

Rapport semestriel d'activité n°X/Y

A. Identification

Programme – année	Blanc 2006
Projet (acronyme)	ALPAGE
Coordonnateur du projet (société/organisme - laboratoire ou entité de rattachement)	Hélène Noizet (LAMOP)
Référence convention/décision	ANR-06-BLAN-0099
Période du projet (date début – date fin)	Novembre 2006-novembre 2009
Période faisant l'objet du rapport d'activité (date début – date fin)	Novembre 2006-juin 2007
Rédacteur de ce rapport : nom	Hélène Noizet
téléphone	03 20 56 11 98
adresse électronique	Helene.noizet@univ-paris1.fr
Date du rapport	XX juillet 2007

B. Pour les projets multi-partenaires, rappel des tâches allouées par partenaire pour l'ensemble du projet (partir du planning généralement fourni dans le projet. Ce document est à remplir par le coordonnateur du projet à partir des informations fournies par les partenaires)

Ce tableau décrit les principales tâches du projet qui ont été définies lors de son démarrage

Tâches du projet	Partenaires concernés par la tâche (2)		2006		2007	2008	Commentaires
			Semestre 1 (1)	Semestre 2 (1)			
Obtention des images rasters Vasserot auprès des Archives de Paris	P1	Prévue/Nouvelle	P				Signature de conventions avec les Archives de Paris et les Archives Nationales
		Réalisée/Abandonnée		R			
Mise en place de la méthode de géoréférencement grâce à l'obtention des fonds de plans de référence	P2 et P1	Prévue/Nouvelle	P				Récupération de fonds de plans géoréférencés (APUR, Berty, atlas foncier 1930) et de plans papiers (plans d'expropriations) ; détermination de l'échelle de valeur de chacun de ses plans de référence
		Réalisée/Abandonnée		R			
Géoréférencement des rasters Vasserot	P2	Prévue/Nouvelle	P		Achever		Actuellement 120 plans d'îlots géoréférencés sur 912 au total
		Réalisée/Abandonnée		Commencé, non achevé			
Analyse de l'espace couleur des rasters Vasserot	P4	Prévue/Nouvelle	P				??
		Réalisée/Abandonnée		R			
Élaboration des métadonnées	P2 et P3	Prévue/Nouvelle	P		À confirmer		À la fois pour le géoréférencement des rasters Vasserot, mais aussi pour les futures couches d'information historique du SIG
		Réalisée/Abandonnée		R			
Définition des données historiques (tables et champs)	P1 et P3, aidés par P2	Prévue/Nouvelle	P		À tester		Travail sur les tables des enceintes, des édifices religieux, des censives, des hôtels aristocratiques, des paroisses, des marchés et des foires
		Réalisée/Abandonnée		R			

Tâches du projet	Partenaires concernés par la		2006		2007	2008	Commentaires
			Semestre 1 ⁽¹⁾	Semestre 2 ⁽¹⁾			
Obtention des images rasters Vasserot auprès des Archives de Paris Construction du modèle conceptuel de données	P1	Prévue/Nouvelle	P				Signature de conventions avec les Archives de Paris et les Archives Nationales
		Prévue/Nouvelle	P		À tester		Choix du MCD historique, à compléter pour aboutir à une couche sur les morceaux d'objets
	Réalisée/Abandonnée		R				
Choix des images sources pour la vectorisation des rasters	P4, aidés par P1, P2 et P3	Prévue/Nouvelle	P				Après de longs débats et des expérimentations pour tenter de réappliquer aux vecteurs les déformations issues du géoréférencement des rasters, il est finalement décidé que la vectorisation se fera à partir des images déjà géoréférencées, et non pas des images sources.
		Réalisée/Abandonnée		R			
Définition des besoins des historiens en matière de vectorisation	P1, P3 et P4	Prévue/Nouvelle	P				Par ordre de priorité décroissant, la vectorisation portera sur les parcelles, les espaces bâtis/non bâtis, les pièces à l'intérieur du bâti, et les murs et les ouvertures.
		Réalisée/Abandonnée		R			
Mise en place de la base de données sur le filaire des voies	P3, aidés par P2 et P1	Prévue/Nouvelle	P				Travail à partir du plan du CNRS de Paris au 14 ^e
		Réalisée/Abandonnée		R	À poursuivre		

