

# Projet GéoTopia

## Réponse au questionnaire ADONIS

### 1) Nom du projet et résumé en une phrase

*GéoTopia* : Géoréférencer pour transmettre, organiser, partager et interpréter des données d'archive.

### 2) Partenaires impliqués au plan scientifique

Il y a deux principaux partenaires scientifiques dans le projet.

L'équipe Imagerie et Systèmes d'Information Géographique (ISIG) de l'UMR Environnement-Ville-Société du CNRS (<http://umr5600.univ-lyon3.fr/isig.html>) est une équipe de recherche spécialisée dans le domaine de l'information géographique. Elle regroupe une vingtaine de chercheurs, ingénieurs et techniciens du CNRS, des Universités Lyon 2, Lyon 3 et Saint-Etienne, de l'INSA, de l'ENTPE et de l'ENS Lyon. L'équipe fonctionne à la fois comme une plate-forme technique pour l'UMR et comme un groupe de recherche à part entière. Les compétences des chercheurs concernent principalement la télédétection, les systèmes d'information géographique et la géomatique, l'analyse et la modélisation spatiales, la cartographie, la géovisualisation 3D ...L'application de ces compétences se fait dans de multiples champs : urbanisme et aménagement, hydrologie, géographie physique, environnement, géographie historique.... C'est dans cette équipe qu'a été développée, avec déjà un partenariat industriel, Geowebexplorer, plate-forme pédagogique pour un apprentissage collaboratif des SIG . Les chercheurs mobilisés seront Thierry Joliveau, Hervé Parmentier, Luc Merchez, Anne-Sophie Cléménçon, Kristell Michel, Marie-Laure Tremelo, Pierre-Olivier Mazagol, Justine Ultsch, Sylvain Genevois,.

Le Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour (LIUPPA : <http://liuppa.univ-pau.fr>) est une Equipe d'Accueil (EA 3000) commune à différents établissements situés à Pau, Bayonne et Mont de Marsan. Le groupe de recherche directement mobilisé est "Document électronique, sémantique et interaction" (DESI : <http://liuppa.univ-pau.fr/DESI/> ). Le domaine de compétences de ce groupe couvre des méthodes et des techniques en Modélisation et Traitements des Contenus (Représentation sémantique de l'information, Extraction et Recherche d'Information) ainsi qu'en Interaction et en Education. A titre d'exemple, un projet représentatif "Pyrénées Itinéraires Virtuels" (PIV : [http://liuppa.univ-pau.fr/spip/article.php3?id\\_article=114](http://liuppa.univ-pau.fr/spip/article.php3?id_article=114)) mené en collaboration avec la Médiathèque de Pau a donné lieu à un prototype d'extraction, d'indexation et de recherche d'informations spatiales dans des fonds documentaires fortement liés au patrimoine pyrénéen. Ces travaux ont été largement publiés et présentés récemment dans le cadre de la soutenance de thèse de Julien Lesbéguries (doctorant co-encadré par Mauro Gaio et Christian Sallaberry).

La société Makina Corpus (<http://makina-corpus.com/>) est le troisième partenaire

principal. Makina Corpus est une société de services en logiciels libres. Elle a développé la plate-forme MapN qui servira d'outil pour ce projet, uniquement à partir de logiciels et composants libres sous Licence GNU/GPL. Cette plate-forme a vocation à faciliter la publication, la diffusion, l'échange de données d'archives numériques à référence spatiale en permettant aussi la rédaction d'articles, de commentaires sur les ressources documentaires publiées afin d'en faciliter la consultation élargie et l'enrichissement sous un mode collaboratif. Makina Corpus est intéressée à suivre l'usage effectif de l'outil qui vise à être développé au sein des différentes communautés d'utilisateurs concernés. Le responsable du projet pour Makina Corpus sera Pierre-André Le Ny, chef de produit SIG, D'autres partenaires interviendront dans le projet en tant qu'utilisateurs de l'outil : les Archives Municipales de Saint-Etienne (Corinne Porte) et la ville de Saint-Etienne (Aline Bruyère), la Maison du Rhône (André Vincent). D'autres partenaires seront invités à s'ajouter au réseau.

### **3) Contribution du projet aux avancées de votre discipline en termes de production, exploitation, valorisation et diffusion des connaissances**

La question du géoréférencement, le renseignement spatial des objets de toute nature, a dépassé depuis longtemps le domaine spécialisé de la géomatique, des SIG et des applications spécifiquement géographiques ou topographiques. Elle concerne maintenant un domaine extrêmement large intégrant tout type d'information, par exemple celle des bibliothèques, des musées, des inventaires. Elle touche aux sciences de la documentation et de l'information en général (Hill 2006). Le géoréférencement ne se fait pas forcément par l'attribution de coordonnées en x,y,(z). Il peut passer aussi par l'association de mots-clés géographiques (géotags), d'une adresse postale, d'une référence à un toponyme ou à une entrée dans un index géographique.

Notre projet vise à explorer au moyen d'un dispositif technique spécifique la manière dont le géoréférencement peut faciliter la publication, la diffusion et l'échange de données d'archives numériques à travers leur consultation élargie et leur enrichissement sous un mode collaboratif. Il vise aussi à favoriser la rédaction de commentaires sur les ressources documentaires publiées et la rédaction d'articles documentaires et scientifiques sur ces ressources..

Le projet propose l'élaboration conjointe et coordonnée *d'un outil logiciel*, sous forme de plate-forme en ligne, qui propose des fonctions de géoréférencement, de description et d'annotation de données d'archives et d'une série *d'applications de cet outil* dans des contextes multiples : démarche partenariale de gestion, production de recherche scientifique, valorisation et utilisation de résultats à des fins pédagogiques.

