

Bilan scientifique 2013 du consortium archiPolis.

S.Duchesne, ISP Nanterre. Coordination scientifique.

S.Cadorel, CDSP. Coordination opérationnelle.

6 décembre 2013

Avant de présenter le bilan scientifique de la première année d'activité d'archiPolis – l'archive des sciences sociales du politique -, il nous semble important de rappeler une de ses spécificités par rapport à d'autres consortia HUMA-NUM : dans notre cas, les corpus ne préexistent pas à l'intervention du consortium : avant de songer à les numériser, il nous faut repérer les enquêtes dont les données n'ont pas été détruites, convaincre leurs auteurs de nous les confier et rassembler les documents. De fait, les documents qui constituent les enquêtes (transcriptions, notes et carnet, photos, enregistrements audio et video, documents administratifs, projets de recherche, rapports d'activités et notes d'avancement, littérature grise, etc.) sont la plupart du temps en possession des chercheurs qui en disposent librement. Leur participation est donc indispensable pour la conduite de toute opération relative au projet. Numériser ces documents pour les archiver suppose au préalable d'en repérer l'existence, ou on pourrait dire, la persistance ! De plus, archiver ces enquêtes et ouvrir la possibilité que les collègues puissent voir de près de quoi elles sont constituées, comment elles ont été conçues et mises en œuvre ne va pas de soi : les chercheurs peuvent arguer de raisons épistémologiques aussi bien qu'éthiques pour justifier le refus d'archiver leurs enquêtes –c'est-à-dire, quand ils les ont conservées, puisque le principal obstacle à notre entreprise demeure la non conservation des documents. archiPolis s'efforce d'apporter des réponses pour limiter ces réserves. En particulier, le consortium a adopté, en lien avec le projet scientifique de l'équipe beQuali (www.bequali.fr), la notion d'enquête comme objet de l'archivage. archiPolis s'efforce de collecter des enquêtes plutôt que des données à archiver, c'est-à-dire

qu'il porte une attention particulière à numériser en même temps que les données les documents qui permettent de comprendre comment ces données ont été construites.

Parce que les enquêtes à archiver ne sont pas déjà disponibles, archiPolis n'a pas encore été amené à numériser des projets. Cette première année de fonctionnement aura permis au consortium de consolider le partenariat sur lequel a été créé, de développer une méthodologie commune dans les 6 équipes qui le composent en matière d'inventaire et de collecte des enquêtes, de mettre en place une organisation stable pour réaliser ce travail d'inventaire et surtout de collecte, de se doter d'un programme d'action lequel consiste notamment à développer en parallèle l'archivage des enquêtes anciennes – en tous cas, celles dont les documents ont été préservés et que les chercheurs acceptent d'archiver – et le développement d'une plate-forme d'auto-archivage qui permettra pour les enquêtes à venir que le stockage des documents se fasse au cours de la recherche et non ex post. Enfin, un travail de mise à plat des différents systèmes de métadonnées, utilisés par les différents réseaux avec lesquels nous voulons garantir l'interopérabilité de nos données, a été engagé.

1. Consolider le consortium :

La consolidation du consortium a été menée dans plusieurs directions. Tout d'abord, en créant des occasions pour que les différents partenaires se rencontrent. Le comité de pilotage s'est réuni à deux reprises : le 27 février à Paris, puis le 14 novembre à Lyon, dans le cadre de deux journées de travail qui ont réuni le consortium archiPolis, les membres de l'ANR réanalyse et l'équipe beQuali. Les deux comptes-rendus de ces journées sont annexés à ce bilan. La deuxième réunion, en particulier, a été organisée sur deux jours de façon à ce que chacun puisse vraiment faire connaissance. À ces réunions s'ajoutent des visites de la coordinatrice auprès des partenaires : le CERAPS et le Centre Emile Durkheim en particulier, afin de présenter le projet dans ces laboratoires.

Au-delà du consortium, les journées de Lyon avaient aussi pour objectif de consolider l'architecture complexe dans laquelle s'insère ce travail d'inventaire et d'archivage des enquêtes : il est pour nous inséparable d'un objectif plus scientifique que patrimonial, celui du partage des enquêtes. archiPolis s'articule donc avec deux autres projets : l'équipex DIME-SHS où nous construisons un outil de mise à disposition des enquêtes ; le projet ANR réAnalyse, dans lequel nous expérimentons l'analyse secondaire de ce type d'enquêtes et nous nous attachons à préciser les conditions d'archivage et de mise à disposition qui favoriseront une réutilisation heuristique des données (représentation graphique de cette organisation ci-dessous).



