

---

## L'accès vu comme une contrainte et un défi : le processus de naissance d'une base de données

Dina Bacalexi

dina.bacalexi@cnrs.fr

National Center for Scientific Research (CNRS), France

Pinelopi Skarsouli

pinelopi.skarsouli@cnrs.fr

National Center for Scientific Research (CNRS), France

---

Une base de données organise la connaissance en vue de son utilisation par un public donné. Cette définition trop rapide et lacunaire ne tient pas compte de la complexité du processus de création, ni de la véritable nature d'une base numérique, qui, pour nous, doit être le fruit d'un dialogue entre la partie scientifique et la partie technique du projet. Nous allons étudier, pour deux bases antiquisantes, le processus de passage d'une approche textuelle, libre de toute contrainte, à une autre basée sur la structuration, c'est-à-dire la transcription en structuration formelle des décisions prises par les analystes des textes antiques.

Dans ces deux cas, un nouvel accès, résultat de la restructuration des outils, vise un public nouveau tout en servant mieux l'ancien :

- 1) l'abandon d'un modèle conçu pour l'imprimé, où la recension annuelle des données pour produire le volume papier de *l'Année Philologique*, accompagné d'une transformation en site web, constituait pendant des décennies le but recherché : l'objectif était un changement total du contenu et de la structure, pour obtenir une véritable base de données, tout à fait différente ([IPhiS](#), la nouvelle base philologique en construction).
- 2) la transformation d'un modèle sans volume papier paru, qui n'était pas pour autant une construction « native numérique » mais un simple export html périodique d'un ensemble

de données thématiques : l'objectif était la reprise des anciennes données avec une modélisation nouvelle permettant d'obtenir une nouvelle base, qui continuerait l'ancienne seulement en ce qui concerne le contenu (le cas du [Répertoire des sources philosophiques antiques](#)).

La question de l'accès libre aux données s'est posée à nous comme une contrainte. Notre travail d'antiquisants analystes de sources grecques et latines philologiques ou philosophiques (au sens large des termes) a croisé celui des développeurs d'applications numériques quand nous avons commencé à transformer les deux outils bibliographiques précités. Il existe plusieurs façons d'envisager la perméabilité et l'articulation scientifique-technique ; nous étudierons comment le dialogue permanent direct permet un meilleur accès aux données d'une base.

Quand un modèle a fait ses preuves depuis près de 80 ans et que sa notoriété est installée dans la communauté, la difficulté est d'imposer un modèle nouveau quant à la pertinence scientifique et le traitement de l'information. La contrainte est double : passer d'un logiciel propriétaire (4D) dont l'adaptation convenait tout à fait à la bibliographie papier et à son export en ligne, à une solution de logiciel libre qui apporterait cependant la même richesse et finesse d'indexation que 4D et créerait une véritable base de données numérique. La question ne concernait pas seulement le développement de l'outil nouveau, mais aussi le type des données à implémenter dans la base. Notre questionnement concernait aussi le fait même de rendre la base accessible librement sur le web : au moment de la prolifération de l'information grâce aux moteurs de recherche généralistes du type *Google scholar*, quelle serait la « plus-value » d'une énième base ? C'est pourquoi nous avons commencé à réfléchir comment les entrées de la base pourraient pointer vers des données textuelles littéraires extérieures en XML/TEI, afin que nos bases ne soient pas simplement des *thesauri* bibliographiques riches et bien ordonnées, mais sans lien avec les sources qu'elles sont censées analyser. Cette question est pour l'instant pour nous au stade de la réflexion théorique.

La liberté d'accès, dans le cadre des disciplines rares comme les nôtres, pose la question de la protection contre le copiage, le plagiat, l'imitation et donc l'appropriation de notre travail. C'est pourquoi, dès le début du projet, il est important de choisir une licence adaptée, par exemple de la famille *Creative Commons*.

Il faut aussi envisager l'*accessibilité effective* des données *via* l'interface utilisateur et le respect des standards du web.

