



HAL
open science

Dix ans de développement du projet international ”Initiative pour un Conservatoire Numérique du Cunéiforme”

Bertrand Lafont

► **To cite this version:**

Bertrand Lafont. Dix ans de développement du projet international ”Initiative pour un Conservatoire Numérique du Cunéiforme”. Cahier des thèmes transversaux ArScAn, 2015, XII, pp.73-79. hal-02278195

HAL Id: hal-02278195

<https://hal.science/hal-02278195>

Submitted on 11 Sep 2019

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.

DIX ANS DE DÉVELOPPEMENT DU PROJET INTERNATIONAL INITIATIVE POUR UN CONSERVATOIRE NUMÉRIQUE DU CUNÉIFORME*

Bertrand LAFONT

CNRS, ArScAn – HAROC

bertrand.lafont@mae.u-paris10.fr

L'entreprise dont il va être ici question concerne les textes cunéiformes qu'ont produits les civilisations de l'ancien Proche-Orient. Il convient d'abord de dire quelques mots de cette documentation dont on peut rappeler d'emblée qu'elle constitue largement, d'un strict point de vue quantitatif, le plus important corpus de textes que nous a légué l'Antiquité (sans doute plus d'un demi-million de documents inscrits, et des millions de lignes de texte, soit largement plus que toutes les sources classiques grecques et latines réunies).

L'écriture cunéiforme est apparue en Mésopotamie, au sud de l'Irak actuel, vers 3300 avant notre ère. C'est le plus ancien système d'écriture connu ; on en attribue la mise au point aux Sumériens. Le système s'est ensuite rapidement et largement diffusé, jusqu'à être adopté dans toute la Mésopotamie et dans les régions adjacentes ; il y a été utilisé pendant trois millénaires entiers, jusqu'à l'hellénisation du Proche-Orient, et il a servi sur la longue durée à noter des langues très différentes les unes des autres, comme l'akkadien (langue sémitique), le hittite (langue indo-européenne), le sumérien, le hourrite ou l'élamite (trois langues n'appartenant à aucune famille linguistique répertoriée).

Le support de cette écriture est le plus souvent constitué (mais pas toujours) d'une tablette d'argile molle dans laquelle les signes d'écriture étaient incisés avec un stylet (= le calame), avant qu'on ne la fasse sécher et donc durcir. Ce sont les incisions en forme de clous faites par le calame qui ont donné son nom de « cunéiforme » à cette écriture. Ce qui importe ainsi pour nous c'est qu'une tablette cunéiforme, contrairement à un papyrus ou un parchemin, est donc un objet en trois dimensions, qui peut être inscrit sur chacun de ses six côtés (face, revers et quatre tranches).

C'est grâce à ces milliers de documents que l'on a pu commencer à redécouvrir, depuis un siècle et demi, l'histoire des Sumériens, des Babyloniens, des Assyriens et autres Hittites. Tout cela a provoqué la naissance d'une discipline scientifique nouvelle : l'assyriologie, qui a donc aujourd'hui environ cent cinquante ans d'âge. Mais l'histoire même de ces recherches (reposant, à ses débuts, sur une vive compétition entre les grandes puissances coloniales) et les aléas des découvertes faites jusqu'à nos jours ont entraîné une extrême dispersion des documents exhumés. Il faut ajouter à cela que, dès l'origine des recherches, les fouilles clandestines se sont multipliées et des milliers d'objets inscrits ont ainsi été dispersés, dont on ignore souvent la provenance exacte.

