

Rapport technique: projet IPhIS

Julie Giovacchini, Bernard Weiss, Laurent Capron, Sébastien Grignon, Pierre Caye

► **To cite this version:**

Julie Giovacchini, Bernard Weiss, Laurent Capron, Sébastien Grignon, Pierre Caye. Rapport technique: projet IPhIS: construire une bibliographie numérique des sources littéraires gréco-latines. [Rapport Technique] CNRS - Centre Jean Pépin (UMR 8230); CNRS - Culture, Langues, Textes (UPS 2259). 2017. hal-01509598

HAL Id: hal-01509598

<https://hal.archives-ouvertes.fr/hal-01509598>

Submitted on 18 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Rapport technique – Projet IphiS : construire une bibliographie numérique des sources littéraires gréco-latines

Pilotage :

Pierre Caye, Centre Jean Pépin – UMR 8230 CNRS – Ecole Normale Supérieure

Maîtrise d'œuvre :

Centre Jean Pépin – UMR 8230 : équipe d'information scientifique (Julie Giovacchini, Laurent Capron, Sébastien Grignon, Dina Bacalexi, Pinelopi Skarsouli)

Cultures, Langues, Textes - UPS 2259 : Bernard Weiss

Rapport rédigé par Julie Giovacchini & Bernard Weiss, avec la collaboration de Laurent Capron et Sébastien Grignon – auteur correspondant : Pierre Caye

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Introduction

Dans un contexte d'explosion de la mise à disposition des ressources classiques en ligne, l'accent est mis, au sein des projets d'Humanités numériques sur l'Antiquité, sur l'accès aux textes et l'on considère que la fouille de données est la voie royale pour explorer ces corpus toujours plus gigantesques. Les Humanités numériques produisent de fait du texte¹ ; dans un article important² G. Crane et al. proposent un état des possibilités ouvertes par l'applications des techniques numériques de 3e et 4e génération aux textes anciens. Celles-ci, en matière d'établissement et de balisage fin des textes, semblent autoriser l'espoir de parvenir à un sorte d'hyper-texte ultime, concrétion finale de l'ensemble du savoir connu sur le texte : ce que Crane et al. nomment "apographeme"³.

C'est un vieux rêve de philologue : reconstruire le texte ultime, retrouver l'archétype, non plus sous forme d'objet unique mais de réseau. Or un tel réseau est extrêmement complexe à établir dans le cas des textes anciens ; en effet les antiquisants travaillent essentiellement sur un Canon qui, contrairement aux apparences, n'est absolument pas fixé, pour des raisons qui tiennent à la fois à l'histoire de la

-
- 1 "Digital technology creates texts further freed from fixed, physical form." in Smith, Neel. "Two Thousand Years of Scholarly Apps." *CHS Research Bulletin* 4, no. 2 (2016), 3&2, http://nrs.harvard.edu/urn-3:hinc.essay:SmithN.Two_Thousand_Years_of_Scholarly_Apps.2016
 - 2 Crane and al. « Classics in the Million Book Library » in DHQ 3/1 (2009) *Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure*, <http://digitalhumanities.org:8081/dhq/vol/3/1/000034/000034.html>
 - 3 "For classics, the most important such project is what we have termed the apographeme of classical Greek and Latin — an analogy to the *genome*, representing the complete record of all Greek and Latin textual knowledge preserved from antiquity, ultimately including every inscription, papyrus, graffito, manuscript, printed edition and any writing bearing medium. This *apographeme* constitutes a superset of the capabilities and data that we inherit from print culture but it is a qualitatively different intellectual space. In the mature *apographeme*, every canonical text is a multitext, with dynamic editions linked to visual representations of the manuscripts, inscriptions, papyri and other sources. In the mature *apographeme*, each source is linked to the background data that we need to understand it — a transcription, information about the particular type of Greek or Latin script and its abbreviations, about the monastery, print shop or Egyptian village that produced it, etc." (Crane and al. 2009, 10).

discipline et à la nature des sources qui la nourrissent.

Il est alors essentiel de conserver une place pour des outils consacrés à ce Canon lui-même, qui permettent une recherche non automatique, capables donc de prendre en compte dans leur démarche l'épaisseur historique du texte (son histoire, sa réception), ses états différents - et pour cela il faut encore des méta-données bibliographiques non exclusivement automatisées ; en d'autres termes, le chercheur doit garder en partie la main et ne pas être exclusivement tributaire d'outils de fouille de données, excellents pour explorer un corpus mais encore insuffisants pour le critiquer et le constituer comme tel.

Malgré cette nécessité, la question bibliographique est rarement prise au sérieux, sinon sous un angle strictement quantitatif. Ainsi la somme de Babeu⁴, qui constitue un remarquable balayage du paysage numérique sur l'Antiquité, signale-t-elle les outils existants sans les critiquer et fait comme si la question de la bibliographie était résolue. Pourtant la bibliographie dans sa méthodologie contemporaine appliquée aux textes antiques est un problème en soi.

C'est pourquoi l'équipe d'information scientifique du centre Jean Pépin, composée d'IR et d'IE spécialisés en analyse de sources anciennes a délibérément choisi de construire un outil bibliographique numérique consacré aux sources textuelles littéraires de l'Antiquité⁵. L'objet du présent rapport est de documenter cet outil, IPhiS, qui se présente actuellement sous la forme d'une base de données (=BDD) avec module de saisie (back office), déposé sous licence CC et développé et maintenu par Bernard Weiss, Ingénieur à l'UPS 2259 du CNRS.

La première partie de ce rapport présente le cadre scientifique de ce travail et les choix faits par les analystes en amont et pendant la constitution du modèle de données. La seconde partie expose les choix techniques consécutifs à ces décisions scientifique et documente la BDD comme telle.

1. Cadre et choix scientifiques

1.1. Quelques définitions

Le projet, né en juin 2014, a d'emblée été orienté autour des textes littéraires de l'Antiquité, laissant de côté d'une part les textes spécifiquement documentaires, d'autre part les artefacts. Plus précisément encore, IPhiS a été pensé comme un outil destiné essentiellement au signalement des éditions et traductions des textes littéraires antiques, le champ des commentaires et de la littérature dite secondaire en général ayant été considéré comme trop abondant et finalement moins pertinent scientifiquement car soulevant moins de difficultés de description.

Il s'agit donc d'un outil signalant des textes (éditions, traductions) portant eux-même sur des textes (sources) ; sa structure générale s'articule de ce fait assez naturellement autour de deux types d'objets principaux, qui constituent la partie la plus importante du dictionnaire de la BDD : d'une part des auteurs (anciens ou modernes) d'autre part des titres (anciens ou modernes).

Cette première approximation du dictionnaire de données présente déjà en elle-même des difficultés, et entraîne des contraintes assez lourdes sur le plan du modèle : en effet, aussi bien du point de vue de la segmentation que de la normalisation finale, les contraintes bibliographiques vont varier énormément selon qu'on va traiter d'auteurs anciens ou modernes, ou de titres anciens ou modernes : en d'autres termes, les standards internationaux et notamment ceux du Dublin Core, dont nous souhaitons rester proches, ne s'adaptent pas avec autant de pertinence et d'exactitude aux sources ou aux éditions humanistes qu'aux éditions critiques modernes.