### **4) Contribution du projet aux missions et besoins du TGE Adonis tels qu'exprimés dans la «charte» du TGE figurant en pièce attachée**

*Rendre accessibles les ressources et données* est évidemment le but principal de notre projet. Les principales collections des archives intéressant directement les Sciences Humaines et Sociales sont en voie de numérisation. La question de leur géoréférencement émerge comme une question fondamentale, mais les services sont encore démunis par rapport à cette question. Le géoréférencement peut pourtant contribuer, au moins en

partie, à la résolution des problèmes de préservation, de stockage, de signalement par des méta-données normalisées et partagées, de réactualisation ou de réutilisation de ces archives. L'élaboration de nouveaux outils documentaires doit permettre *la valorisation des collections en ligne* car la question cruciale devient l'accès de l'ensemble de la communauté des chercheurs à ces documents d'archives numérisés en liaison avec des organismes partenaires représentant des centres de production et d'hébergement de données.

La continuité numérique permet théoriquement d'espérer une mise en relation des corpus de données d'archive brutes désormais numérisées avec les articles publiés dans des revues SHS. Un outil du type de celui que nous proposons devrait donc contribuer à valoriser et *développer l'édition électronique en SHS*, sous forme d'articles originaux et d'autres ressources documentaires originales sous licence Creative Commons qui représente une alternative à la publication traditionnelle y compris sous sa forme électronique. Un système basé sur un CMS est particulièrement adapté pour permettre aux internautes de se constituer rapidement en réseaux par domaines d'intérêts et/ou par types d'archives et d'accroître la rapidité dans la diffusion de leurs travaux. Cela permettra *d'accroître la réactivité et les échanges* autour des ressources documentaires.

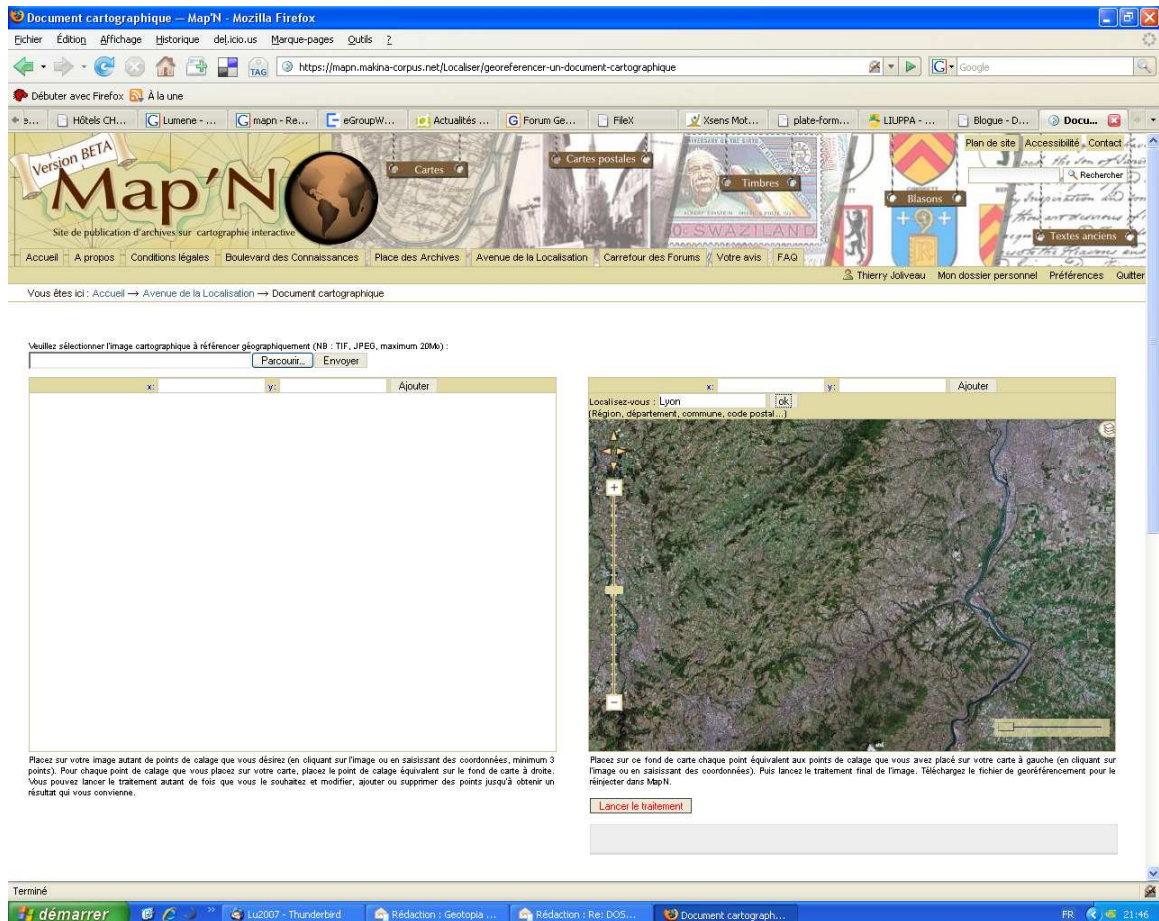
Enfin, notre projet s'appuie sur le développement d'une plate-forme visant à *interconnecter des ressources* documentaires de nature et de forme très diverses et *des acteurs* d'horizons et de disciplines très variés. Pour cela, il convient de s'intéresser aux *transformations dans les usages* de l'information numérique géoréférencée, et de les prendre en compte dans la conception des instruments. Si l'outil que nous entreprenons de construire vise les chercheurs en SHS concernés par les ressources documentaires, notre projet répond aussi à l'exigence du TGE ADONIS de permettre *une diversité de pratiques*. Nous voulons aussi expérimenter comment il est possible de s'adresser au grand public et pas seulement à des publics de professionnels de la recherche en SHS. En ce sens, nous espérons répondre à l'esprit de la Déclaration de Berlin en matière d'ouverture de la recherche vers le reste de la société et d'encouragement d'un débat social indispensable aux sciences humaines et sociales. Nous étudierons les conditions pour qu'un public non spécialisé puisse avoir accès plus directement aux travaux de recherche menés sur les données d'archives et y contribuer en soumettant des documents issus d'archives privées, par exemple.