Tâches du projet	Partenaires concernés par la		2006		2007	2008	Commentaires
			Semestre 1 ⁽¹⁾	Semestre 2 ⁽¹⁾			
Obtention des images rasters Vasserot auprès des Archives de Paris	P1	Prévue/Nouvelle	P				Signature de conventions avec les Archives de Paris et les Archives Nationales
Méthode de mise en relation de la base de données déjà existante sur les métiers avec le SIG	P1, P2 et P3	Prévue/Nouvelle	P		À tester		Liens avec les tables des circonscriptions et géocodage des foyers fiscaux par tronçon de rue
		Réalisée/Abandonnée		R			
Définition de l'architecture physique du SIG et de son administration	P1 et P3	Prévue/Nouvelle	P		À confirmer		Pour régler l'usage et l'accès au SIG lorsqu'il sera disponible au terme du programme
		Réalisée/Abandonnée		R			
Synthèse des connaissances actuelles sur les technologies de travail en ligne	P3	Prévue/Nouvelle	N		À poursuivre		Dans la perspective de travailler un jour en ligne sur le SIG (consultation, analyse par requête sémantique ou attributaire, édition de données nouvelles)
		Réalisée/Abandonnée		Commencée, non achevée			
Élaboration d'un accord de consortium	P4 et P1, P2, P3	Prévue/Nouvelle	N		À poursuivre		Pour déterminer la propriété de chaque partenaire sur les outils produits et les modes d'utilisation des données produites
		Réalisée/Abandonnée		Commencée, non achevée			

⁽¹⁾ P=Prévue, N=Nouvelle, R=Réalisée, A=Abandonnée (dans ce dernier cas, justifier obligatoirement avec un commentaire)

⁽²⁾ Partenaires concernés : ceux-ci sont à expliciter dans un tableau (cf. ci-dessous)

Partenaire 1	LAMOP (UMR 8589)	Hélène Noizet
Partenaire 2	OTELLO (JE 2433)	Frédéric Pouget
Partenaire 3	ARSCAN (UMR 7041)	Brigitte Boissavit-Camus
Partenaire 4	L3i (EA2118)	Jean-Marc Ogier

Eléments qualitatifs

C. Description des travaux effectués pour la période concernée et conformité de l'avancement aux prévisions (15 à 50 lignes maximum suivant le nombre de partenaires)

Les travaux se sont déroulés comme prévu sous la forme de réunions de travail de taille variable, restreinte ou large, et dont le lieu était alternativement Paris et La Rochelle. Entre septembre 2006 et juin 2007, 10 séances de travail se sont ainsi tenues, avec à chaque fois des questions précises à aborder, soit des questions nouvelles, soit la continuation des recherches entamées lors des précédentes réunions. Un certain nombre de thèmes étaient récurrents et ont été développés au fur et à mesure des séances, de plus en plus précisément, grâce notamment au travail mené, entre les réunions, par les 2 personnes recrutées par le programme. Ces thèmes concernent le géoréférencement et la vectorisation des images rasters Vasserot, ainsi que le modèle conceptuel de données du SIG. Les réunions comprenaient parfois un public disciplinairement homogène (historiens et archéologues notamment pour avancer sur l'organisation des données historiques), mais le plus souvent, celles-ci rassemblaient des personnes des 3 grandes disciplines concernées par le programme, c'est-à-dire à la fois des historiens, des géographes et des informaticiens. L'intérêt principal est de faire comprendre aux uns et autres les besoins et les spécificités de chacun. A chaque fois, des comptes-rendus de séances conséquents ont été rédigés par la coordinatrice pour conserver la mémoire du travail réalisé, pour servir de point d'appui aux réunions suivantes et enfin pour communiquer avec les institutions partenaires qui soutiennent le programme d'une manière ou d'une autre (APUR, STDF, Archives de Paris, Archives nationales...).

Durant ce premier semestre, qui visait à fixer les besoins et les méthodes de travail, le déroulement du travail s'est effectué conformément au calendrier prévu. Il est apparu cependant que le géoréférencement des rasters prendrait plus de temps que prévu.