Ainsi trouve-t-on aujourd'hui des collections de textes cunéiformes dans la plupart des grands pays du monde et dans de multiples collections publiques ou privées. Quantitativement, les plus grosses collections sont à Bagdad, Istanbul, Londres, Yale, mais aussi à Paris, Berlin, Saint-Petersbourg, Philadelphie, Chicago, Damas, etc. Une telle dispersion n'a bien sûr en rien facilité les recherches... Et l'on peut citer le cas du grand sumérologue américain S. N. Kramer (l'auteur du fameux « L'Histoire commence à Sumer ») qui, dans les années 1950, s'est vu contraint de reconstituer des textes de littérature sumérienne qu'il déchiffrait en pionnier, à partir de fragments de tablettes que les aléas de l'histoire des recherches avaient dispersés dans les collections de Jena en Allemagne, de Philadelphie aux USA et d'Istanbul en Turquie. Et cela en un temps où l'on ne disposait pas, bien entendu, des moyens de communication actuels...

THÈME VII

* Cuneiform Digital Library Initiative, <http://cdli.ucla.edu>

Outre cet éparpillement, d'autres menaces pèsent sur ce précieux patrimoine : fragilité des objets en argile dont certains s'effritent, se cassent ou se décomposent, notamment sous l'effet de l'humidité ; périls quant à leur conservation dans certaines régions menacées (cf. la situation actuelle des musées de Bagdad ou d'Alep). Et c'est peu dire que l'extraordinaire trésor intellectuel que représente cet héritage culturel demeure aujourd'hui mal connu et largement sous-exploité, alors qu'il s'agit là de sources primaires qui remontent aux origines mêmes de l'écriture et de l'Histoire.

C'est de ces divers constats (dispersion, menaces diverses, sous-exploitation scientifique) qu'est née, il y a une douzaine d'années, l'idée de créer un conservatoire numérique, accessible via internet, de toute la documentation cunéiforme conservée à travers le monde. Ce projet international de bibliothèque numérique du cunéiforme, la *Cuneiform Digital Library Initiative* (CDLI, <http://cdli.ucla.edu>), vise à recenser, cataloguer, documenter et présenter sur le web l'ensemble des collections de textes cunéiformes dispersées dans tous les nombreux musées, institutions et collections présents sur les cinq continents.

À l'origine du projet se trouve un assyriologue de Los Angeles, professeur à UCLA, Robert K. Englund, associé à un historien des sciences et informaticien allemand, de l'Institut Max-Planck de Berlin, Peter Damerow. Leur souhait était d'abord de développer un outil de recherche participatif, mais ils ont vite réalisé l'impact que pouvait avoir leur initiative en termes de conservation du patrimoine, d'éducation et de diffusion des connaissances. Voici comment, à l'origine, ils résumaient eux-mêmes les trois principaux objectifs de leur projet :

CULTURAL HERITAGE PRESERVATION AND ACCESS

- a) *digital capture, storage and presentation of artifacts*
- b) *development of "tool box" for collection curators*
- c) *assist national and international policing agencies*

RESEARCH AND ARCHIVING TOOLS

- a) *act as a permanent repository for primary documents*
- b) *develop tools for graphemic, lexemic, and higher-level linguistic analysis*

EDUCATIONAL TOOLS FOR FORMAL AND INFORMAL LEARNERS

- a) *develop "Wikipedia"-type educational background for informal learners*
- b) *leverage influence on secondary curricula through learning technology partnerships*

Aujourd'hui, malgré le récent décès de P. Damerow, le projet est toujours piloté depuis Los Angeles, Berlin et désormais Oxford, même s'il s'est largement internationalisé depuis son origine, notamment dans le cadre d'une coopération avec quelques grands musées, institutions ou universités d'Europe et d'Amérique. Dans un cadre strictement académique, le CDLI vise principalement aujourd'hui à procurer aux chercheurs et au public intéressé un libre accès, via internet, au contenu des diverses collections de tablettes cunéiformes inventoriées à travers le monde. Étant donné la masse documentaire à traiter, c'est un projet de longue haleine, mais sont d'ores et déjà disponibles en ligne près de 250.000 textes sous forme de catalogue complet, d'images (pour la moitié d'entre eux) et de translittérations (actuellement plus de trois millions de lignes de texte), pour un total de 2 TBytes de données. L'ensemble repose pour l'essentiel sur une *database* au format FileMakerPro, dont les données sont exportées chaque nuit en XML dans une base MySQL qui est utilisée pour l'affichage et la recherche web. Bien entendu, l'*open access* et l'utilisation d'outils en *open source* sont systématiquement privilégiés dans ce projet.

