



HAL
open science

Du catalogage au travail coopératif: nouveaux enjeux numériques pour le patrimoine séculaire d'Éthiopie

Anaïs Wion

► **To cite this version:**

Anaïs Wion. Du catalogage au travail coopératif: nouveaux enjeux numériques pour le patrimoine séculaire d'Éthiopie. "La société de l'information et ses enjeux", Actes du colloque de bilan du programme interdisciplinaire "Société de l'information" 2001-2005, May 2005, Lyon, France. pp.296-299. hal-00450504

HAL Id: hal-00450504

<https://hal.science/hal-00450504>

Submitted on 28 Sep 2022

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.

[WION, A. "Du catalogage au travail coopératif : nouveaux enjeux numériques pour le patrimoine séculaire d'Éthiopie", *La société de l'information et ses enjeux. Actes du colloque de bilan du programme interdisciplinaire "Société de l'Information" 2001-2005. ENS-LHS-Lyon, 19-21 mai 2005*, dir. J-L. Lebrave, CNRS, 2005, pp. 296-99.]



Du catalogage au travail coopératif : nouveaux enjeux numériques pour le patrimoine séculaire d'Éthiopie

Nom du porteur : WION Anaïs

Laboratoire de rattachement :

Mutations Africaines dans la Longue Durée, UMR 8054

Thématique de recherche : manuscrits et traditions orales d'Éthiopie

Noms des partenaires et Laboratoires de rattachements des partenaires :

D. OGILVIE, Archives Nationales, Section du XIX^e siècle.

C. BOSCH-TIESSÉ, Institut d'Études Africaines, UMR 6124.

ATKELT ASSEFA et HELINA SEYUM, National Archives and Library of Ethiopia, (NALE)

Wendy MACKAY et J.D. FEKETE, Institut National de Recherche en Informatique et en Automatique (INRIA Futurs)

O. TOURNY, Langues Musique Société, UMR 8099.

Le patrimoine documentaire éthiopien

L'Éthiopie, pays d'Afrique Orientale de culture chrétienne orthodoxe, possède un patrimoine documentaire d'une grande richesse. La tradition manuscrite, ininterrompue du XIII^e siècle au XX^e siècle, a généré un corpus unique. Ces manuscrits, essentiellement religieux, contiennent aussi les archives de ces institutions politico-religieuses que furent monastères et églises. La plupart des codex sont conservés en Éthiopie, dans les monastères des hauts plateaux ainsi que dans les institutions patrimoniales. En Europe, de nombreuses bibliothèques ont constitué d'importantes collections¹.

Par ailleurs, la tradition orale transmise en langue amharique dans les institutions religieuses vient éclairer et compléter ces écrits. Cette diglossie guèze (langue de l'écrit) / amharique (langue de l'oral) caractérise la diffusion des connaissances dans l'Éthiopie chrétienne, aussi un travail de recherche sur le patrimoine documentaire éthiopien se doit-il d'en tenir compte. L'idée d'éclairer l'une par l'autre les sources écrites contenues dans les manuscrits et les sources orales est relativement nouvelle dans le champ des études éthiopiennes². Pourtant, il existe un *continuum* entre, par exemples, les chartriers anciens et

¹ BEYLOT, R. ET RODINSON, M. *Répertoire des bibliothèques et des catalogues de manuscrits éthiopiens*, C.N.R.S. éd, Paris et Brepols Turnhout, 1995.

² La découverte puis l'exploration de cette méthodologie a été réalisée lors du travail doctoral de A. WION, *Aux confins le feu, au centre le Paradis. Qoma Fasilādas, un monastère royal dans l'Éthiopie du XVII^e siècle*, thèse de doctorat en Histoire de l'Université de Paris-1 Sorbonne, 2003, 2 vol. (à paraître aux Presses Universitaires de la Sorbonne, 2006).

leurs commentaires oraux, ou encore entre des textes hagiographiques et leurs versions locales. Ce rêve du médiéviste est en Éthiopie une réalité.

L'utilisation aujourd'hui incontournable des supports virtuels offre la possibilité de se dégager du compartimentage traditionnel entre les types de sources pour envisager leur traitement de façon transversale.

