



**Titre:** Conception de procédures de gestion des données techniques : le cas de l'Agence Spatiale Canadienne  
**Title:**

**Auteur:** Benoît Bélanger-Roy  
**Author:**

**Date:** 2011

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Bélanger-Roy, B. (2011). Conception de procédures de gestion des données techniques : le cas de l'Agence Spatiale Canadienne [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/770/>  
**Citation:**

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/770/>  
**PolyPublie URL:**

**Directeurs de recherche:** Mario Bourgault  
**Advisors:**

**Programme:** Génie industriel  
**Program:**

UNIVERSITÉ DE MONTRÉAL

CONCEPTION DE PROCÉDURES DE GESTION DES DONNÉES TECHNIQUES :  
LE CAS DE L'AGENCE SPATIALE CANADIENNE

BENOÎT BÉLANGER-ROY

DÉPARTEMENT DE MATHÉMATIQUES ET GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE INDUSTRIEL)

DÉCEMBRE 2011

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

CONCEPTION DE PROCÉDURES DE GESTION DES DONNÉES TECHNIQUES :  
LE CAS DE L'AGENCE SPATIALE CANADIENNE

Présenté par : BÉLANGER-ROY Benoît

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. TRÉPANIÉ Martin, Ph.D., président

M. BOURGAULT Mario, Ph.D., membre et directeur de recherche

Mme SAINT-MARSEILLE Josée, M.B.A., membre

## REMERCIEMENTS

J'aimerais tout d'abord remercier Josée Saint-Marseille, Daniel Marion, Sonia Dupuis et Geneviève Michaud de l'Agence spatiale canadienne. Merci de m'avoir accueilli au sein de votre équipe, de m'avoir accompagné dans mes moindres déplacements à l'Agence et merci pour les nombreuses rencontres qui m'ont permis de mieux comprendre la problématique à laquelle vous êtes confrontés, l'environnement dans lequel vous évoluez et qui ont grandement alimenté ma réflexion.

Je tiens aussi à remercier toutes les personnes rencontrées en entrevue pour leur ouverture et pour le temps qu'ils m'ont accordé. J'espère que ces rencontres ont été aussi bénéfiques pour eux qu'elles l'ont été pour moi.

Merci à M. Martin Trépanier d'avoir accepté de présider mon jury.

Je tiens finalement à offrir mes plus sincères remerciements à M. Mario Bourgault, mon directeur de recherche, pour sa disponibilité, son encadrement et surtout ses judicieux conseils qui m'ont permis de mener à terme cette maîtrise.

## RÉSUMÉ

Le développement incessant de l'Internet et des technologies de l'information engendre une multiplication des ensembles de données techniques provenant de mesures expérimentales, d'observations et de calculs. Comme ces technologies facilitent la collaboration et le partage d'information au sein des communautés scientifique et technique, il devient de plus en plus facile de créer, partager et exploiter des ensembles de données.

Étant donné le potentiel de ces données, la plupart des organisations publiques, en particulier celles à vocation scientifique, se préoccupent de plus en plus de leur saine gestion. Cette recherche s'intéresse à l'une d'entre elles, l'Agence spatiale canadienne (ASC).

Suivant une méthodologie de recherche-intervention, le chercheur s'est intégré au groupe de travail sur la gestion des données, au sein même de l'ASC. Par le biais d'une immersion sur le terrain et d'une longue série d'entrevues, le chercheur a développé un cycle de vie des données spatiales, une liste de rôles et responsabilités ainsi que des analyses et recommandations sur les modes d'implantation et les conditions particulières liées au contexte spécifique de l'ASC. À cette étude sur le terrain s'ajoute une vaste étude comparative des pratiques de gestion de données de multiples pays et agences spatiales étrangères.

Ce mémoire démontre l'importance des rôles de créateur/utilisateur, curateur, mais aussi d'administrateur sur l'ensemble du cycle de vie des données. Il met aussi en opposition des concepts comme les intérêts commerciaux ou scientifiques avec des pratiques enviabiles comme la libre diffusion des données provenant de fonds publics.

## ABSTRACT

The constant evolution of the information technologies and of the Internet causes a multiplication of the technical datasets coming from experimental measurements, observations and calculations. Since these technologies facilitate the collaboration and the sharing of the information beneath the technical and scientific communities, it becomes much easier to create, share and utilize datasets.

Considering the potential of the data, most public organisations are concerned with increasing their stewardship, particularly the organisations that have a scientific vocation. This research is focussing on one of them, the Canadian Space Agency (CSA).

Following a research-intervention methodology, the researcher integrated the working group on data management, inside the CSA. Through an immersion on the field and a long series of interviews, the researcher has developed a life cycle for space data, a list of roles and responsibilities as well as analysis and recommendations on the possible implantation patterns and a description of the specific conditions related to the particular context of the CSA. In addition to this field study, this research includes an extensive comparative study of data management practices from multiple countries and foreign space agencies.

This *mémoire* demonstrate the significance of the creator/user, curator but also of the administrator during the whole data life cycle. It also opposes some concepts like the commercial of scientific interests to some presumably beneficial practices like the free and open access to research data from public founding.

## TABLE DES MATIÈRES

REMERCIEMENTS .....	iii
RÉSUMÉ.....	iv
ABSTRACT .....	v
TABLE DES MATIÈRES .....	vi
LISTE DES FIGURES.....	x
LISTE DES SIGLES ET ABRÉVIATIONS .....	xi
LISTE DES ANNEXES.....	xiii
INTRODUCTION.....	1
CHAPITRE 1    REVUE DE LITTÉRATURE .....	4
1.1    Problématique.....	4
1.1.1    Nouvelles méthodes de recherche .....	4
1.1.2    Caractéristiques des données techniques.....	5
1.1.3    Conservation des données .....	10
1.2    Agence spatiale canadienne .....	16
1.2.1    Description de l'organisation .....	16
1.2.2    Gestions des données spatiales et scientifiques à l'ASC.....	18
1.3    Situation canadienne .....	21
1.3.1    Initiatives.....	21
1.3.2    Ententes internationales.....	24
1.4    Situation internationale .....	25
1.4.1    Royaume-Uni .....	25
1.4.2    États-Unis.....	28

1.4.3	Autres pays .....	30
1.5	Autres agences spatiales .....	30
1.5.1	NASA .....	30
1.5.2	ESA .....	31
1.6	Conclusion .....	32
CHAPITRE 2 DESIGN DE LA RECHERCHE .....		34
2.1	Objectifs de la recherche .....	34
2.1.1	Objectifs spécifiques .....	34
2.2	Justification du type de recherche .....	35
2.3	Contexte de collaboration Polytechnique-ASC .....	37
2.4	Description des étapes .....	37
2.4.1	Définition de la problématique de recherche .....	40
2.4.2	Revue de littérature .....	40
2.4.3	Élaboration du cycle de vie des données spatiales de l'ASC .....	41
2.4.4	Élaboration d'un plan d'entrevue .....	42
2.4.5	Entrevues .....	42
2.4.6	Élaboration du rapport sur la gestion des données spatiales .....	44
2.4.7	Élaboration du gabarit de plan de gestion des données scientifiques .....	45
2.4.8	Analyse .....	45
2.5	Conclusion .....	45
CHAPITRE 3 RÉSULTATS .....		47
3.1	Synthèse de l'information recueillie .....	47
3.1.1	Données de l'ASC .....	48
3.1.2	Conditions particulières .....	53



3.1.3	Critères à respecter .....	58
3.2	Résultats .....	59
3.2.1	Cycle de vie .....	59
3.2.2	Rôles et responsabilités .....	68
3.2.3	Gabarit de plan de gestion .....	77
3.3	Mise en application .....	78
3.3.1	Recommandations .....	78
3.3.2	Mise en œuvre .....	81
3.4	Validation .....	88
3.5	Conclusion.....	88
CHAPITRE 4	DISCUSSION .....	89
4.1	Contributions théoriques .....	89
4.1.1	Gestion selon le cycle de vie .....	89
4.1.2	Intégration de la conservation dans le processus de recherche .....	90
4.1.3	Préservation .....	91
4.1.4	Accessibilité .....	91
4.2	Contributions pratiques .....	92
4.2.1	Comparaison des différents modèles.....	93
4.2.2	Valeur pratique pour l'organisation .....	94
4.3	Limites de la recherche .....	95
4.3.1	Problématiques complémentaires à étudier.....	95
4.3.2	Implantation .....	96
4.4	Avenues de recherche et de développement futur.....	96
4.4.1	Autres types d'actifs.....	96

4.4.2	Autres organismes .....	97
4.4.3	Prochaines étapes .....	97
4.4.4	Sujets connexes .....	98
CONCLUSION .....		99
LISTE DES RÉFÉRENCES .....		101
ANNEXES .....		107

## LISTE DES FIGURES

Figure 1-1 : Exemple de fiche bibliographique créée à l'aide du logiciel EndNote X4.....	7
Figure 1-2 : Mosaïque du Québec réalisé avec les données de RADARSAT-1[11] .....	9
Figure 1-3 : Flux d'information : Troisième niveau de conservation[3] .....	13
Figure 1-4 : Cycle de vie de l'information de BAC et de l'ASC tel que présenté dans la version 2011 du Cadre de gestion de l'information de l'ASC[19] .....	20
Figure 1-5 : Cycle de vie des données de recherche canadiennes du Groupe de travail sur la stratégie des données de recherche[24]. .....	22
Figure 1-6 : Cycle de vie des données scientifiques tel que présenté par le DCC[36] .....	27
Figure 2-1 : Démarche utilisée dans le cadre de cette recherche .....	39
Figure 2-2 : Provenance de données de la présente recherche.....	41
Figure 2-3 : Organigramme de l'ASC indiquant la répartition des personnes rencontrées .....	43
Figure 3-1 : Cycle de vie des données spatiales de l'ASC.....	60
Figure 3-2 : Responsabilités des créateurs et utilisateurs dans le cycle de vie des données .....	69
Figure 3-3 : Responsabilités des curateurs dans le cycle de vie des données .....	72
Figure 3-4 : Responsabilités des administrateurs dans le cycle de vie des données .....	75
Figure 3-5 : Analyse du modèle d'archive des données dans une agence comme la NASA ou l'ESA .....	84
Figure 3-6 : Analyse du modèle d'archive des données dans une université.....	84
Figure 3-7 : Analyse du modèle d'archive des données dans un organisme du gouvernement canadien.....	85
Figure 3-8 : Analyse du modèle d'archive des données à l'ASC.....	87

## LISTE DES SIGLES ET ABRÉVIATIONS

AAP	Architecture d'activités de programmes
ASC	Agence spatiale canadienne
ASCII	<i>American Standard Code for Information Interchange</i>
BAC	Bibliothèque et Archives Canada
CCDA	Centre canadien des données astronomiques
CCT	Centre canadien de télédétection
CNC/CODATA	Comité national canadien pour CODATA
CSA	<i>Canadian Space Agency</i>
CODATA	<i>Committee on Data for Science and Technology</i>
CRSH	Conseil de recherches en sciences humaines du Canada
CRSNG	Conseil de recherches en sciences naturelles et en génie du Canada
DCC	<i>Digital Curation Center</i>
DVD	<i>Digital Versatile Disc</i>
ESA	Agence spatiale européenne ( <i>European Space Agency</i> )
GI	Gestion de l'information
GPS	<i>Global Positioning System</i>
GTSGD	Groupe de travail sur la gestion des données
ICIST-CNRC	Institut canadien de l'information scientifique et technique du Conseil national de recherches Canada
ISO	Organisation internationale de normalisation
JHU	Université John Hopkins
MATMOS	<i>Mars Atmospheric Trace Molecule Occultation Spectrometer</i>
NASA	<i>National Aeronautics and Space Administration</i>

OAIS	Open Archival Information System
OCDE	Organisation de coopération et de développement économiques
PDS	<i>Planetary Data System</i>
RCM	<i>RADARSAT Constellation Mission</i>
TI	Technologies de l'information
URL	<i>Uniform Resource Locator</i>
VHS	<i>Video Home System</i>
XML	<i>Extensible Markup Language</i>

## **LISTE DES ANNEXES**

ANNEXE A – PLAN D’ENTREVUES.....	107
ANNEXE B – GABARIT DE PLAN DE GESTION DES DONNÉES.....	112
ANNEXE C – LISTE DES ENTREVUES.....	113
ANNEXE D – ÉLARGISSEMENT DU CYCLE DE VIE.....	116

## INTRODUCTION

Les données techniques sont la base de toutes sciences empiriques. Les principaux extrants de la plupart des expériences scientifiques et de nombreux projets techniques sont des données brutes et transformées. Leur acquisition se fait souvent à fort prix ce qui amène à se demander si ces données sont exploitées à leur plein potentiel.

De plus, on constate désormais qu'un usage insoupçonné peut apparaître longtemps après la première utilisation des données. Par exemple, personne n'aurait pu prévoir que les données d'observation des glaces prises il y a plusieurs décennies serviraient un jour à quantifier le réchauffement climatique.

Chaque année, il devient de plus en plus facile et financièrement accessible d'enregistrer de larges quantités d'information. De plus, le développement incessant de l'Internet et des réseaux de télécommunication facilite le transfert toujours plus rapide de cette information.

Afin d'éviter les délais et le coût lié à l'acquisition des données, les chercheurs sont de plus en plus poussés à s'échanger des données. Les collaborations multidisciplinaires et le partage de données sont de plus en plus fréquents dans le monde scientifique. Ces collaborations facilitent et accélèrent le progrès scientifique et, au bout de la ligne, tout le monde tire profit de cette ouverture.

De nombreuses institutions ont compris les bénéfices de la conservation et du partage des données techniques et ont emboîté le pas en ce sens. C'est maintenant la classe politique qui entrevoit la nécessité et les avantages d'une ouverture et de la libre diffusion des données acquises à partir de fonds publics. Le Canada n'est pas en reste, il est signataire de nombreuses ententes internationales visant le partage des données publiques et a entrepris divers projets et études afin d'évaluer comment il serait possible, dans son contexte, de faire la meilleure exploitation de son actif informationnel.

En tant qu'agence gouvernementale canadienne, et étant donné sa proximité avec le milieu de la recherche, il est dans l'intérêt de l'Agence spatiale canadienne (ASC) de suivre cette nouvelle tendance et de tenter de faire le meilleur usage de ses données techniques. De plus, l'ASC encourage fortement la conservation et la libre diffusion de ses données. C'est donc dans ce contexte que la présente recherche s'est effectuée.

Suivant les principes de la recherche-intervention, l'auteur de ce mémoire s'est joint à une équipe de travail de l'ASC afin d'analyser les pratiques de gestion des données techniques de cet organisme public et d'évaluer les pratiques enviables pour en faire la gestion ainsi que les éléments pouvant nuire à l'implantation de telles procédures.

Grâce à une vaste revue des publications académiques, des initiatives internationales et des pratiques de gestion des données de nombreux pays, organismes et agences spatiales étrangères et grâce à de nombreuses rencontres avec les intervenants de la gestion des données spatiales de l'ASC, de nombreux outils ont pu être développés pour aider l'ASC à surmonter le défi de taille que représente la gestion des données techniques.

Cette recherche permet tout d'abord de cerner le contexte dans lequel se fait la gestion des données spatiales de l'ASC, les caractéristiques de ces données et les conditions particulières avec lesquelles l'ASC doit composer. Elle propose un cycle de vie spécifique aux données spatiales de l'ASC. Elle expose les rôles et responsabilités des différents intervenants de la gestion des données, suggère un plan de gestion des données pour l'ASC et fournit des recommandations et des analyses afin d'évaluer les possibilités de mise en œuvre de ces nouvelles procédures.

Cette étude pose ensuite un regard critique sur ces outils pratiques et démontre que leur conception contribue bien plus que simplement au projet d'amélioration des procédures de gestion de données de l'ASC. En appliquant les concepts théoriques de la gestion des données dans un contexte complexe et aussi particulier que celui de l'ASC, cette recherche vient confronter les modèles théoriques avec la réalité du terrain. Cette confrontation permet de solidifier plusieurs éléments de la théorie mais aussi d'en améliorer et d'en nuancer d'autres.

Ce mémoire se divise en quatre chapitres qui respectent la structure traditionnelle d'un travail de recherche. Le Chapitre 1 dresse la table en présentant une revue des concepts théoriques présents dans la littérature. S'ajoute à cette revue théorique un recensement des différentes initiatives internationales relatives à la gestion des données scientifiques et techniques et une comparaison des pratiques de gestion des données du gouvernement canadien et de l'ASC par rapport à d'autres pays, organismes et agences spatiales ayant une forte influence sur la scène internationale.



Le Chapitre 2 enchaîne avec une description de la méthodologie et du contexte dans lequel fut effectuée cette recherche. Il énumère tout d'abord les objectifs de la recherche. Il s'appuie ensuite sur différents auteurs pour justifier l'emploi des méthodes de recherche-intervention pour la réalisation de ce travail et termine par une description exhaustive de chacune des étapes de ce processus de recherche.

Le Chapitre 3 présente, pour sa part, les différentes observations effectuées lors de la collecte d'information ainsi que les résultats de la recherche. Ce chapitre se construit autour du cycle de vie des données spatiales. Il commence par une description du contexte dans lequel se fait la gestion des données de l'ASC. Ce contexte comprend les particularités des données et les conditions particulières qui ont été décelées lors des entrevues. Suit une description des outils (cycle de vie, rôles et responsabilités, gabarit de plan de gestion des données), puis une analyse des méthodes de mise en application des nouvelles procédures de gestion accompagnée d'une série de recommandations. Ce chapitre des résultats se termine par une description des étapes de validation des résultats du chapitre.

Ces résultats sont finalement analysés dans le Chapitre 4, qui présente la contribution de la présente recherche, ses limitations ainsi que les différentes avenues de recherche. L'auteur y explique comment les outils et analyses du chapitre précédent sont bénéfiques pour l'ASC mais aussi pour la communauté scientifique et, à un certain niveau, pour l'ensemble des citoyens canadiens. Ce mémoire se termine par une présentation des limites de cette étude et des avenues de recherche qui pourraient être explorées.

## **CHAPITRE 1 REVUE DE LITTÉRATURE**

Depuis la nuit des temps, les organisations sont portées par l'évolution technologique, constamment poussées à revoir leur façon de faire et à s'adapter aux nouvelles réalités qui les entourent. Tout comme l'environnement de l'organisation qui est en perpétuel changement, les mentalités évoluent elles aussi. De nouveaux enjeux apparaissent constamment et viennent changer notre perception du monde et nos priorités.

Les nouvelles technologies de traitement et de partage de l'information amorcent une de ces révolutions. Les communautés politiques et scientifiques du monde entier commencent à constater la richesse que représente l'information et, maintenant qu'ils en ont techniquement les moyens, ils désirent exploiter pleinement cette richesse.

Ce premier chapitre décrit ce nouveau courant de pensée qui promeut la pleine exploitation des données techniques ainsi que leur libre partage. Les concepts théoriques de la gestion des données sont ensuite présentés et le chapitre se termine avec une analyse comparative des procédures de gestion et des initiatives d'amélioration de la conservation et du partage des données de différents pays, agences ou organisations. L'ensemble de cette analyse s'articule autour d'une agence gouvernementale canadienne œuvrant dans le domaine de la science et de la haute technologie, l'Agence spatiale canadienne (ASC).

### **1.1 Problématique**

#### **1.1.1 Nouvelles méthodes de recherche**

Que ce soit dans le domaine des sciences naturelles ou sociales, les données constituent l'élément fondamental du cheminement scientifique. Les sciences empiriques ont comme principe d'observer un sujet ou un environnement, et de tenter d'en expliquer le fonctionnement. La donnée constitue cette observation à partir de laquelle se construit une théorie scientifique.

Dans le domaine scientifique, une donnée est définie comme étant « [...] l'élément fondamental qui permet de caractériser sous une forme quantitative ou qualitative toute forme de manifestation. »[1] Les données n'ont pas de signification en soi, c'est seulement lorsqu'elles sont interprétées, par un être humain, ou par un système, qu'elles prennent un sens et qu'elles deviennent de l'information.

Avec la démocratisation de l'informatique, la majorité des données scientifiques se présentent maintenant sous la forme numérique. La technologie accélère aussi le rythme avec lequel les données scientifiques sont générées à travers le monde. Le développement incessant des technologies de l'information permet d'entreposer et de traiter toujours plus de données. Certains chercheurs estiment que la quantité totale de données scientifiques double presque chaque année[2]. Parallèlement à cette augmentation de la quantité de données, l'Internet permet désormais d'échanger très facilement des données, et ce, sans se soucier de la distance géographique et des frontières.

Peu à peu, la quantité de données partagées et la facilité de partage de ces données permettent l'avènement d'une nouvelle culture de recherche. Les collaborations deviennent de plus en plus fréquentes et multidisciplinaires. Les chercheurs mettent en commun leurs ressources, leur expertise et s'échangent leurs données dans le but de favoriser le progrès scientifique[3]. Cette culture qui émerge dans la communauté scientifique a été baptisée « *E-Science* ». La définition qu'en fait le *Department of Trade and Industry* du Royaume-Uni et celle qui est la plus répandue dans la littérature est la suivante: « *science increasingly done through distributed global collaborations enabled by the Internet, using very large data collections, tera-scale computing resources and high performance visualization.* »[4]

L'*E-science* vient redéfinir l'utilisation qui est faite des données. La communauté scientifique s'ouvre de plus en plus, le partage et la réutilisation des données sont de plus en plus fréquents. Il s'avère donc nécessaire de repenser la gestion des données et d'en favoriser la libre circulation.

### **1.1.2 Caractéristiques des données techniques**

Il est essentiel, avant d'aller plus loin, de bien définir les données dont il sera question dans ce mémoire, quelles sont leurs caractéristiques et quels en sont les différents types. Dans les chapitres qui suivent le terme « données techniques » est employé pour désigner toute donnée provenant de mesures expérimentales, d'observations et de calculs dans tous les champs de la science et des technologies. Ce type de données s'oppose entre autres aux données administratives dont l'utilité est liée uniquement à la gestion. Les données techniques désignent donc les données dont le traitement ou l'analyse engendre une meilleure compréhension d'un phénomène naturel, technologique ou humain.

Les données peuvent prendre de nombreuses formes. Dans la quasi-totalité des cas, il s'agit d'enregistrements numériques qui peuvent avoir de nombreux formats, provenances et contextes de saisie. De plus, les données sont rarement présentées seules. Elles appartiennent généralement à des ensembles de données (*Datasets*) qui rassemblent des données ayant un même format et une même provenance. C'est par l'analyse et la comparaison d'un ou de plusieurs ensembles de données que les scientifiques réussissent à générer de la connaissance.

Tout comme les données, les ensembles de données ont eux aussi des caractéristiques qui leur sont propres[5]. Ils peuvent être de différentes tailles et de différents niveaux de complexité. Un ensemble de données peut être composé d'un seul ou d'une multitude de fichiers reliés entre eux. Il peut être stable, mais il peut aussi être en construction, c'est-à-dire que des données s'y ajoutent constamment. Il est aussi possible que l'ensemble de données soit dynamique c'est-à-dire que sa taille est constante, mais que les données qu'il contient sont remplacées de façon cyclique. Finalement, un ensemble de données peut être dépendant d'un logiciel particulier ou d'un système lorsque vient le temps d'en faire la lecture et l'interprétation.

Pour pouvoir interpréter adéquatement un ensemble de données, il est nécessaire que celui-ci soit accompagné de métadonnées. Les métadonnées sont des données qui portent sur des éléments de données[1]. Ce sont des fichiers numériques qui viennent documenter le format, la provenance, le contexte de saisie et toutes autres informations nécessaires à l'interprétation et à l'utilisation des données. Les métadonnées sont généralement encodées dans un langage pouvant être lu directement par un ordinateur.

Les métadonnées sont utiles à la fois pour décrire un ensemble de données mais aussi pour faciliter sa recherche. Un exemple commun de métadonnées est l'utilisation des fiches bibliographiques dans les bibliothèques pour décrire les livres ou articles contenus dans un répertoire. Ces fiches fournissent de l'information sur la publication (auteur, titre, années, localisation, etc.) mais contiennent aussi des mots clés permettant à des outils de recherche de trouver des ouvrages pertinents.

Une multitude de normes existe pour documenter adéquatement les ensembles de données. L'absence de modèle unique de métadonnées s'explique par la très grande diversité des ensembles de données. Même au sein d'un même domaine scientifique, la forme des ensembles de données peut varier considérablement. Prenons l'exemple des sciences planétaires. Pour

décrire une observation d'un phénomène très lointain dans l'univers, un chercheur pourrait noter des informations comme la date de l'observation, l'heure et la position dans le ciel (latitude et longitude). Par contre, si le phénomène observé se trouve sur la surface de la planète Mars, il serait préférable de noter la position selon les coordonnées polaires et l'heure martiennes et non selon les références terrestres. Le format de ces deux ensembles de données et les métadonnées qui les accompagnent seront donc considérablement différents.

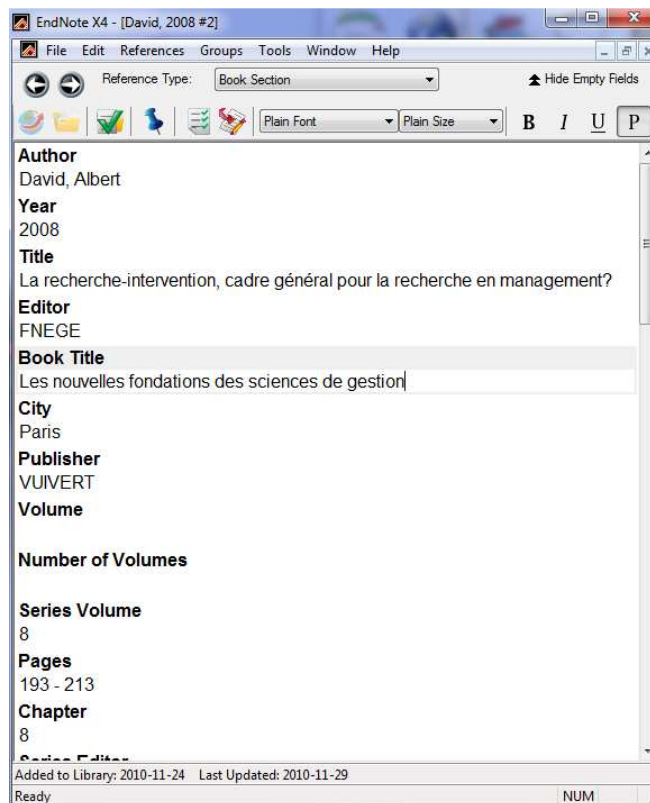


Figure 1-1 : Exemple de fiche bibliographique créée à l'aide du logiciel EndNote X4

Voici trois exemples de normes répandues pour les métadonnées utilisées pour les données scientifiques ou géospatiales:

- *Open Archival Information System (OAIS) reference*[6];
- *ISO 19115, 2003, Geographic information -- metadata*[7];
- *ISO 15836, 2009, Information and documentation -- The Dublin Core metadata element set*[8].

Les ensembles de données peuvent prendre une infinité de formes, mais ils peuvent tout de même être regroupés en familles ou types particuliers. Il est important pour cette recherche de bien distinguer deux types d'ensembles de données techniques : les données d'expérience et les données géospatiales. Malgré cette catégorisation, tous les ensembles de données qui se retrouvent au sein d'un même groupe ne sont pas identiques, loin de là, mais ils présentent certaines caractéristiques communes fondamentales qui en permettent le regroupement.

### **1.1.2.1 Données d'expérience**

Les ensembles de données d'expérience regroupent toutes les données recueillies dans le cadre d'une recherche scientifique. L'OCDE les décrit comme des « *factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.* »[9] Les données d'expériences sont recueillies pour une utilisation précise, généralement pour décrire un phénomène ou pour vérifier une hypothèse. Les ensembles de données d'expérience sont généralement stables dans le temps et relativement faciles à manipuler. Leur partage et leur réutilisation s'avèrent donc assez simples. Leur provenance peut varier grandement. Ces données peuvent être créées expressément aux fins d'une recherche, récupérées d'une expérience précédente ou d'un ancien projet ou elles peuvent être obtenues en combinant ou en comparant des ensembles de données déjà existants.

Les premières archives numériques de ce type de données ont vu le jour dans les années 1960[4]. Elles ont été créées par les universités qui désiraient conserver les résultats des sondages effectués dans le domaine des sciences sociales. Depuis ce temps, l'archivage des données numériques de recherche s'est étendu à l'ensemble des domaines scientifiques.

### **1.1.2.2 Données géospatiales**

Par comparaison au type précédent, les ensembles de données géospatiales sont tous des ensembles comportant un élément géographique. Ces ensembles de données visent à caractériser un territoire. Leur vocation scientifique est moins directe que pour les données d'expérience puisqu'elles ne sont pas recueillies spécifiquement pour réaliser une expérience. Ceux-ci constituent néanmoins des bases de données essentielles à la science et au développement industriel et technologique.

Les données géospatiales n'ont pas de format unique, mais elles possèdent tout de même plusieurs caractéristiques communes[5]. La plupart d'entre elles dépendent d'un logiciel, parfois spécialisé, ou d'un système particulier afin de pouvoir être lues et interprétées. Les ensembles de données géospatiales sont généralement de très grande taille et résident sur une multitude de fichiers inter-reliés ce qui complique leur manipulation[10]. Très souvent, à cause de la longue durée de leur collecte, il s'avère nécessaire de commencer à archiver les données géospatiales avant que l'ensemble de données ne soit complet. Parfois, certains ensembles de données géospatiales peuvent même être dynamiques, c'est-à-dire que leur contenu est constamment renouvelé. Les bases de données météorologiques en sont un bon exemple. Ce sont donc des ensembles de données très complexes et difficiles à manipuler.

Les données géospatiales sont de différents niveaux selon les transformations qui y ont été apportées. Plusieurs normes peuvent être utilisées pour décrire ces niveaux. Les données recueillies sont très nombreuses et elles sont souvent combinées sous forme de produits ayant un intérêt commercial ou scientifique. Ces produits présentent généralement les données sous forme de cartes qui sont plus faciles à lire et à interpréter. La Figure 1-2 présente un exemple de produit réalisé à partir des données du satellite canadien RADARSAT-1.



Figure 1-2 : Mosaïque du Québec réalisé avec les données de RADARSAT-1[11]

Dans ce mémoire, les termes « données de recherche » ou « données d'expériences » sont utilisés pour faire référence aux ensembles de données d'expériences et le terme « données d'observation de la Terre » est utilisé pour les données géospatiales. Le terme « données techniques » regroupe l'ensemble des données numériques utiles à la génération de connaissances scientifiques et techniques, c'est-à-dire les données de recherche et les données géospatiales.

### **1.1.3 Conservation des données**

Les données sont de plusieurs types et ont de nombreuses caractéristiques, mais elles ont aussi une valeur. Il est donc important de les conserver adéquatement afin de pouvoir en exploiter le plein potentiel. La valeur des données peut être définie selon différents facteurs ; voici une brève description de quatre types de valeurs pouvant être attribuées aux données.