Au fil de l'année, au fur et à mesure de la construction de ce partenariat complexe, le consortium lui-même s'est un peu étoffé, essentiellement au sein de Sciences Po Paris. En plus du CEE et du CDSP, deux autres unités ont formellement rejoint le consortium. Le projet de consortium prévoit que le partenariat grandisse au fur et à mesure qu'avancera l'inventaire, la numérisation et l'archivage des enquêtes passées. Mais la proximité des équipes de Sciences Po et les ressources dont elles disposent ont contribué à faire entrer le CEVIPOF et l'OSC dès à présent dans le consortium.

Enfin, le consortium s'est surtout développé par l'élargissement des contributions au sein de chacune des équipes. Les participants directs sont aujourd'hui nettement plus nombreux qu'au moment de la signature du projet :

- ISP (Nanterre) : Sophie Duchesne, Aleksandra Mikanovic, Catherine Guaspere, Julie Testi,

- Sciences Po (Paris) : Anne Both, Sarah Cadorel, Guillaume Garcia, Odile Gaultier-Voituriez, Nonna Mayer, Julien Rault, Bernard Corminboeuf, Goulven Le Brech,
- Ceraps (Lille) : Thomas Soubiran,
- Pacte (Grenoble) : Stéphanie Abrial, Céline Belot, Bernard Denni, Annie-Claude Salomon,
- Triangle (Lyon) : Camille Hamidi, Cécile Laube, Samantha Saidi, Valérie Sala Pala, Thibaut Riouffreyt,
- Centre Emile Durkheim (Bordeaux) : Florence Hallou, Caroline Sagat, Viviane Le Hay.

2. Organiser l'inventaire et la collecte

La première étape du travail pour archiPolis consiste donc à inventorier, dans les équipes partenaires, les enquêtes de sociologie et de science politique¹ utilisant des méthodes d'investigation qualitative et qui ont été effectuées depuis la création de ces équipes. Les premiers échanges donc été consacrés à élaborer une grille d'inventaire et à échanger autour des différentes méthodes pour la remplir. Deux types de sources sont utilisables pour un premier repérage : les rapports d'activités des laboratoires d'un côté, les pages internet des chercheurs de l'autre. En sachant que la grille définitive requiert d'interviewer les chercheurs concernés puisqu'il ne s'agit pas seulement d'inventorier ce qui a été fait mais aussi, d'identifier voire localiser ce qui existe encore. Les premiers inventaires au CERAPS, à PACTE, au CEE et à Triangle ont été réalisés grâce au financement de contrats à court terme par archiPolis. Au CEVIPOF, à l'OSC et au Centre Emile Durkheim, nous avons eu la chance de pouvoir nous appuyer sur des personnels permanents (de statuts divers :

¹ Cette délimitation disciplinaire résulte de l'organisation disciplinaire des consortia et de notre volonté d'éviter à tout prix, dans ce contexte de ressources rares, une guerre des enquêtes que peut laisser augurer les disputes observées autour des archives de grands sociologues.

Sciences Po, CNRS et Éducation nationale). En conclusion de cette première année, il apparaît comme absolument nécessaire de mieux homogénéiser la construction de ces inventaires dont la compilation se révèle, à ce stade, assez difficile. Or un des premiers objectifs d'archiPolis est bien de pouvoir produire un inventaire des enquêtes disponibles (par différence avec d'autres listes d'enquêtes dont celle mise en ligne par la MSH de Dijon mais qui ne permet en rien de savoir si les enquêtes recensées ont été conservées, et où). Dans un premier temps, la grille Excel qui sert à l'inventaire va être complétée pour une partie des champs par des menus déroulants. En parallèle, nous développons un outil informatique de saisie des inventaires avec un masque de saisie le plus complet possible qui permettra de construire directement la liste compilée de tous les labos partenaires.