Les Archives Nationales et la Bibliothèque Nationale d'Éthiopie (NALE)

La NALE (National Archives and Library of Ethiopia) est une agence indépendante au sein du Ministère de la Culture éthiopien, qui bénéficie depuis 1999 de l'appui d'une législation spécifique³ : celle-ci donne une définition très complète du "patrimoine documentaire" d'Éthiopie et des missions de collecte, conservation et communication qu'elle entraîne. De plus, la loi rappelle que ce patrimoine s'étend aux documents produits en Éthiopie et conservés hors du territoire éthiopien (manuscripts appartenant aux bibliothèques européennes, productions musicales, etc.) et prescrit aux deux institutions qui composent la NALE (Archives nationales et Bibliothèque nationale), d'en faciliter l'accès. C'est ce texte fondateur des missions de la NALE qui a suscité, dans le cadre d'une coopération scientifique et technique engagée depuis 2001, la naissance de ce projet orienté selon deux axes :

- créer un outil permettant de cataloguer les ressources manuscrites⁴ et orales⁵ de la NALE
- inventer un espace de travail numérique permettant d'enrichir ce premier outil par le travail des chercheurs

Depuis juillet 2003, la réflexion et le partage d'expérience nous a mené à explorer les diverses facettes de ce champs d'étude : outils informatiques, langages pour l'encodage des données, interopérabilité (possibilité d'échanger des données avec d'autres systèmes)... L'univers des outils de recherche patrimoniaux étant en pleine mutation⁶, cette première phase du projet financé par le programme IDA est une étude de faisabilité et a permis d'établir un cahier des charges. La NALE disposera à partir de l'automne 2005 d'un nouveau bâtiment et c'est à ce moment que le cahier des charges commencera à être utilisé par les équipes compétentes.

1^{er} volet : un catalogue multi-sources pour la NALE

L'encodage des métadonnées : construire la structure du catalogue

Le langage de description utilisé est l'XML (eXtended Markup Language), désormais considéré comme le langage le plus souple et le plus pérenne pour la structuration des métadonnées⁷. C'est un langage de balises (*markup*) qui pour pouvoir échanger les données qu'elles contiennent doivent être définies par une norme : une DTD (Document Type Definition).

³ Voir le site de la NALE : <http://www.nale.gov.et> . Le texte de la proclamation y est téléchargeable.

⁴ La collection de manuscrits compte 800 volumes, ainsi que 10 000 volumes microfilmés.

⁵ Cette collection est encore embryonnaire. Une politique d'acquisition et de collecte sera mise en place lorsque le nouveau bâtiment lui offrira un espace. Elle pourra bénéficier du projet de O. Tourny (LMS-CNRS), financé par l'UNESCO, voir www.vjf.cnrs.fr/lms/actu.htm

⁶ L'excellent weblog Figoblog effectue un travail de veille technologique en matière de bibliothéconomie particulièrement passionnant : <http://www.figoblog.org/>

⁷ Les métadonnées sont l'ensemble des données descriptives qui elles-mêmes structurent les données contenues dans l'instrument de recherche.

L'encodage des descriptions des manuscrits sera effectué selon la DTD EAD (Encoded Archival Description)⁸, mise au point à l'origine pour la structuration et la description des inventaires d'archives et qui s'impose désormais comme le standard de description du patrimoine documentaire, au détriment pour ce qui concerne les manuscrits de la DTD MASTER élaborée pour la description des manuscrits anciens mais qui faute de financement, a été abandonnée⁹. Ce choix obéit à des critères de pérennité et d'interopérabilité organisationnelle : le langage (XML) et sa grammaire (EAD) s'inscrivent dans des pratiques internationales communes. Néanmoins, il est nécessaire d'adapter l'EAD aux caractéristiques des manuscrits éthiopiens. Pour cela, Anaïs Wion participe au groupe de réflexion¹⁰ qui, au sein de la BnF, travaille à l'adaptation de cette DTD aux fonds de manuscrits.

Pour ce qui est des corpus de documents sonores, peu d'expériences ont été menées dans ce domaine jusqu'à maintenant mais rien ne s'oppose à une utilisation de l'EAD pour structurer les descriptions. Un travail comparé avec les recherches du département des archives de l'Université Libre de Bruxelles¹¹ a apporté des pistes de réflexion fécondes.

Une utilisation homogène de l'EAD permet d'établir des passerelles entre ces deux corpus. L'avantage d'un tel catalogue ou outils de recherche c'est de pouvoir mettre sur un même plan sources écrites, souvent anciennes, et sources orales, contemporaines, et ainsi de sortir d'une logique de collection pour faire émerger des logiques d'institutions productrices, par exemple, ou bien encore pour mener des études thématiques. Nous y reviendrons.

Cet encodage et la consultation des catalogues se fait par le biais d'une suite logicielle en OpenSource qui devra avoir les fonctions suivantes, inspirées de l'expérience des applications MidosaXML¹², EdiTEI¹³ et PLEADE¹⁴ : gestion des polices éthiopiennes¹⁵ (unicode) et des polices de transcription (signes diacritiques) ; saisie des données possible à travers deux sortes d'écran : rédaction directe en XML ou masque de saisie (interface proposant une grille de saisie) ; création d'un thésaurus spécifique au contexte éthiopien permettant l'harmonisation des graphies ; outils de validation permettant le contrôle de l'encodage et la validité des liens hyper-texte ; exportation pour l'édition en HTML, en RTF ainsi qu'en XML suivant des feuilles de style ouvertes et adaptables ; exportation d'une sélection de balises des métadonnées en Dublin Core qualifié et simple pour l'intégration de celles-ci à un entrepôt OAI.