- Il y a tout d'abord la valeur monétaire liée au coût d'acquisition de la donnée. Cette valeur représente le coût qui serait nécessaire pour refaire l'acquisition des mêmes données. Cette valeur peut diminuer avec le temps et avec l'évolution des technologies utilisées pour la saisie des données. Cette valeur monétaire incite à préserver et à exploiter pleinement les données afin d'en rentabiliser l'acquisition ;
- Il y a ensuite une valeur historique rattachée à certaines données. C'est le cas des données qui décrivent ou correspondent à un environnement temporel, circonstanciel ou pouvant être identifié dans le moment présent comme étant historique. Ces données d'observation ont une valeur particulière due au fait qu'elles ne peuvent pas être recréées ;
- Aussi, les données peuvent avoir une valeur scientifique. Celle-ci correspond à la quantité de savoir qui peut être généré à partir des données. Il n'est pas possible de connaître à l'avance la valeur scientifique exacte d'un ensemble de données ni de la chiffrer ;
- Finalement, les données peuvent avoir une valeur opérationnelle. Celle-ci ressemble beaucoup à la valeur scientifique en ce sens qu'elle reflète l'ensemble des utilités potentielles des données. La valeur opérationnelle représente l'ensemble des tâches ou opérations qui peuvent ou pourront être réalisées grâce aux données.

Des facteurs comme la rareté ou l'unicité des données ont aussi un impact majeur sur la valeur qui est accordée aux données. De même, la valeur des données est généralement liée à leur degré de lisibilité et d'accessibilité ; une donnée perdue, cachée ou illisible n'aurait plus aucune utilité.



Étant donné leur valeur, il est donc essentiel de faire la conservation des données. Lord et Macdonald définissent la conservation (*curation*) comme étant l'activité consistant à gérer et faire la promotion des données depuis leur création jusqu'à leur élimination, en s'assurant qu'elles sont adaptées aux besoins contemporains et qu'elles sont disponibles pour être réutilisées.[3, 12] La conservation représente donc beaucoup plus que le simple entreposage des données, elle implique de s'assurer qu'elles restent pertinentes avec le temps et de promouvoir et faciliter leur réutilisation et ce, peu importe dans quel domaine et par quel utilisateur.

Quatre éléments clés de la conservation peuvent être tirés des recherches de Yakel[13]: la gestion sur l'ensemble du cycle de vie, l'implication à la fois des curateurs et créateurs des données, la préservation et l'accessibilité.

### **1.1.3.1 Gestion selon le cycle de vie**

Le principe de la gestion des données selon leur cycle de vie est d'identifier toutes les étapes par lesquelles passent les données afin d'en assurer la gestion, depuis la planification de leur création jusqu'à leur élimination. Ce principe suggère une attention portée sur l'ensemble des étapes du cycle de vie des données et non pas uniquement sur les étapes immédiatement utiles à l'utilisateur des données. Ainsi, l'approche selon le cycle de vie redonne de l'importance à la planification de la gestion des données et à leur gestion à long terme.

Différents modèles de cycle de vie des données existent. Ceux-ci varient beaucoup selon l'éventail de données qu'ils considèrent, selon le contexte et les contraintes de l'organisation qui en fait la gestion ou selon le niveau d'abstraction désiré. En effet, certains cycles de vie peuvent être de nature plus opérationnelle et prendre en compte une multitude de cas précis d'utilisation des données, tandis que d'autres restent à un niveau d'abstraction relativement élevé et visent moins une description précise d'actions à prendre. Les Figure 1-4, Figure 1-5 et Figure 1-6 sont des bons exemples de cycles de vie. Ils seront analysés plus loin dans ce chapitre.

Pour bien faire la conservation des données techniques, il est essentiel d'avoir une vision de l'ensemble du cycle de vie de ces données afin de pouvoir anticiper les besoins et de planifier adéquatement chacune des phases de la vie des données.

### 1.1.3.2 Intégration de la conservation dans le processus de recherche

Un autre élément essentiel à la bonne gestion des données scientifiques est l'intégration de la conservation au processus de recherche, c'est-à-dire l'action de lier les objectifs de la conservation à ceux de la recherche.

Dans l'*E-Science Curation Report*[3], Lord et MacDonald présentent trois niveaux distincts d'intégration des activités de conservation dans le processus de recherche classique. Le premier modèle présente simplement le processus de recherche et les rares activités de conservation des données qui s'y retrouvent ne sont là que pour permettre l'atteinte des objectifs de la recherche.

Le deuxième niveau de conservation est marqué par l'apparition d'un nouvel intervenant, l'archiviste. Celui-ci œuvre à l'organisation et à l'archivage des données du chercheur. Les tâches effectuées par l'archiviste nécessitent une certaine participation du chercheur. Elles visent à assurer la bonne préservation des données et à permettre leur réutilisation. Ce deuxième modèle contient des tâches d'organisation et d'archivage qui n'ajoutent rien à la recherche en cours, mais répondent plutôt à une vision à long terme de la vie utile des données.

C'est le troisième niveau de conservation qui présente la vision la plus complète de l'intégration des tâches de conservation au processus de recherche. Dans ce dernier modèle, l'archiviste est remplacé par un curateur. Le rôle de ce dernier va au-delà de la simple préservation des données, il a aussi la responsabilité de diffuser les données et d'en promouvoir la réutilisation. Le curateur doit travailler de concert avec le chercheur afin que ses données puissent être exploitées pleinement. Le troisième niveau de préservation de Lord et MacDonald est présenté à la Figure 1-3. On y distingue le processus de recherche classique et son entrelacement avec le processus de conservation des données. Ce modèle démontre clairement que les objectifs de la conservation des données vont bien au-delà des objectifs de la recherche en cours et que la collaboration du créateur des données et du curateur est indispensable à la bonne conservation de celles-ci.

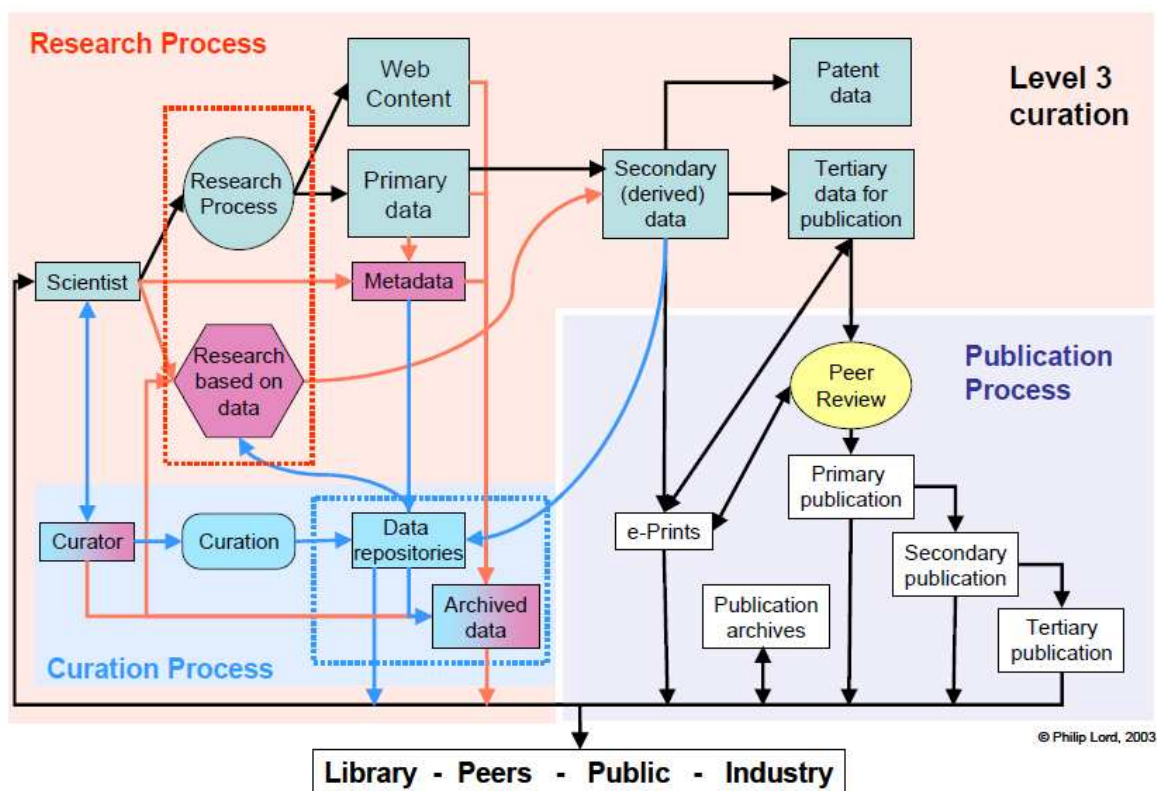


Figure 1-3 : Flux d'information : Troisième niveau de conservation[3]

### 1.1.3.3 Préservation

Afin de pouvoir réutiliser les données, il est essentiel de les archiver c'est-à-dire, les sélectionner et les entreposer adéquatement, mais aussi s'assurer que leur intégrité et leur accès soient maintenus à travers le temps[3]. Il ne suffit pas d'entreposer les données pour en faire l'archivage puisque les données sont à la merci des changements technologiques. Ces changements technologiques, qui se produisent de plus en plus rapidement, guettent à la fois les supports matériels et logiciels[4]. Si ces supports sont perdus ou endommagés, les données deviendront inutilisables. Un bon exemple de changement des supports matériels est la vidéocassette. Depuis l'arrivée du DVD, il est de plus en plus difficile de se procurer un lecteur VHS afin d'en lire le contenu. Pour ce qui est des logiciels nécessaires à la lecture et à l'interprétation des données, rien n'indique qu'ils seront compatibles avec les prochaines générations de systèmes d'exploitation de nos ordinateurs. Si les données ne sont pas adéquatement préservées, elles deviendront donc rapidement inaccessibles.

La définition de la préservation la plus répandue dans la littérature est celle de Lord et MacDonald, reprise par le *Digital Curation Center* (DCC) du Royaume-Uni, qui la décrit comme une sous-activité de l'archivage consistant en le maintien de l'accessibilité et de la compréhensibilité de certains éléments des données et ce malgré les divers changements technologiques.[3, 12]

Plusieurs stratégies existent afin de préserver les données. Les deux plus communes sont l'émulation du système et la migration des données[4]. L'émulation consiste à recréer l'environnement du système d'origine dans l'environnement actuel afin de permettre la lecture des données dans leur format d'origine. Cette technique est exigeante mais elle a l'avantage de ne pas modifier les données d'origine. Les données restent intouchées mais des efforts parfois considérables doivent être entrepris afin de concevoir une application qui simule l'environnement technologique d'origine.

La technique de la migration, elle, consiste à transférer ou convertir les données dans un format ou sur un support moderne. Ce transfert doit être effectué à chaque changement technologique sans quoi les données seront inutilisables. Cette méthode est moins laborieuse que l'émulation, mais à chaque transfert, il y a un risque de perdre des données ou de faire des erreurs dans leur retranscription. Il est possible, afin d'atténuer l'effet de certains changements technologiques, de faire la migration des données vers un format indépendant de tout logiciel comme le code ASCII, Unicode ou XML. Différentes variantes de la migration existent, telles que l'encapsulation[14]. Elles impliquent toutes un transfert des données d'un format à un autre.

L'*E-Science Curation Report*[3] présente lui aussi ces deux techniques, mais il décrit trois autres stratégies de préservation, moins efficaces à long terme mais néanmoins très utilisées. La première, qui est en développement, est la description du format. Elle implique le développement d'un Ordinateur virtuel universel. Il s'agit de créer une technologie de référence qui serait facile à recréer à tout moment et de mettre les données dans un langage lisible par cette technologie « universelle ». L'« *Open format* » s'apparente à cette méthode. Il s'agit de développer un format qui sera toujours lisible et d'y crypter les données.

La deuxième, le musée technologique, implique la conservation de l'entièreté des systèmes, matériel et logiciel, requis pour lire un ensemble de données. Cette technique peut sembler très simple, mais elle s'avère généralement très peu pratique puisque les systèmes risquent de se

détériorer avec le temps. L'équipe de curation se retrouve donc à mettre beaucoup d'énergie à faire la préservation de tout un système.

La troisième stratégie de préservation ajoutée par Lord et MacDonald est l'archéologie numérique. Il s'agit de tenter de retrouver le système nécessaire à la lecture des données au moment où l'on a besoin de lire les données. Généralement, plus le système est rare et plus il s'écoule de temps avant qu'on en ait besoin, plus il est difficile de le retrouver. Par exemple, bien qu'ils aient été très répandus jadis, il peut s'avérer difficile de nos jours de mettre la main sur des ordinateurs capables de lire des disquettes de huit pouces. Cette méthode est donc très risquée. De plus, il est impossible de l'appliquer dans le cas d'un système fait sur mesure.

Certaines méthodes éprouvées existent pour décider quelle méthode de préservation est la plus appropriée pour répondre à nos besoins. Ces méthodes complexes utilisent une multitude de critères pour décider de la méthode la plus adéquate. La *PLANETS Preservation Planning approach*[15] en est un exemple.

Il est donc inutile de garder les données à long terme si aucun effort n'est mis dans leur préservation. La stratégie de préservation choisie dépendra principalement de la durée de conservation souhaitée, des ressources disponibles et de la nécessité d'avoir ou non un accès constant aux données.

#### **1.1.3.4 Accessibilité**

Un dernier enjeu majeur de la conservation concerne l'accessibilité des données. Plusieurs raisons existent pour vouloir conserver les données scientifiques[3]. Ce peut être pour des raisons légales, de validation ou pour des raisons historiques. Leur conservation se fait afin que les données scientifiques soient réutilisées, qu'elles soient utilisées pour l'enseignement et que l'information qu'elles représentent puisse bénéficier au public. Ces dernières raisons nécessitent donc qu'il y ait non seulement une préservation des données, mais qu'elles soient rendues accessibles. Cette accessibilité passe par la diffusion des données. Une bonne accessibilité implique une attitude proactive visant à informer les différents milieux qui peuvent tirer bénéfice des données de l'existence de celles-ci. La diffusion peut passer entre autres par des portails web ou des publications scientifiques.

L'accessibilité des données se reflète dans le concept de données libres (*Open Data*) qui est de plus en plus présent dans la communauté scientifique mondiale. La définition des données libres qui est la plus intéressante pour cette recherche est celle proposée par le gouvernement canadien qui définit ce concept comme une pratique visant à rendre les données brutes accessibles gratuitement, par des portails, des métadonnées et des outils de recherche pour qu'elles soient réutilisées par les différents paliers gouvernementaux, les citoyens, les organismes communautaires et le secteur privé de manière nouvelle et non-anticipée[16]. En plus de s'appliquer aux données brutes, ce concept s'étend souvent aux données transformées et aux produits provenant des données.

Le reste de ce chapitre présente les différentes initiatives mises de l'avant par divers organismes et agences gouvernementales, tant au Canada qu'à l'étranger, pour assurer la conservation des données techniques. Ces initiatives comprennent notamment des modèles de cycles de vie et différentes approches et outils de diffusion des données.

Étant donné que le volet empirique de cette recherche fut réalisé au sein de l'ASC, une brève description de cet organisme est proposée avant la présentation de ses initiatives relatives à la gestion des données.

## **1.2 Agence spatiale canadienne**

### **1.2.1 Description de l'organisation**

#### **1.2.1.1 Historique**

Le Canada participe à l'exploration spatiale depuis les années 1960. Après l'Union Soviétique et les États-Unis, le Canada fut le 3<sup>e</sup> pays à concevoir et construire un satellite. C'est en effet en 1962 que le satellite Alouette 1 fut mis en orbite dans le but d'observer l'ionosphère depuis l'espace. Le Canada a conservé depuis ce temps une place toute particulière dans le milieu de l'exploration spatiale[11].

Depuis le 1<sup>er</sup> mars 1989, c'est l'ASC qui a le devoir de coordonner les efforts canadiens dans le domaine des sciences et de l'exploration spatiale. L'ASC constitue une agence gouvernementale canadienne sous la responsabilité du ministre de l'Industrie et son mandat est inscrit dans la *Loi sur l'Agence spatiale canadienne*[17] qui a été ratifiée par le Parlement du Canada en décembre

1990. Ce mandat consiste à « [...] promouvoir l'exploitation et le développement pacifiques de l'espace, de faire progresser la connaissance de l'espace par la science et de faire en sorte que les Canadiens tirent profit des sciences et technologies spatiales sur les plans tant social qu'économique. » Chaque année, l'ASC reçoit un budget d'environ 300 millions de dollars du gouvernement fédéral pour accomplir ce mandat.

### **1.2.1.2 Réalisations**

Depuis sa création, la présence de l'ASC s'est fait sentir à bien des niveaux. Sa plus célèbre réalisation est sans aucun doute le bras canadien. Avec Canadarm<sup>1</sup>, Canadarm2<sup>2</sup> et Dextre<sup>3</sup>, le Canada a occupé un rôle central dans l'assemblage et l'entretien de la Station spatiale internationale. L'Agence canadienne a aussi trois satellites en orbite, possède des instruments sur plusieurs satellites scientifiques étrangers et participe à une multitude de projets en partenariat avec des agences spatiales étrangères[11]. Sur le plan scientifique, l'ASC effectue et finance de la recherche dans une multitude de domaines liés à l'exploration et aux sciences spatiales[18]. Au niveau des missions spatiales, 8 astronautes canadiens ont participé à 15 missions spatiales internationales. En 2013, un Canadien occupera même, pour la première fois, le poste de commandant de la Station spatiale internationale.

### **1.2.1.3 Retombées**

Les retombées découlant des sciences spatiales et du travail de l'ASC sont multiples. Tout d'abord, il y a les retombées directes liées aux technologies spatiales. La présence de satellites a un impact réel et quotidien sur nos vies que ce soit pour les télécommunications, la localisation par GPS, la télévision, la radio ou tout simplement les prévisions météorologiques. Les satellites opérés par l'ASC, eux, permettent entre autres de faciliter la navigation maritime et de cartographier notre territoire et nos ressources naturelles.

---

<sup>1</sup> Bras robotisé utilisé sur les navettes spatiales américaines de 1981 à 2011

<sup>2</sup> Bras robotisé de la station spatiale internationale

<sup>3</sup> Robot agile utilisé comme une extension de Canadarm2

Ensuite, les technologies spatiales engendrent des retombées indirectes. La plupart des technologies développées expressément pour l'exploration spatiale peuvent être récupérées pour d'autres utilisations. L'exploration spatiale, canadienne ou étrangère, a fait évoluer grandement les technologies de transformation et d'utilisation de l'énergie solaire, les techniques d'épuration de l'eau, la robotique et ceux-ci ne sont que quelques exemples.

Finalement, l'exploration et les sciences spatiales ont des retombées scientifiques importantes. Celles-ci peuvent donner directement naissance à des innovations technologiques comme dans les cas mentionnés précédemment ou peuvent être d'un ordre plus théorique. Les retombées scientifiques qui découlent du travail de l'ASC et de ses partenaires représentent un éventail très large de domaines allant de la cristallographie, à la médecine, jusqu'à l'origine de l'Univers.

## **1.2.2 Gestions des données spatiales et scientifiques à l'ASC**

### **1.2.2.1 Cadre de gestion de l'information**

Depuis quelques années, le gouvernement canadien désire solidifier les pratiques de gestion de l'information. Des efforts tangibles ont été réalisés en ce sens par le gouvernement et par certaines de ses agences gouvernementales. L'ASC fait partie de ces agences qui perfectionnent continuellement leurs pratiques de gestion de l'information. Dans cette optique, le secteur de la gestion de l'information (GI)<sup>4</sup> de l'ASC a émis, en 2009, un Cadre de gestion de l'information afin d'identifier les concepts clés liés à cette activité et les directives devant être respectées dans ce domaine au sein de l'ASC.

Ce cadre définit l'information comme étant une ressource opérationnelle stratégique de l'ASC, un actif et « un renseignement ou une connaissance pouvant être représenté sous des formes adaptées à la communication, à l'entreposage ou au traitement. »[19] Selon cette définition de l'ASC, différents types d'actifs informationnels peuvent constituer de l'information. Ces éléments sont identifiés comme étant les données, autant scientifiques et techniques qu'administratives, le savoir explicite<sup>5</sup> et les informations ministérielles et documents. Tous ces

---

<sup>4</sup> Le rôle du secteur de la GI au sein de l'ASC sera expliqué à la section 3.2.2.2.1

<sup>5</sup> Le savoir explicite regroupe l'ensemble des connaissances ayant été transcrites sur un support écrit



éléments sont donc considérés par l'ASC comme des actifs, au même titre que les biens immobiliers et financiers de l'organisation.

Du côté du Secrétariat du Conseil du Trésor, l'information est plutôt définie comme « une représentation de faits, d'idées ou d'opinions sur des sujets, des événements et/ou des processus peu importe leur support ou leur présentation. »[20] Les trois actifs informationnels identifiés par l'ASC, les données, le savoir explicite et la documentation sont donc conformes à la définition gouvernementale d'information étant donné qu'ils ont tous un support.

La Cadre de gestion de l'information de l'ASC présente aussi un cycle de vie puisque, conformément à la Politique sur la gestion de l'information du Secrétariat du Conseil du Trésor[20], la gestion de l'information doit se faire à partir de la planification jusqu'à la destruction de l'information. Ce cycle de vie, présenté à la Figure 1-4, a été directement calqué du cycle de vie de l'information de Bibliothèque et Archives Canada (BAC). Il représente de manière conceptuelle les étapes par lesquelles passent l'ensemble des actifs informationnels du gouvernement canadien, et du même coup, ceux de l'ASC. Comme l'éventail d'actifs informationnels devant être géré par ce cycle de vie est très diversifié, celui-ci n'a pas la prétention de couvrir l'ensemble des exceptions possibles, mais présente très bien, de façon conceptuelle, les différentes étapes de la gestion de l'information.

### **1.2.2.2 Cas précis des données spatiales**

La gestion de l'information à l'ASC ne se résume pas au Cadre de gestion de l'information. Ce cadre ainsi qu'une politique en gestion de l'information ont été réalisés afin de fournir des lignes directrices pour la gestion de l'ensemble des actifs informationnels de l'ASC. Le groupe de la GI travaille maintenant au développement de procédures précises pour faire la gestion de certains types particuliers d'actifs informationnels. C'est présentement le cas pour les données spatiales.

Le Cadre de gestion de l'information définit les données comme des observations ou des mesures de base. Elles n'ont aucune forme matérielle particulière. Elles n'ont pas de signification en soi, c'est lorsqu'elles sont interprétées par un être humain ou un système qu'elles prennent une signification et deviennent de l'information.



Figure 1-4 : Cycle de vie de l'information de BAC et de l'ASC tel que présenté dans la version 2011 du Cadre de gestion de l'information de l'ASC[19]

Il existe différents types de données techniques à l'ASC. L'Architecture d'activités de programmes 2011-2012 de l'ASC (AAP)[21] regroupe toutes les données générées par les missions et projets de l'ASC sous le qualificatif de données spatiales puisqu'elles sont toutes en lien avec l'espace de par leur provenance, leur sujet ou leur utilité. Il existe deux types de données spatiales à l'ASC, les données d'observation de la Terre et les données scientifiques.

Bien qu'elle ne soit pas centralisée, la gestion des données spatiales s'effectue déjà au sein de l'ASC. Certains projets possèdent leurs propres politiques de données. Par exemple, les missions satellitaires ont chacune une politique qui indique comment doivent être gérées les données de la mission et quelles sont les responsabilités des différents partenaires du projet par rapport aux données. La gestion des données des projets réalisés conjointement avec des partenaires comme la *National Aeronautics and Space Administration* (NASA)<sup>6</sup> ou l'Agence spatiale européenne (ESA) se fait présentement en suivant les procédures de gestion des données spatiales de ces agences étrangères ou des partenaires responsables du projet.

---

<sup>6</sup> Agence spatiale des États-Unis

## 1.3 Situation canadienne

De nombreuses initiatives ont été menées par l'ASC en matière de gestion des données techniques. Au-delà de ces réalisations, différentes initiatives canadiennes ont été entreprises par le gouvernement ou par d'autres agences et ministères fédéraux. Étant donné son caractère public, il est important de connaître ces initiatives et ententes canadiennes puisqu'elles peuvent avoir un impact sur la façon de gérer l'information à l'ASC.

### 1.3.1 Initiatives

Au niveau du Canada, une *Consultation nationale sur l'accès aux données de la recherche scientifique*[22] s'est penchée en 2005 sur la problématique de la diffusion des données de recherche. Ce rapport fait suite à la *déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics*[23] qui a été publiée en 2004 et dont le Canada est signataire. C'est depuis cette signature que le gouvernement canadien s'est graduellement sensibilisé au potentiel des données scientifiques et aux défis liés à leur gestion sur l'ensemble de leur cycle de vie.

Par la suite, un *Groupe de travail sur la stratégie des données de recherche* a été formé afin de « décrire l'état actuel de la gérance des données de recherche au Canada ». Le rapport de ce groupe[24], publié en 2008, présente une analyse détaillée de la gestion des données de recherche selon dix indicateurs. Les indicateurs couvrent l'ensemble des phases du cycle de vie des données de recherche. Ce cycle en quatre étapes, présenté à la Figure 1-5, est très simple mais fournit une vision générale des principales phases de la gestion des données. Ce cycle est intéressant parce qu'il s'applique aux données de certaines agences gouvernementales canadiennes. Sa portée s'apparente donc à celle de la gestion des données spatiales de l'ASC. Malgré la simplicité de ce cycle de vie, ses quatre étapes sont décrites avec précision, ce qui en fait un outil intéressant.

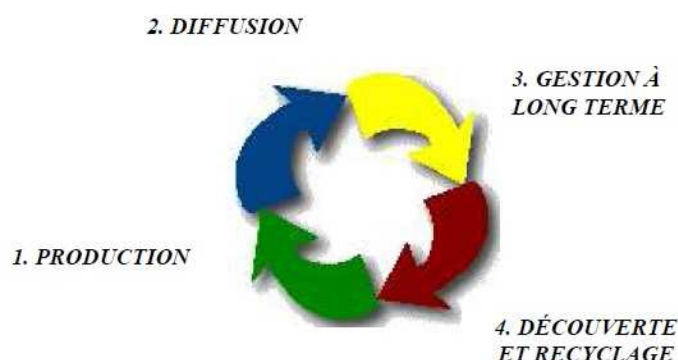


Figure 1-5 : Cycle de vie des données de recherche canadiennes du Groupe de travail sur la stratégie des données de recherche[24].

Au cours de ses travaux, le groupe de travail a déterminé la situation idéale pour faire une bonne gestion des données et a déterminé l'écart entre la situation réelle dans les organismes canadiens et cette situation idéale. Cette étude englobe la grande majorité des organismes canadiens liés à la recherche, que ce soit les universités, les organismes de financement (CRSH, CRSNG, etc.) ou certains ministères canadiens (BAC, Pêches et Océans Canada, Environnement Canada, etc.). Voici quelques-unes des conclusions de ce rapport :

- « La priorité est axée sur l'utilisation immédiate plutôt que sur les possibilités d'exploitation à long terme;
- peu d'institutions de recherche exigent des plans de gestion des données;
- aucun organisme national ne peut fournir des conseils et aider à l'application des normes en matière de données;
- les chercheurs sont réticents à partager les données, parce qu'ils ne disposent pas du temps et des compétences nécessaires;
- les dépôts de données ne disposent pas de la couverture et de la capacité nécessaires;
- la plupart des données sont conservées sur les disques durs des ordinateurs des chercheurs et ne sont pas accessibles à autrui;
- un grand nombre de chercheurs sont réticents à permettre l'accès à leurs données parce qu'ils croient qu'elles font partie de leur propriété intellectuelle. »[24]

Ce rapport conclut que le Canada peut se donner un avantage concurrentiel sur le plan international dans la mesure où il assure une gestion de ses ressources numériques sur l'ensemble de leur cycle de vie. Le rapport indique que les organismes canadiens priorisent l'utilisation immédiate des données et oublient trop souvent la gestion à long terme, que peu d'institutions exigent que les gestionnaires présentent des plans de gestion des données, qu'une grande partie des données sont conservées par les chercheurs sur leur disque dur et que ceux-ci sont souvent réticents à partager leurs données. Ce rapport identifie aussi certaines barrières qui nuisent à l'atteinte d'une gestion idéale des données de recherche.

Cette étude pancanadienne a ouvert la porte à de nombreuses initiatives liées à l'amélioration de la gestion des données gouvernementales. De plus, le *Secrétariat du Conseil du Trésor* a émis la *Directive sur la tenue de documents* en juin 2009[25]. Cette directive oblige les ministères à mettre en place des procédures pour faire la gestion sur l'ensemble de leur cycle de vie des ressources documentaires ayant une valeur opérationnelle. Les différentes agences et ministères canadiens sont donc amenés à réaliser un inventaire de leurs données, en connaître les différents dépôts, gérer l'accès à ces données et élaborer un calendrier de conservation.

Une autre initiative canadienne est la *Passerelle vers les données scientifiques*[26]. Ce portail lancé par l'*Institut canadien de l'information scientifique et technique du Conseil national de recherches Canada* (ICIST-CNRC) a pour objectif d'exploiter au maximum les données scientifiques canadiennes. Ce site contient deux sections principales. La première contient des liens vers une multitude d'ensembles de données, regroupés par domaine scientifique. Un lien vers certaines données du satellite canadien RADARSAT-1 s'y retrouve. La deuxième section présente une sélection de politiques et de pratiques exemplaires de gestion des données. Ce portail n'est donc pas un moyen d'archiver les données, mais plutôt de diffuser un large éventail de données canadiennes sur un même site.

Un projet pilote est présentement en cours afin d'instaurer le principe de l'*Open Data* au sein du gouvernement canadien[27]. Ce projet a pour but d'améliorer la visibilité et de promouvoir la libre utilisation des données appartenant au Gouvernement du Canada.

Des organismes publics canadiens opèrent aussi une multitude de portails spécialisés dont l'accès est plus restreint. C'est le cas par exemple du *Canadian Space Science Data Portal*[28, 29] et du portail du *Canadian Geo-Space Monitoring*[30] qui regroupent respectivement des données

d'expériences et des données géospatiales recueillies par des agences fédérales et des universités canadiennes. L'objectif de ce type de portail est de faire profiter les chercheurs et les entreprises des données spatiales canadiennes.

La gestion des données de recherche est désormais un enjeu prioritaire pour des organismes comme le Conseil national de recherches Canada, l'Association des bibliothèques de recherche du Canada et le Conseil du Trésor. L'importance de cet enjeu a même été soulignée par le Gouverneur général du Canada lors du *Discours du Trône* du 3 juin 2011[31].

Considérant la grande quantité de données qu'elle génère et manipule, il est naturel que l'ASC travaille présentement à la mise en place de procédures de gestion de ses données spatiales selon leur cycle de vie.

### **1.3.2 Ententes internationales**

Le Canada a signé plusieurs ententes internationales visant à améliorer le partage des données de recherche au sein de la communauté scientifique. Une de ces ententes est la *Déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics* signée le 30 janvier 2004[23]. Cette entente ratifiée par 35 pays vise à diffuser publiquement toutes les données de recherche dont l'acquisition a été financée par de l'argent public afin de permettre la réutilisation de ces données par d'autres chercheurs. Cette entente a été suivie en 2006 par une liste de principes et de lignes directrices ayant comme but ultime de renforcer le réseau scientifique global.