##### **5) Définition précise de l'outil ou du dispositif technique que vous souhaitez réaliser**

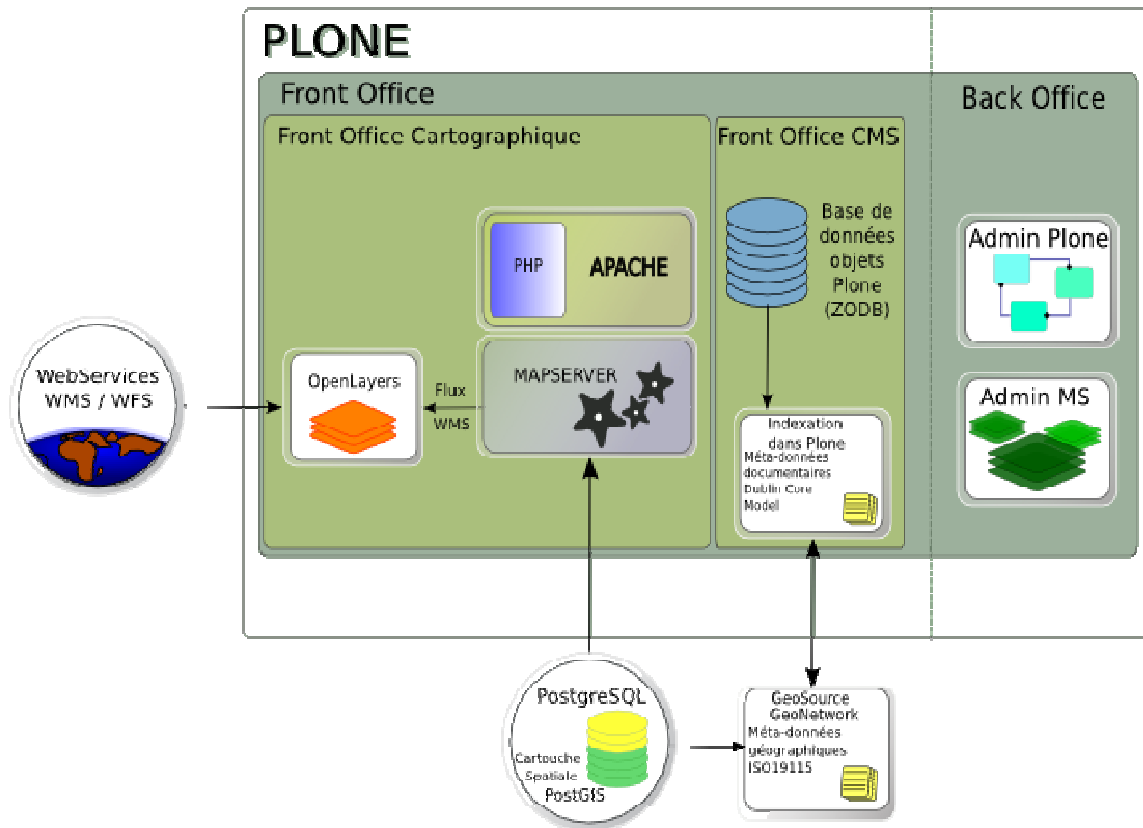
Le projet est construit autour de la plate forme MapN développée par la société Makina Corpus, partenaire industriel du projet. Le prototype de MapN a fait l'objet d'un stage de Master 2 SIG et Gestion de l'espace 2006-2007. Le fait que la société soit fortement impliquée dans le monde du logiciel libre et propose des solutions innovantes dans le domaine des Systèmes d'Information Géographique, de la gestion de contenu (CMS) et des communautés de travail collaboratif (collecticiel) est un facteur très important. Le projet vise à tester la plate-forme MapN en la mettant en oeuvre dans différents contextes, de l'adapter et de la compléter avec d'autres modules, à partir d'usages réels.

## Présentation générale au plan technique

MapN s'appuie sur un système de gestion de contenus (Content Management System-CMS). En l'occurrence MapN utilise Plone (s'appuyant sur un serveur d'application Zope). Ces produits sont des logiciels libres et sont distribués sous licence GNU/GPL. Pour la partie référencement spatial et consultation des documents géoréférencés, le serveur cartographique MapServer est mis en oeuvre du côté serveur et le produit OpenLayers du côté client. Il s'agit d'une application cliente sous licence GPL, qui est paramétrable grâce au langage JavaScript.



**Description des composants de l'outil s'il s'agit d'un système d'information intégré**  
Schéma de l'architecture générale de MapN :



**Description détaillée des fonctionnalités**

Il n'est pas utile de décrire ici de manière détaillée chacune des fonctions de MapN.