La constitution de ce dictionnaire de données historiquement mixtes a donc nécessité au préalable un

4 Alison Babeu, *Rome wasn't digitized in a day : Building a Cyberinfrastructure for Digital Classics*, 2011, <https://www.clir.org/pubs/abstract/pub150abst.html>.

5 Cf. <http://iphi.hypotheses.org/107> Laurent Capron « Une nouvelle base bibliographique : pourquoi, pour qui, comment ? »

important travail de définition et de recherche - travail qui a déjà été en partie documenté dans ses grandes lignes dans le carnet de recherches associé au projet, créé en 2014⁶ (iphi.hypotheses.org).

Voici les définitions des principaux éléments du dictionnaire de données de la BDD :

- **texte** : production écrite littéraire, technique, politique, religieuse etc. transmise par une tradition manuscrite et attribuée ou non à un auteur.
- **titre** : titre associé à un texte soit par la tradition manuscrite directe, soit par une tradition manuscrite indirecte, soit par l'érudition ou l'usage modernes.
- **auteur ancien** : personnage ancien (*terminus ante quem* : Xe s. apr. J.-C.) considéré comme l'auteur d'un texte par la tradition antique ou médiévale, et validé (ou non) comme tel par l'érudition moderne ; dans l'architecture de la BDD, l'auteur antique est par défaut considéré comme l'**auteur créateur** du texte.
- **auteur moderne**: personne moderne ayant contribué à la diffusion imprimée d'une édition ou version de texte ancien par diverses opérations d'édition (établissement du texte, indexation, annotation, traduction) : la notion d'**auteur contributeur** permet de leur adjoindre des auteurs non modernes faisant le même type de travail sur les sources et d'englober des auteurs historiquement anciens mais qui travaillent comme éditeurs, traducteurs et contributeurs d'un texte (pré-humanistes).
- **édition** : diffusion imprimée d'un texte ancien (peut être savante ou de plus grande diffusion ; peut être critique ou plus ou moins diplomatique).
- **source** support écrit (en général mais non exclusivement manuscrit) portant un texte antique (complet ou fragmentaire) ou mentionnant son existence.
- **réception** : histoire des usages, transformations et lectures du texte après sa création et sa première diffusion, en incluant ou non les traductions latines anciennes des textes grecs, et les traductions plus récentes.

Ce travail de définition préalable a débouché sur une première approximation des difficultés spécifiques liées à la normalisation de données par nature mouvantes et ambiguës, et qui sont déjà l'objet d'une forme de prénormalisation traditionnelle au niveau des titres, des attributions (certaines paternités étant contestées ou confirmées au fur et à mesure des avancées de la recherche), de la numérotation au sein de corpus, du découpage des textes, du rapprochement accidentel de textes pourtant différents au sein de recueils ou d'anthologies, etc. La transmission joue ainsi un rôle contraignant dès la conceptualisation du modèle de données, les textes à signaler étant déjà signalés et numérotés (ainsi dans les instruments que sont la CPG/L ou le TLG, qui fournissent des identifiants massivement utilisés aussi bien pour les auteurs que pour les textes⁷) : il est impossible de faire fi de cette tradition bibliographique, qui joue le rôle d'un repère admis par tous les scientifiques mais dont les principes de signalement, tributaires de certains partis-pris philologiquement datés voire totalement accidentels se révèlent des obstacles pour comprendre la vraie nature des textes et par conséquent en fournir des descriptions scientifiquement satisfaisantes. Le signalement établi devait donc idéalement permettre, non seulement de donner des informations à jour et pertinentes scientifiquement, mais aussi de relier les métadonnées ainsi produites à l'information bibliographique traditionnelle sous une forme lisible pour une communauté d'utilisateurs ayant déjà des habitudes dans ce domaine⁸.

Le choix d'un outil de type BDD relationnelle a procédé très naturellement de ces réflexions

6 Notamment la rubrique « cabinet de curiosités » présentée ici : <http://iphi.hypotheses.org/176>

7 <http://iphi.hypotheses.org/345> Sébastien Grignon « Je ne suis pas [qu']un numéro ».

8 <http://iphi.hypotheses.org/41> Julie Giovacchini « De l'usage à la règle ».

autour du dictionnaire de données. La nature protéiforme des objets étudiés, la difficulté, alors qu'un outil informatique suppose des trouver des structures et des descriptions communes, d'en établir dans un contexte où quasiment chaque entrée constitue un cas particulier, nous a conduit à privilégier une forme numérique qui permette de mettre l'accent non pas tant sur la description figée d'une entité que sur la mise en réseau de cette entité avec d'autres, de façon à ce que les notices procèdent de cette mise en réseau, et permettent in fine d'inscrire les textes et les auteurs dans un système d'information complexe qui croise à la fois les sources et leurs différentes présentations historiques. De ce fait, le projet a pris assez rapidement dans sa conception même une dimension supplémentaire, dans la mesure où la production de nos propres données bibliographiques a nécessité de faire une place à la promotion d'autres ressources disponibles. C'est l'avantage d'une BDD en ligne : elle permet de proposer un contenu et de le relier à autre chose, en envisageant la donnée comme une entité complexe et « ouverte » et non comme un contenu simple et « fermé »⁹ ; la donnée est ainsi reliée à d'autres entités de type sources manuscrites, éditions en ligne, fichiers d'autorité déjà existants : la BDD joue en ce sens un rôle de connexion des informations qui épaissit le signalement brut qu'elle propose dans ses propres tables. En amont du travail actuel de nombreuses équipes et chercheurs en Humanités numériques sur l'encodage des textes, notre équipe ayant chois de travailler exclusivement sur les métadonnées et la description exogène des textes (ce que l'on appelle finalement l'analyse de source, qui est la mission principale au CNRS des différents membres de l'équipe du centre Jean Pépin) a constaté une insuffisance manifeste des outils actuels de bibliographie qui ajoutent de l'information sans toujours penser à la relier de façon lisible¹⁰ ou, quand elles le font, abandonnent le travail sur les référentiels d'autorité en alignant leurs données bibliographiques sur des réservoirs de données généralistes - comme par exemple les catalogues des grandes bibliothèques universitaires ou, dans le pire des cas, les plate-formes académiques industrielles alors même que ces catalogues proposent des ressources qui ne sont pas normalisées selon les critères académiques de la recherche scientifique, mais plutôt selon ceux, différents et différemment contraignants, de la bibliothéconomie¹¹.

1. 2. Vers un modèle de données

1. 2. 1. Quelles relations entre les données ?

