

Agnès Vatican – Archives départementales, directrice

Nous allons maintenant passer à la deuxième partie de cette matinée. On a pris un petit peu de retard, mais on va aller vers la synthèse, avec Louis Colombani, de la société Anaphore.

Louis Colombani est bien connu des archivistes puisqu'il est l'éditeur du logiciel métier qui s'appelle Arkheia, qui équipe de très nombreux services d'archives en France, archives départementales, archives communales, et aussi d'autres types de services d'archives. Mais ce n'est pas à ce titre-là que nous vous avons convié aujourd'hui, Louis. C'est plutôt par rapport à de nouveaux projets, innovants et notamment à ce projet d'outil de mise en ligne, auquel vous avez donné ce nom très joli de Bach. Quelques collègues d'autres services d'archives commencent à l'utiliser dès à présent. Ce n'est pas le cas des Archives de la Gironde, mais il nous a paru intéressant d'avoir une présentation de cet outil et des projets que vous avez pour faciliter la mise en ligne des instruments de recherche et des ressources des services d'archives.

Louis Colombani – société Anaphore



Merci de nous avoir invités à présenter une partie de notre travail. Effectivement, nous travaillons dans plusieurs directions. On a évoqué rapidement Arkheia, un outil pour produire des descriptions de ressources archivistiques. Et là vous nous avez invités à présenter le moteur de recherche. Anapahore s'est engagé, depuis environ trois ans, dans le développement d'un nouveau moteur de recherche. Je rejoins un peu Georges Cuer qui disait récemment qu'en fait, avec les sites d'archives actuels, il faut presque être archiviste pour pouvoir se débrouiller. C'est vrai que c'est assez complexe. Et justement le problème que l'on évoquait tout à l'heure, de la description en plusieurs niveaux, cela veut dire qu'il faut pouvoir situer

les réponses et les questions dans leur contexte hiérarchique, et en même temps, il faut pouvoir apporter des réponses immédiates à des questions précises. C'est un peu la quadrature du cercle ou la mission impossible. Il s'agissait d'essayer de relever ce défi. On le relèvera certainement imparfaitement.

Ce qu'on a cherché à faire avec ce moteur qui s'appelle Bach, et de son grand nom Bach Ricercare, c'est essayer d'offrir une étape vers un petit peu plus de simplification de la recherche. Notre objectif est effectivement de permettre aux archivistes de valoriser leur travail ; et pour cela, de s'adresser à l'utilisateur final, donc à l'internaute.

Les premiers à rencontrer Bach ont été les généalogistes. J'expliquerai pourquoi. Je commence par une citation, une présentation de ce qui a été fait pour le Gard. Voilà un généalogiste qui dit « Ouf ! On a eu chaud, le navigateur d'Anaphore fait partie des moins mauvais ». Cela incite à la modestie mais c'est très encourageant, parce que c'est visiblement une personne qui ne doit pas être très tendre en général. D'autres généalogistes disent « Je n'ai jamais vu un moteur de recherche aussi exotique pour les registres matricule, mais on parvient quand même à un résultat ». Puis un autre dit que lui non plus, mais finalement c'est pas mal, « il y a un diagramme des dates et on clique dessus ». Et le troisième dit « Jamais rencontré non plus. Plutôt qu'exotique, je dirais nouveau et original, j'aime assez tout compte fait ».

Après trois ans de travail, pas suivant les méthodes modernes, parce qu'avec les méthodes modernes, on se confronte tout de suite avec l'utilisateur, on a les premières réactions qui sont semble-t-il pas complètement négatives, puisqu'on fait partie des moins mauvais.

Ce qui s'est passé en fait, c'est qu'on a un service client qui, pour un tas de raisons, on ne va pas les citer, mais c'est assez facile à comprendre, a décidé de ne pas ouvrir en ligne officiellement ce travail. Il y a eu un changement de direction dans le service etc. Voyez ce graphique, c'est un graphique de fréquentation. Il y avait un site de pré-production, qui était en ligne depuis assez longtemps avec aucune fréquentation, et puis les généalogistes tout à coup l'ont retrouvé et ont fait ces tweets que je vous ai montrés.

On est passé du jour au lendemain de 3 visites le samedi à 280 ; et le dimanche ? à près de 500 ! Depuis cela se maintient entre 100 et 150 de manière assez régulière.