Par le biais de l'ICIST-CNRC, le Canada est aussi membre de CODATA (*Committee on Data for Science and Technology*)[32]. Ce comité international fut créé en 1966 et regroupe des gestionnaires et des scientifiques du monde entier. Son objectif est d'améliorer la qualité et l'accessibilité des données techniques. CODATA vise à sensibiliser les communautés scientifique et technique sur l'importance d'une gestion adéquate et du partage des données et augmenter les coopérations internationales dans ces domaines. Chaque année, le *Comité national canadien pour CODATA* (CNC/CODATA) publie un rapport qui recense l'ensemble des activités relatives aux données au Canada[33]. Ce rapport aide à diffuser les percées scientifiques canadiennes dans le monde.

Une dernière initiative internationale dont fait partie le Canada est DataCite[34]. Cet organisme fondé en 2009 vise à faciliter l'accès aux données de recherche sur l'Internet, augmenter leur partage et leur réutilisation et faire reconnaître les données de recherche comme des contributions légitimes à la science pouvant être citées au même titre que les publications. L'ICIST-CNRC est membre de DataCite depuis sa fondation et œuvre au Canada à promouvoir la réutilisation des données et à aider les chercheurs à en faire une meilleure gestion[35].

## 1.4 Situation internationale

Tel que démontré, le Canada prend la problématique de la gestion des données techniques très au sérieux et a effectué des études, ratifié des ententes et mis en place des ressources afin d'accroître le bénéfice que ses citoyens tirent des données techniques. Le Canada n'est toutefois pas le premier à s'attaquer à cette problématique. Quelques pays, principalement le Royaume-Uni et les États-Unis, ont étudié la conservation des données et ont mis en place des institutions et des ressources matérielles et informatiques afin de pouvoir exploiter pleinement leurs actifs informationnels. Cette section du mémoire s'attarde à ces pays qui peuvent servir d'exemples en ce qui a trait à la gestion des données.

### 1.4.1 Royaume-Uni

Le Royaume-Uni est parmi les pionniers de la gestion des données scientifiques. Le *Digital Curation Center* (DCC), fondé en 2004, est considéré comme la référence pour tout ce qui concerne la conservation des données.

Dès 2003, le Royaume-Uni a publié une étude faisant état de la situation de la gestion des données de recherche à l'intérieur du pays, *l'E-science Curation Report*[3]. Cette étude s'apparente beaucoup à la consultation nationale que le Canada entreprendra deux ans plus tard[22]. L'étude révèle que les efforts de gestion des données varient beaucoup selon les domaines scientifiques et selon les institutions, que le modèle de financement est très orienté vers les résultats à court terme, que les besoins et bénéfices de la conservation à long terme étaient peu connus des chercheurs et que l'interaction entre ceux-ci et les curateurs de données était très faible. Un des éléments les plus intéressants de cette étude est la façon dont elle intègre la conservation au processus de recherche traditionnel. C'est cette étude qui décrit les trois niveaux de conservation présentés à la section 1.1.3.1 de ce mémoire.

La principale recommandation de cette étude britannique fut de créer un centre national pour étudier les enjeux liés à la gestion des données et fournir de l'aide aux chercheurs dans ce domaine. Ce centre, le DCC, a vu le jour à Édimbourg un an plus tard, le 1<sup>er</sup> mars 2004. Son mandat[36] est de :

- faire de la recherche dans le domaine de la gestion des données;
- développer des outils pour la conservation à long terme des données de recherche;
- développer des outils, des logiciels et de la documentation pour aider les chercheurs britanniques dans la gestion de leurs données;
- documenter les pratiques exemplaires dans le domaine de la gestion des données.

Le DCC a publié de nombreux documents depuis sa création. Le plus important parmi eux est le *Curation Reference Manual*[37]. Avec actuellement 13 chapitres, ce manuel couvre l'ensemble des aspects de la gestion des données. Il fait la synthèse de toutes les recherches du DCC dans ce domaine et il représente un outil clé pour tous les intervenants de la gestion des données. Les définitions de ce manuel sont reprises dans la littérature scientifique par de nombreux auteurs.

Les chercheurs peuvent aussi trouver sur le site du DCC un gabarit de plan de gestion des données qui regroupe tous les éléments devant se retrouver dans un plan de gestion des données. Ce document permet de s'assurer de faire une bonne planification prenant en compte l'ensemble du cycle de vie des données[38]. Bien qu'il soit très complet, le gabarit de plan de gestion proposé par le DCC s'avère peu précis puisqu'il se doit d'être applicable à tous les projets scientifiques, en sciences humaines, naturelles et appliquées au Royaume-Uni.

Pour aider à faire une gestion des données complète, le DCC propose aussi un cycle de vie des données d'expérience très détaillé et très opérationnel (Figure 1-6). Ce cycle de vie très complexe englobe la gestion de l'ensemble des données scientifiques du Royaume-Uni. Un aspect intéressant de ce cycle est la présence de boucles, principalement de celle illustrant la réutilisation des données (*reappraise*).



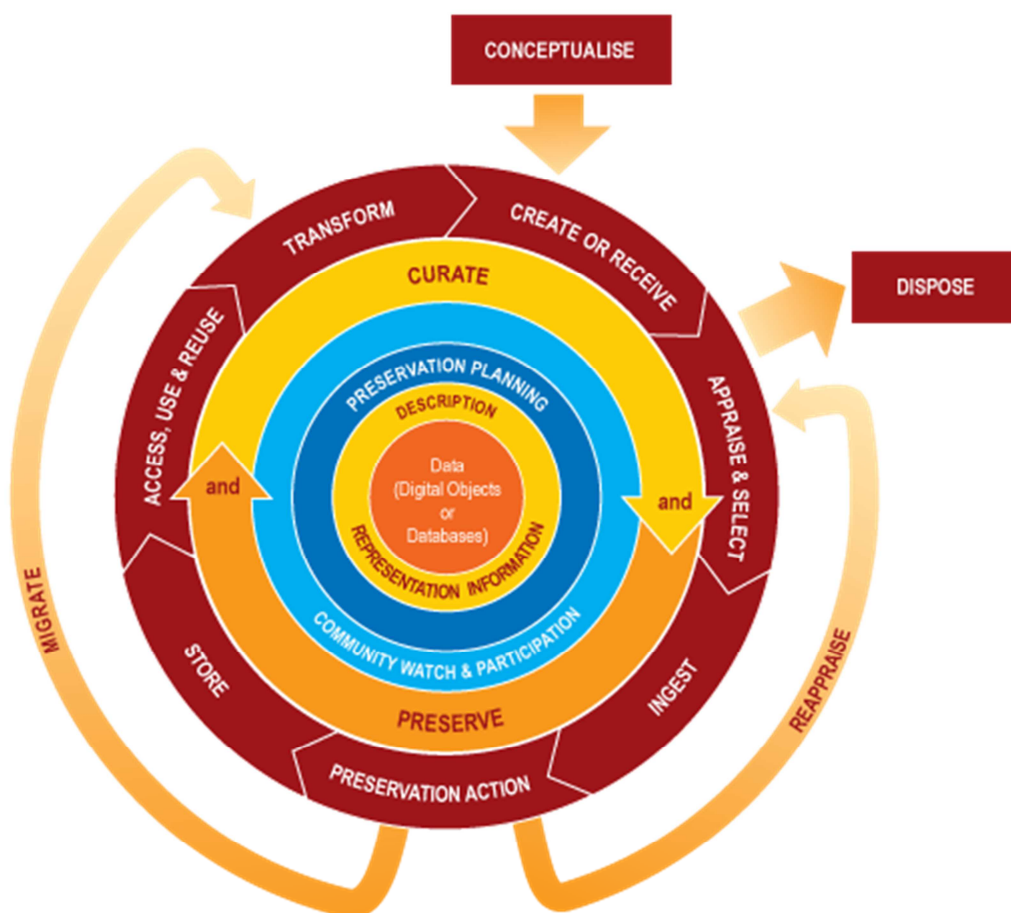


Figure 1-6 : Cycle de vie des données scientifiques tel que présenté par le DCC[36]

Ce cycle de vie utilise une terminologie et une division des tâches distinctes de celles utilisées pour la gestion de l'information de l'ASC, ce qui complique l'intégration directe de ce cycle aux procédures actuelles de l'organisation. Bien qu'il soit très complet et très opérationnel, le cycle de vie du DCC décrit le contexte très large qu'est l'ensemble de la recherche scientifique britannique. Le cycle de vie des données scientifiques du DCC est néanmoins un très bon exemple de cycle de vie des données.

Le DCC constitue une référence incontournable dans tout travail portant sur la gestion des données. Les efforts qui y sont faits visent à uniformiser la gestion des données de recherche et la présenter selon un seul modèle. Cet effort d'uniformisation présente toutefois des limites. En effet, les études du DCC portent sur l'ensemble de la recherche scientifique au Royaume-Uni. Si on tente d'appliquer les modèles qu'il propose dans un autre contexte, ils peuvent s'avérer à la fois

trop larges et trop précis. Trop large en ce sens que les modèles et outils proposés par le DCC couvrent l'ensemble des domaines scientifiques et des contextes de recherche. Trop précis parce qu'ils se limitent aux données d'expériences scientifiques et néglige les autres données techniques. Bien qu'ils représentent une référence incontournable, les outils du DCC ne peuvent pas être appliqués comme tel, dans tous les contextes.

Parallèlement au travail du DCC, la chercheuse britannique Liz Lyon a effectué une étude sur les rôles, droits et responsabilités des différents intervenants de la gestion des données de la recherche scientifique au Royaume-Uni[39]. L'étude dresse une liste des différents intervenants accompagnés de leurs rôles, droits, responsabilités ainsi que des relations qu'ils entretiennent les uns avec les autres. Le rapport final de cette étude recommande à chaque institution de recherche de dresser un inventaire de leurs données, d'élaborer une politique pour la gestion, la préservation et le partage des données de recherche, d'obliger tous les projets de recherche financés par de l'argent public à soumettre un plan de gestion des données et d'améliorer la documentation et les programmes de formation. Comme c'est le cas pour le travail du DCC, il demeure difficile de sortir cette étude de son contexte et de la transposer à une situation qui n'est pas parfaitement similaire.

### 1.4.2 États-Unis

Les initiatives états-uniennes constituent également un exemple intéressant à observer en termes de gestion des données. Celles-ci seront décrites ici à l'exception des réalisations de la NASA qui seront décrites dans la section 1.5.1.

Dès 2007, la *National Science Foundation*, a investi 100 millions de dollars US afin de créer un réseau de dépôts de données inter-reliés afin de favoriser la conservation et le partage des données scientifiques américaines[24].

En 2009, les États-Unis ont également effectué une vaste étude afin de dresser un portrait de la gestion des données effectuée par les agences gouvernementales et d'énoncer des principes directeurs pour améliorer cette gestion. Les conclusions de ce rapport vont dans le même sens que la *Consultation nationale sur l'accès aux données de la recherche scientifique*[22] du Canada et que l'*E-Science Curation Report*[3] du Royaume-Uni. Ainsi, l'*Interagency Group on Digital Data Report*[40] présente un cycle de vie des données de recherche très élémentaire et les lignes

directrices d'une meilleure gestion des données de recherche. La recommandation la plus intéressante de ce rapport suggère au gouvernement de créer un sous-comité du *National Science and Technology Council* dont le rôle serait similaire à celui du DCC au Royaume-Uni, c'est-à-dire d'encadrer les activités de gestion des données de recherche.

Présentement, le gouvernement américain travaille à consolider l'ensemble des données gouvernementales américaines sur un seul et même portail[41] afin de faciliter leur gestion et de favoriser leur libre diffusion et leur réutilisation.

Du côté scientifique, plusieurs universités américaines étudient les problématiques de la conservation des données et créent des portails communs pour faciliter l'accès à leurs données de recherche. C'est le cas, entre autres, de l'université Purdue en Indiana qui a étudié en détail les différents profils de conservation de données[42]. Cette vaste étude visait à analyser et comparer les besoins des chercheurs dans différents domaines de la science en ce qui a trait à la conservation des données.

De son côté, l'Université John Hopkins (JHU) a étudié les rôles des bibliothécaires et des chercheurs et l'importance de leur interaction pour assurer la préservation à long terme des données gérées par l'université[2, 43]. Les chercheurs de JHU ont donc travaillé à mettre en relation des curateurs et les créateurs des données de différents projets d'observations de l'espace comme *le Sloan Digital Sky Survey*[44] et du *National Virtual Observatory*[45].

De même, l'Université Cornell a travaillé sur les dépôts de données scientifiques. Dans le cadre du projet DataStar[46, 47], elle a développé un modèle complexe de conservation des données. Celui-ci contient un dépôt local capable de fournir le support et les ressources nécessaires aux chercheurs durant leur recherche. Une fois la recherche terminée, l'université encourage les chercheurs à transférer leurs données vers des portails plus aptes à faire la conservation à long terme et la diffusion des données. De nombreux outils ont été développés par ces chercheurs afin de faciliter la création de métadonnées au moment du transfert des données.

Pour ce qui est des données géospatiales, plusieurs portails américains rassemblent et conservent ce type de données. C'est le cas, entre autres, du *National Geospatial Data Archives*[5, 10] et du *National Spatial Data Infrastructure*[48]. La mission de ces portails est d'assurer la sécurité, la préservation et la diffusion des données géospatiales américaines.

### 1.4.3 Autres pays

D'autres pays comme la Nouvelle-Zélande travaillent aussi sur des politiques et des normes nationales pour la gestion de leurs données gouvernementales. La Nouvelle-Zélande a, par exemple, formé un groupe de travail en 1999 pour étudier les pratiques de gestion des données des différentes agences gouvernementales. La Nouvelle-Zélande et l'Australie se sont dotées de sites gouvernementaux[49, 50] comparables, bien que moins complets, à ceux du DCC[36] et du gouvernement américain[41]. Le site néo-zélandais regroupe l'ensemble des normes et des politiques nationales du pays. De son côté, le gouvernement australien utilise data.gov.au[50] pour appliquer l'*Open Data* à ses données.

## 1.5 Autres agences spatiales

Après avoir étudié la situation canadienne et internationale, ce chapitre se termine par une analyse des bonnes pratiques de gestion des données techniques de deux agences spatiales majeures, la NASA et l'ESA.

### 1.5.1 NASA

La NASA est l'agence spatiale dont il a été possible d'observer les pratiques de gestion des données techniques les plus avancées. Tel qu'indiqué à la section 1.4.2, les agences gouvernementales américaines appliquent déjà le concept de données libres et diffusent leurs données scientifiques sur un portail gouvernemental[41]. La NASA utilise aussi une multitude d'autres portails pour diffuser et partager ses données avec ses partenaires et avec le reste de la communauté scientifique. Un bon exemple est le *Planetary Data System* (PDS)[51]. Ce portail conserve et diffuse toutes les données des projets de sciences planétaires de la NASA. En plus d'archiver les données, le PDS[51] fournit des outils et des gabarits afin de documenter les données à chaque étape du projet. De cette façon, tous les projets de sciences planétaires américains sont adéquatement documentés à l'aide de plan de gestion des données normalisés reflétant chacune des phases du projet. L'ASC bénéficie des portails, des outils et des pratiques exemplaires de la NASA lorsqu'elle effectue des projets conjointement avec l'agence américaine.

Outre ces portails spécialisés, la NASA possède des documents plus généraux pour assurer une bonne gestion des données. Dès 1993, la NASA a publié un gabarit de plan de gestion des

données scientifiques[52]. Ce document explique pas à pas tous les éléments qui devraient se trouver dans un plan de gestion des données de la NASA. Les gestionnaires n'ont qu'à prendre ce gabarit et remplir chacune de ses sections afin d'obtenir un plan qui explique comment sera faite la gestion des données de leur création à leur élimination.

Dans un autre rapport portant sur la préparation du plan de gestion des données[53] la NASA identifie les trois objectifs de ce type de document comme étant :

- d'identifier et de décrire les données nécessaires à l'atteinte des objectifs scientifiques de la mission;
- d'identifier les produits nécessaires à l'atteinte des objectifs du projet c'est-à-dire les données brutes et transformées, les produits, les données de calibration ainsi que les métadonnées et documents nécessaires à l'interprétation des données;
- d'identifier les ressources nécessaires pour la gestion des données sur l'ensemble de leur cycle de vie.

Finalement, en plus des données et métadonnées, les politiques et plans de gestion des données de la NASA sont parfois rendus eux aussi disponibles au public. La politique de gestion des données des sciences héliophysiques en est un exemple[54].

### **1.5.2 ESA**

De par sa structure supranationale, l'Agence spatiale européenne (ESA) fait face à des défis additionnels en termes de gestion de données, lorsqu'on la compare à d'autres agences comme la NASA ou l'ASC. L'ESA est aujourd'hui financée et opérée par un regroupement de 18 pays européens. Ceux-ci financent tous les programmes principaux de l'Agence mais sont libres de choisir quels programmes optionnels ils désirent réaliser[55]. Dans ce contexte, il demeure donc difficile d'établir des exigences semblables pour tous les programmes.

Peu de documents décrivant les procédures de gestion des données de l'ESA ont été trouvés lors de cette recherche. Les procédures de l'Agence européenne ne sont pas diffusées de manière aussi large que celles du Royaume-Uni ou de la NASA. Certaines observations utiles ont tout de même pu être effectuées.

Étant donnée la nature intergouvernementale de l'ESA, il ne semble pas y avoir de portail central pour la libre diffusion des données à l'image du portail américain[41] ou du projet pilote canadien[27]. L'Agence européenne fait tout de même des efforts pour appliquer le concept de données libres. L'ESA a par exemple décidé, en 2009, via le parlement européen, qu'il y aurait un accès public et gratuit aux données des générations futures de satellites d'observation de la terre Sentinelle[56]. L'ESA travaille présentement à une centralisation de ses archives[57] ce qui va assurément permettre l'amélioration des pratiques de conservation des données et pourrait en faciliter la diffusion.

Bien qu'elle n'ait pas pour le moment de portail centralisé pour ses données, l'ESA semble faire une gestion rigoureuse des données recueillies lors de ses projets. Une multitude de liens sur le site de l'Agence européenne[55] mènent à des données, principalement des produits à valeur ajoutée, et certains plans de gestion des données de projet sont disponibles. Le plan de gestion des données du projet *LISA Pathfinder* en est un bon exemple[58].

## 1.6 Conclusion

La gestion des données techniques constitue un domaine d'études et d'application en plein développement. Quelques chercheurs et institutions commencent à identifier les concepts théoriques de base de cette nouvelle discipline. En plus d'être un enjeu technique, cette gestion est désormais un enjeu politique pour de nombreux pays qui y voient une opportunité de renforcer leur influence sur la scène scientifique et technologique mondiale.

Le Royaume-Uni est un des pionniers dans ce domaine. Par le biais du DCC, il a étudié de nombreuses problématiques liées à la conservation des données et a développé de nombreux outils pour en faciliter la gestion. Ces études se limitent par contre qu'aux données scientifiques et, vu la diversité de ses ensembles de données, les outils du DCC sont très généraux.

Du côté de l'exploration spatiale, c'est de la NASA que proviennent les procédures de gestion les plus avancées recensées dans ce chapitre. Beaucoup de ressources et d'efforts ont été mis par cette organisation dans la création de procédures et d'outils de gestion permettant de faire une bonne conservation et une libre diffusion des données gérées par cette agence. Des portails comme le PDS sont des exemples à suivre en matière d'organisation et de diffusion des données techniques.

Il est naturel qu'une organisation comme l'ASC se penche actuellement sur l'élaboration de procédures de gestion des données techniques selon leur cycle de vie. Cette agence gouvernementale manipule une grande quantité et une grande variété de données techniques. De plus, l'ASC a l'obligation de se conformer à la *Directive sur la tenue de documents*[25] du gouvernement canadien d'ici 2014. Elle travaille donc à élargir les pratiques de gestion de l'information déjà mises en place pour faire la tenue des documents en étendant cette gestion à la situation particulière des données spatiales.

Le défi est qu'étant donné le contexte très particulier dans lequel opère l'ASC, les modèles de gestion des données techniques présentés dans ce chapitre doivent être adaptés si l'on veut les appliquer aux données spatiales de l'ASC. Il s'avère donc nécessaire de faire une analyse détaillée des conditions particulières dans lesquelles opère l'ASC ainsi que des particularités des données qu'elle manipule de manière à développer des outils permettant de faire une bonne gestion des données techniques dans un tel contexte.

## **CHAPITRE 2    DESIGN DE LA RECHERCHE**

Le présent chapitre expose les principaux aspects méthodologiques de la recherche. Il débute par un rappel des objectifs de la recherche et poursuit par une description détaillée de l'approche méthodologique appuyée par la littérature pertinente. Ce chapitre se termine par la démarche opérationnelle de l'auteur, incluant le contexte de travail et l'organisation des tâches.

Très ancrée dans un contexte organisationnel, cette recherche revêt un caractère scientifique particulier, notamment par la stratégie utilisée et décrite à la section 2.2. La démarche globale demeure tout de même fidèle aux exigences habituelles de toute recherche scientifique; elle prend appui dans la littérature pertinente, suit une méthodologie reconnue, propose une solution solidement ancrée et validée par le terrain, et propose une discussion rigoureuse sur la valeur et les limites des résultats.

### **2.1 Objectifs de la recherche**

L'objectif général de la recherche consiste en l'analyse des pratiques de gestion des données techniques d'un organisme public, l'ASC, afin de permettre la création et la mise en application d'un cadre de gestion des données. Prenant appui sur les connaissances scientifiques actuelles, de même que dans la réalité du terrain, cette analyse doit couvrir toutes les étapes du cycle de vie des données techniques provenant de la recherche scientifique et des missions satellitaires de l'ASC. Le cadre de gestion des données qui doit en découler représente un outil essentiel à l'amélioration des pratiques de gestion des données spatiales de l'ASC.

#### **2.1.1 Objectifs spécifiques**

Plusieurs objectifs spécifiques découlent de l'objectif général. Ils sont ici décrits en regard des besoins opérationnels de l'organisme partenaire.

- Analyser le contexte de la gestion des données au Canada et plus précisément à l'ASC et le type de données qui y sont gérées;
- évaluer les enjeux et les conditions particulières liés à la gestion des données dans un contexte comme celui de l'ASC;
- définir les différentes étapes du cycle de vie des données spatiales;



- définir les différentes parties prenantes de la gestion des données spatiales de l'ASC ainsi que les rôles et responsabilités et l'interaction entre chacune d'elles;
- aider à mettre en place les bases pour la mise en application de procédures de gestion des données spatiales et émettre des recommandations visant ces procédures;
- développer des outils de gestion afin d'améliorer et de faciliter la gestion des données spatiales;
- tenter de généraliser les résultats obtenus et de tirer des apprentissages plus larges à partir des conclusions de l'analyse de la situation spécifique de l'ASC.

Bien que formulés en regard des besoins de l'ASC, ces objectifs spécifiques revêtent une grande pertinence sur le plan académique. Tel que l'a démontré la revue de littérature (Chapitre 1), la gestion des données techniques constitue un domaine d'études et d'application en plein développement. Aussi, la réalisation d'un projet de recherche comme celui-ci, directement avec un partenaire, constitue une occasion unique de faire avancer les connaissances à partir d'un cas réel.

## 2.2 Justification du type de recherche

D'un point de vue méthodologique, cette recherche revêt un caractère particulier du fait qu'elle propose une démarche combinant l'observation et la conception d'outils de gestion. Les démarches scientifiques traditionnelles, principalement orientées vers l'observation et l'expérimentation, posent visiblement des limites dans le domaine des sciences de la gestion et de l'organisation, soutient Hatchuel[59]. Au sein des sciences humaines et sociales en général, plusieurs propositions ont été faites pour combler le fossé entre l'objet de recherche observé et les interventions possibles sur le terrain pour modifier une situation organisationnelle. Les approches dites de « recherche-action » ou de « recherche-intervention » sont souvent proposées comme alternatives en ce sens, étant donné leur finalité avouée d'agir dans un milieu donné. Dans leur revue des différentes approches en recherche-action, Dolbec et Prud'homme[60] soutiennent que *« contrairement à d'autres formes de recherche dans lesquelles le chercheur vise, sinon l'objectivité, du moins la neutralité, ce type de recherche s'est donné comme objectif d'influencer directement le monde de la pratique »*. Cette vue est partagée par David[61] lorsqu'il soutient que ces modes de recherche impliquent que le chercheur *« ...intervienne directement, volontairement*

*et dans le cadre d'une relation particulière avec les acteurs de l'organisation étudiée, sur le cours des choses (par exemple en concevant avec les acteurs d'une organisation des outils de gestion nouveaux). »*

La recherche présentée dans ce mémoire adopte résolument les méthodes de recherche-intervention telles que décrites par David[62]. Celles-ci visent à générer de la connaissance en sciences de la gestion en travaillant sur une problématique réelle d'une organisation, et directement avec les intervenants de l'organisation. Dans ce type de recherche, le chercheur s'intègre à une organisation afin d'aider l'implantation de nouveaux modèles, d'outils ou de procédures de gestion venant répondre à un besoin de celle-ci. L'interaction avec le terrain permet de développer à la fois des compétences pratiques pour résoudre le problème de l'organisation, et des connaissances théoriques pouvant être diffusées au sein de la communauté de chercheurs de ce domaine.

Ce type de recherche répond généralement à deux types d'objectifs. D'un côté, il y a les objectifs pratiques de l'organisation. Ceux-ci sont principalement normatifs puisqu'ils visent à améliorer les pratiques de gestion de l'organisation afin de régler un problème concret. D'un autre côté, il y a les objectifs scientifiques qui visent le développement de nouvelles connaissances. Pour le chercheur, il s'agit souvent d'un travail exploratoire qui vise à bien comprendre un secteur professionnel ou une problématique particulière du terrain. Le chercheur tente ensuite de tirer des apprentissages de cette recherche et de proposer des contributions générales sur le plan théorique.

Tel que l'indique David[62], la recherche-intervention mise sur un échange continu entre les différents intervenants. Le chercheur participe à l'action et les sujets participent à la réflexion. Dans ce type de démarche, il y a généralement une alternance entre des phases de travail en groupe et de travail solitaire du chercheur. Ces phases sont généralement des phases d'observation (entrevues et analyse de la documentation), de remue-méninges, d'analyse ou de conception d'outils.

Les méthodes de recherche-interventions se sont vite imposées comme méthodologie à suivre pour la présente recherche, du fait que cette dernière porte sur une problématique très peu documentée et intimement liée au fonctionnement des organisations, la gestion des données techniques selon leur cycle de vie. De plus, étant donné la faible quantité de littérature scientifique portant ce sujet particulier, il s'est avéré pertinent de réaliser une étude sur, *et avec* le

terrain. Cette dernière précision est importante. En effet, les études de cas sont très fréquentes dans le domaine de la gestion[63] mais elles impliquent généralement une collecte de données (ex. : questionnaires, entrevues...) *en dehors* de l'action ou dans une période différente de celle qui intéresse le chercheur. Dans le cas présent, le fait d'intégrer un groupe de travail et de participer à l'analyse et au développement de procédures de gestion permettait au chercheur de prendre part à l'élaboration de dispositifs (outils de gestion) et de valider ces artefacts directement auprès des intervenants du terrain. L'intégration à un groupe de travail et l'application des méthodes de recherche-intervention semblaient donc être la façon idéale d'analyser les problématiques liées à la gestion des données techniques.

## 2.3 Contexte de collaboration Polytechnique-ASC

Cette recherche fut effectuée au sein de l'ASC et plus spécifiquement au Centre spatial John H. Chapman de Saint-Hubert. Le chercheur principal (auteur de ce mémoire) fut un membre actif du *Groupe de travail sur la gestion des données* (GTSGD), pendant neuf mois, soit de septembre 2010 à mai 2011. En plus du chercheur principal, le GTSGD était composé des gestionnaires de la GI et de la TI, de la chef de la gestion des documents et du courrier et d'une analyste en gestion de documents. Ce groupe effectuait un lien permanent entre le chercheur et les autres intervenants internes de l'ASC. Des rencontres hebdomadaires étaient organisées afin d'analyser différents aspects du projet et d'en surveiller le bon déroulement, conformément à la planification.

Comme c'est le cas en recherche-intervention, le groupe de travail a collaboré à la réflexion présentée dans ce mémoire. Cette collaboration s'est faite principalement sous la forme de groupe de discussion et par la révision et la formulation de commentaires par rapport aux livrables techniques du projet (rapport sur la gestion des données spatiales de l'ASC et une ébauche d'un gabarit de plan de gestion des données spatiales). Bien que le groupe de travail ait participé à la réflexion, l'ensemble des documents produits dans le cadre de cette maîtrise l'ont été par l'auteur du mémoire.

## 2.4 Description des étapes

Les différentes étapes du processus de recherche sont illustrées à la Figure 2-1. Ces étapes portent le numéro de la section du mémoire dans laquelle elles sont décrites. Ce processus de recherche

se démarque d'abord par la présence de nombreuses étapes de validation des résultats intermédiaires. Étant donnée l'intégration du chercheur au groupe de travail, cette démarche de validation par le GTSGD s'est avérée possible au fur et à mesure de la recherche, contribuant ainsi à la valeur des résultats, tout en limitant les retours en arrière.

L'autre particularité de la méthodologie se situe au niveau de la provenance de l'information recueillie. La Figure 2-2 de la page 41 présente les diverses sources utilisées, de façon schématique. Cette cueillette des données fut réalisée à plusieurs étapes de la démarche. Une première cueillette s'est faite à l'étape 2.4.2 au cours de laquelle les documents internes de l'ASC furent analysés et comparés avec ceux d'autres organisations et avec la littérature scientifique. Une deuxième cueillette de données fut réalisée plus tard lors des entrevues (étape 2.4.5). En plus de ces deux étapes du processus de recherche, beaucoup d'information a été recueillie lors des rencontres effectuées régulièrement avec le GTSGD, et ce, tout au long de la démarche, de la définition de la problématique à l'analyse.

La méthodologie utilisée pour effectuer cette recherche s'apparente grandement à celle suivie par le projet du développement de profils de conservation menée par l'Université Purdue[42] en ce sens qu'il y a deux étapes de cueillette d'information. Dans les deux cas, un premier modèle théorique fut élaboré à partir de la littérature pour ensuite le confronter à la réalité du terrain lors d'entrevues. Cette recherche vise elle aussi à caractériser une problématique réelle du terrain et à créer des outils de gestion utiles dans le contexte étudié.

Chacune des activités sera maintenant décortiquée afin d'en décrire sa réalisation et d'en expliquer les objectifs.