La base IPhiS a été conçue autour de plusieurs tables de références (auteurs anciens, titres, auteurs modernes) reliées entre elles dans une table principale de références bibliographiques, toutes pouvant être éventuellement associées à d'autres tables (géographie, datation, sources, maisons d'édition). Elle fonctionne donc autour d'une multi-indexation des données. Contrairement à des répertoires de référencement de textes (de type CPG, CPL, BHG, BHL, TLG, TLL), elle ne vise pas à donner une

9 "Books are certainly many other things, too, including artifacts and texts and commodities and objects of desire and vectors for cultural power. But files—digital files—are *also* all of these things. Files, too, are salable, legible, desirable, hegemonic, and even, yes, material. So to say that books are now files is not simply to add another property to a list, but to assert something more fundamental about the way books interact with other cultural and medial forms." in Matthew Kirschenbaum, *Archive Journal* 5 (2017), <http://www.archivejournal.net/issue/5/notes-queries/books-files/>

10 Même les projets soucieux de proposer des données bibliographiques en réseau s'avèrent peu maniables du fait de choix méthodologiques qui ne tiennent pas compte des pratiques scientifiques réelles des philologues : ainsi la Leuven database of Ancient books, associée au remarquable portail Trismegistos : <http://www.trismegistos.org/ldab/search.php> ; la base est d'une grande richesse, puisqu'elle signale 15822 textes en grec, latin, copte, démotique, et syriaque. Mais la recherche y est très difficile. Concernant la recherche par titre de texte, aucune information sur la normalisation des titres et le choix des formes d'autorité n'est donnée clairement ; la recherche par auteur ancien est plus aisée puisque la normalisation latine a été retenue, mais elle fait aboutir à une liste de notices bibliographiques qui ne spécifient pas les textes précis signalés par telle ou telle édition. De fait la table centrale de cette BDD est celle des fragments (sources) liée aux données de Trismégistos ; les éditions mentionnées sont celles reliés à un fragment (on peut donc avoir plusieurs fois le même texte mentionné avec deux numéros identifiants différents, pour peu que le fragment ait été plusieurs fois publié). Or les philologues n'ont pas seulement besoin d'accéder aux données liées à un auteur ou à une source : ils ont aussi besoin de naviguer à partir des textes associés à ces auteurs ou à ces sources. Un spécialiste de Cicéron ou de Platon travaille aussi bien sur un auteur ou sur un manuscrit que sur tel discours précis ou tel dialogue clairement identifié.

11 <http://iphi.hypotheses.org/145> J. Giovacchini « De la bibliographie : qu'elle n'est pas un catalogue »

édition de référence, mais l'ensemble des informations bibliographiques éditoriales relatives à un texte, y compris dans le choix des sources retenues par l'éditeur. Elle est donc amenée à connaître un accroissement permanent de données que seule la base informatisée peut gérer en continu.

1. 2. 2. Le choix des textes comme table centrale

Tous les grands corpus de référence organisent l'information bibliographique ou scientifique selon les auteurs anciens, à l'exception notable des corpus hagiographiques qui les organisent selon le ou les personnages principaux du récit. Il en va ainsi du TLG, de la LLT, de la CPG et de la CPL, mais aussi des outils bibliographiques comme l'APh. Ces bases fonctionnent sur un mode hiérarchique où tout texte dépend d'un auteur, ou d'un groupe d'auteurs, avec, le cas échéant, des renvois vers d'autres auteurs lorsque certaines attributions d'autorité sont remises en cause.

La base IPhiS ayant pour objectif de répertorier les éditions de textes anciens, il a paru plus évident de placer la table des titres d'œuvres au cœur de la base de données, et d'y relier d'une part la table des auteurs anciens, et d'autre part la table des références bibliographiques, chacune de ces deux tables étant associée à diverses tables annexes.

La conséquence première de ce choix fut la nécessité d'établir des titres d'œuvres qui puissent servir de forme de référence. Or, de nombreux textes sont connus sous diverses appellations entre lesquelles il a été nécessaire de choisir, ou, dans certains cas, auxquelles il a été jugé préférable de renoncer au profit d'une nouvelle appellation plus explicite. Toutefois, l'existence d'une tradition, parfois ancrée de longue date, ou de répertoires de référence divergents sur les formes de titres, a nécessité de conserver les formes alternatives associées à leur forme de référence respective dans une table dite d'alias. Ainsi, il sera possible d'effectuer une recherche sur un texte quelle que soit la forme du titre retenue.

1. 2. 3. La place donnée aux ressources extérieures

Au niveau de la BDD elle-même, ces ressources apparaissent sous plusieurs formes :

- association à chaque donnée, lorsque cela est possible, d'au moins un de ses identifiants dans d'autres outils spécialisés : pour les textes et auteurs créateurs, identifiants numériques du TLG, de la LLT, de la CPG ou de la CPL, pour les auteurs contributeurs identifiant ark du data-BnF ou identifiant Viaf.
- création d'une table des BDD extérieures à IPhiS proposant des ressources sur les textes : bases de signalement de manuscrits comme la base Pinakes, bases de textes en lignes comme Perseus, bases d'éditions numérisées comme Gallica ou Internet Archive. Nous avons choisi de privilégier de façon systématique les ressources de ce type en accès libre, afin d'une part de promouvoir l'open access¹² et d'autre part de fluidifier le plus possible la navigation au sein de IPhiS.
- possibilité d'insérer, au sein même des notices textes, éditions ou auteurs, des liens vers ces BDD extérieures pour un accès direct.

1. 3. Vers la BDD réelle : contraintes, renoncements, décisions

1. 3. 1. La question des niveaux de texte

Lors de la préparation de la base IPhiS, il a été souvent question du problème du mode de transmission et du mode d'édition des textes et de ce qui nous en était transmis. Nous avons relevé qu'un texte pouvait nous être parvenu intégralement ou de manière fragmentaire, selon son support ; qu'il pouvait n'avoir que partiellement subsisté dans des citations insérées dans un autre texte ; que tout ou partie de ce texte pouvait n'être parvenu jusqu'à nous que comme composante d'un recueil de textes, ou, à

¹² <http://iphi.hypotheses.org/91> Sébastien Grignon « Quelques notes sur le modèle commercial de l'édition scientifique en Sciences de l'Antiquité ».

l'inverse, qu'il possède une tradition propre et une tradition par une compilation. Partant de ce constat, il s'agissait de trouver le moyen de signaler au chercheur que telle édition donnait tel texte ou partie de texte dans telle ou telle voie de transmission.

Nous avons pensé en premier lieu à une division des données. Certaines de ces divisions relevaient du texte lui-même : le texte est-il un fragment (car le reste est perdu, par exemple), un texte complet, ou une compilation de textes (cas de l'Anthologie grecque, des Apophtegmes des Pères, etc.). D'autres relevaient des choix éditoriaux : s'agit-il de l'édition complète du texte ou d'une partie seulement ? On pense en particulier aux éditions papyrologiques qui ne transmettent quasiment jamais de textes littéraires complets. Ou bien s'agit-il de l'édition d'une collection de textes, comme les *Vitae Patrum*¹³ ? Les limites d'un tel classement sont pourtant rapidement atteintes, puisque une édition de référence comme les *Fragmente der griechischen Historiker* de Jacoby¹⁴ se trouve être la compilation de fragments d'auteurs ou de citations fragmentaires transmises par le texte, parfois incomplet, d'autres auteurs. Comment faire aussi pour les compilations au contenu mouvant, comme les collections d'*Apophtegmata Patrum*, dont on découvre encore de nouveaux éléments dans tel ou tel manuscrit, ou qui reprennent des fragments d'œuvres dont le contenu a paru à certain compilateur assez pertinent pour être intégré à la masse des autres apophtegmes ? Dans ces cas précis, qu'appelle-t-on un texte complet ou un extrait ? Faute d'avoir pu résoudre de manière simple les problèmes techniques posés par ces textes, il a été jugé préférable de s'en remettre, dans un premier temps, à la description précise de l'édition tout en effectuant des renvois entre les différentes œuvres quand certaines compilations intègrent des extraits non négligeables d'œuvres qui ont par ailleurs leur tradition propre.