Ce qu'on a voulu faire, c'est offrir la possibilité de chercher de manière simple. On prend toujours l'exemple de la généalogie. En fait, quand on veut chercher, il y a des matricules militaires ici, on tape ce qui nous vient à l'esprit ou plutôt ce que l'on recherche. Par exemple, je cherche un personnage un peu connu, mais qui peut exister aussi dans la vraie vie. Donc j'ai tapé « Jules César ». Il faut que je mette l'accent quand même. La recherche n'est pas sensible aux majuscules et aux accents, mais malgré tout, pour la pertinence des réponses, il vaut mieux taper avec l'accent. Donc ici j'ai tapé « Jules César ». Effectivement, il y a un monsieur qui s'appelle César et qui a comme prénom Jules. Le principe ici, c'est qu'on va avoir comme réponse toutes celles qui contiennent au moins un des termes qui sont dans la

requête et on va normalement présenter en premier celles qui répondent avec le maximum de critères avec les données.

On va reposer une autre question.... avec Paul.

Vous savez qu'il y a des noms qui sont à la fois des noms et des prénoms. Ce n'est pas toujours facile de retrouver. Ici quand on tape Martin puis Paul, on a d'abord toutes les personnes qui ont à la fois Martin soit comme nom soit comme prénom, et Paul soit comme prénom ou comme nom. On a la possibilité ensuite de filtrer, grâce à un système de facettes, par exemple en choisissant soit Martin comme nom soit Paul comme nom. Ici vous voyez qu'on a d'abord tous les Martin nom, Paul comme prénom. Ensuite, si on va sur les pages suivantes, on aura au bout d'un moment Paul tout seul. Et si on va plus loin encore, on pourra avoir Martin tout seul.

Pourquoi par exemple est-ce qu'on a le nom Paul tout seul qui vient avant le nom Martin ? Tout simplement parce, vous voyez ici une banque de réponses. Avec notre requête, on a 477 personnes qui portent le nom de Martin et 56 qui portent le nom de Paul. En fait, il y a un algorithme de pertinence qui fait que, c'est moins intéressant dans le cas des noms de famille, mais dans le cas d'une recherche dans des instruments de recherche, plus un terme est rare, plus il a de poids dans le nombre de réponses, ce qui explique que Paul tout seul vienne avant Martin. Et si effectivement on veut sélectionner uniquement les Martin comme nom, on utilise la facette, et à ce moment-là, on réduit le nombre de réponses. On peut également, puisque qu'il s'agit de registres matricules, taper tout simplement Martin, Paul, et puis par exemple si on sait que c'est une personne qui avait le registre matricule 833, et là effectivement on a une personne qui s'appelle Martin, qui a Paul dans ses prénoms, et qui porte le registre matricule 833. Et ensuite, il y a une autre personne qui s'appelle Martin, qui est aussi 833, mais qui n'a pas Paul comme prénom. C'est pour montrer comment ça fonctionne. Ensuite on a la possibilité, puisque là on a une dichotomie entre d'une part des recherches nominatives dans une base de données qui est à part des instruments de recherches eux-mêmes, on a également la possibilité de naviguer dans les instruments de recherche et d'interroger les instruments de recherche.

Je vais prendre un autre cas, mais cette fois-ci dans le Gard. Et plutôt que « Jules César », je vais taper « César Auguste ». On a bien un César comme nom et Auguste comme prénom. Là encore, César peut être nom ou prénom et Auguste également. Ici, mettons que ce soit la personne qu'on cherche, on peut visualiser le registre de matricules, éventuellement en qualité supérieure, l'agrandir, c'est classique. Et ici, ce que je voulais vous montrer, c'est que vous avez un fil d'Ariane qui situe l'image non plus dans la base de données nominatives, mais dans l'instrument de recherche lui-même. Donc on a toute la hiérarchie de l'instrument de recherche. Le dernier niveau, c'est le registre dans lequel se trouve cette image. Donc on a la possibilité de visualiser la notice qui décrit ce registre et également de la situer dans le corps de l'instrument de recherche. On retrouve ici notre notice et elle fait partie du Bureau de recrutement du Pont-Saint-Esprit pour la classe 1889.

On a vu ce qui concernait principalement la recherche généalogique qui constitue malgré tout, pour les services d'archives, la clientèle la plus nombreuse. Heureusement il n'y a pas que ça. On a la possibilité d'explorer les instruments de recherche. Voici le premier utilisateur de Bach et un utilisateur assez particulier puisque le service des archives de Meurthe et Moselle ont mis en ligne un inventaire du XVIIIème siècle. L'inventaire des biens lors du rattachement du Duché de Lorraine à la Couronne de France.

On ne va pas regarder cela en détail, mais ce que je voulais vous montrer, c'est une certaine tolérance orthographique, qui fait que si je tape un terme ... il n'y a pas du tout d'indexation pour cet instrument de recherche. En termes de frappe, on a des suggestions, mais pas des suggestions qui sont issues d'un dictionnaire a priori, mais de données qui se trouvent dans la base de données.

