



# Séminaire du LaDHUL

30 septembre 2013



## Nicolas Bugnon, « Bases de données en sciences humaines : création et pérennisation »

Cet exposé de Nicolas Bugnon (UNIL, Faculté de Lettres, Section Sciences du langage et de l'information) est issu de son mémoire de master. Le temps imparti ne lui permettant pas de quadriller l'entier de son enquête, il décide de se limiter ici à 2 grandes thématiques : 1) Les difficultés liées à la conception de bases de données en sciences humaines 2) Les difficultés liées à la pérennisation de ces mêmes bases de données.

### Introduction : de quoi parle-t-on ici ?

De quoi parle-t-on ici ? Qu'entendons-nous par « bases de données en sciences humaines » ? Qui en créent ? Pourquoi ? Et pour qui ?

En se référant – entre autres – au site internet *Calame* qui répertorie les bases de données françaises en sciences humaines et sociales, on remarque que ce sont les disciplines issues des Lettres (histoire de l'art en tête) qui s'engagent le plus dans des projets de bases de données. Ces bases de données se veulent majoritairement *documentaires* et visent le plus souvent à recenser des ressources culturelles (*médiathèques*) dans un format de données textuelles.

En se basant sur un inventaire des projets de bases de données de la faculté des Lettres de l'UNIL (23 projets), Nicolas Bugnon nous indique que leur « taille » varie le plus souvent entre 500 et 10 000 entrées.

En ce qui concerne le public de ces bases de données, il peut aller de 1 (la base de données ne concerne que celui ou celle qui l'a créée) à, potentiellement, plusieurs millions lorsqu'elles sont destinées à être publiquement mises en ligne.

Ces tout premiers éléments – assez secs – permettent tout de même de préciser le sujet de l'exposé. Ainsi donc, lorsque nous parlons, ici, de « bases de données en sciences humaines », nous parlons le plus souvent de **médiathèques** (c'est-à-dire des bases de données documentaires) **s'attachant à recenser des ressources culturelles dans un format textuel**, dont **les tailles variables** oscillent surtout entre 500 et 10 000 entrées et le **public-cible** de 1 à plusieurs millions d'individus. Le décor est ainsi posé.

UNIL | Université de Lausanne

LADHUL - Laboratoire  
de cultures et humanités  
digitales de l'UNIL

## **Des difficultés pratiques de conceptions : choix du système et bonne collaboration**

Après ces éléments d'introduction, Nicolas Bugnon nous plonge dans son enquête qui s'appuie sur des questionnaires envoyés à des projets de création de bases de données au sein de la faculté de Lettres de l'UNIL. Assez vite, un premier constat : le choix du système des bases de données (duquel dépendra ensuite leurs fonctionnalités) est le plus souvent confié à *une seule personne* : l'informaticien de l'équipe. Travers ô combien problématique tant cette décision devrait être prise collectivement ! En effet, le choix du système est trop important, les chemins qu'il suggère trop irréversibles, pour être assumés par une seule personne. D'autant que pour se repérer, celle-ci recourt souvent à la force de l'habitude : « je suis proche de ce logiciel ; je le connais bien ; je ne suis pas certain des fonctionnalités souhaitée par l'équipe mais celles-ci *devraient* pouvoir être accomplies par ce logiciel, d'une manière ou d'une autre ; de toute façon, je n'ai pas vraiment le *temps* de me lancer dans une programmation ; sauf si c'est ce qu'ils souhaitent ; mais le souhaitent-ils ? Nous n'en avons pas *discuté*. Je ne sais pas... ». Ainsi donc, beaucoup de brouillard, peu de réflexions *collectives* de fond alors même que c'est à ce moment-là – précisément ! – qu'elles devraient se déployer. Car il est déjà bientôt trop tard ! L'éventail des fonctionnalités de la base de données documentaires dépendra de ce moment charnière où une direction a été choisie : si celle-ci s'est faite individuellement, à tâtons et dans l'obscurité (comme c'est le plus souvent le cas), il y a de fortes chances que les fonctionnalités manquent et que la base de données déçoivent.

**Le début du projet est ainsi à soigner** : le plus de collaboration en amont ; le plus de réflexions quant aux fonctionnalités souhaitées ; le moins de choix « par défaut » suggérés par l'habitude (même si elle reste importante), le mieux le projet se déroulera.

## **Pérennisation des bases de données documentaires : l'importance des standards de description**

Même si la question est primordiale puisque liée à la survie des bases de données laborieusement créées, le problème de la pérennisation des bases de données est peu discuté. La faute pour beaucoup au statut instable des projets académiques desquels elles sont issues et pour qui les financements tout comme le personnel ne sont pas assurés. Il est ainsi tentant de considérer la question comme non résoluble : étant donnée le climat de vie instable des bases de données documentaires qui rendent leur maintenance systématique quasi impossible, ces créations semblent destinées à vivre peu longtemps. Rien n'est moins vrai pour Nicolas Bugnon : pour lui il n'y a pas de relation logique entre impossibilité de maintenance et obligation de mort prématurée. Il propose ainsi de prendre acte de l'environnement instable dans lequel vivent les bases de données documentaires et d'œuvrer autre part, à un autre niveau.

**Ce niveau est celui de des données.** En effet, au lieu de miser sur les supports logiciels des bases de données qui ne peuvent s'empêcher de vieillir et d'être ainsi dépendants d'une maintenance que l'environnement académique ne peut leur offrir, **il faut rendre les données « conservables » et capables de migrer vers une nouvelle base sans trop de problèmes et d'efforts.** Des données « génétiquement » plus solides *capables* de migrer d'une base logicielle à l'autre, voilà la proposition centrale de Nicolas Bugnon quant à cette problématique de la pérennisation des bases de données.

Mais comment faire ? Comment dissocier au mieux les *données* de leur base afin qu'elles soient capables de s'adapter à d'*autres bases*, plus tard, sans qu'il y ait trop d'efforts à déployer ? C'est

là qu'apparaissent les **standards de description**, éléments centraux au modèle de Nicolas Bugnon. Ceux-ci sont bien connus des experts en documentation (souvent – bien à tort ! – mis à l'écart des projets de création de bases de données) mais assez méconnus par les autres, informaticiens y compris. Et c'est bien dommage ! Car des données en format « texte » établies dans une forme standardisée s'affranchissent des cycles de vie des logiciels ; qu'ils soient vieux, jeunes ou « dans la fleur de l'âge », ceux-ci ne rencontrent pas de difficulté à exploiter ces données standardisées, pour autant qu'ils prennent en charge le standard (ce qui est de plus en plus la cas). Mieux ! Le fait de construire ses données selon un standard technique établi et bien distribué permet également de mieux partager ses données : **la sauvegarde chronologique se joint alors à une sauvegarde géographique** puisque ces mêmes données standardisées pourraient être stockées et exploitées aussi bien à Lausanne qu'à Leeds ou Bruxelles. C'est pourquoi **il semble primordial de solidifier les partenariats avec les bibliothèques lors de projets de création de bases de données en sciences humaines**. En effet, les bibliothèques sont constituées de professionnels de la documentation qui connaissent cette problématique de la pérennisation des données depuis longtemps et sont à l'aise avec le maniement des standards de description.

Florian Jaton



UNIL | Université de Lausanne

LADHUL - Laboratoire  
de cultures et humanités  
digitales de l'UNIL