**Les principales fonctionnalités de la version actuelle:**

- recherche de contenus
- publication de contenus
- géoréférencement en ligne de contenus planimétriques (cartes, photographies aériennes, images satellitaires, etc.) ou ponctuels (autres types de contenus)
- rédaction collaborative de listes de liens vers des ressources sur l'internet]
- récupération en téléchargement de contenus issus du géoréférencement en ligne

**Les fonctionnalités à développer**

La liste des tâches à réaliser durant le projet est indicative. Elle dépendra bien sûr de l'évaluation des utilisateurs. Certaines pistes ont été déjà déterminées. Leur priorité sera validée en fonction des réactions des usagers et des chercheurs:

*Fonctionnalités générales*

- Amélioration de l'interface de géolocalisation de photos : type "drag and drop"
- Indexation littérale des données documentaires (métadonnées documentaires)
- Indexation spatiale (métadonnées géographiques)
- Connexion à des index toponymiques (gazetteers)
- Rédaction collaborative de notices biographiques, de références bibliographiques, d'articles à valeur scientifique, etc.
- Export de fichiers géoréférencés en .kml pour les rendre consultables sous GoogleEarth
- Association d'une suite de documents à un objet de type itinéraire et leur visualisation sous forme d'animations (on retrouve cela pour GoogleEarth aussi).
- Annotation d'éléments à l'intérieur des documents iconographiques associée ou non à une localisation géographique ou à un autre document :
- Renforcement de la sécurité du site

#### *Modules spécifiques*

- Géolocalisation automatique des textes :

Il s'agit de développer un traitement linguistique ciblé interprétant l'information spatiale. L'hypothèse est que des traitements relativement peu coûteux suffisent à dégager l'essentiel de l'information. Une méthode de recherche d'information spatiale multi-niveaux indexant un corpus textuel brut extrait l'information d'un corpus et l'interprète. Elle permet d'améliorer l'efficacité de systèmes de recherche d'information à chaque fois que l'interrogation comporte une connotation spatiale. L'interprétation permet en outre de retrouver le contexte dans lequel l'information spatiale a été utilisée. En particulier, elle permet d'indexer des unités de texte en leur associant des contextes de type itinéraire, description locale ou comparaison de lieux.

- Géolocalisation de photographies au moyen d'un calcul de visibilité théorique : L'objectif est d'interfacer avec MapN une application de description de photographies basée sur l'interrogation spatiale d'une base de données géographiques décrivant la région que concerne l'image. On s'appuie pour cela sur les fonctions de calculs de visibilité disponibles dans les SIG et fondés sur des algorithmes de lancer de rayons.

#### **Modalités de fonctionnement en réseau**

L'hypothèse est qu'un outil du type de celui qui doit être développé peut faciliter une participation effective et active de communautés diverses et aider à la constitution de réseaux d'acteurs ainsi qu'à des échanges de contenus et de contributions à caractère scientifique. Le projet conduira à connecter différents réseaux.

Le développement informatique sur MapN sera principalement pris en charge par les ingénieurs de Makina Corpus. Le Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour (LIUPPA) prendra en charge le développement d'un module de géolocalisation textuel, tandis que ISIG/UMR EVS s'occupera du module de géolocalisation des photographies tangentiels.

Dans le cadre du projet, deux réseaux d'utilisateurs se constitueront autour de l'usage de cet outil.

### ***L'étude des relations ville-rivière, le cas de Saint-Etienne et du Furan***

Dans le cadre de sa thèse en contrat CIFRE, en tant que salariée du CIRIDD mise à disposition de la Ville de Saint-Etienne, Justine Ultsch utilisera l'outil proposé pour construire une plate-forme d'échange collaborative, afin que tous les partenaires disposent d'un outil commun facile d'accès aux données historiques. Cette plate-forme contiendra des informations nécessaires à la gestion de projets sectoriels mis en oeuvre par les services techniques de la Ville de Saint-Etienne (coordination des travaux d'assainissement, identification du patrimoine à protéger, etc.). Mais cet outil peut aussi contribuer à l'ambition de faire travailler ensemble différents services (eau et assainissement, ville d'art et d'histoire, archives, urbanisme prospectif, etc.) autour de la rivière du Furan, qui serait alors appréhendée de manière systémique.

Ainsi, le premier avantage est de regrouper toutes les données historiques relatives à la ville de Saint-Etienne et à la rivière du Furan sur un seul espace virtuel. En effet, ces données sont aujourd'hui dispersées en plusieurs lieux de conservation et/ou d'accès limité (cf. contraintes de communication des documents anciens). Le second intérêt est de permettre le développement de nouvelles fonctionnalités, telles que la recherche par toponyme ou la possibilité de géoréférencement en ligne. La démarche de mise en commun peut également faire émerger de nouveaux besoins ou de nouvelles attentes de la part des différents partenaires jusqu'ici non perçues (démarche d'expérimentation). Enfin, la volonté d'avoir une information géographique et historique sur la rivière à Saint-Etienne entre dans une logique de vulgarisation et de diffusion en direction de divers publics. Pour cela, l'utilisation d'un outil Internet, et en l'occurrence libre, s'avère stratégique. Ce premier réseau testera l'outil dans un contexte de collaboration de chercheurs, d'un site officiel d'Archives (le service des Archives municipales de la Ville de Saint-Etienne) et d'utilisateurs professionnels (Ville de Saint-Etienne). Dans une deuxième phase du projet, les acquis méthodologiques pourront être étendus au projet de valorisation du patrimoine dans l'agglomération stéphanoise envisagé conjointement par la communauté d'agglomération de Saint-Etienne Métropole et l'agence d'urbanisme de la région stéphanoise EPURES.