Lorsqu'on lance la requête, on va voir un certain nombre de réponses avec cheval, mais aussi avec chevaux, donc au singulier ou au pluriel. Si on va plus loin, on va également avoir des réponses avec des termes dérivés de cheval comme chevalerie avec un L ou deux L etc. etc.

Cet instrument de recherche est également mis en ligne, on peut visualiser les images du texte lui-même...

Alors, ce qu'on a voulu faire aussi, c'est simplifier une charte graphique qui est à peu près jolie quand on la voit sur un écran d'ordinateur, mais peu visible quand elle est projetée comme ici.

Ce qu'on a voulu faire, c'est simplifier, pour l'archiviste, les recherches spécialisées. Voici un outil, et c'est ce qu'on trouve dans la plupart des moteurs qui sont proposés actuellement. Vous pouvez paramétrer des formulaires de recherche. Il y a un certain nombre de manipulations qui sont à faire, avec des fichiers xml etc. Ce qu'on a voulu faire, c'est à partir de la possibilité de faire des recherches grâce aux facettes, des facettes supposant quand même de l'indexation en général, la possibilité d'avoir des recherches spécialisées, mais sans qu'il y ait un paramétrage particulier à faire dans le moteur lui-même.

Ici, on a travaillé la semaine dernière avec des archivistes de la ville de Nice, et Myriam a travaillé en particulier sur des fonds d'urbanisme. On a défini un certain nombre de grands domaines de recherche, ces domaines pouvant correspondre en gros aux formulaires de recherche que vous avez dans Pleade... Sauf que là on n'est pas forcément illimités et que là aussi, c'est au moment où vous produisez l'instrument de recherche que vous allez faire une indexation particulière en disant « je veux rattacher tel instrument de recherche ou telle partie d'un instrument de recherche à tel domaine de recherche », ce qui fait qu'ensuite le moteur va pouvoir l'exploiter automatiquement. En fonction de l'indexation que vous avez faite, il y a un domaine « Bâtiments et urbanisme », quand on clique sur « Bâtiments et urbanisme », on voit apparaître d'autres facettes. Il y a une facette « Inventaire ». Ces facettes sont aisément paramétrable. On peut les déplacer, on peut les faire apparaître ou pas. Dans le cadre de l'urbanisme, on a indexé des bâtiments, on a indexé des architectes, on a indexé des pétitionnaires, ceux qui ont demandé des permis de construire ou de lotissements. En fait, notre formulaire de recherche se construit

dynamiquement, en fonction évidemment du travail qu'on a fait au niveau de la constitution de l'instrument de recherche.

La dernière chose que je voulais vous montrer, on est en train de travailler pour un service d'archives départementales, la Haute-Marne, et qui utilise actuellement Pléade. Ce service a un site assez important puisqu'il y a près de 2000 instruments de recherche. Il y a un grand nombre de données nominatives, à peu près 400 000 données nominatives. Et, pour diverses raisons, ce service veut changer d'outil. On a fait le choix de Bach. Tout en voulant quand même conserver un principe que vous voyez ici exposé.

Il y a une distinction entre les archives numérisées, directement accessibles en ligne, et puis les instruments de recherche. On voit ici le cadre de classement, recherche simple, recherche avancée et puis les collections spécialisées. Si on prend par exemple l'état civil, on a des textes de présentation des registres paroissiaux, etc. On a accès à un formulaire de recherche spécifique. Et on a ici la liste des instruments de recherche qui correspondent à l'état civil.

On a réalisé un prototype, attention ce n'est qu'un prototype, ce n'est pas une page telle qu'elle doit être présentée directement au public, mais c'est pour montrer le principe de ce qu'on peut intégrer dans un site internet d'archives, c'est-à-dire des liens vers des parties spécialisées du moteur de recherche. En haut, on a l'accès à Bach, le formulaire standard de recherche comme on l'a vu avec les autres exemples, un accès à la recherche dans les soldats, un accès au cadre de classement, un accès à la liste d'index. Ensuite, pour les recherches spécialisées, on a prévu, soit comme je viens de vous le montrer, pour le cas de Nice et pour le cas de l'urbanisme, l'adaptation automatique des facettes aux recherches spécialisées du domaine, la liste des instruments de recherche et puis la possibilité de faire apparaître, là c'est juste un exemple réduit mais pour reprendre ce que fait actuellement ce service, à savoir des explications sur l'état civil, donc la possibilité [...] et la liste des instruments de recherche, on arrive dans Bach. Là, on a tout un tas de cotes qu'il ne faudrait pas afficher parce que ça gêne la lecture. Et puis, si je reviens en arrière, je vous explique un peu le mécanisme, l'« état civil ». Actuellement c'est assez peu indexé. C'est indexé à la commune, et puis on a des recherches par date, on a la géolocalisation avec la recherche par la commune.