On peut dire, sans exagération, que par son expertise technique, les moyens matériels et humains qu'il met en œuvre et son ancrage international, le CDLI est aujourd'hui devenu un acteur incontournable des études cunéiformes et du champ disciplinaire qu'est l'assyriologie et qu'il demeure en pointe dans le développement de cette nouvelle science qu'est la « science des données ».

Afin d'entrer un peu plus dans le détail de ce projet, regardons en quoi consiste l'essentiel du travail de ceux qui y collaborent. Il repose sur trois actions principales, répétées sans cesse comme une sorte de mantra : *cataloguer, numériser, translittérer*. Reprenons-les dans l'ordre :

CATALOGUER

Le premier travail est une entreprise d'inventaire de tout ce qui existe comme documents inscrits en cunéiforme dans les collections et musées à travers le monde : il s'agit d'abord d'en faire un recensement le plus complet et le plus précis possible. C'est le travail de constitution du *catalogue*. CDLI utilise simplement pour cela une base de données sous FileMakerPro. À chaque objet inscrit est attribuée une cote du type P000000 (ce qui permet d'envisager de pouvoir cataloguer jusqu'à 999 999 objets). Ce « P-number » est la signature de l'objet. Il est à la base du système d'indexation et de recherche des textes du CDLI. Voici comment se présente une fiche descriptive d'un objet inscrit catalogué dans la base (fig. 1).

The image shows a screenshot of the CDLI Catalog form. The form is divided into several sections, each with a title and a question mark icon in the top right corner. The sections are:

- CDLI Catalog**: Includes a "Missing designation" label, a "Check Web" button, and an "aka_name" field.
- Publication:** Fields include primary_publication, author, publication_date, publication_history, electronic_publication, citation, author_remarks, and published_collation.
- Physical Information:** Fields include tablet (with sub-fields for object_type and material), object_remarks, height, width, thickness, Object preservation (list), Surface preservation (list), condition_description, join_information, seal_id, and seal_information.
- Provenience:** Fields include provenienc, provenienc_remarks, excavation_no, findspot_square, elevation, stratigraphic_level, and findspot_remarks.
- Chronology:** Fields include Period (list), period_remarks, date_of_origin, dates_referenced, date_remarks, alternative_years, accounting_period, and writing_phase.
- CDLI data:** Fields include id_text, external_id, cdli_remarks, atf_source, cdli_collation, atf archival, db_source, photo archival, data_entered, lineart archival, public_atf, public_images, 2010-09-10 date_updated, and entry layout.
- Collection Information:** Fields include owner, museum_no, accession_no, acquisition_history, and owner_remarks.
- Text content:** Fields include Genre (list), subgenre, subgenre_remarks, Language (list), SS, and text_remarks.

Fig. 1 : Fiche descriptive d'une entrée type dans la base de données du CDLI.

L'examen de cette fiche type montre que le catalogue enregistre pour chaque document des informations précises concernant :

- les conditions de sa publication et la bibliographie afférente
- son origine et les données archéologiques éventuelles le concernant
- son apparence physique (dimensions, couleur, etc.) et son état de conservation
- sa date et toutes données utiles concernant la chronologie
- ses références dans le catalogue CDLI (dont le fameux « P-number »)
- ses références et numéros d'inventaire dans la collection auquel il appartient
- le contenu du texte (genre, langue, remarques d'ordre linguistique, etc.)