D. Résultats obtenus pour la période concernée, dégager notamment les faits

marquants (15 à 50 lignes maximum) *Décrire les résultats obtenus et préciser éventuellement les livrables déjà réalisés en interne au projet.*

- définition des protocoles de travail entre géographes, informaticiens et historiens :

tout d'abord, la méthode de géoréférencement des rasters Vasserot a été mise en place par les géographes grâce à l'obtention, par les historiens, de fonds de plans de référence, dont 3 sont disponibles en version numérique et géoréférencée, c'est-à-dire qu'ils sont tous superposables (le parcellaire actuel, l'atlas foncier avec le parcellaire 1930, le plan de Berty). Une évaluation des avantages et inconvénients de chacun d'eux a été réalisée afin d'utiliser, selon les plans d'ilot, celui qui produira le géoréférencement le plus satisfaisant aux yeux des historiens. Enfin, la BHVP nous permet d'obtenir des tirages papier des grands plans d'expropriation du 19^e s. qui superposent l'ancien parcellaire remanié et le projet de percement de voie ou d'alignement des îlots : la largeur des voies sur ces plans est mesurée puis reportée dans le logiciel de géoréférencement permettant ensuite de mieux placer les îlots Vasserot. Puis, le protocole de travail entre géographes et informaticiens a été établi pour l'interaction des opérations de géoréférencement et de vectorisation des rasters. Étant donné que les rasters doivent être géoréférencés et qu'ils sont déformés par cette opération, la question se posait de savoir à partir de quelles images la vectorisation sera faite, les images déjà géoréférencées (de moins bonne qualité mais bien positionnées) ou les images sources (de meilleure qualité mais à repositionner). Comme il n'est pas envisageable de faire un 2^{ème} géoréférencement pour les vecteurs, il a été étudié la possibilité de conserver la mémoire des transformations appliquées aux rasters par le géoréférencement, pour les réappliquer ensuite aux vecteurs. Mais cette possibilité n'ayant pas donné satisfaction, il a été finalement décidé que la vectorisation se fera à partir des images déjà géoréférencées. En revanche, afin d'obtenir un

résultat final fiable, la vectorisation du parcellaire produite par les informaticiens sera validée par les historiens. Afin de faciliter ce travail de validation, les informaticiens proposent de guider les historiens vers les zones qui poseront le plus de problèmes à l'aide d'un outil permettant de naviguer facilement au sein des images vecteurs.

- l'analyse des caractéristiques des rasters Vasserot en vue de la vectorisation :

la vectorisation du Parcellaire s'inscrit dans une démarche complète de rétro conversion de document graphique, soit le passage d'un document papier à son interprétation sémantique. Dans notre contexte, nous possédons les plans déjà numérisés, des images brutes composées de pixels. Tout l'enjeu de la vectorisation est alors de donner du sens à ces images qui ne sont à ce stade que des amas de données inertes et de formaliser cette connaissance sémantique sous forme vectorielle. Dans cet objectif, il est nécessaire de mettre en exergue l'information contenue dans les images de plans cadastraux. Nos premières expériences ont montré un affadissement des plans lié au temps. Les pigments de la peinture se sont altérés. Afin de pallier ce problème et de revenir vers la couleur originelle du document, une stratégie de restauration d'image a été mise en place. L'information couleur prend une place prépondérante dans les plans cadastraux de Vasserot car elle caractérise l'entité parcelle. En effet, à une parcelle est associée une couleur. Les techniques de segmentation couleur visent à localiser des ensembles chromatiques afin d'identifier les parcelles présentes sur un plan. L'analyse des espaces couleurs, nous a mené à calculer un modèle de représentation dédié aux images très spécifiques que sont les plans cadastraux de Vasserot. Leurs spécificités sont caractérisées d'une part par une répartition des pixels le long de la diagonale des gris dans le cube Rouge Vert Bleu, et d'autre part une concentration importante des pixels dans une sphère de centre 220,220,220 et de rayon 30. La qualité d'un espace couleur est déterminée par sa capacité de distinction entre les différentes couleurs ainsi que la robustesse aux variations au sein d'une même gamme chromatique. La résultante de cette modélisation est un espace géométrique de représentation de la couleur, un cylindre d'équation $x=y=z$ et de rayon 30 dans le cube RVB.