En fait, comment indexer ? On a la liste des réponses, on a des images, quand on a n'a pas d'images, on a mis Monsieur Bach avec son clavier. Évidemment, on peut l'afficher sans les imageries s'il n'y a pas lieu. Là on a 24 943 réponses, c'est tout ce qu'on peut trouver dans le service actuellement en ligne comme ressources sur l'état civil. Après, on peut taper du texte ou choisir sa commune, choisir ses dates, etc.

On a créé une facette « Niveaux ». C'est plus pour vous, archivistes. Cette facette permet de générer automatiquement la liste des instruments de recherche correspondant à un domaine. On a distingué un niveau qu'on a appelé provisoirement « Ensemble », c'est-à-dire le niveau des instruments de recherche, puis les niveaux intermédiaires, les niveaux hauts et les niveaux bas. Si je clique sur

« Ensemble », je n'aurai que la liste des instruments de recherche. C'est de cette façon-là que tout à l'heure, lorsqu'on a cliqué sur une liste d'instruments de recherche, en fait, on a activé la facette « Domaines de recherche » puis la facette « Niveaux » avec le niveau « Ensemble ». C'est à peu près ce que je voulais vous présenter. Je peux aussi supprimer ma facette.

La fameuse question de l'héritage et du contexte, on la retrouve ici. Il y a toujours un fil d'Ariane qui donne la suite hiérarchique dans laquelle se trouve la réponse. Et puis, tout à l'heure, mais là c'est une version un peu plus ancienne pour situer dans l'arborescence, vous aviez une petite icône qui permettait effectivement d'afficher l'instrument de recherche. Ce n'est pas forcément l'idéal, mais on est face à cette problématique de devoir à la fois répondre à des questions précises par des réponses précises, et puis à la fois de situer la ou les réponses dans leur contexte. Avec la possibilité évidemment de choses classiques, d'afficher le cadre de classement du service. Et puis, à partir du cadre de classement, de naviguer aussi dans les instruments de recherche. Ca, c'est de la navigation html assez classique.

J'ai essayé d'être bref parce que le programme est chargé. S'il y a des questions auxquelles on peut répondre ...

[Applaudissements]

Agnès Vatican :

Merci beaucoup. Donc c'est un moteur de recherche qu'ont commencé à utiliser quelques services d'archives mais qui n'est pas encore très visible sur la toile ?

Louis Colombani :



Actuellement, le premier exemple est celui de la Meurthe-et-Moselle. On avait un cas très spécifique. Il y a 33 volumes sur l'inventaire du 18ème siècle. Pour l'instant, il y en a 5 en ligne. Il faut le temps de finir le travail de ... la retranscription a été faite, vérification etc. côté archives.

Il y a aussi la Fondation Maison des sciences de l'homme, qui a un site, Nabu, avec les problématiques de portails qu'on évoquait tout à l'heure : passer dans Isidore, dans les portails spécialisés, d'où le travail sur le Dublin Core, avec cette question de la gestion de l'héritage etc. Et ce n'est pas si compliqué que ça, cela se fait bien.

Il y a les archives du Vaucluse qui ont mis en ligne les registres matricules. Les archives du Vaucluse ont Pleade également. Pour les matricules, on a choisi Bach. Là aussi, normalement, il y a un appel d'offres qui va être lancé cette année, en 2016.

Les archives du Gard où c'est en ligne, mais ce n'est pas officiel. C'est celui qui est le plus consulté actuellement. L'essentiel des consultations, je dirais 95%, c'est les registres matricules, mais il y a bien 5% sur les instruments de recherche. Il n'y en a pas beaucoup, 130 actuellement. Ce n'est pas un très gros corpus d'instruments de recherche.

On travaille sur la Haute-Marne. Les problématiques, sans vous montrer très en détail, c'est que, comme on géolocalise, il faut que l'indexation lieux soit bien faite. Or, quand on regarde ce qui a été fait, soit par le prestataire soit par le service d'archives, parce que la Haute-Marne est un service qui s'est informatisé il y a très longtemps, depuis le début des années 90, il y a eu des générations d'archivistes, donc il y a de l'hétérogénéité et en fait la nécessité de reprendre cela. Il y a un travail de préparation qui

n'est pas négligeable. Nous, avons aussi développé des outils pour faciliter le contrôle et la correction de cette indexation. Pour la Haute-Marne, c'est un projet assez important et qui est en cours, qui n'est pas encore...