1. 3. 2. Le cas de la géographie

On associe souvent aux noms des auteurs anciens une épithète géographique, généralement dans le but de distinguer des homonymes. Le choix a été ici fait de séparer les informations de noms et de lieux, en particulier pour rendre compte du fait que certains auteurs anciens ont eu plusieurs lieux d'attache (lieu de naissance, lieu de résidence ou d'enseignement, lieu de mort, etc.) et qu'ils sont parfois reconnus et identifiés, dans les nombreux cas d'homonymie, selon ces différences ethniques. Les données géographiques sont aussi importantes dans l'élaboration des listes de contributeurs, postérieurs à l'Antiquité, qu'ils soient copistes, traducteurs, collectionneurs ou imprimeurs.

Ces données permettent d'interroger la BDD sur des critères spatiaux et non seulement individuels et rendent compte de courants de pensée liés à des localités (p. ex. Alexandrie, Syracuse, Antioche, etc.), ou des hauts-lieux de la tradition humaniste lors de l'apparition de l'imprimerie (p. ex. Lyon, Cologne, etc.). L'objectif est, à terme, le développement d'applications qui permettent d'accéder au contenu des tables d'autorité au moyen de cartes interactives. L'un des défis à relever lors du développement d'une application de recherche ou de représentation cartographiques sera la différenciation des données géographiques associées tantôt à des auteurs créateurs (auteurs antiques), tantôt à des contributeurs (érudits du Moyen-Âge, de la Renaissance ou de l'époque moderne, etc.).

Les références géographiques contiennent, lorsqu'elles existent et sont connues, le nom moderne du lieu, le nom antique et les coordonnées géodésiques des lieux. Y sont données aussi les URL stables vers d'autres bases : GeoNames, Pleiades et TrisMegistos Places. GeoNames (<http://www.geonames.org>) est une base qui recense l'ensemble des données liées à un lieu et les hiérarchise : divisions administratives, type de lieu (lieu habité, site archéologique, montagne, rivière, etc.). Ainsi, par exemple, le site de Priène en Turquie, qui n'a pas d'habitat moderne, n'est pas confondu avec le village de Güllübahçe Turun, sur le

13 <http://iphi.hypotheses.org/280> Laurent Capron « Les Vitae Patrum, brève histoire de l'édition d'un corpus » ; voir aussi sur le même sujet <http://iphi.hypotheses.org/562> Laurent Capron « De quoi l'Anthologie grecque est-elle le nom ? »

14 <http://iphi.hypotheses.org/185> Julie Giovacchini « Atthis ou les homonymes »

territoire duquel il se trouve. Pleiades (<http://pleiades.stoa.org>) est une carte interactive des lieux anciens : elle fournit, pour chaque lieu, des données comme l'indexation dans le Barrington Atlas, la référence dans la Realencyklopädie (RE), voire une bibliographie scientifique sur le lieu. C'est une base collaborative et chaque chercheur qui dispose de données vérifiées sur tel ou tel lieu est invité à les partager. Enfin TrisMegistos Places (<http://www.trismegistos.org/geo/>) est une base qui recense les attestations de noms de lieux antiques dans les documents et textes littéraires et les relie à une table de lieux. Ces trois bases, toutes libres d'accès, sont donc complémentaires les unes des autres.

1. 3. 3. Les controverses d'attribution

La notion d'*auteur ancien* étant problématique en soi (comme la définition qui en a été donnée supra le laisse deviner) toute BDD qui se donne pour tâche de faire une description raisonnée des textes anciens doit répondre, si possible adéquatement, à la question de l'attribution du texte à un (ou plus rarement plusieurs) auteur-s ancien-s. Si la BDD (contrairement à ce qui fut jadis la limite du livre imprimé) permet d'assigner sans difficulté plusieurs auteurs à un texte dont l'attribution serait discutée, l'utilisateur gagne beaucoup à ce que la chose lui soit explicitement signalée et, s'agissant d'un outil scientifique, il nous a semblé pertinent de mentionner également l'/les auteur-s contributeur-s à l'origine de la controverse à l'aide d'une commande "controverse d'attribution" ad hoc, et de documenter précisément via un lien interne la controverse en indiquant ses sources (articles ou autres publications scientifiques dans leur description bibliographique complète). Cette approche tente ainsi de respecter les deux principes posés par Groth et al.¹⁵ concernant la provenance des données : « (1) actors must only record propositions that they know to be true, through statements of what they observe; and (2) each statement of provenance must be attributable to a particular actor »¹⁶.

1. 3. 4. Les rubriques collectives : comment gérer les anonymes

Un serpent de mer du signalement de sources anciennes, directement apparenté à la difficulté précédente, est la difficulté à gérer la multitude de textes transmis, souvent sous forme fragmentaire, sans auteur explicite et pour lesquels aucune attribution, même discutable, n'est proposée¹⁷. Les antiquisants travaillent extrêmement souvent sur des données anonymes (traités, commentaires, scholies) qu'il faut pouvoir signaler de manière non ambiguë. Ce problème classique est particulièrement douloureux pour les BDD relationnelles, dans la mesure où la nécessité d'associer un élément de la table "auteurs" à un élément de la table "textes" semble imposer de relier de très nombreux textes à un "anonyme" - la seule solution praticable pour éviter les homonymies étant de numéroter les anonymes en question. L'encombrement qui en résulte dans la table des "auteurs" est très dommageable (on a alors en effet une grande quantité de données inutiles qui surchargent une table déjà bien remplie et la rendent peu lisible

15 Cf. Paul Groth, Mile Simon and Steve Munroe "Principles of High Quality Documentation for Provenance: A Philosophical Discussion" in L. Moreau et I. Foster (éd.) *Provenance and Annotation of Data*, Berlin / Heidelberg 2006 (*Lectures Notes in Computer Science*, 4145) p. 278–86. doi:10.1007/11890850_28

16 Cité notamment par Monica Berti et al., « The Linked Fragment: TEI and the Encoding of Text Reuses of Lost Authors », *Journal of the Text Encoding Initiative* 8 (2014 - December 2015), <http://jtei.revues.org/1218> ; DOI : 10.4000/jtei.1218 aux § 20 et 21, décrivant le projet d'encodage des auteurs et textex fragmentaires qui affrontent le même problème ainsi décrit : "In our use case, we need to be able to (1) reference ancient data that can be identified but that did not literally come into existence as the result of any modern computational interaction (and which may in fact no longer be extant in any preserved source); and (2) identify the role a data item, such as an ancient scholarly assertion, plays as the vehicle for the modern scholarly claims. A third requirement, which results from the second, is that we need to be able to represent the assertions of the ancient scholars, on which our modern assertions depend, in a format that can be included computationally in a common data set with the modern claims."

17 Le problème est redoublé quand l'Anonyme se révèle faux du fait justement d'une lacune dans les outils de signalement : cf. <http://iphi.hypotheses.org/178> J. Giovacchini « L'Anonymus Veronensis »

pour un utilisateur qui souhaiterait naviguer directement dedans) ; plus ennuyeux, l'accès à ces textes anonymes est alors conditionné à la seule requête par "titre", et se fait donc sur un mode isolé, ce qui constitue une perte d'information importante. En effet, ces différents textes n'ont pas pour unique caractéristique d'être anonymes : ils ont un contenu, clairement identifié la plupart du temps, et des traits stylistiques ou historiques qui permettent de les relier à d'autres familles de textes.