- le choix du modèle conceptuel de données du SIG : suite aux besoins exprimés par les historiens, les géographes ont proposé d'utiliser le modèle conceptuel de données par "historique et information différentielle", qui introduit le concept de successeur. Ce modèle a été pensé en interne à l'IGN dans l'objectif de gérer les évolutions de mise à jour pour la BD-CARTO. Chaque objet géographique est décrit dans sa totalité et représente une entité indépendante. Un lien existe entre les objets. Celui-ci décrit sous la forme de deux champs l'objet successeur et l'objet qui le précède. Un objet géographique est un élément qui ne subit aucun changement, ni sémantique ni géométrique sur une période donnée. Cet objet est stocké à l'aide de champs couples début/fin. Enfin, le modèle semble relativement évolutif car il est possible d'envisager pour les historiens qui le souhaitent de basculer vers un modèle topologique "entités spatiales/fonctionnelles" (informatisation des DEPAVF par le CNAU, travail de F. Pirot à Tours...).

- la première formulation du modèle logique de données : en vue de constituer des couches d'information dans le SIG, les caractéristiques géométriques des données historiques suivantes ont été déterminées (point, ligne, surface), et leurs données attributaires ont été formalisées sous formes de champs : enceintes, hôtels aristocratiques, édifices religieux, paroisses, quêtes, censives, marchés et foires. De même, certaines questions que voudront poser les historiens à ces données grâce au SIG ont été formalisées sous forme de requêtes.

E. Difficultés rencontrées et solutions de remplacement envisagées (15 à 50 lignes

maximum) *ex : impasse technique, abandon d'un partenaire ou d'un sous traitant, maîtrise des délais, maîtrise des budgets. Faut-il revoir le contenu du projet ? Faut-il revoir le calendrier du projet ?*

- retrait partiel ou total de certains historiens (P. Benoit et J. Rouillard d'une part, K. Berthier d'autre part) : les couches historiques du SIG sur les équipements hydrauliques dont ils avaient la charge ne seront donc pas construites par ces personnes, même si P. Benoit et J. Rouillard restent des personnes-ressources pouvant aider les personnes qui feront le travail. En accord avec ces derniers, il a été défini et proposé des sujets de masters à des étudiants en histoire médiévale pour réaliser l'intégration de ces données dans le SIG. L'intégration de ces données dépendra donc de la participation des étudiants de master d'histoire.

F. Livrables externes réalisés (15 à 50 lignes maximum)

Pour les articles et communications écrites, préciser s'il s'agit d'articles dans des revues à comité de lecture / d'ouvrages ou chapitres d'ouvrage / d'articles dans d'autres revues / de communications dans des colloques ou des congrès / de dépôt de brevet... Référencer selon les normes habituelles. Mentionner également s'ils peuvent ou non faire l'objet de communications externes par l'ANR et son unité support

Indiquer, *Le cas échéant*, les thèses démarrées, en cours et/ou soutenues en relation directe avec le projet :

Préciser le titre, date de soutenance (prévue ou réelle), soutien financier, devenir des étudiants pour les thèses soutenues

- les comptes-rendus des 8 séances de travail, qui permettent de mesurer l'intensité des échanges scientifiques entre des disciplines différentes, sont accessibles en ligne sur le site du LAMOP : <http://lamop.univ-paris1.fr/lamop/LAMOP/lamopII.Alpage.htm>

Il ne s'agit pas encore de publications en tant que telles, mais le contenu scientifique n'en est pas moins réel. Le travail qui résulte de ces rapports donnera lieu prochainement à des présentations lors de colloques (colloque ATHIS à Rome en décembre prochain).

- Le sujet de thèse de Romain Raveaux se positionne dans le cadre d'une valorisation du patrimoine Français, des plans cadastraux de la ville de Paris issus de l'atlas VASSEROT du 19ème siècle. Afin de rendre exploitable l'information spatiale contenue dans les documents cadastraux, il est nécessaire d'individualiser informatiquement des objets métiers en l'occurrence les rues, les îlots, les parcelles pour ensuite insérer ces éléments dans un système d'information géographique (SIG). Ces étapes d'extraction de connaissance s'appuient sur les avancées de domaines scientifiques tels que la reconnaissance de formes, la vectorisation d'objets linéaires, le traitement d'images couleur : les espaces de représentation, l'analyse d'histogrammes couleur 3D, les invariants colorimétriques et les signatures dynamiques pour le suivi spatial et/ ou temporel d'objets.

