

Groupe de travail sur les fichiers de données numériques

Sous-groupe de travail sur la géomatique
ou les données numériques à référence spatiale

**PROJET POUR
UN PORTAIL UNIVERSITAIRE QUÉBÉCOIS DE
DONNÉES NUMÉRIQUES ET GÉOSPATIALES**

présenté au

Sous-comité des bibliothèques



CREPUQ

Adopté le 14 novembre 2002

TABLE DES MATIÈRES

I-	MISE EN CONTEXTE.....	3
	DÉFINITION D'UN PORTAIL DE DONNÉES NUMÉRIQUES ET GÉOSPATIALES :	4
II-	CONTENU DU PORTAIL.....	4
	PRÉAMBULE	4
	MODULE 1 – RESSOURCES.....	5
	MODULE 2 – AIDE ET FORMATION.....	6
	MODULE 3 – COMMUNICATION	7
III-	STRUCTURE DU PORTAIL	7
IV-	MOTEUR DE RECHERCHE	9
V-	RECOMMANDATIONS.....	9
ANNEXE I		
	SCHÉMA CONCEPTUEL DU CONTENU.....	11
ANNEXE II		
	Liste des membres du Groupe de travail sur les fichiers de données numériques.....	12
	Liste des membres du sous-groupe de travail sur la géomatique ou les données numériques à référence spatiale	12

I- MISE EN CONTEXTE

Dans le cadre de l'Initiative de démocratisation des données - IDD, un consortium désormais permanent auquel adhère la majorité des bibliothèques universitaires québécoises, et des capacités du Web à diffuser l'information, l'accès aux données numériques a connu une croissance considérable au cours des dernières années. C'est pour tirer profit de ces nouvelles possibilités que les bibliothèques membres de la CREPUQ se sont dotées d'outils collectifs pour la consultation des données acquises en consortium : **SHERLOCK** donne maintenant accès aux microdonnées de quelque 140 enquêtes alors que le Site des tableaux de données statistiques agrégées offre des milliers de tableaux, principalement en format IVT.

Alors que les données du Recensement figurent parmi les produits de Statistique Canada les plus utilisés, soulignons que l'IDD a également rendu accessibles les fichiers géographiques du Recensement, tant en format ArcInfo que MapInfo, ouvrant la voie aux nouvelles possibilités de représentation géospatiale des données. Cette thématique a d'ailleurs occupé une place considérable lors de la conférence annuelle conjointe de l'Association canadienne des utilisateurs de données publiques - CAPDU et de l'Association des cartothonèques et des archives cartographiques du Canada - ACACC, démontrant bien les relations étroites qui prévalent entre les données et leur représentation spatiale. L'un des ateliers thématiques de cette conférence faisait d'ailleurs place aux témoignages de certaines bibliothèques qui ont déjà amorcé une fusion de leur service de données et de leur service de géomatique. Par ailleurs, l'un des ateliers les plus importants de la conférence portait sur les produits géographiques du Recensement offerts en consortium par la firme DMTI Spatial Inc., un consortium auquel participent plusieurs bibliothèques québécoises. Parallèlement aux données géographiques de Statistique Canada, mentionnons en outre qu'une entente est également intervenue en mai 2002 entre Ressources naturelles Canada, le Programme fédéral des services de dépôts (PSD) et l'Association des cartothonèques et des archives cartographiques pour l'accès aux cartes topographiques numériques. De plus, des discussions sont actuellement en cours avec la Photocartothonèque québécoise du ministère québécois des ressources naturelles pour l'accès aux cartes topographiques numériques produites par ce ministère.

C'est devant la multitude de ressources présentement disponibles qu'est apparue aux membres du Groupe de travail sur les fichiers de données numériques (GTFDN) l'idée d'une réflexion sur la création d'un guichet d'accès unique aux données numériques et géospatiales. Puisqu'il y a un chevauchement important entre l'univers des données numériques et celui des données géospatiales, il est primordial que ce guichet donne accès aux deux univers. À cette fin, une collaboration entre les membres du Sous-groupe de travail sur la géomatique ou les données numériques à référence spatiale (SGTG) et les membres du GTFDN pour la réalisation de ce projet s'impose.