La solution que nous avons retenue pour contourner cette difficulté s'inspire de celle employée par au moins deux autres instruments bibliographiques plus anciens : l'Année philologique et la base Pinakes. Elle consiste à regrouper ces anonymes à l'intérieur de rubriques collectives thématiques, utilisées en lieu et place du nom de l'auteur dans la table "auteurs" ; de cette façon, l'utilisateur qui souhaite accéder à un texte précis peut toujours formuler sa requête directement à partir du titre, mais il a également toujours la possibilité d'explorer la rubrique collective associée via la table "auteurs", et d'accéder ainsi à un réseau d'autres textes anonymes de contenu proche, ce qui lui permet une approche moins ciblée.

Notre propre liste de rubriques collectives a été constituée selon deux critères : simplicité du nom de la rubrique, toujours inspiré par le genre du texte ("Grammatica" plutôt que "Scriptores de re grammatici") et évitement de l'inflation. Une telle liste n'est en effet exploitable de façon simple que dans la mesure où sa brièveté permet à l'utilisateur de l'outil de la maîtriser facilement, et où elle ne contient pas de rubriques trop proches l'une de l'autre, ce qui entraînerait des confusions et nous ferait retomber dans l'ambiguïté que cette solution cherche justement à éviter. Ainsi, il est absolument inutile de conserver deux rubriques spécifiques pour les hymnes chrétiennes et païennes : ce sont toujours des hymnes, c'est à dire des textes appartenant à un champ littéraire assez strictement défini, et donc tous regroupables sous la rubrique "Hymni". Un autre exemple : les fragments médicaux et pharmaceutiques. Dans bien des cas il sera difficile de savoir, dans le cas de deux rubriques distinctes, laquelle est la plus pertinente ; il est donc judicieux, pour éviter les erreurs de saisies, de les regrouper dans une unique rubrique "Medica et pharmaceutica".

1. 3. 5. Les datations

Notre impératif était de disposer d'un système de datation normalisé qui réponde à deux exigences contradictoires: 1) une saisie aussi simple et normalisée que possible et 2) la nécessité de couvrir les différents cas particuliers imposés par les textes anciens, sachant que la datation telle qu'on la connaît habituellement (année de naissance, de décès d'un auteur, année de publication d'un texte) est bien plutôt l'exception que la règle dans notre domaine. Il fallait donc, en gardant une certaine simplicité de traitement, pouvoir saisir des datations dont les éléments ne sont pas homogènes: siècle vs année, avant vs. après l'ère chrétienne, date connue vs. date inconnue ou conjecturale, etc. Ainsi, pour signaler des intervalles de temps, on pourra avoir recours à des formules aussi variées que:

— v. 315 apr. J.-C. - 387 apr. J.-C.

— 355 av. J.-C. - 323 av. J.-C.

— 1er s. av. J.-C. - 14 apr. J.-C.

La solution technique a consisté à modéliser toutes ces différentes possibilités par une expression régulière très longue (voir plus loin, 2. 1.), afin que la forme saisie puisse être validée ou invalidée automatiquement : on évite ainsi les saisies fantaisistes ou simplement inexacts qui interdiraient par la suite de formuler des requêtes correctes sur ce champ.

1. 3. 6. Contributeurs anciens et humanistes

Un des choix originaux que nous avons faits est la création d'une table unique des *auteurs*

contributeurs. Un tel choix ne va pas forcément de soi quand il s'agit de littérature ancienne : on pense spontanément qu'une édition imprimée de textes anciens se caractérise par un « auteur ancien » et un « éditeur moderne ». Cependant cette opposition binaire ne suffit pas à décrire toutes les entités « auteur » qui peuvent entrer en relation avec un texte ancien : que faire, dans ce cas, de tous ceux qui, dès l'Antiquité, et a fortiori au Moyen-Âge, ont contribué à diffuser ou à enrichir le texte (compilateurs, scholiastes, commentateurs, traducteurs, etc.) ? Où classer tous ceux qui ont contribué, à partir de l'époque moderne, à la (re)découverte des textes anciens, sans forcément être des éditeurs scientifiques au sens où nous l'entendons aujourd'hui ? Plutôt que d'élaborer de périlleuses distinctions disciplinaires et/ou chronologiques, nous avons préféré nous en tenir à une distinction fonctionnelle de base : « auteur créateur » (c'est-à-dire celui qui répond pleinement à la définition de l'auteur ancien énoncée supra en 1.1.) vs « auteur contributeur » (toute personne ayant contribué à enrichir et/ou à diffuser le texte ancien). La table des auteurs dits « contributeurs » est ainsi appelée à contenir aussi bien des éditeurs modernes qu'anciens, et notamment l'ensemble des humanistes ayant eu une activité d'édition et de traduction. Dans la mesure où il peut être pertinent de pouvoir extraire facilement ces humanistes de la liste complète des contributeurs afin d'effectuer des requêtes spécifiques sur ces auteurs, une case à cocher les signale comme tels dès l'étape de saisie.

La précision documentaire ne perd rien à cette décision originale, puisqu'un système de mots-clés permet comme nous allons le voir d'assigner une ou plusieurs disciplines aux « auteurs contributeurs ».

1. 3. 7. Les mots-clefs

Nous avons assez rapidement établi la nécessité d'intégrer à la BDD une table de mots-clefs, reliée à toutes les autres tables principales, permettant d'affiner le signalement et par là de faciliter encore les requêtes. Nous avons en effet considéré que la seule recherche "libre" ou "en plein texte" n'était pas suffisante pour naviguer thématiquement dans le type de données que nous souhaitions signaler. Les références bibliographiques qui constitueront à terme le coeur des données exposées par l'application web d'IPhiS sont en effet, par la force des choses, en différentes langues, anciennes et modernes. Dans ces conditions, il serait extraordinairement complexe de développer un moteur de recherche susceptible de proposer automatiquement des équivalences entre non pas deux ou trois mais sept ou huit langues (latin, grec, français, anglais, italien, espagnol, allemand, syriaque, arabe etc)¹⁸.

La solution d'une liste de mots-clefs, suffisamment brève pour être exploitable facilement (exit donc les mots-clefs trop précis ou trop ponctuels), adaptée aux données bibliographiques (portant donc essentiellement sur des genres de textes, des périodes historiques, des types d'activité éditoriale) et traduite dans différentes langues, a été retenue. Les traductions sont associées au mot-clef français sous forme d'alias, et seront donc ensuite utilisables par spécification préalable de la langue de requête.

1. 4. Une BDD, et après ? Perspectives

Le modèle de la base IPhiS est aujourd'hui stable et utilisable comme tel pour la saisie de données bibliographiques concernant les éditions et traductions de textes anciens. Le travail de modélisation s'est accompagné d'un considérable travail d'alignement portant sur les données destinées à alimenter les

18 Dans les premiers mois de travail sur le modèle de données, cette option n'était pas absolument exclue néanmoins, et une collaboration avait été envisagée avec l'équipe nantaise du LINA, spécialisée dans le Traitement Automatique des Langues, sous la direction de B. Daille. Cette perspective a été abandonnée faute de perspectives de financement solide, dans la mesure où ce genre d'approche suppose, avant tout résultat exploitable en ingénierie, un travail de recherche fondamentale de plusieurs années. Voir le numéro spécial de la revue TAL de 2009 (50/2), consacré au traitement structuré des langues anciennes, et le numéro de la même revue de 2010 concernant le traitement des données multilingues, <https://www.atala.org/IMG/pdf/1-preface-TAL51-2.pdf>

tables principales qui constituent le référentiel d'autorité d'IPhiS : à savoir les tables des auteurs créateurs et contributeurs, des titres, des mots-clefs et des titres de périodiques et de collections monographiques. Si le travail sur les auteurs créateurs, les périodiques, les collections et les mots-clefs est à l'heure actuelle achevé, il est encore en cours concernant les auteurs contributeurs humanistes et modernes et les titres qui posent des problèmes spécifiques et demandent donc un travail de vérification important¹⁹.