G. Autres commentaires

Eléments quantitatifs

H. Liste des réunions/séminaires/colloques organisés durant la période et des missions à l'étranger

(préciser la date, le lieu, l'objet, le nombre des participants)

1. 1^{er} septembre 2006 à Paris : 1^{ère} réunion générale de présentation des personnes et du programme (15 participants)
2. 19 octobre 2006 à La Rochelle : réunion géographes/informaticiens/historiens pour mettre en place les méthodes de travail et les besoins (13 participants)
3. 27 novembre 2006 à Paris : réunion des historiens et archéologues pour déterminer les données attributaires des objets historiques et présenter les types de modèle conceptuel de données (9 participants)
4. 21 décembre 2006 à La Rochelle : réunion géographes/informaticiens/historiens pour faire le point sur les besoins liés au géoréférencement et sur l'analyse de l'espace couleur (7 participants)
5. 18 janvier 2007 : réunion des historiens et archéologues pour la répartition des données historiques à constituer dans le SIG (9 participants)
6. 25 janvier 2007 à La Rochelle : réunion géographes/informaticiens/historiens sur le géoréférencement et les premiers essais de segmentation (16 participants)
7. 21 février 2007 à Paris : mise au point sur le géoréférencement et le MCD (3 participants)
8. 26 mars 2007 à Paris : réunion générale géographes/informaticiens/historiens pour explorer les possibilités de réapplication des déformations dues au géoréférencement (17 participants)
9. 24 mai 2007 à La Rochelle : réunion générale géographes/informaticiens/historiens pour fixer précisément les méthodes de géoréférencement et de vectorisation ainsi que le MCD (22 participants)
10. 4 juin 2007 à Paris : construction du filaire des voies et harmonisation des bases S. Robert et de C. Bourlet (4 participants)

I. Par rubrique et par partenaire, établir la consommation des dépenses financées par l'ANR, depuis le démarrage du projet.

Partenaire	Fonct. (Keuros)	Equip. Nature	Equip. (Keuros)
Lamop (+ ArScAn)	2,62	Imprimante et base de données	2,72
Otelo	1,49	Logiciels et ordinateur	3,9
L3i	3,17	Logiciels et ordinateur	2,1
Total projet	7,28		8,72

Rubriques détaillées pour le Lamop :

achat bases de données (IGN)	1899
missions	1187,12
matériel informatique (imprimante)	830,54
prestation de services (formation)	850
papeterie	355,22
réception	149
poste	87,07
Total	5357,95

Rubriques détaillées pour OTELO

Frais de déplacement - missions	1140
Repas lors de réunions	250,10
Petits logiciels	72,73
Equipement informatique (ordinateur)	1450,79
Logiciel ArcInfo	2451,80
Personnel (CDD)	7826,20
Total	13191,62

Rubriques détaillées pour L3I

Frais de déplacement - missions	2536,8
Repas lors de réunions	634,2
Ordinateur + logiciels	2100
Personnel (CDD)	16324,05
Total	21895,05

J. Le cas échéant et pour les programmes thématiques, préciser les travaux réalisés par les partenaires étrangers associés au projet sans aide de l'ANR

Nota : sans objet pour les programmes « Blanc » et « JCJC »

K. Liste des personnels recrutés en CDD par des établissements publics dans le cadre du projet sur l'aide allouée par l'ANR

Nom	Prénom	Qualifications	Date de recrutement	Durée du contrat (en mois)
Bethe	Anne-Laure	Ingénieur d'études 2 ^e classe	1 ^{er} janvier 2007	31 mois
Raveaux	Romain	Ingénieur d'études 2 ^e classe	1 ^{er} Novembre 2006	36 mois

Indiquer leur devenir postérieur à leur participation au projet : intégration comme chercheur, enseignant-chercheur, ingénieur, emploi dans le privé, chômeur, etc....

- Anne-Laure Bethe a été recrutée comme ingénieur d'études 2^{ème} classe afin de réaliser le géoréférencement des images Vasserot. A l'issue du projet elle sera soit demandeur d'emploi soit ingénieur dans la fonction publique (admise à un concours en 2006).
- Romain Raveaux a été recruté comme ingénieur d'études 2^{ème} classe afin de réaliser la vectorisation des images Vasserot.