Une première rencontre exploratoire a eu lieu à l'Université Laval le 28 février 2001. Outre les membres du GTFDN, Bruno Langlois de l'UQAR, Sébastien Nadeau et Guy Teasdale de l'Université Laval se sont joints aux discussions afin d'élaborer les éléments de base d'un futur portail qui ne devait alors porter que sur les données numériques. Dans le contexte actuel, cependant, il ne fait aucun doute dans notre esprit que le portail doit maintenant

avoir les deux composantes : données numériques et données géospatiales. Une consultation quelque peu tardive s'est donc effectuée auprès du Sous-groupe de travail sur la géomatique dans le but de solliciter leur participation pour la rédaction de l'actuel projet et d'intégrer les membres du groupe solidement en tant qu'acteurs en parts égales dans les différentes étapes du projet à venir.

DÉFINITION D'UN PORTAIL DE DONNÉES NUMÉRIQUES ET GÉOSPATIALES :

À la suite des discussions et afin de rassembler les idées émises, il nous est apparu nécessaire d'adopter certaines définitions permettant d'établir ce qu'est, ou de ce que devrait être, un portail de données numériques et géospatiales. La définition qui suit est une adaptation, dans le contexte des données numériques et géospatiales, de la définition de portail proposée par l'Office de la langue française dans le *Grand dictionnaire terminologique* :

Site Web dont la page d'accueil propose, en plus d'un moteur de recherche, des hyperliens avec une foule d'informations et de services, qui est conçu pour guider les internautes et faciliter leur accès au réseau des données numériques et/ou géospatiales, mais surtout pour les attirer et fidéliser le plus grand nombre d'entre eux, au point de devenir leur porte d'entrée aux données numériques et/ou géospatiales disponibles dans Internet.

L'objectif du portail est donc de devenir la page de démarrage aux données numériques et géospatiales du plus grand nombre d'internautes. Les termes porte d'entrée et site de démarrage sont donc parfois associés à cette notion de portail.

Pour compléter cette mise en contexte, nous estimons que le projet de portail développé ici est pleinement compatible avec le projet de bibliothèque virtuelle actuellement en développement au Sous-comité des bibliothèques, et qu'il pourra logiquement s'intégrer comme l'un des volets spécifiques de cette infrastructure collective.

II- CONTENU DU PORTAIL

PRÉAMBULE

Il ne fait aucun doute que la qualité du contenu du Portail est une condition essentielle pour la réussite de ce projet. En effet, sans un contenu de qualité, qui est à la fois exhaustif et toujours d'actualité, il est fort peu probable que le Portail devienne une ressource intéressante pour les différents usagers de données numériques et géospatiales des établissements universitaires québécois. C'est pour cette raison que la création de ce Portail ne se limite pas à son lancement initial, mais bien à une gestion continue de ce qu'il est et de ce qu'il a à offrir.

À la suite de nombreuses discussions, il est devenu évident que le Portail doit donner accès en priorité aux données acquises par les membres de la CREPUQ grâce à des ententes de consortium. De même, il est primordial que les ressources locales d'importance développées par les institutions universitaires membres de la CREPUQ soient mises en évidence. De plus,

étant donné que le monde des données numériques et des données géospatiales est vaste et très diversifié, il est clair que le Portail ne peut se limiter aux seules données achetées et produites par les membres de la CREPUQ. Par conséquent, il devient nécessaire d'intégrer le maximum de sources de données qui peuvent intéresser les usagers des établissements universitaires québécois concernés par ce projet, et ce, dans la mesure du possible bien sûr.