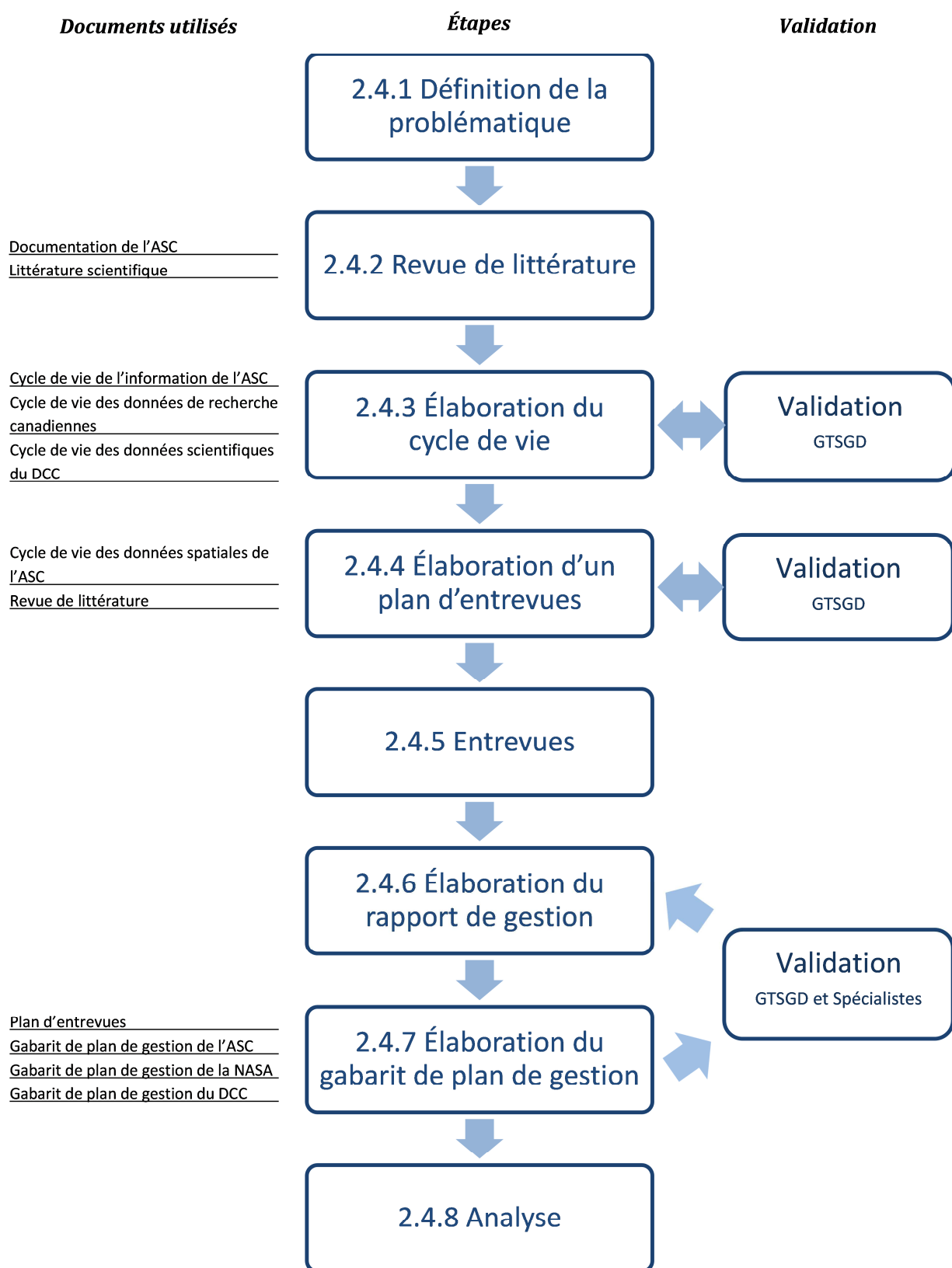


Figure 2-1 : Démarche utilisée dans le cadre de cette recherche

### **2.4.1 Définition de la problématique de recherche**

La première étape fut d'identifier les défis auxquels faisait face l'ASC en termes de gestion des données spatiales et quels étaient les objectifs du projet actuel de celle-ci par rapport à ces défis. Une bonne connaissance du projet et des attentes de l'organisation étudiée a permis de clarifier le contexte et les limites de l'étude. La définition de la problématique de recherche a permis d'établir les bases d'une collaboration permettant de concilier les objectifs pratiques de l'ASC et les objectifs scientifiques de la recherche. Cette conciliation est un élément central de la recherche-intervention[62] puisque les connaissances théoriques générées par ce type de recherche découlent de l'exécution d'un projet technique, sur le terrain. Le projet et la recherche sont donc deux éléments indissociables.

### **2.4.2 Revue de littérature**

Une revue exhaustive des connaissances et pratiques de gestion des données fut ensuite effectuée. Cette revue littéraire comportait deux volets : un volet académique et un plus pratique.

Le volet académique a consisté en une analyse de la littérature scientifique. Il avait pour objectif d'identifier les concepts et enjeux importants de la gestion des données scientifiques et spatiales et de documenter les recherches passées et présentes effectuées dans ce domaine. Parallèlement, le volet pratique comportait une revue des documents, rapports et politiques de l'ASC, du gouvernement canadien et des agences étrangères ayant une mission semblable à celle de l'ASC de manière à connaître la situation actuelle à l'ASC et dans certains organismes canadiens, les enjeux particuliers en présence et, précisément, les objectifs corporatifs de l'organisation. Cette revue a aussi permis de mettre en lumière les façons de faire d'organisations qui gèrent des données comparables à celles de l'ASC, d'identifier les bonnes pratiques et de situer l'ASC par rapport à ces pratiques enviables.

La mise en commun des volets académiques et pratiques a permis d'obtenir une vision d'ensemble de ce qui constitue la gestion des données techniques, du contexte particulier de l'ASC et des pratiques existantes dans le domaine de la gestion des données. Le Chapitre 1 du présent mémoire représente les aboutissants de cette revue de littérature.

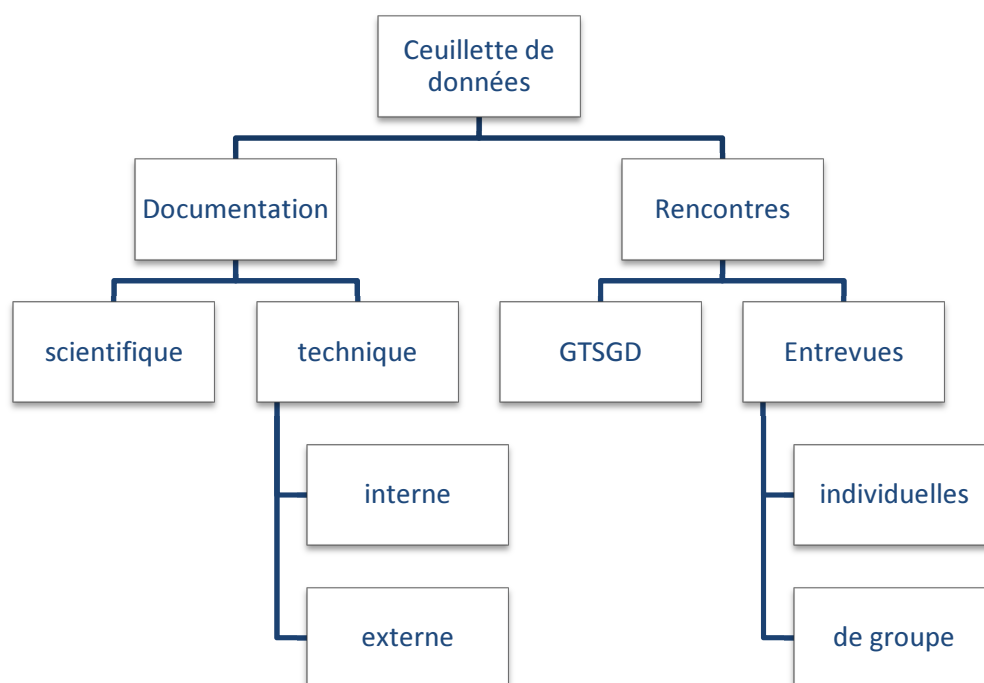


Figure 2-2 : Provenance de données de la présente recherche

### 2.4.3 Élaboration du cycle de vie des données spatiales de l'ASC

Une fois la littérature bien étudiée, un cycle de vie représentant spécifiquement les phases de la gestion des données spatiales de l'ASC fut élaboré. Cette étape a nécessité une analyse des processus par lesquels passent toutes ces données depuis la planification de leur acquisition jusqu'à leur destruction. Pour ce faire, le cycle de vie des données de plusieurs autres organisations et le cycle de vie de l'information de l'ASC furent analysés. Chacun des cycles de vie étudié proposait des éléments intéressants, mais, étant donné la portée, l'objectif et le contexte particulier de la gestion des données spatiales de l'ASC, un cycle de vie particulier a dû être créé.

Le cycle de vie a été retravaillé suite aux nombreuses rencontres avec le GTSGD. La version qui est présentée au Chapitre 3 de ce document est la version ayant été validée par ce groupe. C'est à partir de ce cycle de vie qu'il fut possible d'analyser les enjeux importants liés à la gestion des données spatiales.

## **2.4.4 Élaboration d'un plan d'entrevue**

Afin de pouvoir mettre en place des procédures de gestion des données spatiales selon leur cycle de vie au sein de l'ASC, il est essentiel de connaître les pratiques actuelles de l'organisation, les avantages et les inconvénients de ces pratiques ainsi que les enjeux et défis liées au contexte particulier de l'ASC. Une partie de cette information a pu être recueillie lors de la revue de littérature, mais cette information ne permettait pas d'avoir une vision globale des pratiques de gestion des données. Il fut donc nécessaire de rencontrer les intervenants de la gestion des données spatiales afin d'identifier les façons de faire actuelles, les obstacles auxquels ils font face, les possibilités d'améliorations et les défis et avantages qu'engendrent ces possibles améliorations. Des rencontres ont été organisées avec des employés de l'ASC dont le rôle est lié aux données spatiales.

Pour ces rencontres, un plan général d'entrevue a été élaboré (Annexe 1). Ce plan avait comme objectif d'informer à l'avance les personnes rencontrées de la nature des entrevues. Il reprend les phases principales du cycle de vie des données spatiales et permet ainsi de faciliter le déroulement des rencontres. Certaines questions du plan ont été modifiées en fonction du milieu et du niveau hiérarchique des personnes rencontrées.

## **2.4.5 Entrevues**

### **2.4.5.1 Personnes rencontrées**

Au total, 33 personnes de l'ASC ont été rencontrées lors de cette recherche. Cet échantillon représente 4,9 % des employés de l'ASC et 5,5 % de tous ceux qui travaillent au centre John H. Chapman de Saint-Hubert. Quatre d'entre eux sont membres du GTSGD. Les 29 autres personnes ont été rencontrées de façon plus formelle lors d'entrevues individuelles ou de groupe. La liste complète de ces entrevues se trouve à l'annexe 2.

Les personnes rencontrées avaient des profils très variés tant sur le plan de leur champ d'expertise que de leur niveau hiérarchique. En tout, les entrevues ont touché cinq des principaux secteurs de l'ASC et un total de cinq niveaux hiérarchiques. La Figure 2-3 illustre bien la répartition des personnes rencontrées dans l'organigramme de l'ASC. On y remarque une dispersion autant verticale qu'horizontale des personnes rencontrées.



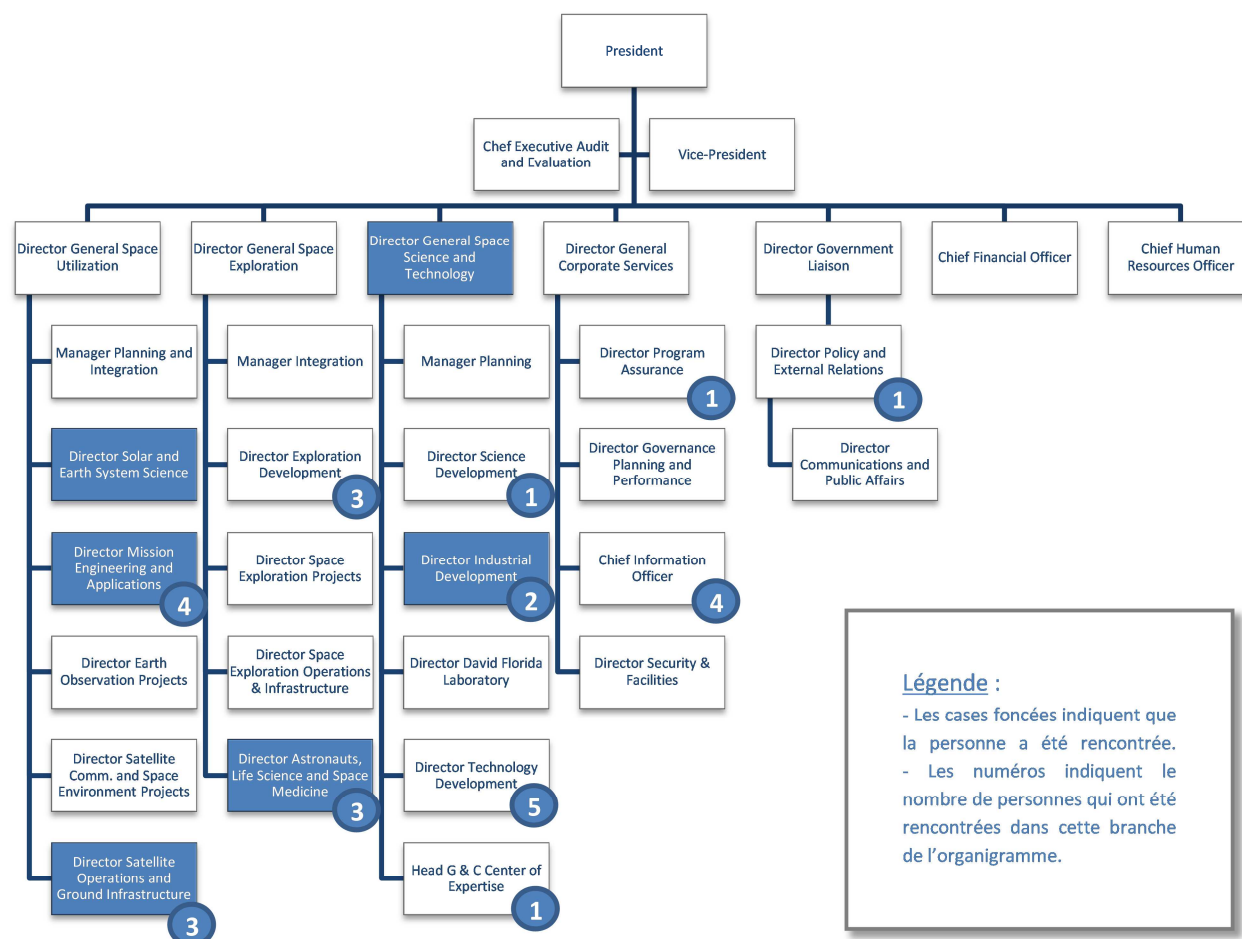


Figure 2-3 : Organigramme de l'ASC indiquant la répartition des personnes rencontrées

### 2.4.5.2 Déroulement

Trois types d'entrevues ont été effectués lors de cette recherche : les rencontres avec le GTSGD, les entrevues de groupes et les entrevues individuelles. Il y a d'abord eu les rencontres avec les quatre membres du GTSGD. Ces rencontres informelles ont eu lieu presque toutes les semaines entre septembre 2010 et mars 2011. Ces rencontres ont principalement pris la forme de remue-ménages ou de réflexions de groupe. Elles ont permis d'avoir une bonne compréhension des objectifs et des contraintes de l'ASC, de connaître les projets passés et futurs liés à la gestion des données et de bien comprendre les défis particuliers touchant les secteurs de la GI et de la TI.

Pour les autres entrevues, le GTSGD a identifié quels étaient les intervenants clés devant être rencontrés et impliqués dans le projet. Le plan d'entrevue fut expédié dans la semaine qui précédait l'entrevue afin de préparer les intervenants et ainsi permettre de tirer plus d'information

pertinente des rencontres. Lors des entrevues, le plan servait de support. Les entrevues se déroulaient sous forme de discussion et les sujets discutés variaient selon le profil des personnes rencontrées et selon l'information qu'ils pouvaient fournir.

En tout, 15 entrevues formelles ont été réalisées, cinq d'entre elles étaient des entrevues de groupe. Ces groupes de discussion ont permis de mettre en lumière et de comparer les façons de faire de différents intervenants provenant d'un même milieu scientifique. Quatre d'entre elles ont été enregistrées de manière à permettre une meilleure implication de l'animateur et une meilleure assimilation de tout ce qui a été dit durant la rencontre.

Pour toutes les entrevues, ce sont les membres du GTSGD qui ont sélectionné et convoqué les participants, et au moins un membre de ce groupe a été présent lors de chacune des rencontres. Les entrevues ont été présentées comme des discussions sur la gestion des données, utiles à la fois pour cette recherche et pour le secteur de la GI qui est responsable du cadre de gestion des données et de son application. Leur durée était généralement d'une heure, mais pouvait parfois varier de quarante minutes à presque deux heures.

## **2.4.6 Élaboration du rapport sur la gestion des données spatiales**

Suite à la cueillette de données, un rapport a été rédigé. Le but de ce rapport est d'aider les secteurs de la GI et de la TI de l'ASC dans la mise en place de pratiques de gestions des données et de faciliter l'élaboration d'un cadre de gestion des données spatiales. Il identifie les enjeux et défis qu'implique la gestion de ce type de données, leur cycle de vie, les rôles et responsabilités des différents intervenants de la gestion des données et fournit des recommandations et des pistes d'analyses qui facilitent la mise en application de ce cadre par l'ASC.

Durant sa création, des versions préliminaires du rapport ont été remises à six reprises au GTSGD afin de discuter de son contenu et de s'assurer qu'il répond bien aux objectifs des secteurs de la GI et de la TI de l'ASC. Une première version complète du rapport a été remise au groupe de travail le 31 mars 2011. Dans les mois qui ont suivi, ce document a été lu, commenté, et approuvé successivement par les quatre membres du GTSGD. Des spécialistes de différents secteurs de l'ASC ont participé à la révision de certaines sections du rapport.

Le contenu du rapport sur la gestion des données spatiales de l'ASC correspond en grande partie au Chapitre 3 de ce mémoire.

### **2.4.7 Élaboration du gabarit de plan de gestion des données scientifiques**

Le gabarit de plan de gestion des données scientifiques est une liste de questions à partir de laquelle les gestionnaires de mission ou les chercheurs de l'ASC construisent le plan de gestion des données de leur projet. Ce plan de gestion devrait idéalement être réalisé avant chacune des missions spatiales et projets de l'ASC. Il devrait être rempli dès la planification du projet et maintenu à jour tout au long du projet puisqu'une partie de l'information ne sera connue que lorsque certaines étapes d'analyse du projet auront été réalisées.

Ce gabarit a été construit à l'aide de différents exemples de gabarit de plan de gestion, principalement celui du DCC, qui a été adapté en fonction de la réalité de l'ASC et à partir du questionnaire construit pour les entrevues. L'objectif des entrevues était en effet partiellement semblable à celui du gabarit de plan de gestion puisque durant les entrevues le chercheur tentait de découvrir comment se faisait la gestion des données spatiales lors des différentes phases de leur cycle de vie. L'objectif du gabarit, lui, est justement d'aider les gestionnaires concernés et les membres de la GI et de la TI à planifier et identifier les différentes étapes de la gestion des données d'un projet. Comme dans le cas du rapport sur la gestion des données, le gabarit a été approuvé par le GTSGD.

### **2.4.8 Analyse**

La dernière étape de cette recherche fut d'analyser la pertinence et la contribution du travail effectué. Cette maîtrise est basée sur l'étude d'un cas bien précis, il est donc essentiel de prendre du recul par rapport aux livrables techniques du projet et d'extrapoler à partir des réflexions qui ont été faites afin de tenter d'en tirer des apprentissages et d'en extraire des généralités applicables dans d'autres contextes. Cette réflexion se retrouve partiellement dans certaines analyses du Chapitre 3 mais elle est principalement présente dans la discussion du Chapitre 4 qui évalue l'impact et les limites de la recherche.

## **2.5 Conclusion**

C'est donc dans un contexte de collaboration avec l'ASC que s'est réalisée cette recherche. Par son accès privilégié au terrain, l'auteur a pu recueillir une multitude d'informations pointues et pertinentes sur la gestion des données spatiales. Tout au long du processus de recherche, la

communication a été fréquente entre le chercheur principal et le GTSGD. Cette communication a permis de réaliser les objectifs de la recherche-intervention c'est-à-dire de concilier les objectifs organisationnels et scientifiques du projet afin que la recherche soit bénéfique à la fois pour la communauté scientifique, mais aussi pour l'organisation étudiée.

Les résultats qui sont présentés dans le Chapitre 3 de ce mémoire découlent de cet accès privilégié à l'ASC et leur valeur repose sur les nombreuses réflexions de groupe et validation du GTSGD. Le Chapitre 3 présente donc les observations et les résultats de la recherche. Ceux-ci seront ensuite discutés et analysés dans le Chapitre 4.

## **CHAPITRE 3    RÉSULTATS**

Le Chapitre 1 a décrit les enjeux liés à la gestion des données techniques ainsi que les différentes initiatives internationales visant à accroître la diffusion et améliorer la gestion de ce type de données. Ce chapitre démontre aussi comment ces enjeux affectent différentes organisations, principalement publiques, et comment celles-ci s’y prennent pour surmonter les défis que représente la gestion de ce type d’actif informationnel.

Pour sa part, le Chapitre 2 a présenté la méthodologie de recherche-intervention utilisée dans le cadre de cette recherche. De façon spécifique, cette approche permet au chercheur de participer à la conception d’outils qui répondent au besoin d’un milieu organisationnel donné, l’ASC, tout en étant impliqué dans l’action. Le chercheur peut, du même coup, en dégager une analyse critique ainsi que des éléments théoriques pertinents à son domaine.

Le Chapitre 3 présente quant à lui les résultats de cette démarche de terrain. Ce chapitre commence par une synthèse de l’information recueillie lors des entrevues. Cette première section décrit quels types de données sont manipulés par l’ASC et dans quelles conditions particulières doivent être gérées ces données. Ensuite, le cycle de vie et les autres outils construits en intégrant ces enjeux particuliers aux éléments théoriques et à l’analyse comparative du Chapitre 1 sont présentés et décortiqués dans la deuxième portion de ce chapitre qui se termine, finalement, par l’explication de la méthode de validation des résultats de la recherche.

### **3.1 Synthèse de l’information recueillie**

Comme l’indique le Chapitre 2 de ce mémoire, en particulier la Figure 2-2, la cueillette d’information a été réalisée en plusieurs étapes : l’analyse de la documentation, les rencontres avec le GTSGD puis les entrevues. De nombreux concepts ont pu être identifiés lors des deux premières étapes. C’est par contre lors des rencontres qu’il a été possible d’avoir un aperçu de la situation réelle de la gestion des données dans certains secteurs de l’ASC. Lors de ces rencontres, il a été possible de mieux comprendre dans quel environnement se faisait la gestion des données de l’ASC et à quels défis faisaient face les personnes concernées.

La première moitié de ce chapitre est dédiée à ces observations tirées des entrevues. Elle est essentiellement descriptive. Tout d’abord, les types de données qui sont produites et utilisées par l’ASC y sont décrits ainsi que la façon selon laquelle ils sont présentement gérés. Ensuite, cette

section comprend une description des défis particuliers qui touchent la gestion des données à l'ASC. Cette première section du Chapitre 3 est donc une synthèse de l'information collectée lors des entrevues et de la réflexion qui s'y est amorcée. Elle décrit la vision que l'auteur s'est fait de la situation.

### **3.1.1 Données de l'ASC**

L'ASC a, entre autres, une vocation scientifique. Elle effectue et finance une multitude de projets de recherche et de missions satellitaires afin de faire progresser les sciences liées à l'espace au Canada. De nombreux projets requièrent l'acquisition et le traitement d'une grande quantité de données brutes et transformées.

De plus, l'ASC fournit des ensembles de données spatiales aux autres ministères afin que ceux-ci accomplissent leur mandat[21]. Plusieurs ministères et agences fédérales qui ont besoin de données provenant de l'espace ou étant liées à la recherche spatiale utilisent de manière significative les services de l'ASC pour les obtenir. L'ASC est donc amenée à collecter et manipuler une très grande quantité de données de types et d'utilités très variés.

L'AAP 2011-2012[21] présente les trois types de missions de l'ASC ainsi que les deux types de données qui en découlent. Ce document clé décrit l'ensemble des programmes de l'ASC, leurs activités et les objectifs de ces activités pour la période couvrant les années 2011 et 2012. Les trois types de missions spatiales effectuées par l'ASC qui y sont présentées sont les missions scientifiques, d'observation de la Terre et de communication. À chacun de ces deux premiers types de missions correspond un type de données spatiales qui a des caractéristiques qui lui sont propres.

Ces caractéristiques sont décrites dans cette section du mémoire. Bien qu'ils ne soient pas étudiés dans le reste de ce document, une description des données autres que spatiales et des échantillons se trouve à la fin de la présente section.

#### **3.1.1.1 Données spatiales**

##### *3.1.1.1.1 Données scientifiques*

Les missions scientifiques regroupent tous les projets de recherche scientifiques menés ou financés par l'ASC. Ces missions englobent la conception, le développement de technologies, et

la collecte de données nécessaires à la réalisation de ces recherches. Ces missions sont souvent effectuées en collaboration avec des universités, d'autres ministères, l'industrie spatiale ou des agences spatiales étrangères et touchent à une multitude de domaines dont, entre autres, la robotique, les sciences atmosphériques, les sciences de la vie, la géologie, l'astronomie et l'astrophysique.

Les données qui découlent de ces missions sont des données d'expérience telle qu'elles ont été décrites à la section 1.1.2.1 du présent mémoire. Ces données sont généralement créées ou rassemblées dans le but de confirmer ou d'infirmer une hypothèse, et elles ont une portée très spécifique.

La gestion qui est présentement faite des données scientifiques à l'ASC varie grandement d'un secteur à l'autre. Dans certains domaines comme ceux des sciences planétaires ou de la robotique, les données sont déjà largement diffusées et partagées tandis que d'autres domaines scientifiques ont plutôt tendance à conserver les données sur les disques durs des chercheurs qui les ont acquises[24]. Au sein d'un même domaine, les pratiques de gestion des données peuvent varier considérablement puisque les données peuvent être recueillies par l'ASC, par des scientifiques de l'extérieur qui collaborent à un projet de l'Agence ou par des scientifiques extérieurs à l'ASC travaillant sur leur propre projet mais recevant du financement de l'ASC.

Les pratiques de gestion des données varient aussi selon le ou les partenaires du projet. Par exemple, lorsque des chercheurs de l'ASC effectuent des projets avec la NASA, les données spatiales de ses projets sont gérées en fonction des exigences de l'agence américaine.

Lors d'une expérience financée par l'ASC, la gestion des données dépend du type d'entente conclue avec le partenaire du projet. Un contrat peut contenir des clauses concernant la gestion des données scientifiques ou peut identifier les données comme étant un livrable du projet. Lorsque l'ASC finance des projets par l'entremise de subventions et de contributions, c'est le bénéficiaire qui a comme responsabilité de faire la gestion de ces données, l'ASC ne fait qu'apporter un soutien financier au projet.

#### *3.1.1.1.2 Données d'observation de la Terre*

La deuxième catégorie de données est celle recueillie lors des missions d'observation de la Terre telles que RADARSAT-1, RADARSAT-2, SCISAT et, bientôt, *RADARSAT Constellation*

*Mission* (RCM). Ces missions ont comme objectif de concevoir, développer et opérer des technologies et des satellites en orbite autour de la Terre afin de produire des données, des informations et des images de la Terre, de son atmosphère et de surveiller les objets et débris qui gravitent autour d'elle[21]. Les données qui en découlent sont en grande majorité des données géospatiales telles que décrites dans la section 1.1.2.2 de ce mémoire. Ces données sont principalement destinées aux autres ministères et aux universités. La réception et le traitement de ces données ne sont pas organisés par un chercheur, mais plutôt par des gestionnaires de projet et des équipes d'opération de satellites à l'ASC ou chez ses partenaires.

Des procédures internes strictes et des politiques de données régissent la gestion qui est faite de ce type de données. Voici les grandes lignes de la gestion des données spatiales des quatre principales missions d'observation de la Terre de l'ASC :

#### 3.1.1.1.2.1 Données de RADARSAT 1 et 2

Les données provenant des missions RADARSAT 1 et 2 sont des données d'observation de la Terre. Ces données sont reçues à Cantley (Qc) (anciennement Gatineau), Prince-Albert (Sk), par une quarantaine de stations étrangères et par des stations mobiles. L'ASC paye pour que les données reçues au Canada (Cantley et Prince-Albert) soient archivées à Ottawa, au *Centre canadien de télédétection* (CCT). Dans le cas des stations mobiles et étrangères, les données sont conservées et transformées sur place. Si une station ferme, l'ASC rapatrie les données brutes de cette station et en effectue la conservation. Les données sont transformées en produits par les stations étrangères et mobiles ainsi que par le partenaire commercial de l'ASC dans les missions RADARSAT.

Les politiques de données de RADARSAT 1 et 2 obligent l'ASC à conserver les données brutes au moins 15 ans après la fin de la mission. Les missions sont encore en cours et les gestionnaires n'ont pas encore pris de décision sur ce qui serait fait avec les données après le délai prescrit de 15 ans.

Ces missions ayant un aspect commercial, leurs données ne peuvent pas toutes être diffusées librement au grand public. Le *Remote Sensing Space System Act* (loi C-25)[64] stipule qu'il est nécessaire de détenir un permis pour opérer un satellite. Le détenteur d'un permis doit savoir où sont les données provenant du satellite et il ne peut les distribuer que dans un cadre défini. L'ASC possède le permis pour RADARSAT-1, mais c'est son partenaire commercial qui détient



celui pour opérer RADARSAT-2. L'Agence spatiale peut diffuser les données en fonction des restrictions liées à la licence, laquelle tient compte des intérêts commerciaux du partenaire de l'ASC.

Le Centre canadien de télédétection (CCT) gère la bibliothèque de données brutes RADARSAT-1 au nom de l'ASC et la bibliothèque des données brutes RADARSAT-2 au nom de son partenaire commercial. À l'initiative du CCT, une bibliothèque des produits RADARSAT-2 du gouvernement a été créée ainsi qu'un catalogue et un système de gestion des accès basé sur le profil des usagers, permettant ainsi de respecter les licences et les droits de partage des produits.

#### 3.1.1.1.2.2 Données de SCISAT

De son côté, SCISAT produit majoritairement des données géospatiales d'observation de l'atmosphère de la Terre, mais produit aussi certaines données scientifiques. Les données de SCISAT sont reçues à l'Université de Waterloo et à l'ASC. Étant donné qu'il n'y a pas d'aspect commercial à cette mission, les archives de données pourraient théoriquement être mises à la disposition du public. Par contre, aucun des deux partenaires n'a présentement l'obligation de conserver les données de SCISAT. Leur conservation est donc effectuée par, et pour, les utilisateurs immédiats des données.

#### 3.1.1.1.2.3 Données de RCM

Finalement, RCM est la suite des missions RADARSAT 1 et 2. Cette mission qui prévoit le lancement de trois satellites a pour but d'assurer la continuation des observations de RADARSAT 1 et 2. Le lancement du premier satellite est prévu pour 2016. La politique de données de RCM est en cours d'élaboration au moment de l'écriture de ce mémoire.

#### 3.1.1.1.3 Données de télécommunication

La troisième catégorie de mission spatiale identifiée dans l'AAP de l'ASC regroupe toutes les missions de télécommunications. Bien que ces missions aient une grande importance stratégique pour le Canada, les données qui en découlent sont souvent opérationnelles et n'ont pas de valeur au-delà de leur utilisation immédiate. Ces missions servent à mettre en place des systèmes de positionnement et de télécommunication et ne fournissent pas de données d'observation. Ce type de donnée ne sera donc pas considéré dans le reste de ce mémoire.