L. Le cas échéant, indiquer les différents types d'aides complémentaires obtenues grâce à ce projet.

(Il peut s'agir de ressources financières, ressources humaines, allocations de recherche,...)

M. Le cas échéant, modalités d'utilisation du complément de financement « pôles de compétitivité » (15 lignes maximum)

Rappel : ceci ne s'applique pas aux entreprises, mais seulement aux laboratoires publics et autres structures non soumises à l'encadrement communautaire des aides d'Etat à la R&D. Le complément de financement est destiné à couvrir des frais supplémentaires liés à la participation aux activités du pôle : ingénierie de projets partenariaux publics-privés, recherche de partenaires ; valorisation de la recherche ; relations inter-pôles et internationales...

N. CADRE RESERVE AU COORDONNATEUR DU PROJET (15 à 50 lignes maximum)

Commentaire général sur l'état d'avancement du projet, les interactions entre les différents partenaires, les efforts particuliers en matière d'interdisciplinarité, l'ouverture internationale, etc.

L'état d'avancement du programme est très encourageant car, au terme de ce premier semestre, nous avons réussi à mettre en place les protocoles de travail nécessaires aux différentes disciplines concernées (histoire, informatique, géographie). Cela s'est traduit par l'intégration des besoins, des contraintes et des possibilités de chaque discipline par les 2 autres, et, au-delà, par la transformation des pratiques disciplinaires habituelles. Par exemple, le fait que les informaticiens acceptent de vectoriser des images de seconde main, dégradées par le géoréférencement, et non pas les images sources comme ils le font habituellement, est certainement une petite révolution qui résulte de la prise en compte de la contrainte géographique de géoréférencer les rasters. De même, former des historiens médiévistes au SIG pour qu'ils puissent utiliser cet outil par la suite de manière autonome reste assez peu fréquent dans le monde de la médiévistique, cette nouveauté les obligeant à faire le grand écart du pixel à la charte du 12^e s.

Nous formons désormais un petit groupe d'une vingtaine de personnes qui participent activement au débat comme le montre l'évolution des effectifs des participants aux réunions et les discussions scientifiques lors de ces journées de travail (cf comptes-rendus). Au-delà de ce groupe, se sont également formés des collaborations de travail plus restreintes, par exemple le groupe sur le filaire des voies (C. Bourlet, S. Robert et F. Pouget) ou encore celui sur le travail en ligne (L. Costa, A. Dallo, P. Chareille).

La synergie SHS/STIC sera pleinement mise à profit dans la suite du programme, lorsque nous ferons le point sur nos besoins technologiques en informatique, par exemple pour le travail en ligne, ou pour l'obtention de cartes de densité spatio-temporelle avec des morceaux d'objets : s'il est nécessaire de fabriquer des outils ad hoc, nous déterminerons avec les informaticiens des sujets de licence ou de masters pour leurs étudiants visant à répondre à nos besoins en développement, et qui pourront être intégrés dans le parcours de formation de ces étudiants.

Enfin, il y a des développements de notre travail au-delà de ce programme, puisque les données sur l'évolution de l'espace parisien qui seront produites par ce programme intéressent d'autres chercheurs informaticiens qui travaillent sur l'espace-temps et qui pourront utiliser les données parisiennes comme matière de base pour alimenter leur process (P. Boursier).

Une collaboration étroite est née entre le L3I et le laboratoire CVC (Centre de visio per computador) de Barcelone autour du projet ALPAGE. Cette coopération s'est matérialisée par un atelier de travail (Workshop) d'une semaine autour de la reconnaissance d'objet graphique dans des images de documents ainsi qu'un co-encadrement de stage L3I-CVC. Stage de 6 mois d'Alexandre Hamon Master 2^{ème} année.

CADRE RESERVE A l'USAR

Nom du coordinateur scientifique de l'USAR :

Date :

Glossaire

Livrable : tout composant matérialisant le résultat de la prestation de réalisation. Toute production émise par le titulaire au cours du projet : document, courrier revêtant un caractère officiel , module de code logiciel, dossiers de tests, application intégrée, objet, dispositif...

Livrable interne : réalisé au sein du programme et non communiqué à l'extérieur du programme.

Livrable externe : élément diffusé ou livré hors de la communauté du projet de recherche..

Faits marquants : élément non nécessairement quantifiable mais significatif pour le projet.