Il est aussi important de préciser que la seule mention des ressources disponibles ne suffit pas pour faire du Portail un outil incontournable. En effet, il est convenu que cet accès doit avoir un côté dynamique et pédagogique pour qu'il puisse remplir sa mission première, c'est-à-dire être un outil facilitant l'accès aux données numériques et géospatiales dans le cadre d'activités académiques. Par conséquent, il est convenu qu'une composante du Portail doit être dédiée à l'aide et à la formation, tandis qu'une autre doit répondre aux besoins de communication entre cette source d'information et ses usagères et usagers. C'est pourquoi le contenu du Portail universitaire québécois de données numériques et géospatiales sera divisé en trois modules conceptuels que l'on peut visualiser dans le **schéma** annexé à ce rapport (voir l'Annexe 1).

Voici les trois modules qui peuvent donner une idée du contenu du Portail :

MODULE 1 – RESSOURCES

Ce module recense les différentes sources de microdonnées, données agrégées et données géospatiales acquises grâce à des ententes de consortium CREPUQ ou grâce à l'IDD (Initiative de démocratisation des données de Statistique Canada), ainsi que celles auxquelles on peut avoir accès autrement (ressources Internet gratuites et autres, autres portails de données numériques, etc.).

En premier lieu, pour ce qui est des données géospatiales, il est évident que ce sera au Sous-groupe de travail sur la géomatique d'identifier les ressources de ce type de données qui devront faire partie du Portail. Voici déjà à titre indicatif une liste non exhaustive et non ordonnée de différentes ressources proposée par les membres du Sous-groupe de travail sur la géomatique :

- Banque nationale de données topographiques (BNDT) et autres produits numériques géospatiaux de Ressources naturelles Canada
- Produits géospatiaux de la Photocartotheque québécoise
- DMTI (accès limité pour certaines institutions)

En second lieu, voici à titre indicatif une liste non exhaustive et non ordonnée de ressources en microdonnées et/ou données agrégées qui, selon les membres du GTFDN, doivent faire partie du Portail :

- **SHERLOCK** (<http://sherlock.crepuq.qc.ca>)
- Site CREPUQ des tableaux de statistiques agrégées (<http://ivt.crepuq.qc.ca>)
 - Données de l'IDD
 - Données de l'Institut de la statistique du Québec (ISQ)
 - Autres sources
- CIQSS et CADRISQ

- SourceOECD de l'OCDE (accès limité pour certaines institutions)
- CANSIM II @ CHASS (accès limité pour certaines institutions)
- Canadian Census Analyser via CHASS (accès limité pour certaines institutions)
- Profil des communautés canadiennes – site public de Statistique Canada
- E-STAT (produit du PSD)
- Bilan du siècle (<http://bilan.usherbrooke.ca>)
- Les recensements canadiens depuis le début (exemples : E-STAT, site de CHASS à U. Toronto : <http://www.chass.utoronto.ca/datalib/major/censusag.htm>, etc.)
- ICPSR (accès limité pour certaines institutions)
- Roper Center (sondages aux États-Unis) (accès limité pour certaines institutions)
- Stat-USA (accès limité pour certaines institutions)
- Sites Internet des gouvernements du Canada et du Québec (Statistique Canada, ISQ, ministères, organismes, etc.)
- Sites Internet du gouvernement des États-Unis
- Sites Internet gouvernementaux de différents pays (agences statistiques nationales, universités, etc.)
- Sites Internet des organisations internationales (OCDE, UNESCO, ONU, Union européenne, FMI, Banque Mondiale, etc.)
- Site Internet de la Ville de Montréal (ex. : données sur l'immigration par quartier) et sites d'autres municipalités au Canada
- Statistiques historiques du Canada (accès gratuit de Statistique Canada)
- Sites locaux des universités du Québec traitant des données numériques
- Autres portails de données numériques du Canada et d'ailleurs (ex. : économie, sciences sociales, etc.)
- Etc.