La numérisation : un vaste chantier

La NALE a mené en parallèle de ce programme une action de coopération avec le Japon qui lui a fourni, en 2004, le laboratoire informatique nécessaire pour mener à bien la chaîne de numérisation de ces documents écrits. Rappelons que l'informatisation des outils de recherche en archivistique et en bibliothéconomie est un phénomène récent et c'est en grande partie la possibilité de numériser les collections qui est, depuis dix ans, au centre de ce renouveau. Or la NALE avait déjà eu une première expérience de cette dissociation entre contenu et support car, durant les décennies 1970 et 80, une vaste campagne de microfilmage

⁸ Voir le site officiel : <http://www.loc.gov/ead/> ainsi que le guide EAD du Centre Historique des Archives Nationales : <http://www.archivesnationales.culture.gouv.fr/chan/chan/infopro/fr/>

⁹ Voir le manuel d'utilisation datant de 2001 : <http://www.tei-c.org.uk/Master/Reference/oldindex.html>

¹⁰ atelier issu du groupe AFNOR CG46/CN357/GE3

¹¹ <http://www.ulb.ac.be/rech/inventaire/unites/ULB510.html>

¹² <http://www.midosa.de/>

¹³ Cet outil a été développé sous la direction de Marjorie Burghart : <http://www.mom.fr/bdd/editor/>

¹⁴ <http://www.pleade.org/>

¹⁵ à propos des recherches faites dans le domaine de l'utilisation informatique du syllabaire éthiopien, voir : Senamirmir project : <http://www.senamirmir.com/> et la Ge'ez Frontier Foundation : <http://www.geez.org/>

a été menée auprès des églises, constituant une collections de dix mille manuscrits microfilmés¹⁶. C'est cette collection, ainsi que quelques huit cents manuscrits, qui va être dorénavant numérisée.

Ces images vont pouvoir être incorporées à notre catalogue, donnant lieu à un deuxième niveau d'encodage.

L'encodage des contenus

Après l'encodage des métadonnées, il est possible d'encoder les données elles-mêmes, si l'on dispose des « équivalents » virtuels des documents sources. Les modalités de cet encodage dépendent bien sûr de ce qui est considéré comme « équivalent ». Si les manuscrits sont disponibles sous forme d'images numériques, celles-ci peuvent permettre la consultation directe des documents et être appelées à partir d'une interrogation du catalogue (comme c'est le cas actuellement de la base Gallica, BnF). Néanmoins, l'image reste un support figé, et pour encoder du texte il faut que celui-ci soit transcrit¹⁷. C'est l'objet d'un type de travail différent de celui du catalogueur. De même, les documents sonores peuvent exister sous format(s) numérique(s), mais là encore l'encodage de contenu ne peut se faire qu'à partir d'une transcription et/ou d'une traduction¹⁸.

L'encodage des contenus peut être effectué avec la DTD TEI (Text Encoding Initiative)¹⁹. La DTD TEI a déjà été appliquée à l'encodage des manuscrits éthiopiens à l'université allemande de Mayence²⁰. C'est là que nous entrons dans le deuxième volet de ce projet.

2^{ème} volet : la coopérative numérique pour les études éthiopiennes

L'instrument de recherche créé pour la NALE doit rester gérable par son personnel, et ne pas dépasser un travail de catalogage déjà conséquent par son nombre d'entrées. Pourtant, il serait possible de dédoubler cet instrument de recherche pour l'ouvrir aux chercheurs, en créant une « coopérative numérique » fonctionnant avec la même base d'outils logiciels. Cette suite d'outils en Open Source peut être utilisée en monoposte, en réseau ou en ligne. Elle permet d'importer les données du catalogue de la NALE ainsi que d'y entrer des descriptions de documents complémentaires ainsi que leurs contenu numérisé. Ces descriptions harmonisées permettent à un chercheur de se créer son propre corpus d'étude de façon à la fois personnalisée, en développant particulièrement les points qui lui sont utiles, tout en restant dans une norme partagée par la NALE et toute personne utilisant la coopérative. Il peut donc échanger ses données avec d'autres personnes utilisant ces outils. Au terme d'une recherche, ces encodages peuvent être confiés à l'entrepôt de la coopérative et devenir accessibles à tous²¹.