### ***Site Métropole Lyon.***

Dans le cadre de son projet Elyse : Espace Lyon / St Etienne, l'équipe ISIG cherche à intégrer les données saisies séparément dans différents projets SIGéohistoriques, ayant mobilisé des données d'archives variées. Elle cherche aussi à les enrichir avec de nouvelles données d'archives, liés à de nouveaux projets de recherche en cours. L'objectif sera alors de tester la capacité de l'outil à aider à structurer un réseau de chercheurs de disciplines variées. Les chercheurs en question sont ceux de l'UMR Environnement-Ville-Société mais aussi d'autres institutions de recherche (on pense aux historiens du LAHRA, avec lesquels des contacts ont été pris ou de la Maison du Rhône qui a donné son accord pour participer au projet. Dans une première phase, l'objectif sera essentiellement de structurer un niveau d'échange scientifique. Dans une seconde phase,

une expérimentation sera conduite pour ouvrir le site à un plus grand public et aux spécialistes (collectionneurs de cartes postales) sur le thème du paysage et pour mettre en oeuvre un projet pédagogique, en liaison avec Géoconfluence (Ecole Normale Supérieure LSH) et l'INRP (Institut National de la Recherche Pédagogique). Le risque inondation à Lyon est un des thèmes envisagés.

### **Liste détaillée des normes utilisées**

Les normes suivantes sont d'ores et déjà utilisées par MapN :

- Normes d'indexation de données documentaires dans Plone : implémentation du Dublin Core Metadata Model. Le Dublin Core Metadata Model a été conçu à partir des travaux réalisés par le W3C (World Wide Web Consortium) en particulier le cadre de description des ressources (Resource Description Framework RDF). Le Dublin Core Metadata Model a été implémenté dans Plone en utilisant les diagrammes de classes UML.
- Implémentation dans Plone des normes d'accessibilité définies par le W3C en particulier les Web Content Accessibility Guidelines - (WCAG v1.0)
- Implémentation des norme XHTML 1.0 et CSS conformément aux spécifications édictées par le W3C
- Implémentation dans PostGIS (cartouche spatiale de la BD PostGreSQL) des spécifications de l'OGC (OpenGIS Consortium) comme par exemple les spécifications « Simple Features for SQL » (SFS)
- implémentation des normes de l'OGC en matière de webservices cartographiques (WMS)
- Normes de conception de site web du W3C

A noter que Plone respecte de nombreuses normes en vigueur en matière de développements et d'affichage. Le respect des « best practices » permet donc de respecter ces normes en vigueur sans surcoût de développement.

- Normes d'accessibilité : W3C
- Normes d'affichage : Accesweb, niveau Silver.
- Compatible avec tout les navigateurs css 2.1 + opéra.
- Séparation des développement / template : XHTML / CSS2

D'autres normes vont être prises en compte, dont la norme des métadonnées géographiques ISO19115 qui devrait être prochainement implémentée à l'aide de l'outil Geosource lui-même développé à partir de l'outil Geonetwork.

**Enumération des corpus d'application utilisés : intitulés et composition (textes, photos, documents audio-visuels, fonds composites...), en précisant les supports et la volumétrie.**

Le projet prévoit de tester la publication de documents d'archives très divers tels que des cartes anciennes, des plans cadastraux anciens, des photographies aériennes verticales, des photographies, des images satellitaires, des cartes postales, des documents scripturaux. MapN peut aussi gérer des objets mutimédias sonores (chants, hymnes,

enregistrements de témoignages) ou visuels (films, animations), etc., que le projet prendra en compte, sans en faire sa priorité.

Chacun des réseaux d'expérimentation s'appuie sur un corpus documentaire spécifique, qu'il pourra être envisageable mais pas obligatoire de mettre en relation. Les fonds en question sont numérisés dans leur quasi-totalité mais pas nécessairement géolocalisés, ce qui justifie pleinement l'usage d'un outil comme MapN

Pour l'étude géohistorique du Furan à Saint-Etienne, le fond principal est celui des Archives municipales de Saint-Etienne. Celles-ci ont déjà numérisé et mis en ligne sur leur site le cadastre napoléonien. Elles planifient en 2008 la numérisation de des plans généraux de la ville (plus de 400, à partir du XVIII<sup>e</sup> s. et du terrier de la seigneurie de Saint-Priest (1767-1773). Les Archives mettront ces documents en ligne sur leur site web et ils seront également à la disposition du projet en utilisant la plate-forme technique développée à cet effet.

L'équipe ISIG a développé sur la ville de Lyon plusieurs applications de SIG géohistoriques à différentes échelles répondant à des thématiques diverses (histoire architecturale et urbaine, urbanisme, analyse environnementale et paysagère, évolution du risque inondation ... Cela correspond à des dizaines de plans anciens numérisés et géoréférencés et à un fond photographique d'environ 7 000 diapositives couleur sur l'architecture et des formes urbaines de l'agglomération lyonnaise depuis 1977 à aujourd'hui, qui témoignent de l'évolution constante de cette grande agglomération. Ce fond sera complété pour le projet par d'autres fonds numérisés, tels que les collections de photographies de la Maison du Rhône (collection Dürrenmatt de 2 000 cartes postales anciennes (fin 19<sup>e</sup> – début 20<sup>e</sup>me), collection Rondeau de 4000 cartes postales) et ceux d'autres partenaires.