MODULE 2 – AIDE ET FORMATION

Ce module recense différentes ressources utiles pour la compréhension de la nature des données numériques, de la nature des données géospatiales, de la statistique (nécessaire pour le traitement des microdonnées), de l'utilisation des différentes ressources mentionnées dans le module 1 et de la compréhension des différents aspects techniques relatifs au monde des données numériques et au monde des données géospatiales. De plus, il donne accès à une banque d'expertises que les usagères et usagers du Portail ont bien voulu partager avec leurs collègues.

Voici à titre indicatif une liste non exhaustive et non ordonnée des principales composantes de ce module :

- Définitions pour les données numériques et les données géospatiales (microdonnées, données agrégées, données géospatiales, etc.)
- La statistique (définitions, méthodes, exemples, etc.)
- Bibliographie : sources imprimées et électroniques utiles pour la compréhension des données numériques et des données géospatiales (complément aux définitions)
- Guides pour l'usage de : banques de données, logiciels de traitement statistique (SPSS, SAS, etc.), logiciels de traitement de données géospatiales, etc.
- Didacticiels
- FAQ (Foire aux questions)
 - Recherches par thèmes
 - Banque d'expertises (Questions-Réponses)
 - Aide technique
- Etc.

MODULE 3 – COMMUNICATION

Ce module est composé de services permettant le développement de liens dynamiques entre l'outil et ses usagères et usagers.

Voici à titre indicatif une liste non exhaustive et non ordonnée des principales composantes de ce module :

- Liste de diffusion (exemples : *listserv* des responsables de **SHERLOCK** dans les établissements universitaires membres de l'entente, *listserv* élargi selon demandes...)
- Annonces d'activités par un bulletin en ligne (ateliers de perfectionnement CREPUQ et IDD, conférences, congrès, nouveaux produits et services, etc.)
- Formulaire Internet pour les suggestions (TRÈS IMPORTANT pour fidéliser les usagers, leur donner le sentiment qu'ils participent au développement du Portail, leur permettre de partager leurs expériences/expertises, etc.)
- Etc.

III- STRUCTURE DU PORTAIL

Dans le chapitre II, nous avons parlé de trois modules : Ressources, Aide et formation, et Communication. Il s'agit là de modules pour l'organisation conceptuelle du contenu et non de la structure physique du portail comme telle. L'utilisateur accédera au contenu du Portail par de multiples points d'accès, tenant compte de son niveau de connaissances des données numériques, de la nature de son besoin d'information et du contexte de sa recherche.

D'abord, le Portail sera entièrement bilingue pour répondre à l'ensemble des clientèles des établissements membres de la CREPUQ. La page d'accueil du Portail dirigera l'utilisateur vers les pages françaises ou anglaises du site, selon le modèle de **SHERLOCK**. La communication entre les pages françaises et anglaises sera assurée par la présence d'un

bouton dans la bannière de chaque page permettant la navigation facile vers la page équivalente dans l'autre langue.

La deuxième page dans chaque langue présentera à l'utilisateur un **choix de points d'accès**, dans le but de répondre aux différents besoins des clientèles de tous les niveaux, allant de l'utilisateur novice qui cherche une simple statistique à l'utilisateur expérimenté dans la manipulation des fichiers de microdonnées. Les points d'accès devront comprendre les suivants :

- Accès thématique : sans tenir compte du type de données
- Accès par type de données : données numériques : données agrégées, microdonnées d'enquêtes; données géospatiales
- Accès géographique : considérant que la recherche de données se fait très souvent en fonction du pays ou de la région; pourrait être intégré avec l'accès thématique
- Accès par moteur de recherche : sujet traité dans le chapitre suivant
- Accès « pédagogique » : pages à l'intention de l'utilisateur novice qui ne comprend pas les concepts de base de la recherche dans le domaine des données numériques (données agrégées, microdonnées, etc.)
- Accès par ordre alphabétique des ressources : pour l'accès rapide à une ressource déjà connue
- Accès par organismes producteurs des données.