Si l'on veut faire un rapide état de la situation actuelle du catalogue du CDLI, en triant les entrées disponibles par exemple par période chronologique, par grande collection, par genre de textes ou par lieu de provenance, on peut obtenir en quelques clics ce genre de récapitulatifs (situation au 15/09/2013) :

Nombre de textes cunéiformes par période chronologique :

Uruk archaïque (3350-3000 av.)	6611
Dynasties archaïques (3000-2350 av.)	13340
Paléo-akkadien (2350-2200 av.)	7946
Gudéa et III ^e dynastie d'Ur (2200-2000 av.)	98702
Paléo-babylonien (2000-1500 av.)	48953
Paléo-assyrien (2000-1800 av.)	5575
Médio-babylonien (1400-1000 av.)	11082
Médio-assyrien (1400-1000 av.)	2731
Néo-assyrien (1000-600 av.)	29165
Néo-babylonien (620-539 av.)	0826

Nombre de textes cunéiformes par collection (exemples de collections et musées) :

British Museum de Londres	65778
Université de Yale	25734
Musée archéologique d'Istanbul	16425
Musée National Irakien de Bagdad	13015
Musée du Louvre	7453
etc.	

Nombre de textes cunéiformes par genre (quelques exemples) :

Textes administratifs	155648
Pièces de correspondance	11075
Inscriptions royales et monumentales	4479
Pièces de littérature	8139
etc.	

Nombre de textes cunéiformes par lieu de provenance (exemples de sites) :

Tello (Irak du sud)	32004
Nippur (Irak du sud)	25956
Ninive (Irak du nord)	24369
Uruk (Irak du sud)	11086
Ebla (Syrie du nord)	6947
Assur (Irak du nord)	3462
Babylone (Irak du centre)	2290
Sippar (Irak du centre)	1664
etc.	

Au total, l'objectif à atteindre est celui qu'affirme et répète régulièrement Robert Englund lors de ses conférences : « We want every single tablet on earth in CDLI catalog ».

NUMÉRISER

La seconde grande priorité du projet est la numérisation des tablettes pour constituer une banque d'images aussi complète que possible. C'est l'étape de la **numérisation** des documents. Bien que les avancées technologiques soient très rapides dans ce domaine, ce n'est pas le lieu de s'étendre ici sur les progrès actuellement réalisés en termes d'imagerie numérique. Le problème est qu'une tablette cunéiforme, contrairement à une feuille de papier, est comme on l'a dit un objet en trois dimensions : les signes sont inscrits en creux dans la tablette d'argile qui est elle-même un objet d'une certaine épaisseur sur les tranches duquel peuvent se trouver des signes d'écriture. Il faut donc pour chaque objet que l'on veut numériser réaliser six images : face, revers et quatre tranches. Diverses expériences ont montré, dès le début du projet, que l'on pouvait réaliser d'assez bonnes images de ces objets avec certains scanners à plat du commerce, guère différents de ceux couramment en usage dans nos bureaux. La lumière rasante du scanner convient bien pour faire ressortir les creux des signes cunéiformes incisés dans l'argile.

À côté d'appareils photos perfectionnés et de dispositifs d'imagerie complexes (images en 3D, dispositifs utilisant le protocole RTI, etc.), le CDLI utilise donc encore la plupart du temps ce type de scanner à plat A4 pour numériser et obtenir de bonnes images utilisables des tablettes que l'on a rassemblées. Il s'agit au total de produire des images au format TIFF en 600 ppp. Pour chaque objet, les six images TIFF obtenues sont ensuite traitées sous Photoshop pour produire une vue composite de chaque texte (*fatcross*). La qualité est alors suffisante pour que l'on puisse procéder ensuite à des agrandissements afin d'observer puis lire les détails des signes cunéiformes inscrits, comme on pourra s'en rendre compte en allant visiter les images disponibles sur le site du CDLI (<http://cdli.ucla.edu>) ou en examinant l'exemple ci-dessous (fig. 2).