Quoiqu'il en soit, les inventaires réalisés ont permis de commencer à localiser quelques enquêtes à archiver et donc d'abord, à collecter. Mais il s'est rapidement révélé impossible de confier la collecte à des personnels temporaires dès lors que ce travail dépend d'abord et avant tout de la participation des auteurs des enquêtes, chercheurs et enseignants chercheurs la plupart du temps, et donc de leur disponibilité (sans même parler de leur bonne volonté). Comme indiqué ci-dessus, des solutions ont pu être mises en place sur certains sites mais d'autres devront être trouvées auprès des autres partenaires. Une des décisions du deuxième comité de pilotage est d'assortir la participation des laboratoires au consortium à la recherche de personnels local (à recherche essentiellement du côté des centres de documentation et bibliothèques) qui puissent consacrer un peu de leur temps au projet. Moyennant quoi, la coordination opérationnelle va être resserrée, Sarah Cadorel prenant dorénavant la responsabilité d'animer le groupe en charge de la collecte des documents.

3. Créer une plate-forme d'auto-archivage.

Les différentes interventions dans les laboratoires partenaires pour sensibiliser les chercheurs au projet d'archiPolis ont rapidement fait apparaître la dualité de l'intérêt que ceux-ci trouvent à l'archivage de leurs enquêtes². S'ils peuvent se laisser convaincre de l'intérêt qu'il y aurait à partager les enquêtes et donc à voir leur travail réutilisé par d'autres, les chercheurs sont encore plus directement sensibles à l'intérêt qu'il y aurait à publier les matériaux d'enquête en même temps que les résultats. C'est d'ailleurs ce que nous avons fait pour l'enquête prototype de la banque que nous développons. Dès lors, il n'est pas possible d'archiver ex post : il faut prévoir de procéder à l'archivage en même temps qu'on avance dans l'enquête. Processus qui aurait sans doute l'avantage d'éviter que des documents se perdent. Mais pour conserver la logique qui est la nôtre – celle d'un archivage piloté scientifiquement au sens où la sélection des documents conservés répond à une exigence de contextualisation justifiée en termes épistémologiques et méthodologiques – il importe que cet auto-archivage soit encadré par un outil qui accompagne le chercheur dans la sélection des documents à conserver – et non, comme dans le cas du *repository* hollandais, DANS ou du système développé à Harvard, Dataverse, des lieux de stockage à plus ou moins long terme dans lequel chacun verse ce qu'il veut. Cette plateforme d'auto-archivage permettrait par ailleurs de répondre à des besoins techniques dans le développement des outils d'archivage et de mise à disposition que nous développons. De manière à optimiser la collecte des documents électroniques et plus largement l'ensemble de leur traitement, notamment en sécurisant et en structurant mieux le processus, il nous cette plate-forme d'auto-archivage pourrait à terme servir de back-office pour la version suivante de la banque enQuêtes. Ce nouveau produit informatique doit pouvoir d'une part ingérer

² Nous laissons de côté d'autres motivations tout aussi importantes comme faire de la place dans leur bureau, organiser leurs archives, rendre leurs travaux plus visibles, etc.

les informations descriptives des enquêtes qui sont non seulement les inventaires cités ci-dessus mais aussi tout un ensemble de métadonnées descriptives et techniques. D'autre part il doit servir d'outil permettant aux chercheurs de déposer eux-mêmes les documents des enquêtes en étant guidés dans la sélection des documents à déposer pour la conservation ou la mise à disposition.

Un cahier des charges a été élaboré pour le *repository* et Alexandre Aazzouz, informaticien au CDSP, a commencé une phase de veille et de test des outils existants, préliminaire au choix de la solution logicielle. Nous commençons par regarder s'il est possible de reprendre et d'adapter une solution déjà existante et utilisée par d'autres avant d'opter pour une solution « maison » entièrement développée par nous. Le premier outil testé, Alfresco (outil de gestion documentaire), s'est révélé riche en fonctionnalités mais peu paramétrable. Fedora Commons (dépôt de fichiers et de métadonnées, outil très utilisé notamment par beaucoup d'institutions dont TGE-Adonis, labo néerlandais, interface Hydra Project, etc.) est en cours de test et sera suivi par Data Verse, E-Prints et Silo (développé par Human-Num). Si aucun des outils existants ne correspond à ce que nous cherchons, il faudra construire notre propre solution à partir de morceaux d'outils développés à Sciences Po.

En termes de calendrier, l'objectif serait de pouvoir choisir notre solution en janvier afin que les tous premiers tests puissent commencer à la fin du printemps 2014. Toutes les solutions testées sont des logiciels libres. Le coût n'est donc pas dans le prix du logiciel mais dans le temps et la main d'œuvre pour l'adapter aux besoins.