Sur la page Web des points d'accès aux ressources, une présentation graphique mettant en relief les différents accès pourrait faciliter le choix de celui qui conviendrait le mieux à l'utilisateur et à sa recherche. Cette page devra comprendre également des liens vers les modules « Aide et formation » et « Communication ».

Pour l'accès thématique, il faudra établir un plan de classification, un travail déjà entamé par les membres du Groupe de travail sur les fichiers de données numériques dans le contexte des travaux d'amélioration de **SHERLOCK**. Nous préconisons un plan de classification non hiérarchique selon le modèle du site Web de l'University of Toronto Data Library (<http://www.chass.utoronto.ca/datalib/other/findcans.htm>). On pourra s'inspirer également d'autres modèles déjà en place sur le Web, tel le plan de classification de CANSIM sur le site de Statistique Canada. En fonction des enquêtes disponibles, ce plan de classification, ou un sous-ensemble de ce plan, devrait s'appliquer intégralement à **SHERLOCK** et devrait être élaboré dans les plus brefs délais en vue de son implantation immédiate dans **SHERLOCK**.

Pour faciliter la navigation dans l'ensemble des pages du portail, y compris le passage d'un type d'accès à l'autre, un bandeau de navigation qui prendra sa place sur toutes les pages s'impose. Les éléments qui devront s'y retrouver, entre autres, sont les suivants : les différents points d'accès précédemment mentionnés, un lien vers le module « Aide et formation » et un lien vers le module « Communication ». Finalement, un plan complet du site (« site map ») sera offert pour permettre aux usagères et usagers d'avoir un aperçu global du Portail.

IV- MOTEUR DE RECHERCHE

L'implantation d'un moteur de recherche efficace pour le Portail soulève un certain nombre de problèmes. Ainsi, il y aura dans le Portail des ressources de nature différente comme des pages statiques .html, des enquêtes de **SHERLOCK** présentées dans un mode dynamique avec des scripts Perl et possiblement des notices issues d'un logiciel de gestion de bases de données. Le moteur devra prendre en considération toutes ces ressources.

Le moteur devrait permettre à l'utilisateur de restreindre sa recherche sur une portion du Portail (p.ex. **SHERLOCK**) ou au contraire de faire une requête sur l'ensemble du Portail. Il devrait aussi permettre de consulter des moteurs déjà existants (**SHERLOCK**, tableaux ivt, etc.). On devra se pencher sur le problème de l'affichage de résultats lorsque ceux-ci proviennent de l'interrogation de moteurs différents.

Il faudra voir si le moteur de recherche pourrait prendre en compte, d'une façon à définir, la grille thématique permettant à l'utilisateur un accès aux ressources par grands thèmes. On pourrait penser par exemple à restreindre une interrogation sur un thème donné.

Enfin, à titre d'exemple parmi d'autres et sans présumer de la valeur d'un tel exercice, le Groupe considère que le système développé récemment à l'Université Western Ontario¹ mérite d'être évalué. Ce système prend notamment en compte des fichiers de différentes natures et utilise des métadonnées, un moteur de recherche et une base de données SQL.

V- RECOMMANDATIONS

Sur la base des éléments établis précédemment, nous recommandons la mise en place d'un portail universitaire québécois de données numériques et géospatiales.

