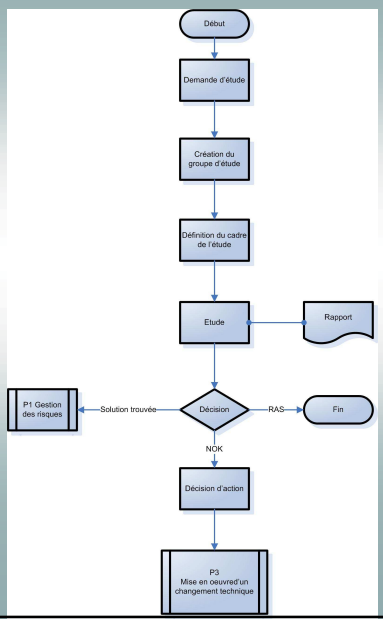
Dans le modèle OAIS

- Définition de la « planification de la préservation » :
  - «cette entité assure les fonctions et services relatifs à la **surveillance de l'environnement** de l'OAIS et à la production de recommandations (…) Les fonctions de l'Entité « Planification de la pérennisation » incluent **l'évaluation du contenu de l'Archive** et la recommandation périodique de mises à jour de l'information archivée pour migrer les fonds courants, le développement de **recommandations** dans le domaine des normes et règles d'archivage, ainsi que la **surveillance** des évolutions à la fois de **l'environnement technologique** et des **exigences de service** de la Communauté d'utilisateurs cible, et enfin de sa Base de connaissance. »
- Définition plus concrète des « processus »

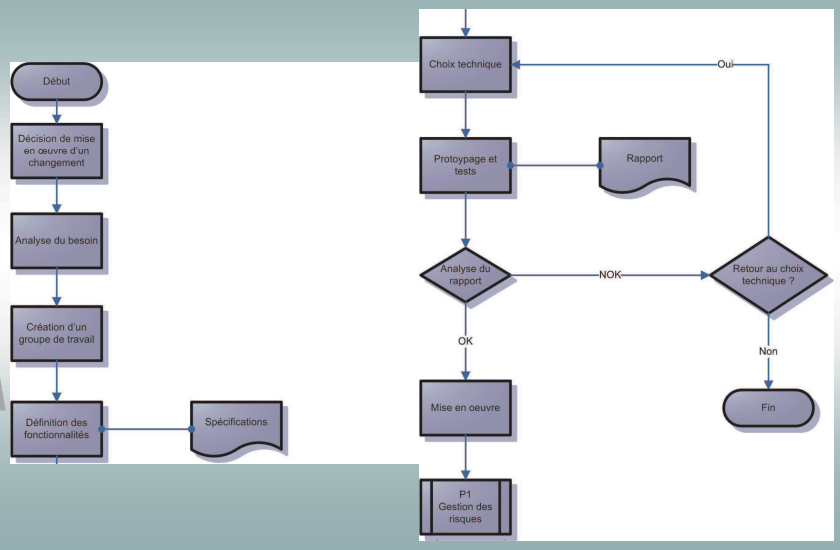
Processus 1 : gestion des risques



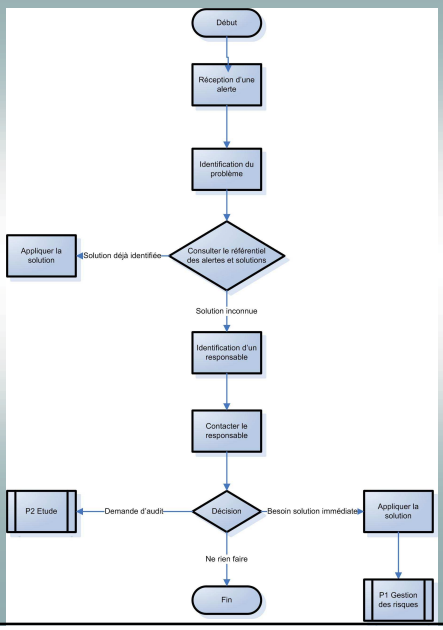
Processus 2 : étude



Processus 3 : changement technique



Processus 4 : résolution d'un problème



Aspects organisationnels de la gestion des risques

- **Suivi de la communauté des utilisateurs** :
  - définition des besoins, modalités, outils et fonctions qui permettent d'assurer ce suivi
  - Ex. études, enquêtes, ateliers etc. menés de façon continue dans le temps.
- La **veille technologique :**
  - Fait partie des maîtrises existantes et recommandées
  - organisée de façon plus formelle : une mention explicite dans les fiches de postes des agents concernés, une répartition avec des correspondants ou des comités de suivis, ou encore des outils de partage d'information (base de connaissance, wiki, système de publication de contenus Web).
- La fonction « **développer les stratégies et les standards de pérennisation** » :
  - collecte d'information internes et externes à l'archive, synthétisées pour déboucher sur des prises de décision, y compris en cas de problème pour trouver une solution.
  - Peut être identifié au processus itératif de gestion des risques.
- La fonction « **développer des modèles d'empaquetage et des plans de migration** » :
  - définir des modèles d'empaquetage n'est pas seulement une fonction initiale mais continue.
  - elle est à itérer à chaque fois que de nouveaux types de paquets se présentent.

Conclusion sur la méthodologie

- Pour gérer les risques
  - Il faut réunir de nombreuses données concernant les risques et les moyens de les maîtriser
  - D'autant qu'un même risque n'a pas les même probabilité/impact/temps suivant le type de matériau/document
- Il faut générer plusieurs « vues » sur ces données :
  - Un référentiel pour les protocoles de négociation avec les producteurs
  - Un référentiel pour faire le suivi des documents présents dans l'archive (revues, contrôles etc.)
  - Un référentiel pour pouvoir résoudre rapidement les problèmes quand ils surviennent.