... Il y a aussi un travail pour le gouvernement d'Haïti, pour une bibliothèque numérique. Pas mal de choses sont en cours.

Agnès Vatican :

Merci. Est-ce qu'il y a des questions sur ce nouveau moteur de recherche ?

Georges Cuer :

J'en ai une qui me travaille beaucoup en ce moment et qui concerne la géolocalisation, éventuelle, par exemple du fonds des archives des notaires. Comment faire, ce serait un énorme travail de le faire, mais une fois qu'on l'aurait fait, comment faire pour ne pas ignorer le notaire de Bordeaux qui instrumente dans le Médoc par exemple ? Ou l'ancien régime où le notaire de Sainte-Foy-la-Grande passe son temps à changer de département puisqu'il n'y a pas encore de département, et qui est gyrovague, et qui peut détenir l'acte qui va permettre de retrouver le chaînon manquant, etc. C'est l'expérience quotidienne de nos lecteurs en salle de lecture qui le font d'ailleurs de manière très sereine parce qu'ils sont très fiers d'avoir trouvé là où ils ne cherchaient pas. Est-ce que le moteur ne serait pas une façon de figer, en tout cas à distance, cette souplesse ?

Louis Colombani :

Pour la géolocalisation, il y a deux façons de procéder. Celle qu'on a mise en place jusqu'à présent qui consiste simplement à indiquer le nom actuel du lieu, département, pays, ce qui est recommandé par le SIAF. Ensuite, il y a des outils qui, automatiquement, vont chercher les coordonnées géographiques de ces lieux. C'est une façon relativement facile. En général, elle est utilisée à la main quand même, mais ce n'est valable que si on utilise les noms de lieu actuels et si on écrit ça bien. D'où le travail qu'on doit faire pour normaliser. La deuxième étape va consister à intégrer les coordonnées géographiques du lieu dans l'indexation elle-même. L'EAD3 le permet, alors que l'EAD2 ne le permettait pas. Si on a un nom de lieu ancien, mais avec les coordonnées géographiques actuelles, il sera géolocalisé par exemple. Je ne sais pas si ça répond à votre question ?

Georges Cuer :

Ca répond à la moitié de sa question.

Louis Colombani :

Mais on ne répond jamais complètement aux questions de Georges Cuer.

Georges Cuer :

L'autre moitié, c'est celui qui instrumente ailleurs.

Louis Colombani :

En fait, c'est une question de fond. Comment est-ce que je vais indexer une étude notariale ? Est-ce que je vais indexer le lieu de résidence ou est-ce que je vais indexer les lieux d'exercice ?

Georges Cuer :

Justement. Il est impossible en fait, matériellement et intellectuellement, d'indexer les lieux d'exercice.

Louis Colombani :

Donc ça veut dire qu'il n'y a pas de réponse ...

Georges Cuer :

En tout cas sur 5 siècles de notariat. Donc on peut difficilement enfermer ces 5 siècles de notariat, c'est quand même 7 km d'archives ici, dans une grille qui les simplifierait peut-être, et qui donnerait l'impression d'une rigueur qui en fait n'est pas réelle puisque si on ouvre les liasses, on trouve autre chose.

Agnès Vatican :

Oui, une autre question ?

Une personne du public :

Moi, je suis responsable, en Indre-et-Loire, des 12 km d'archives notariales.

Agnès Vatican :

Donc dans l'Indre-et-Loire : 12 km. D'accord. Tu vois, Georges, on est bien, on est bien !

[rires]

La même personne :

En fait, l'Indre-et-Loire, c'est la moitié du dépôt des archives notariales. Et pour répondre à votre question, je me la suis posée il y a plusieurs années, et on indexait différents lieux d'exercice suivant l'indication de la charge notariale. En fait, on a relevé toutes les charges du notaire et on avait pris parti d'indexer à tous les lieux de relevé. Donc on faisait de la double, voire de la triple indexation. Enfin, pour un même notaire, on pouvait avoir trois ou quatre lieux géographiques. Effectivement, comme vous le dites, ça ne couvre pas l'ensemble du territoire, mais ça permettait effectivement, comme on le voit, quand il y a un notaire qui est situé dans le nord du département, il peut très bien exercer une charge au sud du département.

Pour répondre partiellement à votre question, nous, on avait pris ce parti-là puisque ce fonds-là représentant la moitié du dépôt, ça allait porter une importance particulière pour les chercheurs, qu'ils soient généalogistes ou autres.

Agnès Vatican :

Merci beaucoup.