À cette fin, nous proposons les étapes suivantes pour la réalisation du projet:

- 1) Définition du concept et du contenu pour le projet du Portail universitaire québécois des données numériques et géospatiales.
- 2) Concertation entre les membres du Groupe de travail sur les fichiers de données numériques et ceux du Sous-groupe de travail sur la géomatique en vue d'établir une collaboration étroite pour la conception du projet.
- 3) Analyse du projet par le Groupe de travail responsable de la gestion de **SHERLOCK** (GTGS).
- 4) Soumission du projet à l'approbation du Sous-comité des bibliothèques et obtention des ressources nécessaires à la réalisation d'un devis.
- 5) Réalisation du devis par un-une bibliothécaire spécialiste des données numériques et/ou géospatiales et un-une spécialiste en informatique (ayant de l'expérience dans la

¹ Vincent, S. « Système de données sur Internet de la bibliothèque (SDIB) de l'Université Western Ontario ». *Bulletin de l'IDD* [En ligne]. Vol. 5, n° 1 (printemps-été 2002), p. 2-9.
http://www.statcan.ca/francais/Dli/Document/updatev51_f.pdf (Page consultée le 27 août 2002)

réalisation de portails) qui travailleront en tandem pour une période d'environ un mois. Les personnes choisies pour la préparation du devis devront consulter tout au long du processus les membres du Groupe de travail sur les fichiers de données numériques et du Sous-groupe de travail sur la géomatique afin de s'assurer que le devis répond bien aux attentes des institutions membres de la CREPUQ. De même, il serait souhaitable, dans la mesure du possible, d'associer d'une façon ou d'une autre (ex.: sondage par courriel) l'ensemble des responsables locaux de données numériques du Québec pour la réalisation du devis.

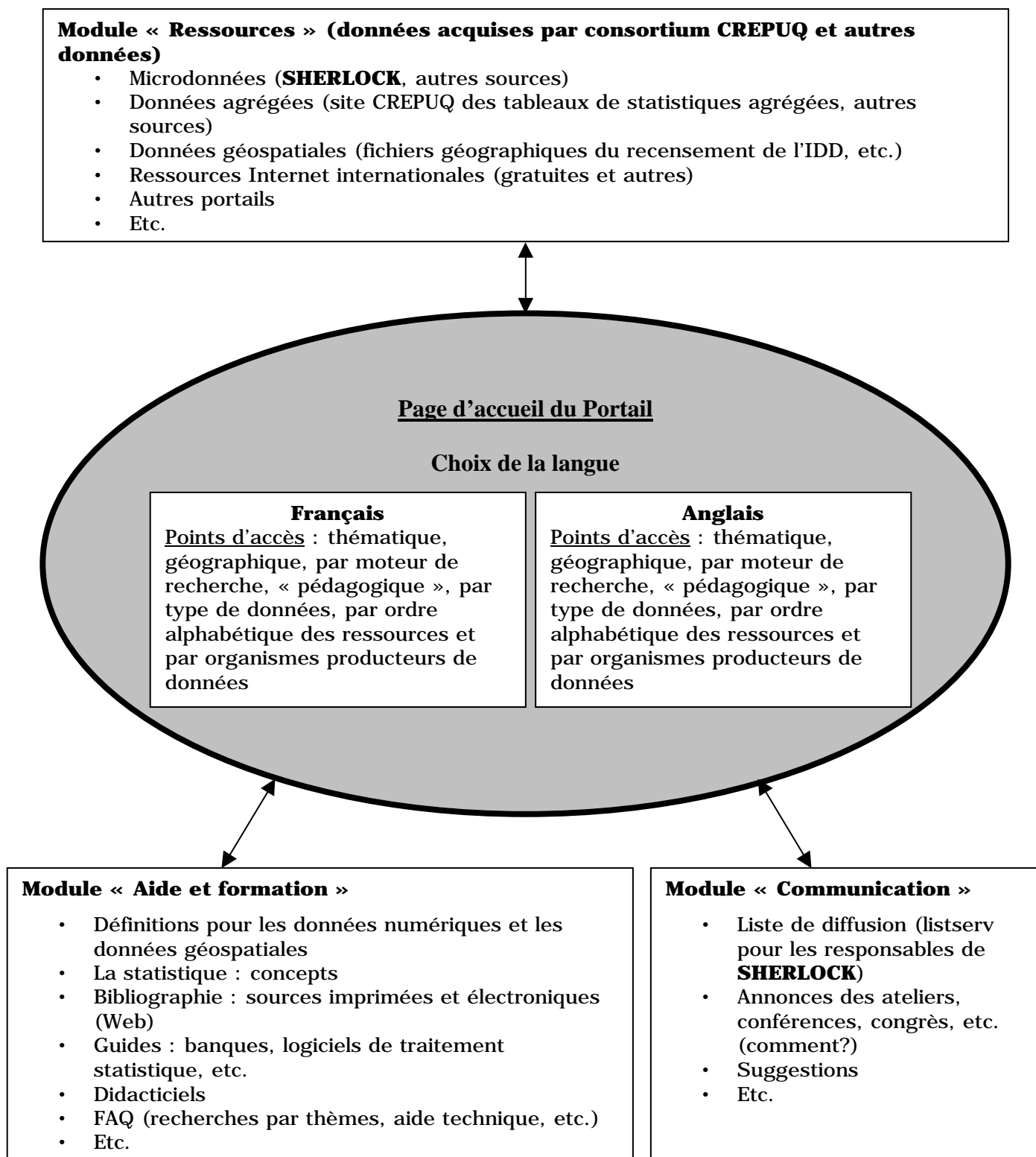
Le GTFDN évalue qu'une somme d'environ 10 000 \$ à 12 000 \$ serait nécessaire pour rémunérer le travail des deux personnes (environ 5 000 \$/personne selon le salaire moyen d'un professionnel).