THÈME VII

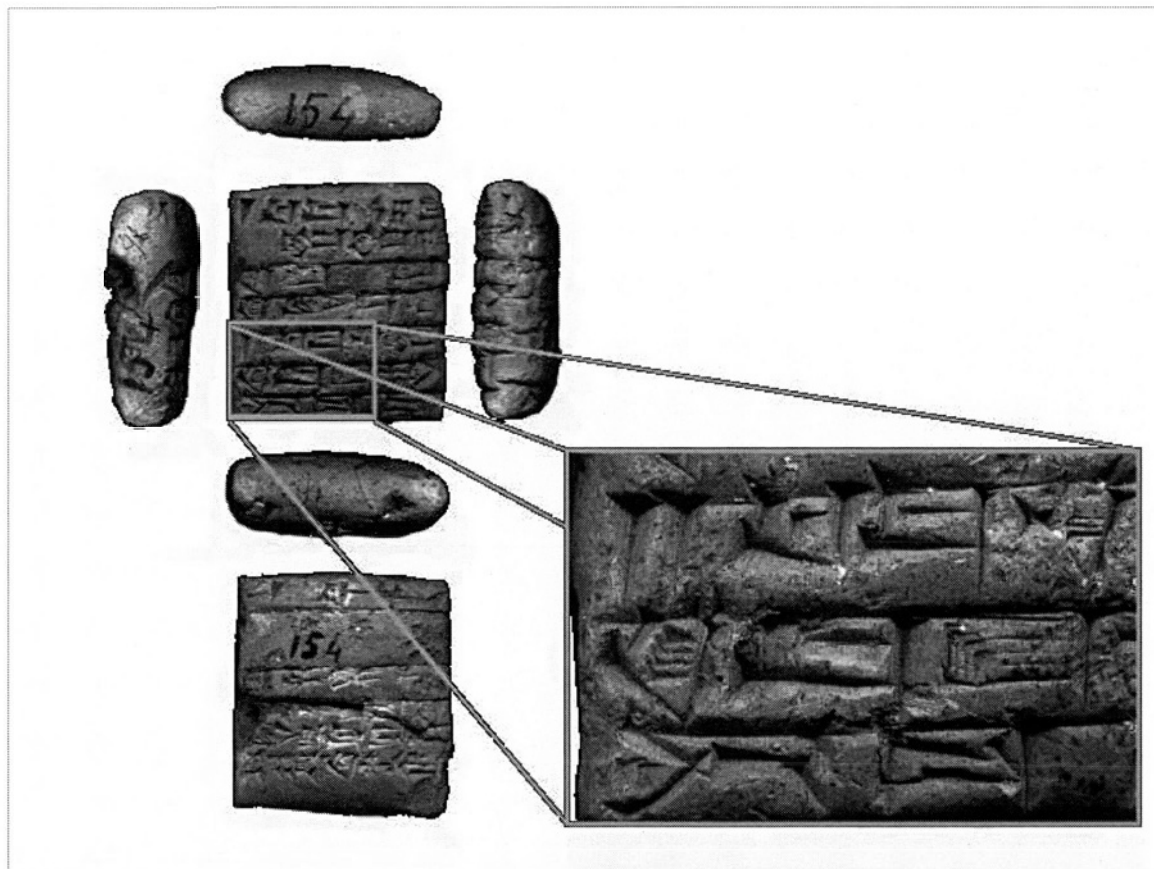


Fig. 3 : Exemple d'image composite et agrandissement en pleine résolution (600 ppi).
Tablette administrative sumérienne de l'époque de la III^e dynastie d'Ur (XXI^e siècle av.),
provenant du site d'Umma (ICP 453 = P128724, image CDLI).