### 3.1.1.2 Données non spatiales

L'ASC possède aussi des données non spatiales comme des données de programmation, des données de test ou des données utilisées pour la gestion de l'Agence. Ces données constituent elles aussi des actifs informationnels et comme elles ont souvent une valeur opérationnelle, elles doivent respecter la *Directive sur la tenue de documents*[25] émise par le gouvernement fédéral.

Par contre, ces données n'ont pas de valeur scientifique et n'ont généralement pas de nouvelles utilisations possibles. Leur conservation peut être importante mais seulement pour des raisons de validation ou de contre-vérification. La gestion qui doit en être faite est donc plus proche de la gestion des documents que de la gestion des données spatiales telle que décrite dans ce mémoire. Les données non spatiales de l'ASC ne sont donc pas considérées dans les analyses qui suivent.

### 3.1.1.3 Échantillons

Finalement, dans certains domaines de recherche comme les sciences de la vie ou les sciences planétaires, l'élément brut découlant d'une expérience peut être une série d'échantillons. Au lieu de recueillir des observations qui sont encodées et conservées sur des supports numériques, certaines missions recueillent des échantillons physiques, parfois vivants, sur lesquels seront ensuite effectuées de multiples observations. Ces échantillons ne sont pas des données, les données brutes sont les observations qui en découlent. L'échantillon est donc le sujet d'observation duquel proviennent les données brutes. Les échantillons ne sont donc pas pris en compte par le présent mémoire.

Plusieurs principes énoncés dans ce mémoire pourraient tout de même s'appliquer aux échantillons. La préservation et la gestion selon le cycle de vie sont des exemples de concepts qui s'appliquent aussi aux échantillons. Mais bien que les objectifs soient souvent semblables, la gestion des échantillons entraîne des problématiques très complexes et très différentes de celles liées à la gestion de données digitales. L'aspect périssable de certains échantillons vivants ou inertes et l'impossibilité de dupliquer les échantillons ne sont que quelques exemples de ces éléments particuliers qui empêchent l'application des conclusions de ce rapport à la gestion des échantillons.

### **3.1.2 Conditions particulières**

Durant les entrevues, de nombreux enjeux de la gestion des données ont pu être identifiés. Certains d'entre eux sont spécifiques à la situation de l'ASC, d'autres sont plus généraux. L'identification des conditions particulières dans lesquelles doit opérer l'ASC est un apport important de cette recherche puisqu'elles influenceront le cycle de vie des données et les responsabilités de chacun des intervenants de la gestion des données spatiales. Ces éléments ne représentent pas des problématiques mais plutôt des contraintes propres à l'ASC, qui caractérisent cette organisation lorsqu'il s'agit de la gestion des données. Ces éléments sont des observations qui reflètent le point de vue de l'auteur.

#### **3.1.2.1 Fonctionnement actuel**

Les entrevues ont permis d'observer le fonctionnement actuel de l'ASC en ce qui concerne la gestion des données spatiales. Bien qu'il n'y ait pas encore de lignes directrices et d'objectifs corporatifs pour la gestion de ces données, celles-ci sont manipulées par certains employés de l'ASC dans le cadre de leurs activités et fonctions. La gestion qui en est faite n'est pas uniforme mais elle est parfois plus qu'adéquate. Il a donc fallu identifier comment se faisait cette gestion et quels sont les défis qui en découlent.

##### *3.1.2.1.1 Procédures*

Un premier élément noté lors des entrevues est l'absence de procédures corporatives de gestion des données. Pour réussir à faire une gestion optimale des données, il est important que tous les secteurs de l'organisation suivent les mêmes procédures, soient soumis aux mêmes exigences et visent les mêmes objectifs. L'uniformité facilite la coordination des efforts de gestion et permet de mieux communiquer les attentes de l'organisation.

Cette constatation n'est pas surprenante puisque le but de l'actuel projet de l'ASC est justement de créer des procédures de gestion des données spatiales uniformes qui considèrent l'ensemble du cycle de vie des données. L'ASC a bien identifié cet élément et met les efforts nécessaires pour bonifier ces pratiques.

Il a tout de même été intéressant de constater que la grande majorité des personnes rencontrées ont identifié le besoin d'uniformiser les procédures de gestion des données et accueillent positivement les initiatives de l'organisation dans ce domaine.

### *3.1.2.1.2 Culture*

Les entrevues ont aussi permis d'observer des divergences dans la vision de la gestion des données des intervenants des différents domaines scientifiques. Comme c'est probablement le cas dans l'ensemble de la communauté scientifique, les chercheurs et gestionnaires de certains domaines scientifiques sont plus ouverts à la diffusion et au partage de leurs données que le sont certains de leurs collègues œuvrant dans d'autres domaines de la science. Certains domaines ont une forte culture de préservation et de partage de leurs données tandis que d'autres, plus compétitifs, travaillent dans le secret. Comme pour l'élément précédent, le défi repose ici dans l'uniformisation des pratiques qui permettent de respecter les principes sous-jacents à une bonne gestion des données.

Il est aussi important de constater qu'un changement de cette ampleur constitue bien plus qu'un changement des procédures, il introduit un changement de culture organisationnelle. Il demande à l'organisation de revoir ses pratiques par rapport à la gestion des données spatiales.

### **3.1.2.2 Financement**

Plusieurs enjeux importants de la gestion des données spatiales de l'ASC découlent de son mode de financement. Un enjeu important de la situation de l'ASC est la provenance de son financement. Étant une agence gouvernementale, l'ensemble du budget de l'ASC provient des fonds publics. L'ASC utilise ensuite l'argent qui lui a été attribué pour financer de nombreux projets techniques et scientifiques au sein de l'Agence, mais aussi à l'extérieur sous la forme contrats, partenariats, subventions et contributions diverses.

#### *3.1.2.2.1 Provenance du financement*

Comme l'entièreté du budget de l'ASC provient des fonds publics canadiens, l'Agence spatiale rend des comptes quant à l'utilisation de cet argent aux contribuables canadiens. L'ASC doit démontrer aux Canadiens qu'elle fait bon usage de l'argent qui lui est confié.

L'ASC désire par contre faire plus que de simplement rendre des comptes aux citoyens. Une importante partie du budget de l'ASC est utilisée de façon directe ou indirecte pour la création, l'acquisition ou la transformation des données spatiales. Il s'avère donc enviable de rendre accessible l'actif informationnel que représentent les données spatiales aux citoyens, chercheurs et entreprises qui les ont financés, lorsqu'il est possible de le faire. La mission de l'ASC consiste justement à faire bénéficier les Canadiens de l'exploration et des sciences et technologies spatiales. L'ASC permet aux industries aérospatiales de se développer, les technologies aident les entreprises à être plus profitables. La diffusion des données spatiales est donc une autre façon pour l'ASC de remplir son mandat.

Toutefois, il n'est ni possible ni enviable, dans le contexte actuel, que la diffusion des données se fasse de façon totalement libre. Certaines mesures doivent être mises en place afin de protéger les intérêts des chercheurs, des partenaires commerciaux de l'ASC et des sujets de recherche comme il sera expliqué à la section 3.1.2.3.

#### *3.1.2.2.2 Destination du financement*

L'utilisation du budget de l'ASC est elle aussi un enjeu important. En effet, un des rôles de cette agence gouvernementale canadienne est de financer la recherche liée aux sciences et technologies spatiales au Canada. Ce financement se fait à l'aide de contrats, de subventions et de contributions. Le problème est qu'il peut s'avérer difficile d'imposer aux partenaires qu'ils conservent et diffusent les données de leur projet même si celles-ci ont été acquises grâce à de l'argent public.

Lorsque l'ASC procède par contrat, il serait envisageable de tenter d'inclure l'accès aux données ou les données elles-mêmes en tant que livrables du projet de manière à pouvoir les récupérer et en faire la gestion dans un but de conservation. C'est par contre différent lorsqu'il s'agit de subventions et de contributions. Ces modes de financement sont régis par la *Politique sur les paiements de transfert*[65]. Ce mode de financement est différent de celui des contrats et ne permet pas à l'ASC d'exiger des livrables de la part du bénéficiaire de la subvention ou de la contribution.

### 3.1.2.3 Contraintes de diffusion

Tel qu'expliqué à la section 1.1.3.4 portant sur l'accessibilité des données, la diffusion est un élément important de la conservation. Par contre, cette diffusion ne peut, ni ne doit, se faire de n'importe quelle façon. Il n'est pas bénéfique de diffuser n'importe quel ensemble de données n'importe quand et pour n'importe qui. Voici trois éléments ayant été soulevés lors des entrevues qui expliquent la nécessité d'encadrer la diffusion des données.

#### 3.1.2.3.1 *Reconnaissance des chercheurs*

Le mode de reconnaissance des chercheurs nous force à réfléchir aux conséquences de la diffusion des données et à encadrer adéquatement cette pratique. La plupart des chercheurs qui travaillent pour ou avec l'ASC obtiennent leur reconnaissance en grande partie de la publication des résultats de leurs recherches. Si les données qu'ils utilisent sont rendues accessibles trop tôt et de façon trop ouverte, il est possible qu'un autre chercheur travaillant dans le même domaine batte de vitesse le premier chercheur et publie des résultats provenant de ces données avant lui (*scoop*). Le premier chercheur aura donc travaillé très longtemps afin de développer des outils et de collecter les données pour finalement ne pas obtenir de reconnaissance pour ces efforts.

#### 3.1.2.3.2 *Aspect commercial*

Aussi, étant donné l'ampleur de certaines missions d'observation de la Terre, l'ASC collabore parfois avec des entreprises du secteur privé. Dans le cas de RADARSAT-1, l'ASC a une entente commerciale pour la distribution des données de son satellite. Pour RADARSAT-2, l'ASC a plutôt prépayé une allocation des données à l'entreprise propriétaire du satellite. L'objectif de ce mémoire n'est pas d'entrer dans les termes spécifiques de chacune de missions satellitaires, mais plutôt d'identifier les défis que soulève ce type de collaboration à volet commercial.

Tout d'abord, selon les ententes, certaines données recueillies sont la propriété de l'ASC, d'autres de l'entreprise privée. D'un côté l'ASC pense à mettre en place les principes de données libres au sein de son organisation. De son côté, le partenaire de l'ASC cherche des clients afin de leur vendre les données saisies par le satellite. Si l'ASC en venait à diffuser l'ensemble de ses données sans restriction, elle viendrait nuire aux activités commerciales de son partenaire. L'aspect commercial oblige donc l'ASC à se questionner, comme elle le fait présentement, sur les limites de la libre diffusion de ses différents ensembles de données.

Cet enjeu très particulier est très peu, sinon pas du tout, documenté dans la littérature puisqu'il est engendré par la nature spécifique des missions RADARSAT. La présence d'un partenaire commercial et la politique de données qui en découle limitent la diffusion des données et empêche des initiatives comme l'« Open Data » d'être applicable à ce type de projet.

#### *3.1.2.3.3 Respect de la vie privée*

Le respect de la vie privée est lui aussi une contrainte à la diffusion de certaines données. C'est particulièrement le cas dans le domaine des sciences de la vie qui effectue, par exemple, des expériences afin d'assurer la bonne santé des astronautes durant et après leurs séjours dans l'espace.

Il est essentiel de garder secrètes l'identité et l'information privée relative aux sujets humains en vertu de la *Loi sur la protection des renseignements personnels*[66]. Avant de diffuser un ensemble de données il est important d'y retirer toute information pouvant servir à identifier le ou les sujets humains de la recherche. Ceci peut restreindre partiellement ou entièrement la diffusion et la réutilisation des données par d'autres utilisateurs.

#### **3.1.2.4 Propriété des données**

Un autre élément qui caractérise la situation de l'ASC est la grande complexité liée à la propriété des ensembles de données. L'ASC effectue une multitude de projets en coopération avec de nombreux partenaires. Ces partenariats ont chacun leurs arrangements. Du point de vue de la GI-TI il est difficile de savoir si un ensemble de données appartient ou appartiendra à l'ASC. Afin d'encadrer la gestion des données de l'ASC, il est essentiel que l'organisation soit en mesure de savoir quels ensembles de données appartiennent à l'organisation.

#### **3.1.2.5 Aspect politique**

Finalement, la mise en place de nouvelles procédures et infrastructures peut être affectée par des contraintes politiques ou budgétaires. Toutes les organisations doivent se donner des objectifs et des priorités. Les projets sont ensuite sélectionnés de façon à répondre le mieux possible à ces objectifs tout en respectant le budget de l'organisation.

L'ASC doit elle aussi sélectionner les projets qui répondent le mieux à ses priorités mais, étant une agence gouvernementale, elle doit effectuer ces choix dans un contexte gouvernemental et

aligner autant que possible ses objectifs et stratégies avec les priorités du gouvernement qui sont énoncées dans le Discours du Trône. Ces priorités dépendent du gouvernement en place et peuvent changer suite au dépôt d'un budget gouvernemental, d'un Discours du Trône ou d'une élection.

Cet aspect politique peut affecter un projet de grande ampleur comme la refonte de certaines pratiques administratives. Un bon exemple de contrainte politique est la *Directive sur la tenue de documents*[25]. Cette directive émise par le gouvernement fédéral en 2009 oblige les ministères et agences gouvernementales à gérer les ressources documentaires à valeur opérationnelle tout au long de leur cycle de vie.

### 3.1.3 Critères à respecter

Cette première partie du Chapitre 3 fait l'énumération des observations qui ont été faites lors des entrevues. Celles-ci sont d'une grande importance pour le reste du chapitre. Ce sont ces conclusions tirées des entrevues qui ont permis de bien cerner les conditions particulières de la gestion des données spatiales à l'ASC. La simple identification de ces enjeux est de grande valeur puisqu'elle permet de réaliser des outils qui répondent adéquatement aux besoins spécifiques et au contexte de l'ASC. Plusieurs de ces particularités peuvent aussi être applicables à d'autres contextes précis ayant des similitudes avec le contexte de l'ASC.

Afin de répondre adéquatement à ces contraintes particulières, les outils développés dans le cadre de gestion des données et à la suite de cette recherche doivent et devront respecter les critères suivants :

- englober à la fois les données d'expérience et les données d'observation de la Terre, et ce dans tous les domaines touchés par l'ASC;
- encourager la diffusion des données tout en respectant les contraintes liées aux sujets de recherche, aux chercheurs et aux partenaires commerciaux de l'ASC;
- respecter les lois et politiques auxquelles est soumise l'ASC ainsi que les procédures déjà en place.



## 3.2 Résultats

Cette portion du chapitre est consacrée aux résultats purs de la recherche. Ces résultats regroupent l'ensemble des analyses et des outils qui ont été effectués afin de décrire ou de normaliser la gestion des données spatiales. Ces résultats sont :

- le cycle de vie des données;
- les rôles et responsabilités des différents intervenants de la gestion des données spatiales;
- le gabarit de plan de gestion des données spatiales de l'ASC;
- les analyses et recommandations liées à la mise en œuvre d'un cadre de gestion des données à l'ASC.

### 3.2.1 Cycle de vie

Afin d'effectuer une bonne gestion des données spatiales, il s'avère essentiel de bien comprendre l'ensemble des étapes par lesquelles passent ces données. La conception d'un cycle de vie représente donc une tâche incontournable dans l'amélioration des pratiques de gestion des données.

Le cycle de vie des données spatiales est l'élément central de cette recherche. Il intègre les éléments théoriques présentés au Chapitre 1 et sert de fondement pour les outils et analyses des Chapitre 3 et Chapitre 4. C'est par le cycle de vie que les éléments conceptuels sont mis en pratique. C'est aussi le cycle de vie qui permet d'élaborer la liste de tâches et responsabilités qui doivent être effectuées en lien avec les données.

#### 3.2.1.1 Conception

Le cycle de vie des données spatiales de l'ASC fut élaboré principalement à l'aide des trois cycles de vie déjà existants, présentés au Chapitre 1. Ces trois cycles de vie sont le cycle de vie de l'information de l'ASC[19], le cycle de vie des données de recherche canadiennes[24] ainsi que le cycle de vie des données scientifiques du DCC[36].

La gestion des données spatiales représentant un cas particulier de la gestion de l'information de l'ASC, il s'avère essentiel que les deux cycles de vie soient en lien l'un avec l'autre et que ces

deux cycles utilisent le même vocabulaire pour les mêmes activités lorsqu'il est possible de le faire.

Le cycle de vie des données spatiales de l'ASC est en quelque sorte une représentation détaillée du cycle de vie de l'information puisqu'il représente un cas précis d'application de ce processus. Il ne peut donc pas contredire les procédures de gestion de l'information déjà en place. Le cycle de vie des données spatiales doit plutôt présenter les particularités de la gestion d'un type précis d'actif informationnel de l'ASC.

Étant donné sa portée très large, le cycle de vie de l'information de l'ASC s'avère peu précis quand vient le temps d'encadrer la gestion des données techniques. Certaines étapes ont donc dû être déplacées ou scindées en activités plus précises et des boucles supplémentaires y ont été ajoutées afin de représenter le cas précis des données spatiales. Le vocabulaire a été majoritairement préservé afin de marquer une continuité entre ces deux cycles de l'ASC.

Les trois cycles de vie étudiés diffèrent du cycle de vie des données spatiales de par le type d'actif informationnel qu'ils caractérisent, par le niveau de profondeur qu'ils présentent ou par le contexte de gestion qu'ils illustrent. Ils apportent néanmoins chacun des éléments intéressants devant être intégrés à la gestion des données spatiales de l'ASC. C'est donc en combinant ces trois cycles de vie qu'a été élaborée la première ébauche du cycle de vie des données spatiales. Celle-ci a été raffinée suite à des discussions avec le GTSGD et a ensuite été approuvée par le GTSGD comme l'indique la Figure 2-1 de la méthodologie.

### 3.2.1.2 Description du cycle de vie

#### 3.2.1.2.1 Description générale

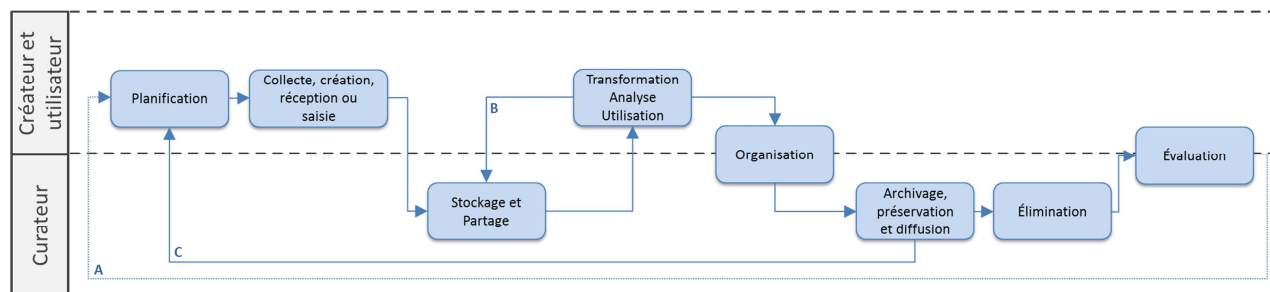


Figure 3-1 : Cycle de vie des données spatiales de l'ASC

La Figure 3-1 présente le cycle de vie des données spatiales de l'ASC. Ce cycle illustre la séquence des activités de gestion et de manipulation des données spatiales depuis la planification de leur acquisition jusqu'à une phase d'évaluation du processus.

En plus de la boucle « A » qui vient fermer le cycle et illustrer sa continuelle répétition, deux boucles de retour ont été insérées à l'intérieur du cycle de vie des données. La plus petite, la boucle « B », sert à illustrer la possibilité de transformer, d'analyser ou d'utiliser les ensembles de données plus d'une fois. La boucle indique que durant toutes ces phases d'utilisation et de transformation, il est nécessaire d'entreposer et de faire le partage, si nécessaire, des différents ensembles de données et/ou produits générés.

La boucle « C » illustre quant à elle la réutilisation des ensembles de données. Le but de l'archivage, de la préservation et de la diffusion est de permettre que les données soient réutilisées. Il est donc normal que cette activité mène à la planification d'un nouveau projet lors duquel la collecte des données se fera en partie ou entièrement en récupérant des données archivées. Le désir de l'ASC, du Gouvernement du Canada et d'une multitude d'autres pays, agences ou organisations d'accroître la réutilisation des données techniques a été démontré dans le premier chapitre de ce mémoire. Il est donc essentiel que la réutilisation soit représentée dans le cycle de vie des données spatiales de l'ASC.

En plus de ces boucles, les rôles de curateur et de créateur/utilisateur ont été illustrés dans le cycle de vie. Le cycle représente visuellement le partage des tâches entre ces deux principaux intervenants. Ces deux rôles ont été choisis puisque ce sont eux qui manipulent les données. Les autres intervenants agissent principalement lors des phases de planification et d'évaluation et ne manipulent que rarement et indirectement les données.

Ce cycle de vie a donc l'avantage d'illustrer directement l'essentielle interaction entre les curateurs et créateurs/utilisateurs des données. Cette interaction avait été identifiée à la section 1.1.3.1 comme étant un élément clé de la conservation des données. Un élargissement du cycle de vie des données spatiales de l'ASC se trouve à l'annexe 4.

#### *3.2.1.2.2 Activités*

Chacune des étapes du cycle de vie se trouve ici explicitée. Ces descriptions présentent l'objectif visé par l'activité, sa raison d'être et comment elle est ou doit être réalisée par l'ASC. Dans

certaines situations, la distinction existe entre la gestion des données d'observation de la Terre et celle des données scientifiques. Lorsque c'est le cas, ces distinctions sont adéquatement documentées.

#### 3.2.1.2.2.1 Planification

Cette activité du cycle de vie est primordiale puisque toutes les autres seront basées sur le plan qui y est fait. Avant d'entreprendre un projet, le gestionnaire doit planifier la gestion des données liées à ce projet. Il faut construire un plan de gestion des données qui décrit comment sera effectuée la gestion des données tout au long de leur cycle de vie. Ce plan décrit la méthode d'acquisition, la quantité, le format, la valeur des données ainsi que les différents intervenants qui auront accès aux données et comment le partage sera effectué entre ceux-ci. Si la planification est bien faite, il sera facile pour tous les intervenants de préparer et de réaliser chacune des autres activités du cycle de vie. L'importance de la planification n'empêche pas que la conception du plan de projet est un processus évolutif. La planification doit être faite le plus tôt possible mais doit inévitablement être complétée et raffinée au fur et à mesure que l'information pertinente au projet est acquise.

#### 3.2.1.2.2.2 Collecte, création, réception ou saisie

Une fois le projet planifié, il faut procéder à l'acquisition des données brutes du projet. Cette deuxième activité regroupe donc toute situation où une personne fait l'acquisition de données qu'il n'avait pas en sa possession auparavant. Cette acquisition peut prendre différentes formes selon le type de projet et la provenance des données. Les données brutes peuvent être créées ou récupérées de façon manuelle ou automatisée par des employés de l'ASC ou par ses partenaires.

Il est aussi possible, lors de projets scientifiques, de faire une recherche en utilisant les données brutes ou transformées d'une autre expérience ou d'une mission précédente pourvu que celles-ci aient été adéquatement conservées et documentées. Dans ce cas, l'acquisition des données se fait par la réception des données d'une autre équipe de recherche ou en allant les récupérer dans une base de données ou un portail.

Dans le cas des missions d'observation de la Terre, les satellites sont programmés ou opérés dans le but de faire des acquisitions d'images. Ces données sont ensuite transmises au sol.

Si les données sont créées, il est important de décrire de façon détaillée le contexte ainsi que les méthodes de saisie utilisées. Si les données sont acquises à partir d'une source extérieure à l'ASC, il faut s'assurer de se conformer aux conditions d'utilisation lors de l'usage de ces données.

#### 3.2.1.2.2.3 Stockage et partage

Cette troisième activité est effectuée pendant le projet. L'activité vise à entreposer les données recueillies ou acquises par l'ASC ou au nom de l'ASC et à les rendre accessibles aux utilisateurs pour qu'ils puissent en faire l'analyse et/ou la transformation. Cette activité peut impliquer une première organisation des données afin d'en faciliter la compréhension et le partage. L'entreposage doit assurer un accès aux différents collaborateurs du projet, si collaboration il y a, mais doit aussi tenir compte de toutes les contraintes potentielles d'accès. Ces contraintes peuvent concerner un droit de premier regard (embargo) des chercheurs, l'entente de collaboration entre différents organismes ou chercheur ou simplement des raisons commerciales, de sécurité, de confidentialité ou de respect de la vie privée. Il est important que ces conditions particulières soient inscrites dans le plan de gestion de données du projet.

Les données des missions d'observation de la Terre sont généralement archivées par des partenaires de l'ASC ou des stations étrangères et la distribution est effectuée selon les termes et conditions de la politique de données et la licence.

#### 3.2.1.2.2.4 Transformation, analyse et utilisation

Cette activité regroupe toutes les étapes de manipulation des données effectuées durant le projet ou la mission, qu'elles soient réalisées manuellement ou automatiquement, par l'ASC ou par ses partenaires. L'objectif de cette phase est la réalisation des objectifs du projet ou de la recherche pour laquelle les données ont été acquises en premier lieu.

Dans le cas d'une recherche, cette utilisation consiste généralement en la transformation et en l'analyse des données afin de tester différentes hypothèses de recherche. Le chercheur doit bien documenter les étapes de transformation de ses données sans quoi les ensembles de données transformées ne pourront pas être interprétés adéquatement. La description des données devrait idéalement avoir été inscrite dans le plan de gestion des données rédigé lors de l'étape de planification.

Dans le cas des missions d'observation de la Terre, les données acquises sont généralement destinées à la recherche ou à une utilisation opérationnelle, comme par exemple, la cartographie du couvert de glace et l'aide à la navigation maritime par le *Service canadien des glaces*. Un premier traitement des données est effectué par le processeur. Des produits sont ensuite créés à partir des données brutes et transformées. Ces transformations doivent être adéquatement documentées.

Une fois ces transformations effectuées, il existe plusieurs ensembles de données : des données brutes, des données transformées de différents niveaux et des produits à valeur ajoutée réalisés avec les données. Ces ensembles de données et produits n'ont pas tous la même valeur aux yeux des différents utilisateurs de données (communauté scientifique, entreprises, citoyens canadiens, etc.). Il sera donc important durant la phase d'organisation d'identifier quelles données représentent un intérêt pour quelle communauté afin de conserver les bons ensembles de données et d'en effectuer la diffusion de la façon la plus efficace et ciblée possible.

Après chaque transformation, il est essentiel de réévaluer les besoins d'entreposage et de partage du nouvel ensemble de données résultant de la transformation. Celui-ci peut avoir besoin d'être partagé avec les partenaires du projet.

#### 3.2.1.2.2.5 Organisation

Cette étape a pour but de rassembler, trier, organiser et documenter les ensembles de données afin d'être prêt à en faire l'archivage, la préservation et la diffusion. Cette étape vise l'atteinte des objectifs à long terme de la conservation et de la réutilisation des données du projet. Bien qu'elle demande la participation du créateur des données, cette étape s'effectue généralement après le projet ou en parallèle avec celui-ci.

Dans le cas des données scientifiques, il s'agit de déterminer quelles données seront conservées et, si nécessaire, de les modifier en vue de leur conservation. Cette nouvelle transformation consiste généralement en un tri des données pertinentes, une censure des données sensibles, et, potentiellement, une agrégation des données afin de les rendre plus attrayantes et utiles pour d'autres utilisateurs potentiels.

Dans les cas des missions d'observation de la Terre, les données sont organisées dès leur réception afin d'être archivées et distribuées. Il en va de même pour les produits.

L'organisation des données inclut aussi le développement de métadonnées détaillées et complètes qui permettront l'interprétation et la recherche des ensembles de données ainsi que le développement d'un calendrier de conservation pour ces ensembles de données. Il est primordial de documenter la mission, les instruments, les logiciels, les étapes de transformation et les articles et publications en lien avec les données. Il faut déterminer la criticité des ensembles de données afin de déterminer l'effort de préservation qui devra être fournie. La criticité reflète la facilité d'accès dont ont besoin les différents utilisateurs durant la phase d'archivage du cycle de vie. Un ensemble de données qui doit être consulté tous les jours sera jugé extrêmement critique tandis qu'un autre ensemble, conservé pour sa valeur historique, et dont aucune utilisation immédiate n'a été identifiée au moment de son organisation, ne sera pas jugé critique.

Finalement, il est important lors de l'organisation de déterminer la durée de conservation requise pour chacun des ensembles de données et d'inscrire ces durées dans un calendrier de conservation qui permettra d'en faire le suivi.

Afin que l'organisation des données soit réalisée adéquatement, il faut qu'elle soit effectuée conjointement par le curateur et le créateur des données. Le plan de gestion des données est un élément important pour assurer l'organisation des données puisqu'il décrit les différents ensembles de données brutes et transformées en plus d'identifier la valeur potentielle de ces ensembles de données.

#### 3.2.1.2.2.6 Archivage, préservation et diffusion

Le but de cette étape du cycle de vie est d'archiver les données spatiales de l'ASC, de maintenir leur accessibilité sur de longues périodes et de faciliter et promouvoir leur réutilisation. Il ne s'agit donc pas seulement d'entreposer les données, il faut s'assurer de préserver l'intégrité des données jusqu'à leur élimination sans quoi elles ne pourront pas être réutilisées. L'intégrité inclut à la fois l'intégrité physique des données mais aussi la conservation du contexte de saisie des données et des métadonnées. Sans le contexte de saisie (lieu, date, heure, marges d'erreur, type de capteurs, etc.) les données perdent toute leur signification et deviennent inutilisables.

Cette étape rassemble les concepts de préservation et de diffusion qui ont été identifiés comme deux éléments clés de la conservation dans la section 1.1.3 de ce mémoire.

Plusieurs méthodes de préservation des données numériques ont été présentées dans la section 1.1.3.3 de ce mémoire. La technique de préservation sera choisie en fonction de la criticité des données et des ressources disponibles. Un ensemble de données jugé critique devra être accessible en tout temps et une action de préservation devra donc être effectuée à la suite de chaque changement technologique. Pour un ensemble de données jugé non-critique, il pourrait être suffisant de simplement documenter le format d'origine de manière à être en mesure de faire l'émulation des données au moment où celles-ci devront être utilisées, évitant ainsi de nombreuses migrations inutiles.

Pour s'assurer que la préservation puisse être faite adéquatement, il faut que le créateur documente bien le format des données et le système par lequel elles ont été créées dans le plan de gestion des données parce qu'il ne sera peut-être plus là au moment de leur remise à jour. Il faut aussi que le curateur fasse un suivi des changements technologiques et qu'il ait identifié, avec l'aide des créateurs et des utilisateurs de données, la criticité de chaque ensemble de données.

En plus de maintenir les données intègres et lisibles, il est souhaitable de rendre accessibles certains types de données à un public ciblé et d'en promouvoir la réutilisation. Afin de promouvoir leur réutilisation il s'avère nécessaire, dans une certaine mesure, de « publiciser » les ensembles de données disponibles pour leur donner la plus grande visibilité possible. Cette diffusion des données de l'Agence répond à certaines de ses activités qui sont de faire bénéficier l'ensemble des Canadiens des sciences et technologies spatiales. Les principes de l'« *Open Data* » intègrent des concepts liés à cette phase du cycle de vie des données spatiales.