Ce travail d'alignement est indispensable, d'une part dans la perspective d'une mutualisation des données saisies par un modèle de ce type avec d'autres instruments numériques (ce qui est déjà en cours de réalisation avec la base Pinakes via le système d'identifiants numériques communs Diktyon²⁰), d'autre part pour proposer des données vérifiées de première main, afin d'apporter une réelle plus-value face aux données moissonnées automatiquement via les moteurs et plate-formes industriels. Comme nous l'avons expérimenté ces dernières années à l'occasion de la réalisation de ce modèle de données, la question cruciale de la bibliographie suppose, pour être résolue correctement, non pas tant une agrégation toujours plus importante d'éléments que le lien entre ces éléments et des référentiels d'autorité fiables et mis à jours – des métadonnées sans lesquelles aucun travail un tant soit peu précis n'est possible.

Un autre élément pour pouvoir passer à la phase d'exploitation et de publication de la BDD est la mise au point d'une application web de consultation, permettant idéalement trois choses :

- une navigation fluide et potentiellement multilingues ;
- l'extraction de données bibliographiques compatibles dans leur format avec des logiciels de type Zotero ;
- l'extraction de données géographiques sous forme de cartes.

Dans la perspective qui est la nôtre, l'outil IPhiS, en tant qu'il est un outil de saisie et de signalement de données bibliographiques, est par nature collaboratif. Notre équipe l'a conçu avec pour horizon une publication des données qu'il contient mais aussi et surtout un appel aux contributions extérieures. La communauté des "digital classicists" est féconde et portée naturellement vers la collaboration et l'appropriation des nouveaux outils. Les antiquisants ont été parmi les premiers à réfléchir à l'informatisation de leur discipline, et ont souvent joué un rôle pionnier en la matière²¹. Nous espérons donc une appropriation de cet outil. Son format ouvert rend possible son application à différents usages bibliographiques plus ou moins spécialisés, aussi bien qu'à un usage mutualisé de la future application web afin d'obtenir progressivement un panel important d'auteurs et de textes traités, à partir des référentiels fournis et mis à jour par le centre Jean Pépin selon les principes décrits ci-dessus.

Julie Giovacchini (Centre Jean Pépin – UMR 8230 CNRS - ENS)

19 <http://iphi.hypotheses.org/372> Giovacchini « Qu'est-ce que c'est que cela s'il vous plaît »

20 <http://iphi.hypotheses.org/531>

21 cf Terras, http://www.ucl.ac.uk/infostudies/melissa-terras/research/Chapter_10_Terras.pdf

2. IPhIS - Retour sur quelques aspects techniques

Pour différentes raisons exposées plus haut, complexité des notions de texte antique, d'auteur antique, complexité de l'histoire de la réception de ces textes, précision variable de certaines connaissances (ex : les dates, dont la précision varie du jour au siècle), le modèle de données du projet IPhIS pouvait potentiellement devenir très 'touffu', avec tous les inconvénients de cette complexité tels une maintenance malaisée, des problèmes prévisibles de performances de l'implémentation web, sans garantie de couvrir réellement et finement la totalité du domaine.

Ce constat nous a amené à faire le choix de mener simultanément la conception du modèle de données et la conception-développement de l'interface de saisie associée à ce modèle (ou à son implémentation en base de données, pour être plus exact).

2. 1. La base de donnée : le modèle de données et son implémentation

Le modèle actuel (voir schéma) est un compromis entre la représentation fidèle des notions décrites par les philologues du projet, la complexité du modèle (donc les performances de son implémentation) et la nécessité de faciliter la saisie pour les rédacteurs (plus le modèle est complexe, plus la saisie l'est aussi).

Parmi les simplifications dont la pertinence est apparue en cours de conception, on peut mentionner (liste non-exhaustive) :

- la notion de texte antique, regroupés dans une table unique quelque soit leur nature et leur structure (fragment, florilège,...),

- la structure (et la normalisation du format) des dates, stockées dans un champ chaîne de caractère protégé par l'expression régulière suivante : `"/^(avant)?(vers)?(après)?(Inc\ .)?[0-9]{1,4}(e)?(r)?(s\.)?(a[pv]r?\ . J\.-C\.)? ?\-(avant)?(vers)?(après)?[0-9]{0,4}(e)?(r)?(s\.)?(a[pv]r?\ . J\.-C\.)?$/`

- les auteurs humanistes de la Renaissance, conservés avec les auteurs modernes et identifiés par un simple drapeau,

- les controverses d'attribution, représentés par un booléen (oui/non) et l'identité de l'attributeur