TRANSLITTÉRER

La dernière tâche est celle qui permet de rendre disponible le contenu même de chaque texte et d'entreprendre des recherches sur les informations qu'il contient : il s'agit du travail de *translittération*. Il consiste à mettre en ligne une transposition dans notre système d'écriture alphabétique de ce qui est inscrit sur la tablette, quelle que soit la langue utilisée (sumérien, babylonien, hittite, hourrite, etc.). Cette opération repose bien sûr essentiellement sur les travaux préalables de déchiffrement et d'édition des documents cunéiformes, tels qu'ils sont effectués chaque année dans les nombreuses publications scientifiques à travers le monde. Mais l'une des difficultés à laquelle le CDLI a été confrontée d'emblée est venue de la complexité et de l'absence de normes standardisées parmi les divers systèmes de transcriptions qui ont été mis au point par les assyriologues au cours des années, et qui reposent sur l'usage de nombreux signes diacritiques, parfois complexes.

Pour s'affranchir de ces contraintes et pour permettre le portage sur n'importe quelle plateforme informatique présente ou future, on a finalement opté pour un système de translittération des plus simples, n'utilisant strictement que les 127 signes ASCII élémentaires. C'est le système dit ATF (*ASCII Text Format*), qui fait par exemple que la lettre [š], dont ont besoin les assyriologues mais qui n'existe pas en ASCII, est conventionnellement transcrite [sz]. On pourra se reporter à ce sujet et pour une description générale du système ATF à <http://oracc.museum.upenn.edu/doc/builder/cdli>.

On ne saurait développer ici tout ce qui touche au marquage XML de description linguistique du contenu des textes et à la mise en place d'un certain nombre de métadonnées qui accompagnent la transcription de chaque document (travail de lemmatisation avec marquage des lexèmes et morphèmes permettant de faire des recherches lexicographiques et grammaticales) : tout cela se développe concomitamment au travail d'alimentation de la base.

Mais ce que l'on espère au final, c'est que l'étape ultime de présentation des textes ira jusqu'à leur traduction, potentiellement dans de multiples langues, comme on tente actuellement de l'expérimenter (voir à titre d'exemple le texte <http://cdli.ucla.edu/P103956>).

Pour se faire une idée du travail déjà réalisé, on pourra se reporter en fin de compte au mode de présentation de quelques grandes collections conservant des textes cunéiformes et désormais disponibles sur CDLI, soit les quelques exemples suivants :

- British Museum de Londres (<http://cdli.ucla.edu/collections/bm/bm.html>)
- Oriental Institute de l'université de Chicago (<http://cdli.ucla.edu/collections/oi/oi.html>)
- Vorderasiatisches Museum de Berlin (http://cdli.ucla.edu/collections/vam/vam_de.html)
- Abbaye catalane de Montserrat (http://cdli.ucla.edu/collections/montserrat/montserrat_en.html)
- Musée de l'Hermitage à St-Petersbourg (http://cdli.ucla.edu/collections/hermitage/hermitage_en.html)
- Bibliothèque Universitaire de Strasbourg (http://cdli.ucla.edu/collections/strasbourg/bnus_fr.html)
- Institut Catholique de Paris (http://cdli.ucla.edu/collections/icp/icp_fr.html)
- etc.

Quoi qu'il en soit, pour l'utilisateur final, l'outil le plus important est aujourd'hui la fenêtre de recherche du CDLI (<http://cdli.ucla.edu/search>), qui permet d'interroger la base de données, son catalogue, les divers jeux de translittérations et de retrouver des textes et tous types d'informations et contenus souhaités, parmi les dizaines de milliers de documents désormais en ligne.

Bien entendu, des liens étroits sont entretenus par le CDLI avec certains autres grands projets assyriologiques en ligne : ainsi par exemple la *Base de Données des Textes Néo-Sumériens* BDTNS (<http://bdtns.filol.csic.es>), le projet *Open Richly Annotated Cuneiform Corpus* ORACC (<http://oracc.museum.upenn.edu>), le *Dictionnaire Sumérien de Philadelphie* ePSD (<http://psd.museum.upenn.edu>), le *Corpus Électronique des Textes de la Littérature Sumérienne* ETCSL (<http://etcsl.orinst.ox.ac.uk>), etc. Toutes ces entreprises interagissent ou sont destinées à interagir entre elles.