Le volumes de ces contenus n'a pas été précisément évalué, sachant que le principe est que des documents non encore répertoriés puissent être associés aux documents répertoriés. L'idée dans le projet est de distinguer ce qui est de l'ordre des tests techniques visant à calibrer un site grand public et ce qui relève de l'expérimentation sur les usages. Dans ce dernier cas, qui nous intéresse ici, il conviendra de définir les tailles minimales et optimales des fichiers à télécharger pour les utilisateurs de façon à permettre un bon fonctionnement des serveurs et limiter les risques d'une inefficacité du site en cas de chargement de fichiers trop lourds, tant du côté des serveurs que du côté des postes clients. Par ailleurs, les fichiers vidéos et sonores qui sont les plus volumineux, et dont le géoréférencement pose des problèmes spécifiques, ne sont pas ceux visés en priorité par le projet.

**6) Si l'outil ou le dispositif technique que vous proposez existe déjà sous une forme ou une autre, en quoi votre projet induit-il une nouveauté ou une amélioration probante ? Veuillez citer les noms des outils ou dispositifs techniques existants ainsi que des références utilisateurs**

Il existe plusieurs sites de consultation de documents d'archives géoréférencés en France ou à l'étranger. L'intense travail en cours de numérisation des archives va contribuer à les

multiplier. On peut citer ainsi :

- la consultation de la carte de Cassini géoréférencée : <http://cassini.ehess.fr/>
- la consultation de certaines collections de la BNF : [http://www.bnf.fr/visiterichelieu/collections/cp\\_cp.htm](http://www.bnf.fr/visiterichelieu/collections/cp_cp.htm)
- la consultation des fonds d'archives numérisées du Département de Loire-Atlantique : [http://www.cg44.fr/cg44/display.jsp?id=a27sr\\_15187](http://www.cg44.fr/cg44/display.jsp?id=a27sr_15187)
- la consultation sur fond cartographique de documents d'archives sur la ville de Toulouse : <http://urban-hist.toulouse.fr/>
- la consultation de photographies anciennes sur le littoral breton : <http://www.mairie-brest.fr/archives/photos-aeriennes-anciennes.htm>
- la consultation d'archives numérisées sur le département de la Mayenne : <http://www.lamayenne.fr/?SectionId=315>
- la consultation d'archives numérisées sur le département de l'Indre : <http://www.cg36.fr/jahia/Jahia/portail/archives/pid/5862>
- un site permettant le géoréférencement en ligne de documents planimétriques numérisés de type cartes et plans en utilisant le site GoogleMaps comme MapN <http://labs.metacarta.com/rectifier/>
- un site de consultation d'archives numérisées sur les villes historiques (Historic Cities Center of the Department of Geography, the Hebrew University of Jerusalem and the Jewish National and University Library) : [http://historic-cities.huji.ac.il/hc\\_general\\_links.html](http://historic-cities.huji.ac.il/hc_general_links.html)
- un site de consultation de cartes anciennes numérisées Odden Bookmarks (Map Library of the Faculty of GeoSciences, University of Utrecht) : <http://geoweb.venezia.sbn.it/geoweb2004/gwindex.html>
- un site de consultation de cartes anciennes numérisées de la Perry-Castañeda Library Map Collection de l'University of Texas at Austin : [http://www.lib.utexas.edu/maps/historical/history\\_europe.html](http://www.lib.utexas.edu/maps/historical/history_europe.html)
- un site de consultation de cartes anciennes numérisées de l'University of California Berkeley Library : <http://www.lib.berkeley.edu/EART/browse.html>
- un site de consultation de documents cartographiques et graphiques numérisés de l'Ufficio Carte Geografiche (Office des Cartes Géographiques) de la Biblioteca Nazionale Marciana di Venezia : <http://geoweb.venezia.sbn.it/geoweb2004/gwindex.html>
- le site de consultation de quelques unes des 4,5 millions de cartes de la collection des cartes de The Geography and Map Division of the Library of Congress : <http://memory.loc.gov/ammem/gmdhtml/gmdhome.html>
- la consultation de la collection de cartes anciennes de David Rumsey : <http://www.davidrumsey.com/>

**Quels sont à vos yeux les points faibles auxquels vous comptez précisément répondre par un développement innovant ?**

Parmi tous les sites mentionnés ci-dessus, aucun d'eux n'est consacré à la publication en

ligne de documents d'archives à référence spatiale. Il s'agit dans tous les cas de sites de consultation de ressources documentaires sans possibilité d'interagir sur ces ressources. Il s'agit de sites de dépôt d'archives numérisées et non pas de sites collaboratifs de publication d'archives numérisées et encore moins de sites offrant des services interactifs comme par exemple le géoréférencement de ressources documentaires. La plate-forme MapN et l'extension que le projet vise à lui donner, présente donc de réels avantages. Elle est destinée à une publication massive et facilitée de documents d'archives très divers et multicontenus. Nous visons un outil volontairement orienté vers la publication collaborative d'archives numériques.

**7) Citez les partenariats dont vous allez vous entourer pour réaliser l'outil ou le dispositif technique relevant d'institutions scientifiques (pour la partie technique d'une part, s'agissant des contenus d'autre part) : lesquels et pour quelles tâches ? Précisez les coopérations internationales s'il y a lieu. Prestataires externes, lesquels et pour quelles prestations ?**

En France, le projet mobilisera essentiellement les chercheurs de ISIG - UMR EVS et du LIAPPA, plus ceux qui s'impliqueront dans les réseaux d'usage. Le développement informatique s'intégrera dans les travaux du GDR Sigma-Cassini, dont font partie les deux équipes de recherche et contribuera à développer les liens entre centres de recherche et industriels, à travers le rapprochement SIGMA-AFIGEO, association professionnelle à laquelle appartient Makina Corpus.