4. Gérer les métadonnées

En prolongement de la rédaction du cahier des charges pour le *repository*, un travail de mise à plat des métadonnées a été engagé. En effet, les différents travaux relatifs à l'inventaire, à la mise à disposition et à l'archivage pérenne des enquêtes, ont conduit à une utilisation assez hétérogène des informations de description. Si les

systèmes liés à ces différentes finalités doivent communiquer, il est indispensable que les informations qui circulent de l'un à l'autre soient structurées. D'autre part nos différents partenariats nous obligent à garantir l'interopérabilité avec leurs infrastructures. C'est le cas par exemple des services d'archives avec lesquels nous devons être en mesure de dialoguer via l'utilisation de l'EAD, avec le réseau Quetelet dans lequel nous devons être référencés en utilisant la norme DDI, avec le CINES pour lequel nous devons être en mesure de générer le SIP etc.

Nous avons donc entrepris d'établir un tableau de correspondance entre les différents standards de métadonnées qui nous sont utiles de manière à pouvoir passer nos informations d'un système à l'autre. Avec ce modèle de données, notre système interne doit être en mesure d'importer et d'exporter des informations qui suivent ces différents standards. De plus il nous a permis de corriger et d'uniformiser les métadonnées utilisées dans l'ensemble de notre chaîne de traitement des enquêtes. Ainsi nous sommes en train de restructurer nos grilles d'inventaires en suivant mieux les standards et les vocabulaires contrôlés. Ceci afin de fluidifier leur exploitation pour d'une part la mise à disposition et d'autre part l'archivage pérenne. Cela nous a notamment servi de base pour le développement d'un formulaire de saisie qui pourra être utilisé par les personnes chargées de l'inventaire et de la collecte mais aussi par les chercheurs eux-mêmes pour le dépôt des enquêtes dans le *repository*.

Conclusion :

On trouvera dans les deux comptes-rendus des comités de pilotage annexés à ce rapport d'activités plus de détails sur la façon dont nous avons avancé au cours de cette première année de travail. L'impression que nous en avons est très positive. Le choix d'avancer sur deux jambes – inventorier et collecter les enquêtes passées tout en développant un outil qui permettra d'archiver en temps réel les enquêtes à venir – a bien sûr pour conséquence un certain ralentissement du projet d'origine, celui de la

numérisation des enquêtes anciennes. Mais cela nous semble une meilleure solution d'autant que cela permet d'intéresser au projet, parallèlement, deux générations : celles des chercheurs prêts à partir à la retraite et qui se déclarent les plus disposés à nous confier le fruit de leur travail ; et les jeunes chercheurs qui voient dans la plate-forme d'archivage une façon d'organiser autrement leur travail.

Pour l'année à venir, nous comptons une douzaine d'enquêtes à collecter et numériser :

- La boutique contre la gauche. N. Mayer (Sciences Po Paris)
- Représentations du champ social, attitudes politiques et changements socio-économiques. J.-M. Donegani, G. Michelat et M. Simon (Sciences Po Paris)
- La formation des couples, M. Bozon et F. Héran (INED)
- Les intellectuels communistes et la Nouvelle Critique, F. Matonti (Paris 1),
- Les revues de la gauche non communiste des années 1960 aux années 1980. É. Bréhier ;
- Favoritisme et corruption à la française, P. Lascoumes (Sciences Po Paris)
- Sur les syndicats, D. Labbé et D. Andolfatto (Pacte).
- Les réseaux du parti socialiste, F. Sawicki (Ceraps, Paris 1=.
- Les inégalités à l'école, A. Van Zanten (Sciences Po Paris).
- Enquêter en milieu populaire, G. Mauger (Paris 1).

Et en parallèle, comme indiqué plus haut, nous développerons et commencerons à tester la plate-forme d'auto-archivage. Nous savons d'ores et déjà que fin 2015, nous serons loin d'avoir un outil totalement fonctionnel mais nous avançons par ailleurs sur l'expérimentation de l'usage de l'outil et des méthodes d'analyse secondaire (le projet réAnalyse se conclura en décembre 2014 par un colloque sur le sujet). Par ailleurs, nous savons que l'objectif de partage des données qui est le nôtre aura, si nous réussissons, beaucoup de conséquences sur les façons de faire des sciences sociales. Nous avons tous conscience qu'il s'agit donc non pas « seulement » d'une entreprise patrimoniale mais bien d'un projet scientifique qui nous engage pour de nombreuses années.