En liaison avec le projet, a également été créé un journal en ligne, le *Cuneiform Digital Library Journal*, CDLJ (<http://cdli.ucla.edu/?q=publications>), publication numérique dont les articles de recherche scientifique renvoient systématiquement, via des hyperliens, à la source même d'information sur le site du CDLI dès qu'il est question d'un texte cunéiforme.

Le CDLI propose aussi désormais aux assyriologues, étudiants et chercheurs (dans sa section « tools and resources ») divers outils téléchargeables : liste des abréviations bibliographiques courantes utilisées en assyriologie, répertoires de signes cunéiformes en usage à diverses périodes, listes chronologiques des souverains mésopotamiens et de leurs « noms d'année », listes des noms de mois des divers calendriers mésopotamiens, etc.

Et l'on se permettra d'insister ici, dans le contexte d'une actualité tragique, sur l'intérêt de ce vaste projet de numérisation de la documentation cunéiforme, quand on voit ce qu'il en est des menaces qui pèsent sur des pans entiers de ce fragile patrimoine : dans le cadre d'un accord de coopération conclu en 2007 avec la Direction Générale des Antiquités et Musées de Syrie (DGAM), avait été lancée une entreprise de numérisation de toutes les tablettes conservées dans les musées syriens. Elle a dû être interrompue par la guerre. Qui sait ce que sont devenus aujourd'hui les fragiles documents d'argile conservés dans les musées d'Alep, de Deir ez-Zor ou de Raqqa ? Si l'on avait pu au moins en conserver à temps la trace numérique...

Terminons en évoquant quelques perspectives pour le projet dont il a été question ici. Bien sûr, le travail d'inventaire et de traitement des informations est loin d'être achevé. Mais après avoir activement contribué au cours des années passées à la mise en ligne, via CDLI, des fonds documentaires cunéiformes et à leur utilisation dans les pratiques de recherche, il nous semble désormais intéressant de travailler aussi sur l'impact provoqué par l'usage de ces ressources nouvelles, notamment pour l'écriture de l'histoire, mais également pour améliorer leur intelligibilité auprès des « usagers » (enseignants, chercheurs, étudiants, etc.) dans le contexte des technologies numériques.

Nous sommes donc en train d'explorer les possibilités de nouveaux outils à développer à partir de pages web dont l'embryon existe déjà : ce sont celles du site *cdli:wiki* (<http://cdli.ox.ac.uk/wiki>), créé il y a quelques années déjà, adossé au CDLI et désormais hébergé à l'Université d'Oxford. Sur le modèle formel offert par le projet Wikipedia (<http://wikipedia.org>), ces pages de *cdli:wiki* ont commencé à rassembler divers outils essentiels pour l'assyriologie comme des listes d'abréviations (largement utilisées dorénavant dans la littérature assyriologique secondaire), des listes chronologiques ou quelques outils bibliographiques.

L'idée que nous avons est de produire, dans ce cadre, des ressources numériques nouvelles qui soient en quelque sorte dans la ligne de ce que sont les « usuels » que l'on trouve en accès direct sur les rayonnages de nos bibliothèques universitaires, l'objectif étant de démocratiser et faciliter, pour tous types de publics, un accès électronique aux sources, aux outils de travail et aux savoirs accumulés sur l'histoire du Proche-Orient ancien. Ce que nous visons en définitive avec ce nouveau projet (qui vient d'être retenu pour intégrer le LabEx « Les Passés dans le Présent » mis en place à l'université Paris-Ouest de Nanterre), c'est de proposer un contexte et les outils d'analyse qui sont désormais ceux du XXI^e siècle, pour les dizaines de milliers de textes cunéiformes enregistrés sous forme « brute » dans la base du CDLI, tant nous sommes persuadés qu'un accès direct à ces données triées et mises en perspective peut modifier notre compréhension de l'histoire ou notre façon de l'écrire.