- 6) Présentation du devis au Groupe de travail responsable de la gestion de **SHERLOCK**.
- 7) Révision du devis en fonction des remarques, commentaires et suggestions d'ajustements des membres du Groupe de travail responsable de la gestion de **SHERLOCK**.
- 8) Soumission du devis de réalisation révisé à l'approbation des membres du Sous-comité des bibliothèques et obtention des ressources nécessaires à la réalisation du projet du portail.
- 9) Réalisation et implantation du projet du portail selon le devis. Mise au point et validation du fonctionnement, en collaboration avec différents groupes intéressés et en consultation avec des usagers éventuels.
- 10) Entretien et développement continu. Nomination d'un webmestre/responsable du portail et création d'un comité aviseur du portail composé en partie ou en totalité des membres du Groupe de travail sur les fichiers de données numériques et du Sous-groupe de travail sur la géomatique.

Conditions essentielles pour la réussite du projet : une mise à jour et un développement constants.

- 11) Lancement officiel.
- 12) Évaluation du projet après une année.

ANNEXE I - SCHEMA CONCEPTUEL DU CONTENU



ANNEXE II

LISTE DES MEMBRES DU GROUPE DE TRAVAIL SUR LES FICHIERS DE DONNÉES NUMÉRIQUES (GTFDN)

Président

Jerry Bull (Université de Montréal)

Agent de liaison au Sous-comité des bibliothèques

Claude Bonnelly (Université Laval)

Membres

Richard Boily (Université du Québec à Rimouski)

Philippe Feredj (Université de Sherbrooke)

Guy Julien (Université Laval)

Anastassia Khouri (Université McGill)

LISTE DES MEMBRES DU SOUS-GROUPE DE TRAVAIL SUR LA GÉOMATIQUE OU LES DONNÉES NUMÉRIQUES À RÉFÉRENCE SPATIALE (SGTG)

Présidente

Hélène Genest (Université Laval)

Agent de liaison au Groupe de travail sur les fichiers de données numériques

Guy Julien (Université Laval)

Membres

Julie Gendron (Université de Sherbrooke)

Anastassia Khouri (Université McGill)

Françoise Lange (Université du Québec à Chicoutimi)

Marie Lefebvre (Université du Québec à Trois-Rivières)

Rosa Orlandini (Université McGill)

Jean-François Palomino (Bibliothèque nationale du Québec)

Pierre Roy (Université du Québec à Montréal)