¹⁶ C'est la collection EMMML (Ethiopian Manuscripts Microfilm Library). La moitié de ce corpus microfilmé a été catalogué et est aussi déposé aux USA, à la Hill Monastic Manuscript Library : <http://www.hmml.org/coll/collection.asp?coll=eth> L'autre moitié n'est déposé qu'à la NALE est reste à décrire.

¹⁷ Les recherches portant sur l'océrisation (Reconnaissance Optique des Caractères) du syllabaire éthiopien laissent prévoir un aboutissement prochain dans ce domaine.

¹⁸ Des recherches sur la reconnaissance vocale sont en cours et l'amharique, se prononçant comme il s'écrit, est utilisé comme démonstrateur par les équipes de recherche dans ce domaine.

¹⁹ <http://www.tei-c.org/>

²⁰ sous la direction du Pr. Dr. Manfred Kropp : <http://www.uni-mainz.de/Organisationen/TLA/info/about.html>

²¹ Dans le champs des études éthiopiennes, un projet d'inspiration similaire mais constituant plus une banque de données qu'un espace de travail est la base *Mäzgbä Seelät*, conçue et gérée par M. Gervers et E. Balicka-Witakowska : <http://128.100.218.174:8080/ethiopia/index.jsp>

Une deuxième chaîne d'outils logiciels pourrait de plus permettre de travailler sur ces documents. Citons en guise d'exemples les fonctionnalités suivantes :

- visualiser et comparer des documents, par exemple pour des études paléographiques ou pour établir une recension
- reconstituer la structure des bibliothèques monastiques, en réunissant les descriptions de manuscrits qui appartiennent à un même lieu et qui furent ensuite dispersés
- annoter des documents ; conserver et publier ses notes
- établir un lexique de termes dans un domaine donné
- construire une bibliographie raisonnée, etc.

Chacun des postes de travail, dont aucun n'a de caractère systématique, trouve d'autant plus d'intérêt s'il est partagé avec une communauté de chercheurs. Une chaîne d'habilitation doit réguler le fonctionnement du *workflow*, ou "flux de travail"²². Le multilinguisme est aussi un facteur à prendre en compte pour une utilisation par la communauté internationale des chercheurs. Anglais, français et amharique seront donc les langues disponibles pour l'affichage de l'interface et des guides d'utilisation.

La tâche de construction de la suite logicielle et de la plate forme collaborative peut sembler ardue, mais il ne s'agit pas forcément d'inventer des outils *ex nihilo*, car de nombreuses recherches sont menées dans ce domaine et aboutissent la plupart du temps à la création d'outils en OpenSource. Ainsi, même si les outils sont souvent créés pour un type de corpus particulier, il est possible d'adapter les outils aux spécificités du corpus éthiopien.

Conclusion

Le financement IDA (2003-2005) a permis de mesurer la faisabilité de ce projet. Pour son premier volet, un cahier des charges a été établi. Pour le deuxième volet, il s'agissait d'étudier les possibilités d'échanges de données numériques et d'intégrer plus étroitement les problématiques des chercheurs à celles des conservateurs. Les chercheurs qui ne sont souvent que destinataires de l'instrument de recherche peuvent (doivent) aussi en être les acteurs. L'ambition de ce projet est ainsi de considérer l'espace virtuel comme un espace neutre et d'y réunir un patrimoine documentaire aussi dispersé que le sont les acteurs qui le conserve et l'étudie. Les enjeux patrimoniaux et de recherche sont alors confluents : tendre vers un meilleur partage des sources et des ressources.

Interventions publiques :

mai 2005 : *Décrire les manuscrits éthiopiens et constituer une base de données sur ces manuscrits*,

Séminaire « **Livre, Manuscrit et Société** », LAMOP (UMR 8589).

novembre 2003 : *La conception d'une base de données pour le patrimoine documentaire éthiopien :*

*un observatoire des liens complexes entre les sources, **Identités composées. Dynamiques linguistiques et religieuses en Afrique***, 24-26 nov. 2003, Centre de Recherches Africaines.

août 2003 : *Les manuscrits éthiopiens et leurs commentaires oraux : une culture millénaire à l'heure de la révolution numérique*, **24^{ème} Université d'été de la communication d'Hourtin, Les fractures numériques Nord/Sud en question : Quels enjeux ? quels partenariats ?**, 25-28

août 2003, Hourtin.

²² C'est un ensemble de procédures ou de tâches exécutées en parallèle ou en série par au moins deux membres d'un groupe pour atteindre un but commun. Le Workflow décrit le circuit et les modes de validation, les tâches à accomplir entre les différents acteurs d'un processus et fournit à chacun des acteurs les informations nécessaires.