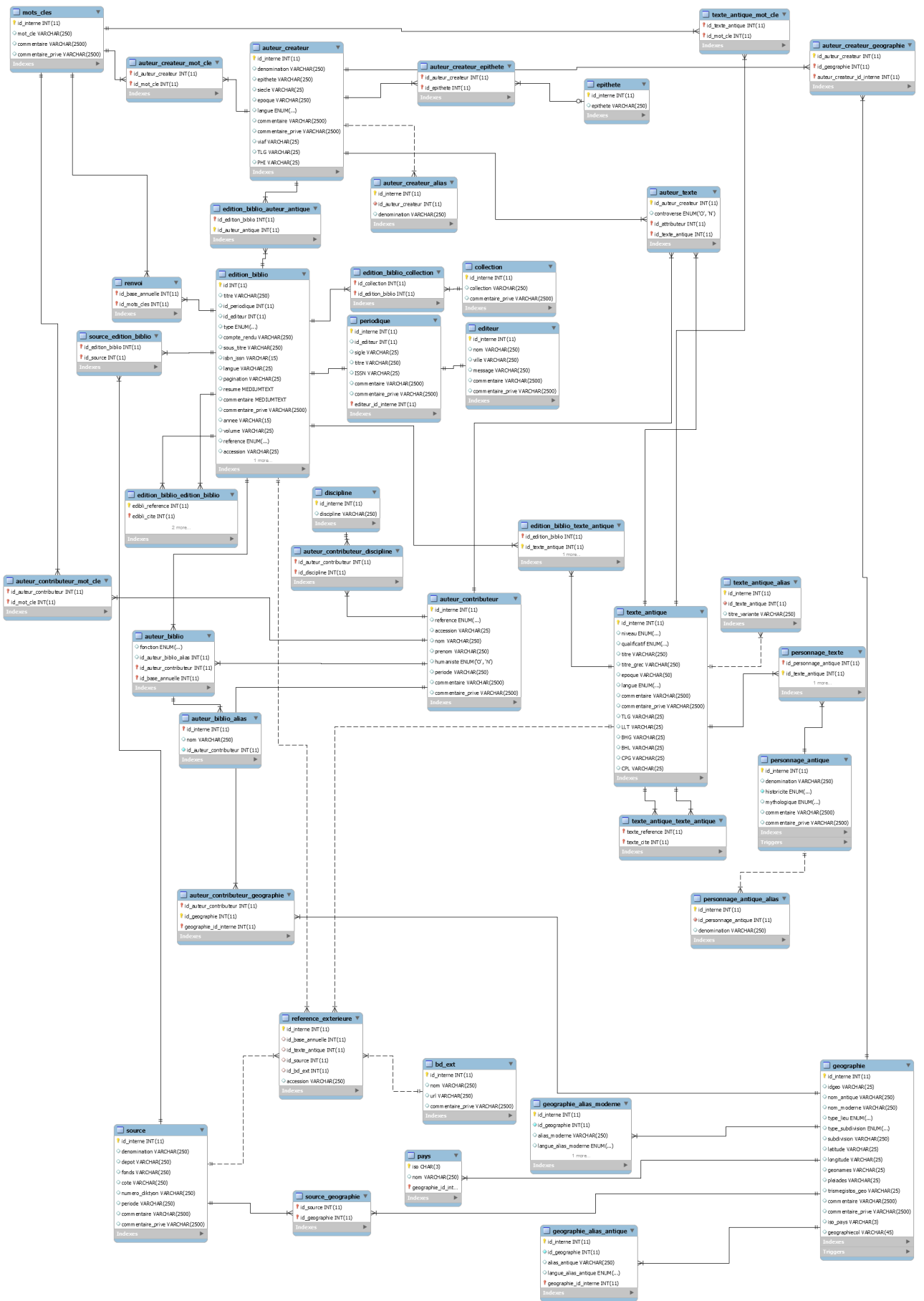


Fig. 1 - Modèle de données IPHIS

Le schéma comporte 42 tables, appartenant à 4 grands domaines : l'auteur antique, l'auteur moderne, le texte antique, la réception moderne (ou édition) de ce texte. 5 ensembles auxiliaires : personnage antique (lié au texte antique), les sources (supports physique, ie manuscrits, des textes anciens), la géographie (lieux d'exercice des auteurs antique et modernes, ainsi que localisation des sources), les mots-clés (indexation des différents domaines) et les références extérieures (bases de données liées) complètent l'ensemble.

Auteur antique :

Il s'agit de l'auteur du texte antique. Antique est à prendre au sens large, les plus récents remontent aux IXe-Xe siècles (Byzantins). C'est la raison pour laquelle la table a été nommée **auteur_createur**.

Outre ses attributs représentés dans sa table (**auteur_createur**), l'auteur peut être affublé de une ou plusieurs **épithètes** (exemple : *Agathangelus* dit *historicus*).

Il peut également avoir des alias (exemple : *Alphuis Auitus* se trouve également sous la forme : *Alfius Auitus*).

On le localise – **géographie** - (*Acacius Caesariensis* est d'*Alexandrie*) et il est indexé par des **mot-clés**. Nous verrons sa relation avec la réception des ses textes (**edition_biblio**) dans ce dernier domaine

Texte antique :

Sans surprise, il s'agit du texte ancien.

Sa classification, fort complexe, est représentée dans sa table **texte_antique**, dans les champs niveau (ex : *rassemblement*) et qualificatif (ex : *varia*) par simplification. Il peut y avoir relation entre deux lignes de cette table, si un texte est inclus dans un autre (**texte_antique_texte_antique**). Par exemple *De sensu et sensibilitas* est un des traités (de *Aristoteles Stagiritis*) regroupés dans *Parua naturalia*, les deux textes étant présents dans la base.

Les liens vers les grandes bases de références, telles par exemple le TLG (Thesaurus Linguae Graecae) ou la BHG (Bibliotheca Hagiographica Graeca) ont été également regroupés dans la table **texte_antique** par simplification (au lieu de figurer dans **reference_exterieure**) d'une part, mais aussi pour tenir compte des références qui ne sont pas en ligne (cas de la BHG).

Comme les auteurs, les textes antiques peuvent être connus sous plusieurs titres (**alias**).

Hormis les grandes bases présentes dans la table **texte_antique**, ils peuvent être référencés dans d'autres bases en ligne (**reference_exterieure**).

Les textes antiques peuvent faire mention de **personnage(s) antique(s)**.

Auteur moderne :

Il s'agit de l'auteur d'une édition, d'une réception ou d'une étude bibliographique d'un ou de plusieurs

textes anciens. Ces auteurs 'modernes' peuvent remonter au VIe ou VIIe siècles pour les cas extrêmes. C'est la raison pour laquelle sa table est nommée **auteur_contributeur** (et pas auteur_moderne).

Ces auteurs pratiquent une (ou des) **discipline(s)**, sont localisés - **géographie** - et sont indexés par **mot-clés**.

Ils sont auteurs d'ouvrages (**auteur_biblio**), pour lesquels ils peuvent avoir des **alias** et avoir des **fonctions** particulières (*éditeur* ou *traducteur* par exemple).

Edition_biblio, la réception moderne :

Il s'agit de l'ensemble des ouvrages présents dans la base IPhis, écrits par des auteurs contributeurs (modernes), traitant des textes antiques.

Les informations bibliographiques classiques (titre, isbn, pagination,...) sont dans **edition_biblio**.

Les ouvrages peuvent appartenir à des **collections** ; il peut également s'agir d'articles de **périodiques**.

Une réception peut faire partie d'une autre réception (**edition_biblio_edition_biblio**), une contribution d'ouvrage collectif par exemple

Les réceptions peuvent être liées à des **sources**, ainsi qu'à des **références extérieures** (liens vers d'autres bases de données).

On peut les indexer par **mot-clés** (**renvoi** mot_cles)

L(es) auteur(s) sont accessibles par **auteur_biblio**

Le(s) textes anciens objets de la publication sont liés par **edition_biblio_texte_antique**

Le(s) auteurs anciens sont liés par **edition_biblio_auteur_antique**. Pourquoi ce lien ? Pour limiter la taille des jointures pour des requêtes que l'on peut supposer fréquentes d'une part, et pour pouvoir modéliser des publications portant sur un auteur ancien, sans porter sur un de ses textes en particulier.

On passera plus rapidement sur les listes auxiliaires :

Le caractère historique ou mythologique des **personnages antiques** sont précisé, ainsi que la nature du personnage mythologique (divinité, héros,...).

Les **sources** sont les supports physiques (manuscrits,...) connus des textes anciens, avec leurs références archivistiques.

La **géographie** référence les lieux d'intérêt pour les auteurs (anciens et modernes) et pour les sources. On conserve entre autre les noms antique et moderne, les coordonnées GPS (latitude et longitude), les numéros d'accession des grandes bases géographiques geonames, trismegistos geo (pour les lieux anciens), pleiades. Un numéro d'accession unique est créé pour chaque lieu.

Les **mot-clés** sont déterminés par les rédacteurs d'IPhis. L'utilisation d'autorités normalisées seraient naturellement souhaitable (en particulier pour préparer une interopérabilité vraisemblablement incontournable à terme, comme nous le verrons dans les perspectives) mais, sauf erreur, il n'existe pas aujourd'hui de liste d'autorités suffisamment détaillée ou spécialisée dans le domaine de la philologie

classique.