Certains éléments peuvent toutefois limiter la diffusion des données. Il peut s'agir d'un droit de premier regard, du respect de la vie privée, de raisons de sécurité ou d'ententes contractuelles. C'est le cas des données RADARSAT-2 dont la licence ne permet pas la libre diffusion des données.

L'utilisation d'outils comme *La passerelle vers les données scientifiques*[26] pour diriger les chercheurs vers les données de l'ASC permet de faire gratuitement et efficacement la diffusion des ensembles de données jugés les plus pertinents.

Un autre outil de diffusion est le *dataset paper*. Ce type de document décrit un ensemble de données, explique la méthode et le contexte de sa saisie et indique comment il est possible de



l'obtenir. Cet article est diffusé comme les autres publications scientifiques et est donc facilement accessible à la communauté scientifique.

#### 3.2.1.2.2.7 Élimination

Cette phase consiste à déterminer si les données doivent être conservées et si ce n'est pas le cas, d'en faire la suppression. Il est important d'avoir un processus clair et rigoureux pour faire la sélection des données devant être conservées. Dans la mesure où toutes les données ne peuvent être conservées, ce processus de sélection s'avère un élément crucial pour limiter la perte de données importantes.

La date d'élimination se base sur le calendrier de conservation qui a été créé lors de l'organisation des données à partir du plan de gestion des données du projet, lequel est conforme aux lois et règlements du Gouvernement du Canada et à la politique de données.

Avant de procéder à leur élimination, l'équipe de curation doit faire une vérification finale afin de s'assurer que les données que l'ASC s'apprête à détruire n'ont vraiment plus aucune valeur. Cette vérification est importante puisqu'il est possible que la valeur de certaines données ait grandement changée entre le moment où leur délais de conservation a été défini et le moment prévu de leur élimination. Cette vérification est commandée par le curateur mais nécessite l'implication de spécialistes capables d'établir la pertinence et la valeur des ensembles de données en question.

BAC, qui suggère la durée de préservation et approuve la disposition des actifs informationnels canadiens, a délégué aux ministères la responsabilité de déterminer la durée de préservation de leurs données. C'est donc l'ASC qui a la responsabilité de définir quelles sont les durées de conservation raisonnables pour les différents ensembles de données qu'elle possède.

#### 3.2.1.2.2.8 Évaluation

Il s'agit principalement d'une évaluation corrective du processus de gestion des données. L'objectif est d'évaluer si le processus est optimal et s'il répond adéquatement aux objectifs énoncés dans ce cadre. Des améliorations seront apportées au processus lorsqu'un problème sera décelé. En plus de l'évaluation corrective, il est important d'effectuer une évaluation cyclique du processus afin de trouver des façons de constamment le bonifier.

Les évaluations sont pilotées principalement par le curateur mais toutes les parties prenantes de la gestion des données doivent y contribuer afin que le processus soit optimal et bien appliqué.

Le cycle de vie des données spatiales de l'ASC est très opérationnel. Il respecte les concepts et le vocabulaire du cycle de vie de l'information de l'ASC, mais présente le cas d'un type précis d'actif organisationnel. Il s'inscrit donc dans la continuité des procédures de gestion de l'information à l'ASC, mais peut être directement utilisé par les créateurs, utilisateurs et curateurs de données spatiales.

De plus, le cycle de vie présente déjà une partie des responsabilités des principaux rôles de la gestion des données. Il illustre bien l'interaction entre ceux-ci tout au long de la vie des données.

### **3.2.2 Rôles et responsabilités**

Le cycle de vie des données techniques implique un grand nombre de parties prenantes dont les intérêts, les connaissances et les champs d'expertise varient considérablement. Par ailleurs, même si elles interviennent à des moments différents, elles sont appelées à interagir les unes avec les autres. Et bien que certains intervenants aient des tâches plus spécifiquement liées aux données, la gestion de celles-ci demeure la responsabilité de tous. Dans ce contexte, il devient nécessaire de prévoir des mécanismes qui facilitent l'interaction nécessaire entre ces intervenants. Les parties prenantes de la gestion des données spatiales sont présentées ici selon le rôle qu'elles occupent par rapport aux données : créateurs/utilisateurs, curateurs ou administrateurs. Ces trois rôles ont chacun une importance cruciale dans la gestion des données.

Le rôle le plus évident est celui de créateur/utilisateur puisqu'il regroupe toute personne qui se sert de données lors de ses projets. Les intervenants qui accomplissent ce rôle génèrent, rassemblent, utilisent et analysent les données afin de créer des outils, de faire des découvertes ou pour répondre à un quelconque besoin de leur organisation ou d'eux-mêmes.

À ce rôle vient s'ajouter celui du curateur. Cet intervenant pense au-delà de l'atteinte des objectifs de la mission en cours. Le curateur tente d'éviter de restreindre l'utilisation des données à un seul projet afin d'en maximiser les retombées. Les curateurs assurent le suivi de la gestion des données afin que celles-ci soient disponibles pour d'autres utilisateurs, présents et futurs et pour que leur gestion soit complète.

Le dernier rôle n'agit pas directement sur les données. Les administrateurs n'ont pas comme tâche de manipuler les données, mais plutôt d'agir sur le processus de gestion des données et sur leur cycle de vie. Ce rôle consiste en l'intégration des préoccupations de la gestion des données sur l'ensemble du cycle de vie à la planification des projets. L'administrateur œuvre à la création et l'implantation de cadres et de politiques afin de normaliser les processus de gestion des données, à faire évoluer et maintenir la pertinence de ces procédures. Il est essentiel d'avoir des intervenants qui conservent une certaine distance par rapport aux données afin d'observer, de normaliser et d'évaluer les procédures de gestion des données spatiales.

Il est important de comprendre qu'il s'agit de rôles et non d'individus. Il est possible qu'un même intervenant effectue plus d'un rôle, par exemple, qu'un chercheur qui crée et utilise des données effectue lui-même la conservation et la diffusion de ses données afin d'en permettre la réutilisation.

La présente section propose une liste détaillée des responsabilités des intervenants de l'ASC et de l'extérieur et illustre à quel moment ceux-ci interviennent dans le cycle de vie des données spatiales. Nonobstant cette division des groupes selon les trois principaux rôles, il est possible que certains individus aient des responsabilités qui aillent au-delà de celles typiquement liées au rôle qu'ils accomplissent.

### 3.2.2.1 Créateurs et utilisateurs

#### 3.2.2.1.1 Projets financés par l'ASC

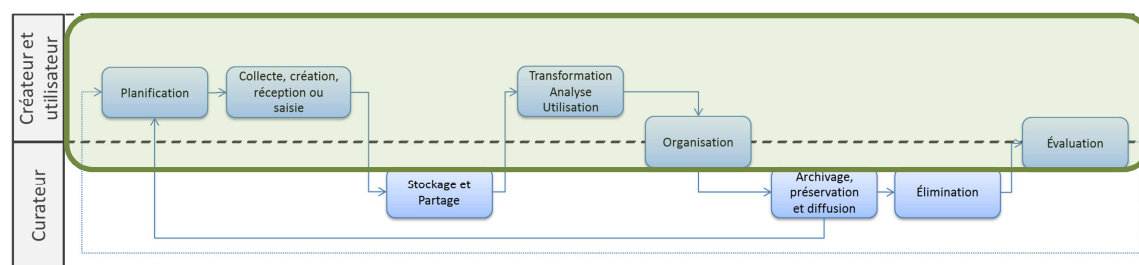


Figure 3-2 : Responsabilités des créateurs et utilisateurs dans le cycle de vie des données

Les créateurs et utilisateurs internes des données sont toute personne ou institution qui manipule les données dans le cadre d'un projet effectué ou financé par l'ASC. Il est essentiel que ces intervenants collaborent étroitement avec les curateurs afin que l'utilisation des données spatiales puisse être optimale et qu'elle dépasse les simples objectifs du projet en cours.

Le rôle de créateur est rempli par toute personne ou institution qui procède à la création ou l'acquisition de données spatiales tandis que le rôle d'utilisateur est attribué aux personnes ou institutions qui font ou gèrent la transformation et l'utilisation des données spatiales de l'ASC. Ces rôles peuvent être remplis par la même personne, mais peuvent aussi l'être par des intervenants distincts.

Ces intervenants peuvent à la fois être des employés de l'ASC ou des intervenants externes. S'ils sont de l'extérieur de l'ASC, les responsabilités des créateurs et utilisateurs des données devraient être inscrites dans l'entente ou le contrat que l'ASC a avec ces personnes ou ces institutions.

Voici donc les responsabilités des créateurs et utilisateurs des données de projets financés par l'ASC :

- remplir le gabarit du plan de gestion des données au début du projet et le mettre à jour à toutes les étapes pertinentes du projet;
  - identifier la quantité de données, le format, l'utilité présente et future, les besoins de conservation, etc.;
  - identifier clairement qui aura accès aux données et quand;
  - tenir le plan à jour pour permettre au groupe GI-TI de bien faire son travail;
- conserver et protéger l'intégrité physique de l'ensemble des données recueillies tant que la permission de les détruire n'a pas été émise ou qu'elles n'ont pas été mises sous la responsabilité d'un autre intervenant;
- filtrer les éléments confidentiels des données avant que celles-ci ne soient diffusées;
- documenter adéquatement l'environnement et le contexte dans lequel les données ont été saisies ainsi que les méthodes utilisées pour leur saisie sans quoi les données ne pourront être réutilisées adéquatement par d'autres chercheurs;
- collaborer avec le secteur de la GI-TI à l'élaboration de métadonnées fiables et utiles;
- aider le secteur de la GI-TI à déterminer quelles données sont pertinentes pour la diffusion;

- rendre les données pertinentes accessibles;
- évaluer la pertinence des données et les détruire si elles ne représentent plus d'intérêt et si la permission leur est donnée au moment où le secteur de la GI-TI indique que les données ont atteint leur fin de vie;
- rapporter toutes lacunes ou toutes possibilités d'amélioration du processus de gestion des données au secteur de la GI-TI afin de constamment l'améliorer;
- aider la GI-TI à déterminer les ressources nécessaires à la conservation et à la diffusion des données;
- participer à l'évaluation du processus de gestion des données spatiales de l'ASC.

### *3.2.2.1.2 Projets n'étant pas financés par l'ASC*

#### *3.2.2.1.2.1 Créateurs de données n'étant pas financés par l'ASC*

Ce groupe représente tout organisme ou individu extérieur à l'ASC et n'étant pas financé par celle-ci qui crée ou recueille des données qui seront par la suite utilisées par des projets de l'ASC. Ce peut être une agence étrangère, une université ou tout autre organisme de recherche. Cet intervenant n'est pas lié à l'ASC. Ces données n'ont pas été créées pour répondre aux besoins de l'ASC, leur créateur ne suit donc pas les lignes directrices du cadre de gestion des données spatiales de l'ASC.

Par contre, pour que les données puissent être utilisées par l'ASC, il est essentiel que l'organisme qui les a créées ait documenté l'environnement et le contexte dans lequel elles ont été saisies ainsi que les méthodes utilisées pour effectuer leur saisie.

#### *3.2.2.1.2.2 Utilisateurs des données de l'ASC n'étant pas financés par celle-ci*

Lors de la phase de diffusion, les gens qui utilisent des données appartenant à l'ASC doivent se plier à certaines conditions d'utilisation. Ils doivent :

- respecter la propriété intellectuelle de l'ASC et de ses chercheurs, principalement en citant la provenance des données;
- respecter la confidentialité des données jugées sensibles qui leur sont confiées;

- respecter la licence d'utilisation, s'il y a lieu.

Ces utilisateurs externes sont aussi invités à émettre des commentaires et suggestions visant l'amélioration du processus de gestion des données spatiales de l'ASC.

### 3.2.2.2 Curateurs

Étant donné sa complexité, le travail de curateur des données de l'ASC doit être partagé entre différents groupes d'intervenants. À l'interne, ces intervenants sont les secteurs de la GI et de la TI. Ceux-ci s'occupent respectivement de l'application et de l'encadrement des bonnes pratiques de conservation des données, et de l'aspect matériel et technique lié à cette conservation. Le rôle de curateur peut aussi être réalisé entièrement ou partiellement par des organismes externes.

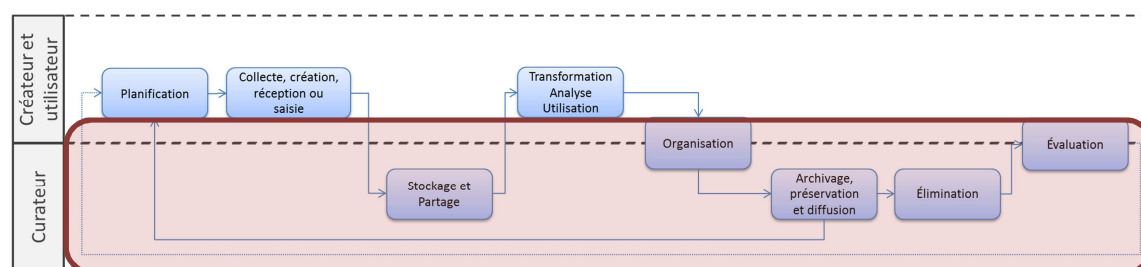


Figure 3-3 : Responsabilités des curateurs dans le cycle de vie des données

#### 3.2.2.2.1 Secteur de la Gestion de l'information (GI)

Le rôle principal de ce secteur de l'ASC est de chapeauter toutes les activités de gestion du cycle de vie des données. Le rôle du groupe de la GI est de soutenir les créateurs et utilisateurs de données dans la documentation et l'application du cycle de vie des données et de mettre en place des mécanismes pour tenir informés les différents intervenants de la gestion des données, autant à l'interne qu'à l'externe, des pratiques de l'ASC. Les différentes responsabilités de ce groupe sont les suivantes :

- évaluer avec les différentes parties prenantes et à l'aide du plan de conservation quelles données sont pertinentes pour la diffusion;
- faire le suivi du calendrier de conservation des différentes données de l'ASC et veiller à son application;
- établir, en collaboration avec les créateurs de données, la criticité des ensembles de données afin de déterminer leurs besoins de conservation;

- s'assurer de la mise en place d'un processus pour que les données soient lisibles et compréhensibles au moment où elles doivent l'être malgré les changements technologiques (préservation);
- collaborer avec les parties prenantes à l'élaboration de métadonnées fiables et utiles;
- évaluer le processus et le cadre de gestion des données de façon cyclique de manière à l'améliorer continuellement;
- appuyer l'ensemble des employés de l'ASC dans leurs activités de gestion de données et mettre à leur disposition des outils pour faciliter leur travail;
- rester informé des nouvelles directives internes, fédérales et internationales concernant la gestion des données spatiales et diffuser ces directives au sein de l'ASC;
- sensibiliser les différents intervenants à l'importance des bonnes pratiques de gestion des données entre autres en les tenant informés de ces pratiques et en participant aux réunions de lancement (*kick-off*) des projets spatiaux de l'ASC;
- fournir les outils et information requises aux intervenants externes de la gestion des données afin qu'ils connaissent les exigences de l'ASC en matière de gestion des données;
- mettre en place des mécanismes de suivi afin de s'assurer que les données financées par l'ASC sont effectivement gérées sur l'ensemble de leur cycle de vie.

#### 3.2.2.2.2 Secteur de la Technologies de l'information (TI)

La TI est le secteur de l'ASC responsable de la mise en place et de la gestion de toutes les ressources technologiques utiles à l'entreposage, à la circulation et à l'accessibilité de l'information. Son rôle est de :

- s'assurer d'avoir suffisamment d'espace de stockage pour toutes les données dont la conservation est nécessaire;
- s'assurer que les données sont facilement accessibles tout au long de leur cycle de vie à partir du site ou des services web de l'ASC peu importe si ces données sont archivées à

l'ASC ou ailleurs. S'assurer que les données et leur logiciel de traitement traversent les changements technologiques;

- s'assurer que l'accès aux données est restreint aux utilisateurs autorisés lorsque les données ne peuvent être diffusées à tous pour des raisons de sécurité ou des raisons commerciales;
- administrer le ou les portails nécessaires à la diffusion des données;
- évaluer les besoins futurs en matériel de l'ASC en ce qui concerne la conservation de données;
- participer à l'évaluation du processus de gestion des données spatiales de l'ASC.

#### *3.2.2.2.3 Organismes externes conservant des données qui appartiennent à l'ASC*

Il est possible que certaines données appartenant à l'ASC soient conservées par d'autres organismes. C'est présentement le cas, entre autres, avec le *Centre canadien des données astronomiques* (CCDA), le CCT, plusieurs universités, les stations satellitaires étrangères, et certaines agences spatiales étrangères. Ces organismes ont les responsabilités suivantes :

- respecter la propriété intellectuelle de l'ASC et de ses chercheurs;
- assurer la préservation de l'intégrité des données tout au long de leur cycle de vie en respectant les délais de préservation prescrits par l'ASC;
- respecter la confidentialité des données jugées sensibles en leur possession;
- s'assurer que les données sont lisibles et compréhensibles au moment où elles doivent l'être malgré les changements technologiques (préservation);
- faire part de toutes suggestions ou commentaires pouvant aider l'ASC à améliorer sa gestion des données spatiales.



### 3.2.2.3 Administrateurs

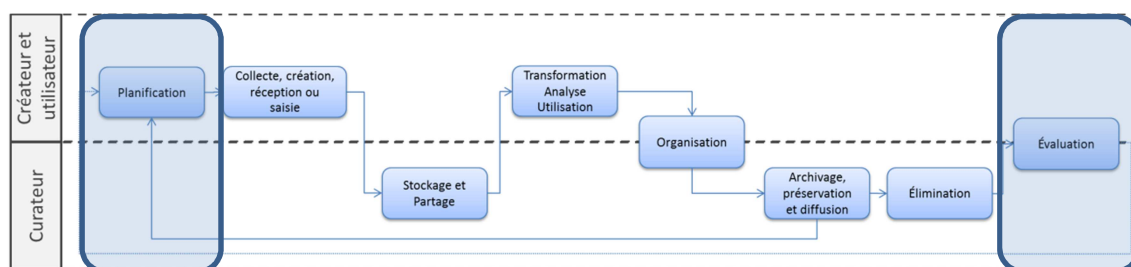


Figure 3-4 : Responsabilités des administrateurs dans le cycle de vie des données

Ces deux premiers intervenants remplissant le rôle d'administrateurs agissent principalement lors de la planification. Le troisième groupe d'intervenants, le Comité de gestion des données spatiales, s'occupe quant à lui de chapeauter la planification et l'évaluation des procédures de gestion.

#### 3.2.2.3.1 Gestionnaires de portefeuille de projets de recherche

Ce groupe est responsable de la sélection, de l'évaluation, de l'approbation et du suivi des projets de recherche de l'ASC. Il a donc comme responsabilité d'inclure à ces exigences l'ensemble des exigences de la gestion des données. Il est responsable de :

- demander aux gestionnaires de remplir le gabarit de plan de gestion des données dès la proposition de leur projet;
- considérer dans les budgets de recherche les frais de gestion des données sur l'ensemble de leur cycle de vie et non seulement pour leur création;
- faire un suivi de la gestion des données avec les chercheurs;
- participer à l'évaluation du processus de gestion des données spatiales de l'ASC.

#### 3.2.2.3.2 Soutien juridique et contractuel

Ce groupe occupe une position normative importante dans la gestion des données puisqu'il vérifie la légalité des ententes de collaboration et rédige les contrats de partenariat des missions de l'ASC. Il regroupe l'ensemble des intervenants responsables des contrats, de la propriété intellectuelle et des services légaux de l'ASC. Voici ses responsabilités :

- indiquer dans les contrats qui est responsable de conserver les données, qui peut avoir accès aux données et à qui elles appartiennent à chacune des phases du projet;
- déterminer à qui revient la propriété de chaque ensemble de données;
- participer à l'évaluation du processus de gestion des données spatiales de l'ASC;
- indiquer au secteur de la GI les contrats qui contiennent des clauses de propriété intellectuelle et qui doivent être conservés plus longtemps que la durée généralement prescrite de 6 ans;
- s'assurer de la conformité légale des ententes et contrats.

#### 3.2.2.3.3 *Comité de gestion des données spatiales (Space Data Management Committee)*

Ce groupe est celui qui aura pour mission de mettre en œuvre le cadre de gestion des données. Ce comité succédera au GTSGD actuellement responsable de réaliser le cadre. Il devra coordonner la mise en application des pratiques présentées dans le cadre, mais aussi faire l'évaluation et l'amélioration continue du processus de gestion des données spatiales. Il devra aussi :

- fournir des commentaires sur les politiques de données spatiales de l'ASC afin que les politiques et le cadre soient en phase les uns avec les autres;
- coordonner l'application du cadre de gestion des données spatiales à l'ASC;
- évaluer au fur et à mesure que la gestion des données progresse à l'ASC, les nouvelles problématiques et enjeux importants et maintenir le cadre de gestion des données à jour;
- informer la haute direction de l'ASC des enjeux liés à la gestion des données spatiales;
- travailler à l'intégration de l'*Open Data* à la gestion des données spatiales de l'ASC;
- développer des outils de planification et de suivi pour aider la gestion des données spatiales au sein des projets et missions de l'ASC.

Cette liste présente donc l'ensemble des responsabilités des intervenants de la gestion des données spatiales de l'ASC. Au-delà de l'énumération de responsabilité, il est intéressant d'y voir quels intervenants remplissent chacun des trois principaux rôles.

La complexité des rôles et le constant échange de responsabilité qu'implique le cycle de vie démontrent bien l'importance qu'il y ait une bonne communication et une bonne coordination entre les intervenants. L'ajout du rôle majeur d'administrateur vient justement renforcer cette coordination des efforts en venant chapeauter et encadrer le processus de gestion des données spatiales.

### **3.2.3 Gabarit de plan de gestion**

Un autre outil ayant été développé au cours de cette recherche est un gabarit de plan de gestion des données spatiales. L'objectif d'un tel document est de faciliter le travail de planification des créateurs de données. Le gabarit de plan de gestion identifie tous les éléments qui devront être inscrits par le créateur dans le plan de gestion des données du projet afin de permettre une bonne gestion des données tout au long de leur cycle de vie. Le plan de gestion des données est un document clé pour permettre au curateur de réaliser son travail.

Le gabarit de plan de gestion des données spatiales qui a été construit dans le cadre de cette recherche (Annexe 3) a été réalisé principalement à partir du plan d'entrevues, mais a aussi été inspiré par un gabarit de plan de gestion utilisé par certains chercheurs travaillant sur les relations Soleil-Terre à l'ASC et celui du DCC[38]. C'est en joignant les questionnements pertinents amenés par ces gabarits, respectivement trop et trop peu précis, à la structure et au contexte du questionnaire des entrevues qu'a été créé le gabarit de plan de gestion des données de l'ASC.

Ce gabarit contient cinq sections : Information générale, Aspects légal et éthique, Collecte et utilisation, Conservation et diffusion et Ressources. Sous ces cinq thèmes sont regroupées l'ensemble des questions qu'un gestionnaire doit se poser avant d'entreprendre un projet générant des données spatiales. Ces cinq sections couvrent l'ensemble des huit étapes du cycle de vie des données spatiales. Cette division a été choisie afin de correspondre davantage aux préoccupations des chercheurs et gestionnaires de mission.

Il est envisageable que ce plan ne puisse être rempli entièrement lors de la planification du projet. Il est possible, par exemple, de ne pas connaître à l'avance la durée de vie d'un satellite et donc de ne pas savoir la quantité totale de données qu'il pourra générer. Il est tout de même nécessaire que le créateur de données se pose toutes ces questions de manière à au moins connaître le niveau d'incertitude et les risques qui en découlent.

Le plan de gestion présenté en annexe a été construit pour les besoins spécifiques de l'ASC et a été proposé au GTSGD le 1<sup>er</sup> avril 2011.

### **3.3 Mise en application**

Le but de cette section est d'évaluer comment l'ASC pourrait arriver à implanter adéquatement des procédures de gestion des données respectant les principes de gestion selon le cycle de vie des données spatiales.

Cette section contient une liste de neuf recommandations du chercheur permettant de répondre aux défis particuliers identifiés à la section 3.1.2. Suit ensuite une analyse de trois modèles possibles d'implantation de nouvelles procédures de gestion des données spatiales avec les avantages et inconvénients de chacun d'eux.

#### **3.3.1 Recommandations**

Suite à l'évaluation des différents modèles possibles d'implantation des pratiques de conservation des données spatiales, une série de recommandations fut remise à l'ASC afin d'orienter l'application des principes énoncés dans le présent chapitre du mémoire et dans le rapport qui a été remis au GTSGD. Les neuf recommandations qui suivent émanent du point de vue du chercheur et concernent la mise en place de mesures de gestion des données spatiales selon leur cycle de vie.

##### **3.3.1.1 Prise en charge des données**

Peu importe le modèle qui sera choisi et appliqué, il importe que l'ASC soit en mesure de prendre en charge les données à tout moment, possiblement par l'intermédiaire d'organismes extérieurs, lorsque ses partenaires ne voudront plus ou ne seront plus en mesure d'en faire la conservation. Les partenaires extérieurs de l'ASC peuvent avoir un intérêt pour les données qui décroît dans le temps. Tant qu'ils sont intéressés par les données ils vont les conserver, mais quand leur intérêt pour ces données diminuera, leurs efforts de conservation risquent de diminuer tout autant. L'ASC, elle, conserve un intérêt pour la conservation des données, elle doit donc être en mesure de rapatrier les données en cas de besoin.

### **3.3.1.2 Stabilité du catalogue des données**

Malgré un modèle de conservation mixte et dynamique, l'ASC devrait tenter de maintenir un catalogue stable et complet de ses données. Il est important que, peu importe où sont conservées les données, celles-ci soient toutes accessibles à un seul et même endroit ou, du moins, qu'il y ait un catalogue central des données de l'ASC. Il est nécessaire que, même si les données changent de responsable ou de lieu physique l'entreposage, elles soient toujours accessibles à la même place (même URL). La stabilité du catalogue permettra alors de citer plus facilement les ensembles de données de l'ASC dans des publications et de créer des applications pouvant aller consulter les données automatiquement tandis que l'existence d'un catalogue complet des données de l'ASC faciliterait leur recherche et permettrait de démontrer la contribution de l'ASC à la recherche scientifique.

### **3.3.1.3 Évaluation cyclique des données**

L'ASC devrait aussi procéder à une évaluation cyclique des données archivées. Cette évaluation devrait être faite à la fois par des membres des secteurs de la GI, de la TI et des spécialistes des différents domaines afin de déterminer la pertinence et la criticité des données plusieurs fois durant leur conservation. Différents outils et mécanismes restent à être construits afin de pouvoir évaluer, durant la phase d'archivage, si les données doivent être conservées ou éliminées, quelle méthode de conservation doit être utilisée (criticité) et quelle est la valeur de ces données. Ces outils peuvent être des questionnaires, des grilles d'évaluation, mais pourraient aussi être des outils informatisés qui comptent le nombre de fois par mois qu'un ensemble de données est consulté.

### **3.3.1.4 Contrôle de la qualité du processus**

Il serait nécessaire de mettre en place des procédures de contrôle de la qualité afin de s'assurer que la gestion des données spatiales se fasse adéquatement autant à l'intérieur de l'ASC que dans les organismes partenaires. Ce contrôle revêt une grande importance pour la mise en application des principes énoncés dans ce mémoire. Il est d'autant plus important de mettre en place une procédure rigoureuse de suivi si l'ASC favorise un modèle de conservation décentralisé.

### 3.3.1.5 Institutionnalisation de la culture de partage

Cette recommandation touche autant l'ASC que l'ensemble de la communauté scientifique. Afin de favoriser le progrès scientifique et les bénéfices qui en découlent pour les citoyens, il serait bénéfique que les intervenants de l'ASC et la communauté scientifique adhèrent à une culture de partage et d'ouverture. L'utilisation accrue d'outils de diffusion comme les *datasets papers* permettrait de faire un usage plus complet des ensembles de données et accélérerait le progrès scientifique.

Pour ce faire, il faudrait mettre en place des mesures pour empêcher que les chercheurs ne se fassent « voler » leur chance de publier leurs résultats comme, par exemple, en leur donnant plus de temps et de ressources avant la collecte des données afin de permettre aux chercheurs de se préparer davantage et de réduire le temps de traitement menant à la publication. Il est aussi possible de réserver temporairement les données, comme c'est présentement le cas dans certains secteurs de l'ASC, afin de retarder la diffusion de certaines données ou d'instaurer des règles de fonctionnement (*rules of the road*) régissant l'utilisation des données d'un autre chercheur et obligeant le chercheur à citer le créateur initial de l'ensemble de données.

Idéalement, il faudrait aussi repenser le mode d'évaluation des chercheurs afin d'éliminer ou de réduire l'aspect de compétition et, ainsi, favoriser la collaboration et le partage des connaissances. En redonnant aux chercheurs le mérite de leur travail, on les incite à diffuser davantage leurs données. La grande difficulté ici est que cet aspect ne s'applique pas uniquement à l'ASC mais à l'ensemble de la communauté scientifique.

### 3.3.1.6 Réalisation de plans de gestion des données spatiales

Au sein de l'ASC, il faudrait encourager les gestionnaires de projet à réaliser un plan de gestion des données spatiales à l'aide du gabarit qui leur est proposé et qui est décrit à la section 3.2.3. Ce plan est effectué par le gestionnaire du projet, avec la collaboration d'intervenants de la GI-TI, de la propriété intellectuelle ainsi que d'autres secteurs de l'ASC. Il faudrait par contre que les exigences relatives à ce plan de gestion soient réalistes. Si on demande trop d'information trop tôt dans le projet, il s'avèrera difficile d'obtenir un plan de gestion des données qui représentent la réalité du projet. Un plan préliminaire de gestion des données, devant être détaillé successivement lors des différentes phases du projet, peut être envisageable.

### **3.3.1.7 Méthodes de financement**

Lorsque l'ASC finance des projets extérieurs, elle procède par contrats, par subventions ou par contributions. Lorsqu'il y a financement par contrat, elle pourrait identifier les données comme étant un livrable des projets à même le contrat. Lors de subventions ou de contributions, l'ASC pourrait ajouter la gestion des données sur l'ensemble de leur cycle de vie et leur diffusion aux critères de sélection des projets à financer, communiquer clairement les exigences de ce cadre aux organismes de recherche et prioriser les projets qui présentent un plan de gestion des données conforme aux exigences de l'ASC.

### **3.3.1.8 Uniformisation du vocabulaire**

Une autre suggestion serait d'uniformiser les définitions de niveaux de données au sein de l'ASC. Présentement, plusieurs normes différentes (celles de la NASA, de l'ESA, etc.) sont utilisées par les gestionnaires de l'Agence pour décrire leurs données. Pour faciliter l'organisation, la documentation et la conservation des données, il serait important d'avoir des descriptions claires des différents niveaux de transformation des ensembles de données. Le but n'est pas de contraindre mais plutôt d'encadrer le choix des normes utilisées afin de favoriser la communication entre les secteurs.