A l'étranger une collaboration est prévue avec les chercheurs de l'Université de Londres (Sanjay Rana) et de l'École Polytechnique de Lausanne (François Golay) qui travaillent sur le thème de la géolocalisation des photographies en liaison avec ISIG.

Pour ce qui est des contenus, les Archives Municipales de la Loire, la Ville de Saint-Etienne et la maison du Fleuve Rhône ont donné un accord ferme pour participer au projet. D'autres partenaires, comme l'Agence d'urbanisme EPURES par exemple sont potentiellement intéressés à s'y joindre.

En prestation, la société Images et Créations s'occupera de l'hébergement du site

**8) Quelle organisation comptez-vous mettre en place pour mettre en œuvre la réalisation et selon quel calendrier ?**

La coordination du projet du point de vue scientifique sera réalisée par ISIG/UMR EVS, sous forme d'un coordinateur général et d'un responsable par réseau, accompagnés d'un chargé d'étude à mi-temps qui veillera au suivi du projet et à la mise en cohérence des deux réseaux en relation avec l'avancée du développement informatique. Ce chargé d'étude s'occupera aussi de l'appui méthodologique à chacun des réseaux. La chefferie du projet de développement logiciel sera assurée par Makina Corpus, qui animera tous les 3 mois un comité technique auquel participeront les différents partenaires scientifiques. Un comité technique sur deux fera un point sur l'état de l'utilisation par chacun des réseaux. Un séminaire final regroupera tous les utilisateurs, y compris ceux qui se seront agréés dans la deuxième phase au projet.

Un système d'observation des pratiques des utilisateurs dans les réseaux sera conçu en début de projet et mis en place en s'appuyant sur des travaux d'étudiants sous forme de stages. La valorisation scientifique se fera aussi bien dans les journaux et conférences informatiques et géomatiques, que dans ceux relevant des sciences de l'information et de la documentation et, bien sûr, des sciences sociales.

Le projet sera mis en oeuvre sur une période de 24 mois, selon le planning suivant :

*1ère phase:*

*Mois 0-3 :* Lancement du projet. Test de la plate-forme MapN par les partenaires des deux réseaux. Définition des spécifications de la nouvelle application et organisation des travaux de développement en deux phases.

*Mois 4-12 :* Développement informatique de la première phase, intégration de contenu sur la plate-forme originelle. Application simple et prise en main de l'outil par les utilisateurs. Bilan intermédiaire.

*2ème phase:*

*Mois 12-18 :* Maintenance des applications de la première phase et développement des applications de la seconde phase et des modules

*Mois 18-24 :* Mise en oeuvre des applications de la deuxième phase. Clôture et séminaire final du projet avec tous les partenaires. Discussion sur la mise à disposition des outils et données.

**9) Précisez le périmètre d'usage de l'outil ou du dispositif technique visé par votre projet, en dehors des besoins propres de votre laboratoire**

La plate-forme a vocation à concerner un très grand nombre de documents d'archives et un grand nombre d'usages potentiels tant dans le secteur scientifique que professionnel ou grand public. Son utilisation dans le projet concerne des données internes aux laboratoires impliqués mais aussi des données classiques des sites d'Archives partenaires du projet. Dans tous les cas, les questions abordées, qu'elles soient techniques ou méthodologiques, sont générales à la problématique du géoréférencement et du partage de données d'archives.

**10) L'outil ou le dispositif technique pourra-t-il de par ses fonctionnalités être utilisé dans une ou plusieurs autres communautés que celle à laquelle vous appartenez? communauté(s) scientifique(s)**

L'usage envisagé dans le projet relève de la géographie historique, avec différentes sous-thématiques (paysage, urbanisme, hydrologie). Il est évident qu'un tel outil trouve un intérêt dans nombre d'autres domaines de l'histoire, mais aussi en archéologie.

Plus largement l'outil pourra être réinvesti dans différentes communautés (de recherche ou de pratiques) ayant pour objectif de partager l'information géographique et de mutualiser des ressources géoréférencées.

**11) Disposerez-vous – ou votre institution de rattachement - des droits de propriété attachés à l'outil ou au dispositif technique, tel que défini plus haut ? Citez s'il y a lieu les autres ayants-droit au plan technique. Citez les autres ayants-droit éventuels en ce qui concerne les corpus d'application. Si les droits attachés restent à établir, veuillez préciser les éléments concernés, les titulaires potentiels et les démarches**

## **entreprises ou à entreprendre**

Pour l'architecture technique, Makina Corpus ne mettra en oeuvre que des logiciels libres pour lesquels aucune redevance n'est à acquitter. En outre, les logiciels libres utilisés sont diffusés sous licence GNU/GPL ce qui permet d'en étudier le code, puisque celui-ci est ouvert, de faire évoluer celui-ci et de le redistribuer avec les modifications apportées. En outre, Makina Corpus souhaite pouvoir reverser certains des développements réalisés pour MapN à des communautés de développeurs de certains des logiciels libres utilisés pour le projet.