Nous constatons trop souvent que les bases de données, sites collaboratifs, plateformes, éditions etc. émanent de programmes limités dans le temps, au financement et au personnel précaire, ce qui impacte la réalisation des projets dans les délais impartis. Or, si ces bases répondent à une demande d'accès à des données fiables et de qualité contrairement à la masse incontrôlée offerte par les moteurs généralistes du web, elles ont vocation à durer et ont besoin d'une mise à jour régulière de leur contenu et de leur infrastructure logicielle pour remplir leur mission. Les changements de technologie (internes ou imposés par l'écosystème) peuvent comporter des risques pour les projets et leurs utilisateurs finaux. La disparition des outils entraîne souvent celle du savoir et du savoir-faire des personnels.

La nécessité d'obéir aux injonctions du court terme impacte le travail de renouvellement des anciennes bases : les délais de transformation sont plus longs que prévu, puisqu'il faut d'abord passer par une expression de nouveaux besoins et ensuite par une modélisation tenant compte des choix scientifiques et traduisant ceux-ci en solutions techniques ; il faut ensuite expérimenter, adapter, affiner. Un tel fonctionnement est à l'opposé d'une demande de solution rapide clés en main.

L'implémentation des notices dans une base en flux continu sur le web (en vue de leur moissonnage par les moteurs de recherche en temps réel), le rêve d'exhaustivité, contribuent à alimenter l'idée de liberté totale individuelle qui va à l'encontre du mode de travail collectif. Le risque est d'instaurer une nouvelle forme de dépendance et de contrôle généralisé, voire d'accélération des cadences et de fin programmée de toute vérification de la pertinence scientifique. C'est tout à fait différent d'une base de données participative et basée sur la contribution de la communauté.

Le travail sur le contenu d'une base dépend aussi de l'élaboration d'un cahier de charges, de la modélisation des données (qui définit à son tour le type d'accès aux données dans l'immédiat mais aussi dans l'avenir), des formulaires permettant de la remplir, de son interface de consultation et de la constitution d'une communauté d'utilisateurs. La question est la (re)structuration et la (re)présentation des données sous une nouvelle forme en ligne, qui tienne compte des standards actuels (libre accès, interopérabilité etc.). La normalisation des données et la réduction de la redondance de l'information sont des gages d'une base de

données pleinement accessible et exploitable par ses utilisateurs.

## Bibliography

- Bénel, A.** 2014. "Quelle interdisciplinarité pour les 'Humanités numériques'?" *Les Cahiers du numérique* 10/4: 103-132. <http://www.cairn.info/revue-les-cahiers-du-numerique-2014-4-page-103.htm>
- Berra, A.** 2015. "Pour une histoire des humanités numériques", *Critique* 8 (n° 819-820): 613-626. <http://www.cairn.info/revue-critique-2015-8-page-613.htm>
- Burnard, L.** 2012. "Du *literary and linguistic computing* aux *digital humanities*: retour sur 40 ans de relations entre sciences humaines et informatique" in *Read/Write Book: une introduction aux humanités numériques*, Pierre Mounier (dir.), Marseille. <http://books.openedition.org/oep/242>
- Clivaz, Cl., Dilley, P. and Hamidovic, D.** ed. 2016. (in collaboration with A. Thomas), *Ancient Worlds in Digital Culture*, Leiden: Boston.
- Crane, G., Seales, B. and Terras, M.** 2009. "Cyberinfrastructure for Classical Philology". *Digital Humanities Quarterly* 3/1 [www.digitalhumanities.org/dhq/vol/3/1/000023/000023.html](http://www.digitalhumanities.org/dhq/vol/3/1/000023/000023.html).
- Flanders, J.** 2009. "The Productive Unease of 21st-century Digital Scholarship." *Digital Humanities Quarterly* 3/3 [www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html](http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html).
- Le Deuff, O. ed.** 2014. *Le temps des humanités digitales. La mutation des sciences humaines et sociales*, Limoges.