### **3.3.1.9 Implantation graduelle des nouvelles pratiques**

La mise en application des principes de gestion présentés dans ce mémoire est un projet d'une grande ampleur. Pour s'assurer de sa réussite il est suggéré de procéder de façon graduelle, une étape à la fois, et de commencer ponctuellement par des projets simples pour se déplacer ensuite vers des projets et des secteurs de plus en plus complexes.

## **3.3.2 Mise en œuvre**

Certains défis se posent lorsque vient le temps de mettre en application les concepts et outils présentés dans ce chapitre. Il est essentiel d'avoir une vision éclairée des différents modes d'application possibles ainsi que les impacts de chacun d'entre eux.

Voici trois modèles d'implantation applicables pour faire la conservation des données. Ceux-ci correspondent parfois au mode de gestion des données implanté par certains secteurs particuliers de l'ASC. Ces modèles identifient les avantages et les inconvénients de confier l'archivage et la

préservation des données à différents organismes. Ces modèles ne peuvent pas nécessairement tous être appliqués à l'ensemble des données spatiales de l'ASC. Il est aussi très peu probable qu'un modèle unique puisse servir à la gestion de toutes les données de l'Agence. Il est par contre intéressant de comparer les caractéristiques de ces modèles pour en déterminer les domaines où leur application est la plus intéressante.

L'analyse comparative qui a été faite se base sur des critères de gestion. Les pratiques désirables ainsi que les rôles de chacun des intervenants ont été établis précédemment, l'analyse qui suit représente la facilité d'application de ces bonnes pratiques de gestion selon trois modèles d'application différents. Trois paramètres de gestion sont étudiés pour chacun de ces modèles :

- Tout d'abord il y a le contrôle que l'ASC a sur ses données. Ce contrôle représente la facilité avec laquelle l'ASC peut gérer, accéder, partager, modifier ou éliminer les données qu'elle possède. Plus les données sont gérées par des intervenants externes, et plus ceux-ci sont distants de l'ASC, moins celle-ci a de contrôle direct sur ses données. Un manque de contrôle peut compliquer des éléments comme la gestion de l'accès aux données ou l'application de changements dans la durée de conservation. Un manque de contrôle peut aussi engendrer un non-respect des règles de conservation et de diffusion de l'ASC.
- Ensuite, il y a la sécurité des données. Lors de la conservation, il y a toujours des risques de perte de données. L'intégrité de celles-ci peut être compromise si la gestion n'est pas bien faite. Pour que l'archivage soit fait de façon sécuritaire, il faut que l'organisme qui l'effectue comprenne bien tous les risques, reconnaisse l'importance de protéger et maintenir les données et mette en place des mesures pour protéger les données et assurer leur support. La sécurité est donc un indicateur du risque de mauvaise gestion ou de perte des données.
- Finalement, il y a l'aspect économique qui reflète les coûts d'implantation et de maintien que devra déboursier l'ASC. La conservation des données coûte cher en main-d'œuvre et en infrastructure. L'ASC devra défrayer ces coûts ou devra financer l'organisme qui assumera ces coûts. Certaines alternatives devraient s'avérer plus économiques que d'autres surtout si les organismes qui font l'archivage ont déjà l'expertise et les infrastructures en place. Plus le modèle est économique plus il est attrayant.



Les paramètres qui ont été choisis visent à évaluer les avantages et inconvénients, sur le plan de la gestion, de l'atteinte des objectifs d'accessibilité et de préservation dans différents contextes d'application. C'est l'atteinte de ces objectifs, communs à chacun des contextes, qui engendrent des variations de coûts. La sécurité et le contrôle représentent quant à eux les risques de ne pas pouvoir atteindre les objectifs d'accessibilité et de préservation dans un contexte donné.

Des graphiques ont été construits pour démontrer les différents modèles. Ils illustrent de façon qualitative les forces et les faiblesses des différents modèles selon les trois paramètres choisis. Chacun des trois axes représente un paramètre. Pour chacun des axes, plus le modèle s'éloigne du centre du graphique, plus il est avantageux.

Les valeurs de ces trois paramètres ne sont pas nécessairement égales entre elles, il est peut-être plus acceptable pour l'ASC de faire, dans une certaine mesure, des compromis sur le contrôle que sur la sécurité des données.

### **3.3.2.1 Archivage à l'externe**

Dans ce premier modèle, les données sont préservées dans des universités et organisations externes comme elles le sont présentement pour plusieurs projets de l'ASC. Par contre dans ce modèle, l'Agence, par l'entremise du cadre de gestion des données spatiales, de ses politiques et des accords de collaboration et contrats, demande, lorsqu'il est possible de le faire, de conserver et de diffuser les données dont elle a financé l'acquisition. Ce premier modèle exclut les organismes qui font partie du Gouvernement du Canada.

#### *3.3.2.1.1 Avantages*

- simple et peu coûteux à mettre en œuvre;
- déjà appliqué, entre autres avec la NASA dans les missions planétaires.

#### *3.3.2.1.2 Inconvénients*

- les données sont décentralisées, ce qui demande beaucoup d'effort pour mettre en place des pratiques normalisées et acceptables, ainsi que pour faire le suivi de tous les organismes manipulant les données;

- la sécurité des données et la qualité des activités de conservation et de diffusion varient en fonction de l'organisme qui gère les données. L'ASC peut émettre des recommandations, mais la vérification de l'application de ces exigences se complique plus il y a d'organismes qui archivent des données pour le compte de l'ASC;
- la communication avec les nombreux intervenants peut s'avérer difficile, ce qui peut nuire au contrôle qu'exerce l'ASC sur ses données;
- il y a un risque de diffusion de données sensibles;
- il faut financer une multitude d'organismes.

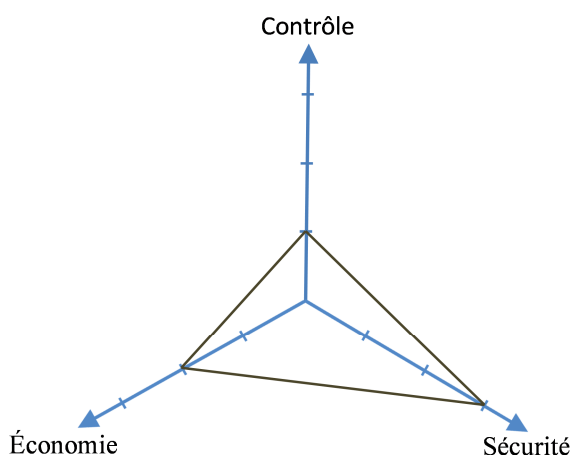


Figure 3-5 : Analyse du modèle d'archive des données dans une agence comme la NASA ou l'ESA

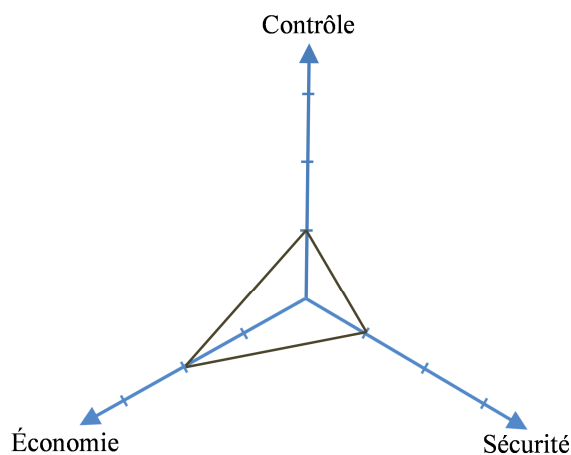


Figure 3-6 : Analyse du modèle d'archive des données dans une université

L'archivage à l'externe est déjà très répandu à l'Agence, principalement pour les données scientifiques. Ce modèle est souvent employé lorsque l'ASC collabore avec une agence étrangère ayant des exigences plus strictes concernant la gestion des données. C'est le cas des missions planétaires comme MATMOS qui est réalisée en collaboration avec la NASA et d'autres agences étrangères. Les données spatiales de cette mission seront archivées sur un portail international comme le PDS[51] qui respecte les normes de gestion des données spatiales de la NASA.

Ce modèle est aussi appliqué lors de collaborations avec des universités canadiennes et étrangères. Par contre, comme l'ASC n'avait pas, jusqu'à ce jour, de processus corporatifs

concernant la gestion des données spatiales, leur conservation et leur diffusion se font présentement en suivant les procédures des universités qui les archivent. Si ce modèle est conservé, il faudra beaucoup d'efforts pour informer les universités et autres organismes qui archivent présentement des données de l'ASC de ses attentes et pour faire un suivi de leur application.

### 3.3.2.2 Archivage dans un organisme du Gouvernement du Canada

Ce deuxième modèle propose lui aussi un archivage des données à l'extérieur de l'ASC mais cette fois-ci les organisations à qui sont confiées les données sont des organismes faisant partie du Gouvernement du Canada.

#### 3.3.2.2.1 Avantages

- l'archivage des données par un organisme gouvernemental canadien donne une certaine assurance quant à la sécurité des données étant donné les normes de gestion strictes du gouvernement;
- la proximité de ces organismes par rapport à l'ASC assure une bonne communication, ce qui renforce le contrôle;
- les frais initiaux pour la mise en place d'un nouveau centre d'archivage sont généralement moins élevés que pour une nouvelle installation;
- on profite du savoir-faire et de l'expérience d'organismes qui œuvrent depuis longtemps à la conservation de données spatiales.

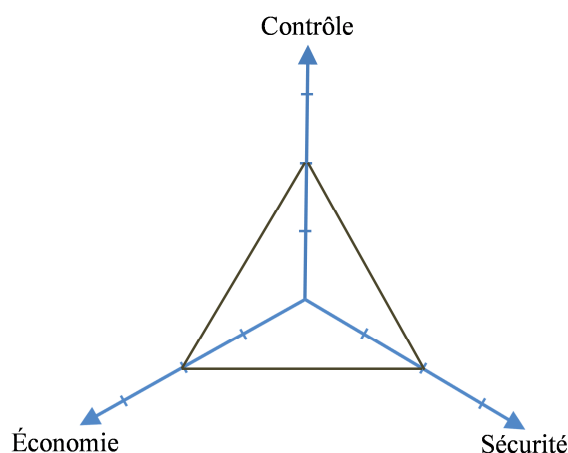


Figure 3-7 : Analyse du modèle d'archive des données dans un organisme du gouvernement canadien

#### 3.3.2.2.2 Inconvénients

- bien que le contrôle soit renforcé, il n'est pas complet. L'accès aux données doit passer par un intervenant extérieur;
- les données sont réparties dans plusieurs organismes;
- il faut déboursier des fonds aux organismes de conservation.

Certaines données scientifiques sont archivées dans des organismes fédéraux canadiens. C'est le cas des données astrophysiques qui sont archivées au CCDA ou du CCT qui archive les données brutes de RADARSAT 1 et 2 ainsi que les produits de RADARSAT-2. De telles collaborations avec des organismes canadiens permettent à l'ASC de profiter de l'expertise d'organisations qui œuvrent dans la gestion des données depuis longtemps.

Ce modèle réduit les coûts et les efforts considérables liés à la mise en place de nouvelles infrastructures et au développement d'une nouvelle expertise puisque ces deux éléments sont déjà en place dans les organismes fédéraux canadiens. Par contre, une fois les frais initiaux absorbés, les coûts de gestion sont semblables. Avec ce modèle, ces coûts de gestion sont simplement versés à des organisations extérieures.

### **3.3.2.3 Archivage à l'ASC**

Ce modèle requiert la création d'un centre d'archivage des données spatiales au sein même de l'ASC. Ce mode de gestion assurerait à l'Agence spatiale d'avoir un contrôle et une sécurité totale sur ses données pourvu qu'elle investisse les sommes nécessaires dans ce centre d'archivage. La mise en place d'une telle infrastructure impliquerait par contre des coûts initiaux élevés.

### 3.3.2.3.1 Avantages

- toutes les données sont gérées de façon uniforme, à un seul endroit;
- l'ASC garde un contrôle total sur l'ensemble de l'actif informationnel que représentent les données;
- l'ASC peut s'assurer elle-même de la sécurité des données;
- l'ASC obtient une plus grande visibilité. Il est facile de voir l'impact global de l'ASC et toute la valeur ajoutée qu'apporte cette organisation;
- la gestion des applications dérivées est simplifiée.

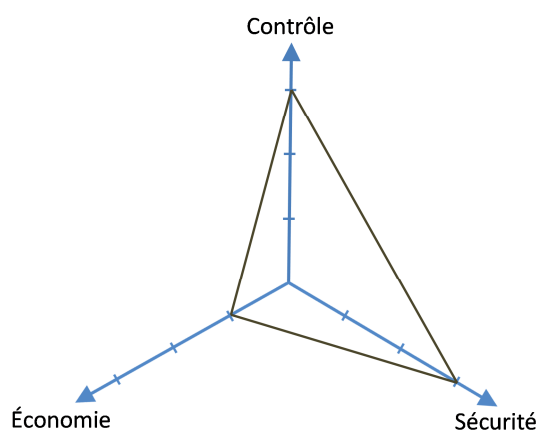


Figure 3-8 : Analyse du modèle d'archive des données à l'ASC

### 3.3.2.3.2 Inconvénients

- nécessite des investissements initiaux majeurs pour les infrastructures;
- nécessite l'embauche et la formation de spécialistes pour développer et administrer ce centre.

Le choix du modèle qui sera mis en place à l'ASC dépend en grande partie des ressources disponibles. Ce choix revient à l'ASC et, à ce jour, la décision n'a pas encore été prise. Il est par contre très probable que le modèle choisi soit un modèle mixte qui incorpore plusieurs des options présentées précédemment. Ce modèle risque aussi d'être dynamique c'est-à-dire que la responsabilité des données peut passer d'un organisme à un autre au cours de leur cycle de vie, comme c'est le cas présentement avec les données des stations étrangères de RADARSAT. Il faudra donc que l'ASC se dote de procédures lui permettant de gérer adéquatement ces transferts de responsabilités.

### 3.4 Validation

Tel que présenté dans la Figure 2-1 du Chapitre 2, plusieurs validations ont été réalisées au cours du processus de recherche. Il y a tout d'abord eu une validation du cycle de vie des données spatiales puis du plan d'entrevues par l'équipe de travail sur les données spatiales de l'ASC.

La plus importante validation du travail effectué est l'approbation du *Rapport sur la gestion des données spatiales* par le GTSGD et certains spécialistes des données d'observation de la Terre et d'expériences. Ce rapport, dont il est mention à la section 2.4.6, regroupe l'ensemble des outils et analyses de ce chapitre et a pour but d'exposer les résultats de la recherche aux membres de l'ASC. La dernière version de ce rapport (version 1.5) a été remise au GTSGD le 17 juin 2011.

### 3.5 Conclusion

La situation de l'ASC est particulière en constatant le type de données qu'elle produit, manipule ou gère, le caractère public de sa mission et le type de collaboration qu'elle entretient avec ses partenaires. Cette situation et ces enjeux particuliers nécessitent le développement d'outils et de méthode d'analyse qui répondent expressément à cette réalité.

C'est dans cette optique que le cycle de vie adapte les concepts théoriques vus au Chapitre 1 à la réalité opérationnelle de l'ASC. À ce cycle de vie s'ajoute une description des rôles et responsabilités des intervenants de la gestion des données spatiales, un gabarit de plan de gestion, une analyse des modèles d'implantation possibles et une liste de recommandations afin de réussir cette implantation.

## **CHAPITRE 4 DISCUSSION**

Ce chapitre propose une discussion des principaux résultats présentés au chapitre précédent. Bien que les outils et analyses développés dans le cadre de ce projet de recherche aient été réalisés dans un contexte spécifique, celui de l'ASC, ces résultats revêtent un intérêt à la fois pour cette agence gouvernementale, mais aussi pour d'autres organisations, et pour la communauté scientifique canadienne.

Concrètement, ce chapitre récapitule les résultats du Chapitre 3 en expliquant leur impact et leur importance; il propose une liste de contributions théoriques et pratiques de la recherche, et se termine par une liste de limitations et de possibles avenues de recherche.

### **4.1 Contributions théoriques**

La contribution théorique de cette recherche consiste en la mise en commun des concepts théoriques de la gestion des données techniques et en l'analyse de leur application dans un cas particulier. Cette identification des concepts passe principalement par les quatre éléments clés de la conservation tirés des écrits de Yakel[13]. Chacun de ces éléments a été décrit de façon conceptuelle dans la section 1.1.3 mais leur portée et leur importance ont été approfondies tout au long de ce mémoire. Dans les pages qui suivent, la contribution de cette recherche sera expliquée tour à tour, pour chacun de ces quatre éléments clés.

#### **4.1.1 Gestion selon le cycle de vie**

Pour commencer, la recherche a fait l'étude du concept de gestion selon le cycle de vie des données. Il y a tout d'abord eu la recherche, l'analyse et la comparaison entre eux de différents cycles de vie liés aux données techniques puis la création d'un nouveau cycle de vie pour le contexte particulier des données spatiales de l'ASC. Ce cycle de vie très opérationnel permet d'illustrer des éléments importants comme la réutilisation des données et leurs multiples transformations.

Ce cycle est opérationnel puisqu'il reprend les concepts et le vocabulaire déjà utilisés par l'ASC dans sa gestion de l'information, mais l'adapte à la gestion plus particulière des données spatiales. La présence de boucles vient illustrer les nombreuses transformations et réutilisations des données. De plus, ce cycle innove en illustrant les deux principaux rôles de la manipulation

des données directement dans le cycle. Cette division du cycle de vie en corridor expose de façon évidente l'échange de responsabilités qui se fait entre ces deux rôles et démontre l'importance d'une bonne collaboration entre ceux-ci.

Cette recherche présente un cycle de vie plus spécifique et plus facilement applicable que ceux proposés dans la littérature[19, 24, 36]. Ce cycle présente à la fois la répartition séquentielle des activités, mais aussi le partage de ces tâches, ce qui n'est accompli par aucun des cycles proposés.

#### **4.1.2 Intégration de la conservation dans le processus de recherche**

Cette recherche pousse aussi plus loin le concept de l'intégration du rôle de curateur au processus de recherche traditionnel. Tout d'abord, elle met en application le 3<sup>e</sup> niveau de conservation présenté par Lord[3] et décrit dans la section 1.1.3.2 de ce mémoire. Les corridors du cycle de vie constituent une application directe de ce modèle théorique. Les rôles proposés par Lord sont même attribués aux intervenants réels du terrain au lieu de s'en tenir simplement au niveau conceptuel.

En plus du curateur et du créateur/utilisateur, cette recherche ajoute un troisième rôle fondamental à la conservation, celui d'administrateur. Ce rôle s'avère essentiel étant donné la longue durée et la complexité du cycle de vie des données. En effet, la responsabilité des données s'échange tout au long du cycle de vie entre les différents intervenants. De plus, comme la vie utile des données est souvent plus longue que la carrière de ces intervenants, plusieurs individus sont donc appelés à se relayer pour une même tâche. Il est donc important que le processus soit bien documenté, organisé et coordonné afin que ces transferts de responsabilités s'effectuent adéquatement.

Le rôle d'administrateur consiste en l'élaboration, au maintien et à l'évaluation de procédures de gestion des données. Il ne manipule pas lui-même les données mais assure la cohésion et la continuité du cycle. C'est l'administrateur qui assure la présence de procédures claires facilitant ces nombreux échanges de responsabilités. Ce troisième rôle majeur de la conservation est responsable de la création, de la mise en application et de la constante amélioration de ces procédures.

Le modèle proposé dans ce mémoire constitue donc un prolongement du 3<sup>e</sup> niveau de conservation de Lord[3]. Ce nouveau modèle considère l'institutionnalisation du processus de



conservation et la nécessité d'une présence normative pour encadrer ces procédures. Ce 4<sup>e</sup> niveau de conservation se caractérise par l'ajout d'un méta-regard sur les tâches de conservation et d'une évaluation et amélioration constante du processus.

### **4.1.3 Préservation**

La préservation des données constitue l'élément le plus traité dans la littérature parmi les quatre. De par sa nature plus technique, ce volet de la gestion des données a été longuement étudié par les gestionnaires de données et les informaticiens[14, 15]. En plus de regrouper et décrire les principales méthodes de préservation présentées dans la littérature, cette recherche réaffirme son importance et souligne son caractère très distinct de l'entreposage. Cette distinction est reprise tout au long du mémoire et la nécessité de déployer des ressources et d'identifier les besoins de préservation y est bien expliquée.

La présente recherche introduit également le concept de criticité des ensembles de données. En étant opposé à la quantité de ressources disponibles pour faire la préservation, la criticité permet de choisir la méthode de préservation la plus appropriée à chaque ensemble de données. Cet élément n'est pas identifié explicitement dans la littérature étudiée malgré qu'il soit un paramètre majeur dans le choix des méthodes de conservation.

### **4.1.4 Accessibilité**

Un quatrième aspect intéressant de la présente recherche est qu'elle confronte des initiatives réelles au concept théorique d'accessibilité. En plus d'analyser les bénéfices et les enjeux de la diffusion des données, de nombreuses initiatives canadiennes et internationales liées à l'accessibilité des données, et décrites dans ce mémoire[9, 26, 27, 41, 49, 50], ont été intégrées dans la conception des outils et méthodes. Il y a donc une analyse des enjeux liés à l'accessibilité mais aussi une observation d'un mouvement politique favorisant le libre accès aux données.

Cette recherche introduit aussi de nouvelles contraintes pouvant affecter l'accessibilité des données. Différents enjeux présentés dans ce mémoire comme la sécurité ou la propriété intellectuelle sont traités dans la littérature, mais d'autres y sont absents. C'est le cas du potentiel commercial de certaines données. Des intérêts commerciaux peuvent être intimement liés à la production de données ce qui en limite la diffusion. Cet aspect commercial est important puisqu'il est en opposition avec les initiatives de données libres.

La contrainte commerciale peut s'appliquer aux projets de recherche universitaires effectués en partenariat avec le secteur privé ou à toute situation où le gouvernement subventionne la recherche et développement (R&D) d'une entreprise. Dans ce type de situations, il est impossible d'appliquer un concept comme l'« Open Data » sur les données du projet sans nuire au partenaire commercial. La diffusion de ses données s'effectue donc uniquement dans le cadre des activités commerciales du projet.

En résumé, cette recherche aura permis de pousser un peu plus loin quatre éléments clés de la conservation des données. L'étude d'un cas réel fournit un exemple concret d'application des concepts de la gestion de données qui pourra être analysé et comparé avec d'autres modèles par d'autres chercheurs. Cet exemple vient s'ajouter aux modèles d'applications étudiés lors de cette recherche. C'est en multipliant ces exemples d'application qu'il sera possible de mieux documenter et de tirer des généralisations dans le nouveau champ de recherche que représente la conservation des données.

## **4.2 Contributions pratiques**

Le fait d'avoir pu travailler sur une problématique de recherche réelle et concrète, avec les intervenants du terrain, ajoute au mémoire une grande valeur pratique. La documentation de l'ASC et des autres organisations a permis d'adapter les modèles génériques et conceptuels présentés dans la littérature scientifique. Les discussions avec le GTSGD et les entrevues ont ensuite permis de confronter ces modèles entre eux et d'identifier les bénéfices de la mise en application de chacun d'entre eux. Les analyses et outils qui en découlent répondent donc aux réalités du terrain et pourront servir de base à la mise en place de processus de gestion des données spatiales.

Cette recherche répond à plusieurs interrogations des communautés de chercheurs et de praticiens œuvrant dans la gestion des données techniques canadiennes. Elle soulève des défis vécus par ces différents milieux. Ce mémoire fournit donc des pistes afin d'améliorer la gestion des données techniques dans une multitude d'organisations.

### 4.2.1 Comparaison des différents modèles

Une première contribution pratique découle de la transformation des modèles de gestion des données existants afin de les rendre applicables dans une situation précise. Le recensement et la comparaison de ces différents modèles et des publications de différents auteurs permettent d'identifier les qualités mais aussi les limitations des différents outils et des différents cas présentés dans la littérature.

Certains modèles conceptuels très répandus, comme celui du DCC, demeurent limités en termes d'applicabilité à l'ASC et ce, malgré le fait que de nombreuses publications qui y sont consacrées [3, 12, 67] lui confèrent un caractère universel. En effet, le DCC couvre l'ensemble des données de recherche du Royaume-Uni. Quand vient le temps d'utiliser ses outils dans un contexte plus restreint comme celui des données d'un seul organisme public, l'application est peu efficace. Certains outils traitent en long et en large des éléments sans intérêt pour l'ASC pendant que d'autres aspects cruciaux y sont presque absents. Par exemple, le gabarit de plan de gestion du DCC[38] s'attarde longuement à la provenance du financement et aux exigences de l'organisme de financement par rapport aux données. Cet élément est critique et varie beaucoup dans le milieu scientifique britannique; en revanche, il est plutôt stable à l'ASC et n'a pas besoin d'être traité au cas par cas puisque ce sont toujours les mêmes situations qui reviennent. En contrepartie, l'ASC doit mettre plus d'énergie dans l'établissement de la portée des différentes ententes de partenariats.

Des outils comme le PDS[51] de la NASA sont quant à eux directement applicables au contexte de l'ASC. Cet outil est justement utilisé par des chercheurs de l'ASC et le sera encore probablement pour des décennies. La particularité de ce type de modèle est qu'il s'oriente non pas sur une organisation mais plutôt sur un domaine scientifique particulier, dans ce cas, les sciences planétaires. Par contre, le défi pour l'ASC est de développer des procédures globales et des lignes directrices qui englobent l'ensemble des domaines techniques et scientifiques. Ce mémoire énonce justement des principes généraux pour la gestion de l'ensemble des données de l'ASC. C'est à partir de ces principes et des exigences qui en découleront que l'ASC pourra décider quels outils pourront être utilisés pour la conservation et la diffusion de ses données.

### 4.2.2 Valeur pratique pour l'organisation

Ce projet de recherche, résolument axé sur l'intervention en milieu de pratique[62], procure à l'organisation d'accueil (l'ASC) une valeur pratique sur plusieurs plans. D'une part, elle a permis d'explicitier les enjeux et défis entourant la gestion des données techniques dans le contexte particulier de l'ASC, de mettre en lumière les différentes valeurs des données et d'identifier les différents scénarios d'implantation possibles. D'autre part, les outils présentés dans ce mémoire proposent des bases concrètes pour répondre à ces défis qui peuvent être utilisés par les acteurs du milieu et aident l'ASC à répondre à la *Directive sur la tenue des documents*[25]. Cette recherche fournit des pistes de solution permettant à l'ASC de mettre en œuvre des procédures répondant aux défis modernes de la gestion des données selon leur cycle de vie. Le cas échéant, l'ASC sera en mesure de bien se positionner et peut être de servir de modèle parmi les organisations désirant améliorer leur gestion de leurs données techniques.

Lors de ses partenariats avec des agences étrangères, des industries ou des universités, l'ASC sera en mesure de faire bénéficier ses partenaires de ses bonnes pratiques de gestion des données. Étant donné le rôle et la position de l'ASC, la mise en application des principes de ce rapport est en mesure d'engendrer un effet positif dans l'ensemble du milieu scientifique et de l'industrie spatiale canadienne.

En plus de ces contributions tangibles au projet de l'ASC, les activités de recherche sur le terrain (ex. : entrevues) auront contribué à la mise en place d'un processus corporatif qui doit soutenir les nouvelles initiatives de gestion de données. Ainsi, tout en permettant de recueillir de l'information, les entrevues ont permis d'informer les employés de l'ASC des objectifs du projet du secteur de la GI-TI, des bénéfices d'une bonne gestion des données techniques ainsi que des problématiques devant encore être surmontées. En retour, ces informations se sont avérées utiles pour bien faire connaître la réalité du terrain aux principaux acteurs du projet, de les intéresser au projet et d'obtenir leur support.

Au niveau gouvernemental, la présente recherche et le projet de l'ASC qui y est lié ont contribué au projet beaucoup plus vaste du gouvernement[27] visant à gérer les données techniques, conformément aux tendances internationales, afin d'en retirer le maximum de bénéfices pour les Canadiens. L'étude qui est présentée ici s'inscrit au début de ce processus de changement et contribue à l'identification des bénéfices et défis de la conservation des données techniques. Ce

mémoire explore ce qui peut être fait pour améliorer ces pratiques de gestion dans un secteur particulier du gouvernement canadien.

Finalement, ce mémoire met en lumière certains défis modernes de gestion de l'information et de la gestion des données techniques selon leur cycle de vie.

### **4.3 Limites de la recherche**

Comme tout projet de maîtrise, celui-ci présente des limites qu'il faut identifier. En effet, malgré les contributions très concrètes que ce projet a pu apporter à l'ASC, la réalisation s'est limitée à un cadre organisationnel bien précis et a dû être réalisé selon un calendrier strict. Certains volets, jugés secondaires à ce stage, n'ont pu être analysés en profondeur. En outre, il faut tenir compte du fait que cette recherche ne portait que sur la première phase d'un projet technique de très grande ampleur.

Ces quelques limites méritent qu'on s'y attarde puisqu'elles permettent de fixer le cadre à l'intérieur duquel les conclusions sont applicables.

#### **4.3.1 Problématiques complémentaires à étudier**

Parmi les problématiques liées à la gestion des données spatiales identifiées dans ce mémoire, quelques-unes devront être abordées afin de rendre possibles l'application des recommandations du mémoire et l'implantation de nouvelles procédures de gestion des données spatiales à l'ASC.

- Le choix du modèle d'archivage à implanter n'a pas été fait. La section 3.3.2 propose différents modèles d'archivage et de préservation ainsi que leurs avantages et inconvénients. Au moment d'écrire ce mémoire, aucune décision n'avait été prise quant au choix du modèle à utiliser. Cette décision représente l'élément central de la mise en œuvre des recommandations de ce mémoire;
- Les données qui seront soumises à l'*Open Data* n'ont pas été identifiées. Certains éléments nécessaires à cette décision sont énoncés dans ce mémoire mais la décision concernant les types de données qui seront diffusées librement au public reste à prendre. Pour ce faire, il faut identifier l'intérêt que les différentes communautés (scientifique, militaire, civile, commerciale) portent aux différentes données brutes, données transformées et produits que possède l'ASC;

- Aucune évaluation n'a été effectuée et aucune décision n'a été prise concernant les durées de conservation appropriées pour les différents ensembles de données. Les durées de conservation minimales qui répondent aux besoins de l'ASC et de ces partenaires ne sont pas connues à ce jour. Il n'y a pas non plus de critères établis afin d'attribuer une durée de conservation à un ensemble de données;
- L'impact de la gestion des données sur la charge de travail de la GI-TI reste aussi un élément devant être étudié davantage. L'intégration de responsabilités supplémentaires devra être tenue en compte par l'organisation. Celle-ci devra s'assurer que chaque secteur a les ressources et l'expertise nécessaires pour accomplir les nouvelles tâches qui lui seront confiées.

### **4.3.2 Implantation**

La principale limitation de la présente recherche se situe au niveau de la mise en œuvre des procédures et des recommandations qu'elle contient. L'implantation de ces dernières exigeant beaucoup de temps et d'efforts sur le plan organisationnel, elle est manifestement hors de portée du projet de maîtrise. Concrètement, l'implantation s'échelonnera sur de nombreuses années et le chercheur n'a aucun contrôle sur le calendrier d'exécution. Malgré cette limitation, un long processus de validation fut entrepris dès le départ du projet pour garantir la validité des résultats. Ce processus, décrit à la section 3.4, assure le réalisme des résultats, en plus d'en assurer la représentativité et l'acceptabilité auprès de la plupart des intervenants impliqués dans la démarche.