La gestion des risques comme outil de pilotage

Gestion de risques pour planifier la préservation

- Il s'agit d'utiliser les données de gestion des risques pour générer un tableau de bord de pilotage de la préservation des objets numériques d'un secteur (« plan de conservation »).
- **C'est un outil de pilotage** qui permet de :
  - planifier les revues : surveiller l'évolution des données, des applications et des supports. La revue permet de s'assurer que les choix techniques ou organisationnels sont toujours valides et que leur qualité est bonne. La revue produit un compte-rendu de revue qui déclenche une alerte en cas de situation anormale
  - prévoir les alertes et les solutions : un signal déclenché par la détection d'un problème. L'alerte déclenche un processus de résolution des problèmes préparé à l'avance (plan d'urgence…)
  - déterminer les risques spécifiques à différents types d'objets, et pondérer l'évaluation des risques en fonction de leurs particularités techniques ou autres.
- Toutes ces informations permettent d'avoir une vue sur la préservation des objets à long terme, et de faire évoluer les conditions de préservation en demandant des études ou des changements techniques. Elles permettent de surveiller les objets et d'appliquer le plan de préservation.

Gestion des risques en cas d'alerte

- On tire des indicateurs qui identifient les risques et leur maîtrise une liste de toutes les alertes connues (ex. inondation, incendie, tentative d'intrusion physique en salle machine, destruction d'un support, indisponibilité du système d'accès, etc.) et des solutions qu'il convient d'appliquer.
- Ces solutions peuvent se trouver dans des documents existants (le plan d'urgence, la liste des n° à appeler en cas d'ur gence…). Dans ce cas un **référentiel des alertes et des solutions** doit aider à les localiser rapidement.
- Ce référentiel doit être facile d'accès et rapide de consultation pour toutes les personnes qui peuvent en avoir besoin.
- Le processus de résolution des problèmes est dans la mesure du possible décrit à l'avance et identifie un responsable du risque pour prendre les décisions en urgence si nécessaire.

Communication sur la gestion des risques

- La gestion des risques peut aider à la négociation entre l'archive et les producteurs, en faisant apparaître de façon quantifiée les risques qui concernent leurs données et les moyens qu'ils pourraient mettre en œuvre pour faire baisser ce niveau de risque.
- Elle permet aussi à l'archive de limiter ses engagements en termes de préservation si elle considère que les moyens mis à sa disposition par le producteur sont insuffisants pour garantir la préservation à long terme (ex. métadonnées insuffisantes, formats propriétaires, mesures de protection technique, etc.)
- Communiquer sur les risques est un gage de transparence pour établir la confiance avec les producteurs et les utilisateurs ; cette communication sera exigée dans les processus de certification
- Le document de sensibilisation des producteurs ne doit pas contenir d'informations confidentielles dont la diffusion pourrait faire augmenter le niveau de risque (par ex. il ne doit pas détailler les vulnérabilités du système de sécurité car les failles seraient plus faciles à exploiter).

Gestion des risques et audit

- On peut auditer le système en vue de la certification, ou en vue de prioriser les actions.
- La gestion des risques donne une vue d'ensemble des actions de préservation menées sur l'ensemble des objets numériques.
- Trois initiatives de certification/audit fortement convergentes mais qui restent distinctes pour des raisons géopolitiques :
  - DRAMBORA
  - RLG Checklist
  - NESTOR checklist
- Travaux du CCSDS : *ISO standard for Audit and Certification of repositories of digital information*

http://www.repositoryaudit.eu/

DRAMBORA

- "Digital Repository Audit Method Based on Risk Assessment"
  - Réalisé par le Digital Curation Centre (UK) et le projet européen DPE (Digital Preservation Europe)
  - Première version parue fin février 2007
- DRAMBORA est un tutoriel détaillé :
  - Pour l'auto-évaluation d'un entrepôt numérique
  - Méthodologie basée sur la gestion des risques
  - Fournit à la fois un guide rédigé et des formulaires pour procéder à l'évaluation
  - Prend en compte les acquis des deux autres initiatives

http://www.crl.edu/PDF/trac.pdf

TRAC

- Trustworthy Repositories Audit & certification (TRAC): Criteria and Checklist
  - fruit d'un travail de RLG (Research Liraries Group) et du CRL (Center for Research Libraries
  - Objectif = certification d'un entrepôt de préservation conforme à la norme OAIS
- une liste de critères
  - Notion de confiance – *trust* – qui est garantie par la transparence
  - Trois parties :
    - aspects organisationnels
    - gestion des objets numériques
    - infrastructure technique et sécurité

CCSDS

- Dans la lignée de l'OAIS, le CCSDS travaille sur des normes complémentaires :
  - PAIMAS pour les protocoles de versement entre les producteurs et les archives
  - Projet de norme pour la certification des archives en conformité avec l'OAIS
  - L'objectif est de mettre en place l'ensemble des normes ISO concourrant à la préservation à long terme des objets numériques
- Le projet est à l'état d'ébauche, un premier *draft* de livre blanc est prévu pour mars 2008

Conclusion

- La théorie :
  - La méthode de gestion des risques, en particulier appliquée à la préservation numérique, est bien définie et a déjà fait ses preuves
- La pratique :
  - Pour compléter ce modèle théorique, il faut une organisation :
    - Qui va procéder aux audits ?
    - Qui va itérer la gestion des risques ?
    - Comment la gestion des risques influera-t-elle sur le management ?
    - Quels seront les moyens pour gérer les risques ?