Enfin les **références extérieures** répertorient des liens sur une dizaine de bases extérieures, en plus des liens déjà hébergés par les tables **auteur_antique** (3 bases), **geographie** (3 bases), **sources** (1 base), **texte_antique** (6 bases),... On arrive au total à plus de vingt bases extérieures référencées.

Implémentation :

Nous souhaitions à l'origine utiliser un SGBD libre et standard, principalement pour assurer la pérennité du projet. Notre choix s'est porté sur mysql (bien que son caractère libre soit discutable) pour deux raisons : mysql est un des principaux standards et le framework choisi pour l'interface rédactionnelle (voir plus loin) supporte nativement mysql. Nous avons donc implémenté le modèle dans une base de données mysql.

La volumétrie des tables principales à ce jour (mars 2017) est présentée dans le tableau ci-dessous :

table	Nombre d'items
auteur_contributeur	1574
auteur_createur	2298
edition_biblio	54
texte_antique	315

On peut observer que les rédacteurs se sont concentrés sur les listes d'auteurs (anciens et contributeurs) pour l'instant. Les réceptions (edition_biblio) et textes anciens (texte_antique) sont en cours de constitution.

Volumétrie des listes auxiliaires :

table	Nombre d'items
geographie	22
mot_cle	486
personnage_antique	3
source	12
reference_exterieure	132

Seuls les mot-clés sont constitués, les autres listes auxiliaires sont eux aussi en cours de constitution.

2. 2. L'interface

2. 2. 1. L'interface rédactionnelle (ou back-office)

Le choix, dès le départ du projet, a été fait de privilégier des technologies web et des solutions ouvertes, afin d'assurer la pérennité d'IPhiS.

L'évolution simultanée du modèle de données avec le développement de l'interface de saisie (certains problèmes de représentation n'ont été identifiés qu'à la lumière des données réelles) a amené des contraintes fortes sur le choix de l'outil de développement. L'idéal (peut-être le graal...) était de pouvoir développer en un temps raisonnable une interface capable de s'adapter automatiquement à un changement du modèle.

Nous avons donc choisi un framework générateur d'interface (xataface) dont l'ergonomie, plus 'user-friendly' que les outils d'administration type phpmyadmin également envisagés un temps, était toutefois largement insuffisante pour les besoins identifiés et exprimés par l'équipe. Un important travail de personnalisation de l'interface générée (par paramétrage et surcharge, deux mécanismes proposés par xataface) a donc été réalisé.

Xataface (<http://xataface.com> et <https://github.com/shannah/xataface>) est un framework php basé sur mysql permettant de générer des interfaces de mise à jour et de consultation de bases de données. Quatre modes de personnalisation sont offerts pour adapter l'interface générée nativement par le framework aux besoins spécifiques du projet : la paramétrisation, la surcharge des feuilles de styles, la surcharge des fonctionnalités, via les classes déléguées (delegate classes) et la mise en place de modules spécifiques.

Paramétrisation :

Les paramètres s'établissent globalement pour tout le site par la mise en place à la racine du site des fichiers conf.ini (base de données, tables, gestion des autorisations d'accès,...) et actions.ini (fonctionnalités à activer, ajouter, retirer, ex : import en masse ou suppression de records – ce fichier peut également être présent au niveau des tables, pour une action spécifique à une table).

Il est également indispensable de décrire la base de données, en particulier les relations entre tables et les formulaires d'accès aux données (fichiers relationships.ini, fields.ini,...).

Pour plus de détails, se reporter au wiki de xataface : <http://xataface.com/wiki/>

Le projet IPhis a nécessité la description explicite des relations de 27 tables (celles accessibles directement aux rédacteurs). Les formulaires de saisie ont été adaptés pour 16 tables.

Surcharge fonctionnelle :

Des adaptations plus importantes peuvent être obtenues par la création de fonctions php (fonction d'import csv pour des tables spécifiques par exemple).

Les 16 principales tables ont nécessité la mise en place de fonctions déléguées.

Module spécifique :

Le module export_ris a été développé et mis en place pour les exports zotero des bibliographies.

Surcharge des feuilles de style :

Pour obtenir la présentation souhaitée du site, il suffit de créer (s'il n'existe pas déjà) un dossier css à la racine du site où on place les feuilles de style locales. Les personnalisations pour IPhIS représentent de l'ordre de quelques centaines de lignes css.

Dernier point, les pages sont générées par des templates (smarty). Ces templates peuvent être modifiés ou remplacés par des templates spécifiques. Pour IPhIS, une quinzaine de templates spécifiques sont nécessaires.

Retour d'utilisation du framework xataface :

Avantages :

Cette interface s'adapte remarquablement bien aux modifications mineures du modèle. En général, seules quelques retouches aux feuilles de style doivent être apportées, avec éventuellement le paramétrage des modifications des relations du modèle.

Inconvénients :

Les principaux inconvénients de cet outil sont la faible communauté d'utilisateurs et la petite taille de l'équipe de développement (une seule personne semble-t-il). De plus, la documentation est rare et déficiente, ce qui entraîne parfois des difficultés de développement. D'où la question de la pérennité de cet outil. Ce dernier problème est toutefois moins grave qu'il n'y paraît : xataface est un outil totalement open-source, ce qui offre la possibilité de maintenir en local une application reposant sur ce framework, telle que Iphis, même en cas d'arrêt de support et de développement communautaire (malgré quelques difficultés prévisibles liés au volume de travail...).

2. 2. 2. L'interface rédactionnelle (ou front-office) - perspectives

L'étape suivante du projet IPhIS va être la conception, la réalisation et le déploiement de l'interface publique. Les contraintes sont de fournir des possibilités de navigation multiples permettant des expériences utilisateur différenciées selon le profil de ce dernier : accès par auteurs, requêtes à facettes,

navigation géographique, multilinguisme...

Concernant l'outil de développement, après avoir envisagé l'utilisation de CMS (type Drupal), nous nous orientons plutôt vers un développement de site dynamique (php/mysql), plus souple et plus pérenne.

2. 3. Conclusion et perspectives

Le modèle de données d' IPhiS est désormais robuste et stabilisé, et son interface rédactionnelle en passe d'être mise en production, après la correction de quelques bogues résiduels.

Les perspectives à court terme sont la mise en place d'une gestion des droits d'accès plus fine et mieux sécurisée pour permettre un éventuel usage collaboratif de l'interface rédactionnelle. Parallèlement, nous allons commencer à concevoir l'interface publique d'accès aux données.

À plus long terme se profile l'enjeu de l'insertion d' IPhiS dans le réseau des bases de données ouvertes du domaine des disciplines des antiquisants. Plusieurs voies méritent d'être explorées, telles que le monde des données liées du web sémantique, l'échange automatisé de liens inter-bases par la normalisation des numéros accessions et des indexations, ou encore l'extension du modèle de données afin qu'il puisse supporter des applications aux thématiques proches. Ce dernier point présente un intérêt particulier pour le centre Jean Pépin : plusieurs projets similaires du Centre, reposant sur la bibliographie des sources antiques, gagneraient à utiliser à terme un modèle générique commun. Cela permettrait d'offrir à la communauté un outil puissant au service de la recherche sur l'antiquité.

Bernard Weiss (CLT – UPS 2259 CNRS)