Pour ce qui est des droits attachés aux documents d'archives (corpus d'application) trois cas de figure peuvent se présenter :

- les données d'archives numériques sont libres de droit : dans ce cas aucun droit n'est à acquitter; c'est d'ailleurs ce cas de figure que nous souhaitons encourager dans ce projet en invitant à publier sur le site des archives numériques et des contributions scientifiques originales sous licence Creative Commons;

- les données d'archives ont des droits réservés, qui sont levés temporairement pour un ensemble d'utilisateurs déterminés dans le cadre du projet : ce cas peut se produire pour certaines données appartenant à des partenaires, sites d'archives ou chercheurs, qui sont intéressés à tester le principe du partage permis par le site mais qui ne souhaitent pas s'engager sur une libre diffusion à la fin du projet. La possibilité de limiter par mot de passe l'accès au site à certains utilisateurs sera garantie, sachant que l'hypothèse est que les partenaires seront convaincus, à l'issue de l'expérience de l'intérêt de laisser un accès libre à une partie significative des données, sans aucune obligation.

- les données d'archives sont payantes : ce n'est pas le cas privilégié dans le projet même s'il paraît difficile d'exclure totalement ce cas de figure (NB : une ligne budgétaire est prévue à cet effet).

## **12) Comptez-vous participer à une communauté d'utilisateurs aux fins de retours d'expériences et de maintenance corrective ou évolutive, ou en créer une ? Laquelle ?**

C'est le principe même du projet que de susciter une communauté d'utilisateurs chargée d'aider à élaborer et préciser les spécifications de la plate-forme puis de la mettre en oeuvre de manière à la faire évoluer dans le temps même du projet.

Des synergies seront aussi recherchées avec des communautés existantes ou en cours de création dans le domaine des archives et de la publication de documents scientifiques dans les disciplines concernées par le projet et aussi dans le domaine du développement d'outils géomatiques participatifs. C'est un des atouts du projet de s'appuyer sur la plate-forme MapN, qui s'est donné comme vocation de devenir un site majeur de diffusion et de publication de données d'archives numériques à référence spatiale.

## **13) Sur ce plan, comment comptez-vous faire maintenir et faire évoluer l'outil ou le dispositif technique envisagé ? Indiquez les moyens techniques et financiers sur lesquels vous comptez vous appuyer.**

La plate-forme développée s'appuyant sur une architecture technique et des composants logiciels très évolutifs, il apparaît d'emblée aisé et nécessaire de faire évoluer le site pour tenir compte de l'évolution des besoins fonctionnels et des retours d'expériences ou encore de la prise en compte de nouvelles formes d'archives numériques. L'ajout de nouveaux modules et de développements spécifiques à Plone, à MapServer et à OpenLayers doivent permettre de répondre à toutes les évolutions nécessaires identifiées au cours du projet et après sa mise en production.

On peut envisager qu'une partie des services proposés par le site après le projet soient financés sur les revenus générés par le site si est mis en oeuvre un modèle économique de génération de revenus de type abonnement ou paiement à l'acte (par Webservice utilisé) ou par d'autres moyens. Cette question fera l'objet d'une étude et d'une discussion entre partenaires lors de la conclusion du projet.

### **Références citées:**

Hill L., 2006, Georeferencing, The geographic associations of Information, MIT Press, 260 pages.

### **Références de l'équipe**

Joliveau T., Calcagni Y. et Mayoud R. (2006), "Géowebexplorer, un outil géomatique collaboratif au service des enseignants et des élèves", Géographes associés, n° spécial. Savoir penser et partager l'information géographique : les SIG, n° 30, p. 169-178.

Gaio, M., Sallaberry, C., Etcheverry, P., Marquesuzaà, C., and Lesbegueries, J. « A Global Process to Access Documents' Contents from a Geographical Point of View ». Journal Of Visual Languages And Computing, 18(6), Special Issue on Spatial and Image-Based Information Systems - Elsevier, 25 pages, (In Press) December 2007.

Genevois S., Sanchez E. et Prieur M. (2006), Usages et enjeux de la géomatique dans l'enseignement, Observatoire de pratiques géomatiques (INRP) <http://eductice.inrp.fr/EducTice/projets/geomatique>

Rana S. et Joliveau T., 2007. Using Urban Viewsheds for Embedding Geographical Context in Photograph Databases of Urban Areas, in 2007 ESRI International User Conference, San Diego.

Sallaberry, C., Baziz, M., Lesbegueries, J., Gaio, M. « Towards an IE and IR system dealing with spatial information in digital libraries - Evaluation Case Study ». 9th International Conference on Enterprise Information Systems - Human-Computer Interaction Area / Geographical Information Systems, ICEIS 2007 Proceedings, Human-Computer Interaction Volume, INSTICC, pp.190-197, ISBN 978-972-8865-92-4

Sallaberry, C., Baziz, M., Lesbegueries, J., Gaio, M. « Une approche d'extraction et de recherche d'information spatiale dans les documents textuels - Evaluation ». Actes de la Conférence en Recherche d'Informations et Applications, pp. 53-64 , CORIA 2007

Sallaberry, C., Baziz, M., Lesbegueries, J., Gaio, M. « An IE and IR Approach to deal with Geographic Information Scope in Textual Documents ». Poster in the 8th Conference on Large-Scale Semantic Access to Content (Text, Image, Video and Sound), RIAO 2007.

Sallaberry, C., Gaio, M., Lesbegueries, J., and Loustau, P. « A Semantic Approach for Geospatial Information Extraction from Unstructured Documents ». Chapter in The Geospatial Web book, published by Springer in the Advanced Information and Knowledge Processing Series. ISBN 1-84628-826-6. pp. 93-105, 2007.