## **4.4 Avenues de recherche et de développement futur**

Le développement des idées, concepts et outils proposés dans ce mémoire pourront se poursuivre dans les prochaines années tant d'un point de vue académique qu'organisationnel. Cette section propose quelques avenues en ce sens.

### **4.4.1 Autres types d'actifs**

La première avenue de recherche pertinente serait d'élargir la gestion des données de l'ASC à d'autres types de données ou d'actifs de l'ASC. Il serait par exemple intéressant d'identifier le potentiel et les enjeux de la conservation des données non spatiales et des données de

télécommunication. Bien qu'elle soit moins évidente, la valeur à long terme de ces types de données peut aller au-delà de leur utilisation principale.

De même, il serait aussi intéressant d'étudier la gestion des échantillons de l'ASC. Tel qu'il est indiqué dans la section 3.1.1.3 du présent mémoire, ces actifs ne constituent pas des données, mais plusieurs enjeux de la conservation se posent du fait de leur grande valeur historique et scientifique. Il serait donc bénéfique de les conserver et d'en promouvoir la réutilisation. La gestion des échantillons entraîne par contre des difficultés propres à ce type d'actif. La préservation de prélèvements vivants ou périssables en est un bon exemple. La diffusion et plus particulièrement la distribution d'échantillons matériels, non reproductibles, en est un autre.

L'étude de la gestion de données non spatiales et d'échantillons peut être faite dans le contexte spécifique de l'ASC mais peut aussi être réalisée de façon globale et conceptuelle.

#### **4.4.2 Autres organismes**

La gestion des données techniques devenant graduellement une préoccupation pour plusieurs organisations, il conviendrait d'étendre les résultats de cette recherche ou du moins, d'en vérifier la transférabilité vers d'autres entités à vocation publique ne s'étant pas attaqué à ce défi (ministères, agences, universités, centres de recherche, etc.) mais aussi vers des entreprises privées. En effet, les organismes manipulant des données techniques ou scientifiques auraient grand intérêt à réaliser une étude semblable à celle-ci afin d'identifier les étapes de gestion de leurs données, les enjeux présents et les rôles et responsabilités des différents intervenants de la gestion de leurs données.

Sur le plan théorique, il serait intéressant de comparer les enjeux et les rôles de la gestion des données dans différents secteurs comme, par exemples, les secteurs public, académique et privé.

Ces deux premiers types d'avenue de recherche visent à répéter sensiblement la même étude mais en modifiant sa portée ou le contexte dans lequel elle est réalisée. Les deux suivants s'attarderont aux possibilités de compléter ou d'approfondir certains aspects de la présente recherche.

#### **4.4.3 Prochaines étapes**

Certaines recherches pourraient être effectuées dans la continuité de celle présentée dans ce mémoire. Il serait par exemple intéressant d'étudier l'implantation de nouvelles procédures et

infrastructures de gestion des données spatiales au sein de l'ASC. La mise en œuvre des concepts et des recommandations de ce mémoire soulèvera de nombreux questionnements dont plusieurs pourraient faire l'objet de recherches universitaires dans le domaine de la GI, des TI ou du génie industriel. En continuant à observer et à travailler avec les intervenants du terrain, il serait possible d'étudier les problématiques rencontrées lors de l'implantation de nouvelles procédures ou d'étudier les besoins de l'ASC par rapport aux ressources ou aux infrastructures.

Il serait aussi intéressant de travailler à résoudre les problématiques identifiées dans la section 4.3.1. Des études ou des outils d'aide à la décision pourraient entre autres être créés afin de choisir un modèle d'implantation, déterminer des durées de préservation, identifier les données les plus pertinentes pour la diffusion ou approfondir notre connaissance des impacts liés à l'aspect commercial de certaines missions satellitaires.

#### **4.4.4 Sujets connexes**

Finalement, bien que la présente recherche décortique plusieurs facettes de la gestion des données techniques, certains sujets connexes n'ont pu être approfondis. Certains d'entre eux représentent un grand intérêt et pourraient faire l'objet d'études plus poussées. C'est le cas de la valeur des données.

Cette notion, brièvement décrite à la section 3.1.1 du mémoire, fait référence au caractère monétaire, scientifique, historique ou opérationnel qu'on peut attribuer aux données. Or, ces différents types de valeur restent difficiles à mesurer et à comparer. De plus, les données numériques présentent probablement d'autres attributs (d'autres « valeurs ») qui restent à définir. Un travail reste donc à faire à ce niveau, tant d'un point de vue théorique que pratique. À terme, ce développement des connaissances pourra permettre aux organisations comme l'Agence spatiale, d'être mieux outillées pour effectuer la conservation des données techniques.



## CONCLUSION

Plus que jamais, les données techniques représentent une richesse. Les technologies de l'information ne cessent d'évoluer, la quantité de données continuera d'augmenter exponentiellement et la capacité de traiter ces données va croître tout autant. Il est donc essentiel pour des organisations comme l'ASC de mettre en place les procédures nécessaires pour en exploiter le plein potentiel. Et comme l'innovation est un levier de plus en plus important sur la scène économique mondiale, le Canada doit se donner les moyens d'en profiter.

Ce mémoire a tout d'abord présenté une revue des concepts fondamentaux de la gestion des données techniques ainsi qu'une comparaison de ces pratiques de gestion au sein de différents pays et organismes. Une méthodologie de recherche-intervention a ensuite été élaborée afin d'identifier et d'adresser les problématiques liées à la gestion des données spatiales à l'ASC. Ces problématiques particulières, le cycle de vie des données spatiales ainsi que les autres outils et analyses ayant été réalisés afin de surmonter les problématiques identifiées ont été présentés dans le Chapitre 3. Le dernier chapitre de ce mémoire a finalement situé ces résultats dans les contextes théoriques et pratiques présentés au début de ce mémoire.

En participant au projet d'amélioration des procédures de gestion des données spatiales, cette recherche lève le voile sur plusieurs éléments nouveaux dans ce domaine de la science. Cette étude illustre la complexité de la gestion de ces types d'actif et tente de l'atténuer par l'ajout d'intervenants œuvrant comme administrateur des processus de gestion. Elle vient aussi identifier des nouveaux éléments pouvant contraindre la diffusion des données comme par exemple le potentiel commercial de certains ensembles de données. L'identification de cet élément est importante puisque la tendance semble pousser les institutions publiques à la fois à augmenter la libre diffusion des données dont elles financent l'acquisition mais aussi à accroître leurs partenariats avec le secteur privé.

Plus que tout autre chose, cette recherche affirme l'importance d'une gestion des données techniques sur l'ensemble de leur cycle de vie, d'une conservation adéquate et d'une valorisation de leur réutilisation.

Considérant le courant politique qui ne cesse de pousser dans le sens de la conservation et de la libre diffusion des données, il est facile de prédire que la transformation du milieu scientifique au

profit de l'*E-science* ne fait que commencer et que la conservation et le libre partage des données techniques risquent de s'accroître considérablement au cours des prochaines décennies.

## LISTE DE RÉFÉRENCES

- [1] Bureau de la traduction, *Termium Plus*. [En ligne]. Disponible: <http://www.btb.termiumplus.gc.ca/>. [consulté le 12 mai 2011].
- [2] J. Raloff, "Preserving digital data for the future of eScience," *Science News*, vol. 174, no. 5, p. 32, 2008.
- [3] P. Lord et A. Macdonald, "e-Science Curation Report," The Digital Archiving Consultancy Limited, Twickenham, Royaume-Uni, Rapport technique, 2003. [En ligne]. Disponible: [http://www.jisc.ac.uk/uploaded\\_documents/e-ScienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf). [Consulté le 29 novembre 2010].
- [4] P. Doorn et H. Tjalsma, "Introduction: archiving research data," *Archival Science*, vol. 7, no. 1, p. 1, 2007.
- [5] G. Janée, "Preserving Geospatial Data: The National Geospatial Digital Archive's Approach," in *Archiving 2009, Final Program and Proceedings*, Springfield, 2009, pp. 25-29. [En ligne]. Disponible: <http://www.alexandria.ucsb.edu/~gjanee/archive/2009/archiving-2009-paper.pdf>. [Consulté le 6 décembre 2010].
- [6] C. A. Lee, "Open Archival Information System (OAIS) Reference Model," dans *Encyclopedia of Library and Information Sciences, Third Edition*. Chapel Hill, North Carolina, U.S.A., 2010, pp. 4020 - 4030.
- [7] ISO, "Geographic information -- Metadata," *ISO, 19115*, 2003.
- [8] ISO, "Information and documentation -- The Dublin Core metadata element set," *ISO, 15836*, 2009.
- [9] OCDE, "OECD Principles and Guidelines for Access to Research Data from Public founding," OECD, 2007. [En ligne]. Disponible: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>. [Consulté le 29 novembre 2010].
- [10] T. Erwin and J. Sweetkind-Singer, "The National Geospatial Digital Archive: A Collaborative Project to Archive Geospatial Data," *Journal of Map & Geography Libraries*, vol. 6, no. 1, pp. 6-25, 2010.

- [11] ASC, *Agence spatiale canadienne*. [En ligne]. Disponible: <http://www.asc-csa.gc.ca/fra/default.asp>. [Consulté le 28 avril 2011].
- [12] P. Lord, A. Macdonald, L. Lyon et D. Giaretta. "From Data Deluge to Data Curation," *The Digital Archiving Consultancy Limited and the Digital Curation Centre*. [En ligne]. Disponible: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.7425&rep=rep1&type=pdf>. [Consulté le 29 novembre 2010].
- [13] E. Yakel, "Digital curation," *OCLC Systems and Services*, vol. 23, no. 4, p. 335, 2007.
- [14] K.-H. Lee, O. Slattery, R. Lu, X. Tang et V. McCrary, "The State of the Art and Practice in Digital Preservation," *Journal of Research of the National Institute of Standards and Technology*, vol. 107, no. 1, pp. 93–106, 2002.
- [15] S. Strodl, C. becker, R. Neumayer & A. rauber, "How to Choose a Digital Preservation Strategy: Evaluating a Preservation Planning Procedure," presented at the 7th ACM/IEEE-CS joint conference on Digital libraries New York, 2007.
- [16] Secrétariat du Conseil du Trésor du Canada, "GC Open Data Exploring the Way Forward," CIOC, Rapport technique, 2010.
- [17] "Loi sur l'Agence spatiale canadienne," dans *L.C. 1990, ch. 13*, 1990.
- [18] ASC, "Rapport de l'Agence spatiale canadienne 37e assemblée scientifique de COSPAR," ASC, Montréal, Canada, 2006-2007.
- [19] J. Saint-Marseille, "Cadre de gestion de l'information V1.1," ASC, Saint-Hubert, Canada, 2011.
- [20] Secrétariat du Conseil du Trésor du Canada, "Politique sur la gestion de l'information," Gouvernement du Canada, 2007.
- [21] ASC, "Architecture d'activités de programmes (AAP) 2011-2012," Agence spatiale canadienne, 2010.
- [22] D. F. Strong et P. B. Leach, "Consultation nationale sur l'accès aux données de la recherche scientifique," Gouvernement du Canada, 2005.

- [23] OCDE, "Déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics." Paris, France, 2004.
- [24] Groupe de travail sur la stratégie des données de recherche, "Gérance des données de recherche au Canada : une analyse des écarts," Gouvernement du Canada, 2008.
- [25] Secrétariat du Conseil du Trésor du Canada, "Directive sur la tenue de documents," Gouvernement du Canada, 2009.
- [26] Conseil national de recherches Canada, "Passerelle vers les données scientifiques," *Conseil national de recherches Canada*, 2010. [En ligne]. Disponible: <http://data-donnees.cisti-icist.nrc-cnrc.gc.ca/gsi/ctrl?lang=fr>. [Consulté le 31 janvier 2011].
- [27] Gouvernement du Canada, "Open data Pilot Project," *Open Data*. [En ligne]. Disponible: <http://www.data.gc.ca/>. [Consulté le 25 mai 2011].
- [28] CSSDP, *Canadian Space Science Data Portal*, 2011. [En ligne]. Disponible: <https://cssdp.ca/ssdp/app/home>. [Consulté le 6 juillet 2011].
- [29] SSDP, *Canadian Space Science Data Portal*, 2006. [En ligne]. Disponible: <http://129.128.234.110:8080/ssdp/jsp/logon.jsp>. [Consulté le 6 juillet 2011].
- [30] CGSM, *Canadian Geo-Space Monitoring*. [En ligne]. Disponible: <http://cgsm.ca/index.html>. [Consulté le 6 juillet 2011].
- [31] Gouvernement du Canada, "Discours du Trône," 41e législature, 1<sup>er</sup> session, 2011. [En ligne]. Disponible: <http://www.speech.gc.ca/fra/media.asp?id=1390>. [Consulté le 17 juillet 2011].
- [32] International Council for Science : Committee on Data for Science and Technology, *CODATA*. [En ligne]. Disponible: <http://www.codata.org/index.html>. [Consulté le 6 juillet 2011].
- [33] Comité national pour CODATA, "Rapport sur les activités relatives aux données au Canada," ICIST-CNRC, Canada, 2008.
- [34] Consortium DataCite, *DataCite*, 2009. [En ligne]. Disponible: <http://datacite.org/>. [Consulté le 15 mars 2011].

- [35] M. Zborowski, "Cisti's Activities in Support of Scientific Data Management in Canada 2008-2010," *Data science journal*, vol. 8, pp. 27-33, 2009.
- [36] DCC, *Digital Curation Centre*, 2010. [En ligne]. Disponible: <http://www.dcc.ac.uk/>. [Consulté le 2 février 2011].
- [37] DCC, "Curation Reference Manual," *Digital Curation Centre*, 2010. [En ligne]. Disponible: <http://www.dcc.ac.uk/resources/curation-reference-manual>. [Consulté le 2 février 2011].
- [38] M. Donnelly et S. Jones, "Template for a Data Management Plan " DCC, Royaume-Uni, Rapport technique, 2010. [En ligne]. Disponible: [www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMP\\_template\\_v1.2\\_100106.doc](http://www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMP_template_v1.2_100106.doc). [Consulté le 21 avril 2011].
- [39] L. Lyon, "Dealing with Data: Roles, Rights, Responsibilities and Relationships," University of Bath, Bath, Royaume-Uni, 2007.
- [40] C. Greer, "Harvesting the Power of Digital Data for Science and Society," Interagency Working Group on Digital Data, Washington, U.S.A., Rapport technique, 2009.
- [41] United States Government, *Data.gov*. [En ligne]. Disponible: [www.data.gov](http://www.data.gov). [Consulté le 10 mai 2011].
- [42] M. Witt, J. Carlson, S.D. Brandt et C. H. Cragin, "Constructing Data Curation Profiles," *The International Journal of Digital Curation*, vol. 4, no. 3, 2009.
- [43] G. Choudhury, "Case Study in Data Curation at Johns Hopkins University," *Library Trends*, vol. 57, no. 2, p. 211, 2008.
- [44] SDSS, *Sloan Digital Sky survey : Mapping the Universe*, 2000. [En ligne]. Disponible: <http://www.sdss.org/>. [Consulté le 6 juillet 2011].
- [45] National Science Foundation, *National Virtual Observatory ...the universe at your fingertips*. [En ligne] Disponible: <http://www.us-vo.org/>. [Consulté le 6 juillet 2011].
- [46] Cornell University Librar, *DataStaR*, 2010. [En ligne]. Disponible: <http://datastar.mannlib.cornell.edu/about?home=1&login=none>. [Consulté le 29 novembre 2010].

- [47] J. A. McCue, "SGER: Planning Information Infrastructure Through a New Library-Research Partnership," Cornell University State, Ithaca, NY, États-Unis, 2006.
- [48] Federal Geographic Data Committee, *The Federal Geographic Data Committee*, 1990. [En ligne]. Disponible: <http://www.fgdc.gov/>. [Consulté le 6 juillet 2011].
- [49] Networking government in New Zealand, *E-government in New Zealand - Data Management Policy and Standards*. [En ligne]. Disponible: <http://www.e.govt.nz/standards/data-management>. [Consulté le 9 mars 2011].
- [50] Gouvernement Australien, *data.gov.au*. [En ligne]. Disponible: <http://data.gov.au/>. [Consulté le 2 août 2011].
- [51] NASA, *PDS: The Planetary Data System*. [En ligne]. Disponible: <http://pds.nasa.gov/index.shtml>. [consulté le 22 février 2011].
- [52] Nasa Office of Space Science and Applications, "Guidelines for Development of a Project Data Management Plan," National Aeronautics and Space Administration, 1993.
- [53] NASA, "Preparation of a Data Management Plan, development of the Data Pipeline, and Efficient Archiving," NASA.
- [54] R. R. Fisher, "NASA Heliophysics Science Data Management Policy," NASA. V1.0, 2007.
- [55] European Space Agency, *ESA space for europe*, 2000. [En ligne]. Disponible: <http://www.esa.int/esaCP/index.html>. [Consulté le 12 juillet 2011].
- [56] European Space Agency, "ESA Member States approve full and open Sentinel data policy principles," *ESA*, 2009. [En ligne]. Disponible: [http://www.esa.int/esaEO/SEMXX570A2G\\_environment\\_0.html](http://www.esa.int/esaEO/SEMXX570A2G_environment_0.html). [Consulté le 4 août 2011].
- [57] I. Leon, C. Arviset, D. Baines, I. Barbarisi, J. Castellanos, N. Cheek, H. Costa, N. Fajersztejn, M. Fernandez, J. Gonzalez, A. Laruelo, I. Ortiz, P. Osuna, J. Salgado, A. Stebe, D. Tapiador, "ESA New Generation Science Archives: State of the art data management techniques for SOHO and EXOSAT Science Archives," presented at the

- Astronomical Data Analysis Software and Systems XIX ASP Conference Series, Vol. XXX, Madrid, Spain, 2009.
- [58] O. Jennrich et P. MacNamara, "LISA Pathfinder Data Management Plan," ESA, Keplerlaan, Pays-Bas, Rapport S2-EST-PL-1004, 2005.
  - [59] A. Hatchuel, "The Two Pillars of New Management Research," *British Journal of Management*, vol. 12, special issue, pp. S33-S39, 2001.
  - [60] A. Dolbec et L. Prud'homme, "La recherche-action," dans *Recherche sociale : de la problématique à la collecte de données*, B. Gauthier, Ed., Québec: Presses de l'Université du Québec, 2009, p. 767.
  - [61] A. David, "La recherche-intervention, cadre général pour la recherche en management ?," dans *Actes de la 9<sup>e</sup> conférence internationale de management stratégique*, Montpellier, France, 2000.
  - [62] A. David, "La recherche-intervention, cadre général pour la recherche en management?," dans *Les nouvelles fondations des sciences de gestion*, FNEGE, Ed., Paris: VUIVERT, 2008, pp. 193 - 213.
  - [63] Y.-C. Gagnon, *L'étude de cas comme méthode de recherche*, Presse de l'Université du Québec ed. Québec, 2005.
  - [64] "Projet de loi C-25 : Loi régissant l'exploitation des systèmes de télédétection spatiale," dans *LS-499F*, 2005.
  - [65] *Politique sur les paiements de transfert*, Gouvernement du Canada, 2008.
  - [66] "Loi sur la protection des renseignements personnels," dans *L.R.C. 1985, ch. P-21*, 1985.
  - [67] T. Hey et A. Trefethen, "The Data Deluge: An e-Science Perspective," dans *Grid Computing: Making the Global Infrastructure a Reality*, L. John Wiley & Sons, Ed., Chichester (UK), 2003.





- Provenance
- Lieu de stockage

Étapes de transformation : - Comment les données sont transformées?

Utilisation : - Pour quelle utilisation?  
- Par qui?

Collaboration?: - Qui est propriétaire des données?  
- Qu'est ce qui est partagé?  
- Qui y a accès? Quand?  
- Comment se fait le partage des données? (moyen de partage)

Si les données sont conservées à l'extérieur, avez-vous un droit de regard sur les données?

SI RECHERCHE : Partagez-vous vos données avant la publication de vos résultats? (embargo?) Quelles données? Avec qui? Sous quelles conditions? Si non, pourquoi?

### ***Collection, Use, Sharing***

*Describe the data:*

- Quantity
- Format
- Type (observation, ...)
- Origin
- Storage place

*Describe the processing steps*

*Who transforms the data? Who uses it? How?*

*Is there collaboration? :- Who owns the data?*

- Which data is shared?
- Who has access? When?
- How are data shared? (Methods)

*If the data is stored outside, do you have a right to look at the data?*

*IF RESHEARCH: Do you share your data before your results are published? (Embargo?) Which data? With whom? Under what conditions? If not, why?*

## **Organisation**

Comment les données sont-elles classées?  
Sont-elles centralisées?

Utilisez-vous des métadonnées? Si oui, lesquelles? (standards)

Quelles informations particulières seraient nécessaires pour :

- Lire les données?
- Comprendre les données?
- Utiliser les données?
- Rechercher les données?

## **Organisation**

*How data is classify?*

*Is it centralized?*

*Do you use Metadata? What kind? (standards?)*

*What information would be needed to:*

- *Read the data?*
- *Understand the data?*
- *Use the data?*
- *Research the data?*

## **Conservation et Diffusion**

Présentement, Est-ce que les données sont conservées après l'utilisation? – Lesquelles? Quel format?

- Pourquoi?
- Combien de temps?
- Où?
- Comment sont-elles organisées?
- Sont-elles diffusées? Partagées?

Avez-vous déjà transféré/migré des données pour leur conservation? Lesquelles? Où? Pourquoi?

Est-ce que vos données pourraient avoir une utilité future?

- Laquelle?
- Quelles données (quel niveau)?
- Pour qui?

Quelles données pourraient intéresser : (Valeur historique, opérationnelle, utilisation)

- l'ASC?
- La communauté scientifique?
- Les entreprises canadiennes?
- Les citoyens canadiens?
- Autres ministères?

Seriez-vous prêt à rendre publiques vos données après utilisation? Quand (embargo?) ? Sous quelles conditions?

Qui devrait être responsable des données? À quel moment?

Vos données peuvent-elles être rendues publiques telles qu'elles sont? Comment doit-on les modifier :

- Pour des raisons de sécurité? (respect de la vie privée, ...)

- Pour permettre leur lisibilité/interprétation/utilisation
- Pour les rendre intéressantes pour d'autres utilisateurs

Est-ce qu'il y a des risques de ne plus être capable de lire les données : - perte du contexte des données (description)  
 -format/logiciel/support/techno. désuet  
 - hardware inexistant ou désuet

## ***Conservation and Distribution***

*Actually, is data kept after it has been used?*

- Which ones? What format?
- Why?
- How long?
- Where?
- How is it organized?
- Is it accessible to other people? Is it shared?

*Do you ever transferred/migrated data for its conservation? Which ones? Where? Why?*

*Do your data could have future utilizations?*

- Which ones?
- Which level of data (primary or processed)?
- For whom?

*Which data could be interesting for : (historical value, operational value, utilisation)*

- CSA?
- The scientific community?
- Canadian companies?
- Canadian citizens?
- Other departments?

*Would you be willing to publish your data after use? When (embargo?)? Under what conditions?*

*Who should be responsible for data? When?*

*Can your data be made public as they are? How should they be modify:*

- For security reasons (privacy, ...)
- To ensure their readability/understandability/usability
- To make them interesting for other potential users

*Is there any risk of not being able to read data:*

- by losing the context around the data
- from an obsolete format/software/support/technology
- obsolete or lost hardware

---

## Élimination

Présentement : - Comment sont-elles détruites?

- Qui autorise la destruction?
- Combien de temps les données sont conservées?
- Pourquoi elles sont supprimées?

## *Elimination*

Currently : - *How is data eliminated?*

- *Who authorize its destruction?*
  - *How long is data kept before its elimination?*
  - *Why is it deleted?*
- 

## Évaluation

Utilisez-vous actuellement des outils ou obtenez-vous de l'aide pour la gestion de vos données?

Idéalement, de quoi auriez-vous besoin pour mieux gérer vos données?

- Outils
- Support
- Processus/plan

Y a-t-il une évaluation entre les phases des projets?

Est-ce que vous verriez l'importance de gérer les données de l'ASC? (comité)

## *Evaluation*

*Do you currently use tools or do you receive some help/support to manage your data?*

*Ideally, what would you need to manage your data better?*

- *Tools*
- *Support*
- *Process/Plan*

*Is there an assessment between project phases regarding data management?*

*Do you consider it important to manage space data from the CSA? (comity)*

## ANNEXE B – LISTE DES ENTREVUES

date	durée	enregistrement	participants
07-févr-11	1h	non	Directeur satellites operations and ground infrastructure Mission manager (sat)
08-févr-11	1h40	non	Program Scientist, Solar Terrestrial Sciences
08-févr-11	40min	oui	Manager, Power and digital electronic Manager, Control and analysis Manager, Material and thermal Director, Technology Development Management* Program Advisor* Head, ESA Program*
09-févr-11	1h15	oui	Head earth observation applications and utilization Project officer EOAU Head exploration robotics
09-févr-11	1h	oui	Program scientist physical science Program scientist space astronomy science manager space astronomy Director solar and earth systems science
10-févr-11	30min	non	Head, IP Management & Technology Transfer
14-févr-11	40min	non	Manager, Policies and Relations with Stakeholders
14-févr-11	50min	oui	Program scientist space life science Chief of the astronautes Senior program scientist life and physical sciences Project officer operation space medicine
14-févr-11	1h	non	Directeur Général, Space Science and Technologies
17-févr-11	30min	non	Manager, Satellite Data
22-févr-11	1h10	non	Planetary scientist Post-Doc Student
24-févr-11	1h	non	Manager, operations planning
28-févr-11	30min	non	Chief Scientist, Life Sciences and ISS utilization
28-févr-11	15min	non	Head, Astronomy and Planetary Mission
10-mars-11	45min	non	Program Officer (Grants & Contributions)

\* ces personnes n'ont pas assisté à l'ensemble de la rencontre

## ANNEXE C – GABARIT DE PLAN DE GESTION DES DONNÉES

### 1 Information générale

#### 1.1 Informations de base sur le plan de gestion des données

Date, Version, Auteur

#### 1.2 Informations sur le projet (référence au PAD)

1.2.1 Nom du projet

1.2.2 Secteur responsable du projet

1.2.3 Durée (durée, date de début prévue, date de fin prévue)

1.2.4 Objectifs principaux du projet

1.2.5 Partenaires

### 2 Aspects légal et éthique

#### 2.1 Politiques

2.1.1 Est-ce qu'une politique spécifique de gestion des données existe pour le projet?

2.1.2 Est-ce que les partenaires ont des exigences en termes de gestion des données?

#### 2.2 Éthique et confidentialité

2.2.1 Y a-t-il des sujets humains ou des éléments de confidentialité dans le projet?

2.2.2 Si c'est le cas, comment la confidentialité des données sera-elle assurée?

#### 2.3 Propriété

2.3.1 Qui sera propriétaire des données?

2.3.2 Y a-t-il un copyright sur les ensembles de données? Si c'est le cas, qui possède ce copyright?

2.3.3 Quelle utilisation l'ASC a-t-elle le droit de faire des données? Y a-t-il des limitations commerciales, légales ou autres?

### 3 Collecte et utilisation

#### 3.1 Description des ensembles de données (données brutes, données transformées, produits)

Niveaux	Description	Format	Quantité	Responsable
Brutes				
Niveau 1				
Niveau 2				
...				

#### 3.2 Collecte

3.2.1 D'où viennent les données? Comment sont-elles créées ou acquises?

3.2.2 Les données sont collectées pour quel usage? Pour qui?

#### 3.3 Transformation

3.3.1 Décrire les différents ensembles de données et les différents produits du projet.

3.3.2 Décrire les étapes de transformation.

3.3.3 Est-ce que les données ou des produits qui en découlent sont des livrables du projet? Lesquels?

- 3.3.4 Est-ce qu'il faut des outils ou logiciels particuliers pour lire les données? Si oui, comment seront-ils conservés? Qui en sera propriétaire? L'ASC aura-t-elle une licence d'utilisation?
- 3.4 Organisation
  - 3.4.1 Quelle norme de métadonnées sera utilisée pour documenter les ensembles de données?
  - 3.4.2 Comment seront créées les métadonnées? Automatiquement ou manuellement? Par qui?
  - 3.4.3 Où seront décrits les instruments et logiciels utilisés, le contexte de saisie, la méthodologie et toutes les informations nécessaires pour réutiliser les données?
- 4 Conservation et diffusion
  - 4.1 Entreposage pendant le projet
    - 4.1.1 Qui est responsable du stockage des données?
    - 4.1.2 Où seront entreposées les données?
    - 4.1.3 Comment sont gérées les sauvegardes? Quoi? Où? À quelle fréquence? Par qui?
    - 4.1.4 Qui a accès aux données durant le projet et comment cet accès est-il protégé?
    - 4.1.5 Y a-t-il des enjeux de sécurité des données durant le projet et comment ces enjeux vont-ils être gérés?
  - 4.2 Conservation après le projet
    - 4.2.1 Les partenaires du projet ont-ils des exigences par rapport à la conservation des données?
    - 4.2.2 Quelles données seront conservées après le projet?
    - 4.2.3 Où seront-elles conservées?
    - 4.2.4 Qui sera responsable de l'archivage et de la conservation des données?
    - 4.2.5 Quelles métadonnées et documents devront être conservés avec les données?
    - 4.2.6 Y a-t-il un délai minimal obligatoire de conservation?
    - 4.2.7 Combien de temps, selon vous, devraient être conservées les données? Pourquoi?
  - 4.3 Diffusion
    - 4.3.1 Qui pourrait être intéressé par les différents ensembles de données du projet? (citoyens canadiens, entreprises, ministères, communauté scientifique, ...) Quelles utilisations futures pourraient être faites de ces données?
    - 4.3.2 Quelles données seront rendues publiques?
    - 4.3.3 Quand ces données seront-elles rendues publiques?
    - 4.3.4 Doit-on modifier les données avant leur diffusion? Pourquoi? Comment? Qui le fera?
    - 4.3.5 Où les données seront-elles rendues accessibles? Qui y aura accès?
    - 4.3.6 Qui sera responsable de la diffusion des données?
    - 4.3.7 Y a-t-il des contraintes qui empêchent ou qui limitent la diffusion des données?
    - 4.3.8 Doit-il y avoir un embargo sur les données? Si oui, pour quelle raison?



## 5 Ressources

### 5.1 Rôles et responsabilités pour l'application de ce plan

- 5.1.1 Qui est responsable de la création et de la transformation des données?
- 5.1.2 Qui est responsable de la documentation?
- 5.1.3 Qui est responsable à long terme? Comment et quand se fera le transfert de responsabilités?
- 5.1.4 À long terme, qui sera en mesure d'évaluer la pertinence des données?

### 5.2 Aspect financier

- 5.2.1 Quel budget sera dédié à la gestion des données?
- 5.2.2 Qui assumera les frais de conservation à long terme?

ANNEXE D – ÉLARGISSEMENT DU CYCLE DE VIE

