



**HAL**  
open science

## Synergie enseignement-recherche pour l'aménagement numérique structuré (TEI) de patrimoines littéraires multilingues et multiculturels

Henri Hudrisier, Ghislaine Azemard, Mokhtar Ben Henda, Sascha Diwersy, Anne Lehmans, Vincent Liquete, Laurent Romary

### ► To cite this version:

Henri Hudrisier, Ghislaine Azemard, Mokhtar Ben Henda, Sascha Diwersy, Anne Lehmans, et al.. Synergie enseignement-recherche pour l'aménagement numérique structuré (TEI) de patrimoines littéraires multilingues et multiculturels. Communication, Information et savoir: quel management pour une organisation durable?, Université de Montréal, Jun 2015, Moncton, Canada. pp.280. hal-04507263

**HAL Id: hal-04507263**

**<https://hal.science/hal-04507263v1>**

Submitted on 16 Mar 2024

**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.



Distributed under a Creative Commons Attribution 4.0 International License



**ACTES**  
*du 7<sup>e</sup> Colloque COSSI*  
*Communication, Organisation, Société du Savoir et Information*

Communication, information et savoir  
**Quel management pour une  
organisation durable?**

10 au 12 juin 2015  
École de bibliothéconomie et des sciences de l'information  
Université de Montréal  
Montréal (Québec), Canada

Sous la direction de  
Monica Mallowan  
Université de Moncton, Campus de Shippagan  
Shippagan (Nouveau-Brunswick), Canada



## Synergie enseignement-recherche pour l'aménagement numérique structuré (TEI) de patrimoines littéraires multilingues et multiculturels

Henri HUDRISIER  
Maitre de conférences HDR, Chaire Unesco-ITEN,  
Université Paris 8

Ghislaine AZEMARD  
Professeure titulaire, Chaire Unesco-ITEN,  
Université Paris 8

Mokhtar BEN HENDA  
Maitre de conférences HDR, Chaire Unesco-ITEN  
MICA, Université de Bordeaux Montaigne

Sascha DIWERSY  
Maitre de conférences en linguistique française et espagnole  
Institut des Langues Romanes, Université de Cologne

Anne LEHMANS  
Maitre de conférences,  
ESPE Aquitaine, Université de Bordeaux

Vincent LIQUETE  
Professeur des universités,  
ESPE Aquitaine, Université de Bordeaux

Laurent ROMARY  
Directeur de Recherches  
INRIA, France

**Résumé** : HD-Muren est un projet d'enseignement/recherche dans des disciplines appartenant aux domaines des Arts (la musique), des SHS (sciences de l'information et de la communication, Sciences de l'éducation, recherche littéraire) et prioritairement des Lettres (littérature ou apprentissage des langues étrangères). En prenant pour cadre princeps la TEI (Text Encoding Initiative), nous mettons en place une expérimentation pilote de pédagogie active pour travailler au plus près des textes (comme en dissertation littéraire ou en commentaire de texte), dans une synergie avec la construction participative de patrimoines littéraires numériques destinés à la recherche. Cette production participative est destinée à être gagnant-gagnant entre d'une part les chercheurs et les acteurs du patrimoine et d'autre part les professeurs, les étudiants et les élèves du secondaire. Au niveau euro-méditerranéen, le projet vise aussi à rééquilibrer le volume relatif des corpus littéraires ou philosophiques donnant à voir les Lumières européennes et l'Islam des Lumières sous-représenté comme l'est la Renaissance

arabe. Nous prétendons que ce projet s'inscrit dans une écologie et une économie durable des patrimoines, de l'enseignement, de la recherche en SHS et de ces biens communs.

**Mots-clés** : enseignement, recherche littéraire, patrimoines littéraires, documents numériques structurés, TEI (Text Encoding Initiative), recherche en SHS, multiculturalisme Nord-Sud, écologie informationnelle

**Abstract**: HD-Muren is an educational/research project in disciplines from the fields of Arts (music), humanities and social sciences (Information, educational sciences, literary research) and mainly Letters (literature or foreign language learning). While taking TEI (Text Encoding Initiative) as a framework, we implement a leading experiment of active learning in order to work closely with texts (such as in literary essay or text commentary) in a synergy with a participatory building of a digital literature legacy for research and heritage conservation. This participatory production is intended to be of mutual interest between researchers and patrimony stakeholders from one side and teachers, students and high school students from the other. At a Euro-Mediterranean level, the project also aims to rebalance the volume of literary or philosophical corpora permitting to see the European Enlightenment and the Islam Enlightenment underrepresented as is the Arab Renaissance. We claim that this project is part of an ecology and a sustainable economy of cultural heritage, education, research in social sciences and of these common legacies.

**Keywords**: education, literary research, structured documents, literary heritage, TEI (Text Encoding Initiative), humanities research, North-South multiculturalism, informational ecology

**Avertissement** : En n'hésitant pas à multiplier le nombre des auteurs de cette communication nous avons conscience que : 1. Nous heurtons frontalement (surtout en SHS) des habitus d'unicité auctoriale de l'écriture scientifique. 2. Nous nous inscrivons dans de nouveaux potentiels de l'écriture scientifique collective en réseaux. 3. Nous refusons d'attribuer la propriété intellectuelle de tout un projet de recherche au seul « scribe compilateur », porte-parole au 7e COSSI donc placé comme premier auteur, les autres auteurs étant par ordre alphabétique. 4. La posture de pluralité auctoriale en SHS semble être une tendance émergente importante de la nouvelle écologie du savoir. Le collectif R. Pédaque (Pédaque, 2007) le proclame très tôt.

### **Contexte et objectifs d'HD-Muren**

La présente contribution fait état de mise en place du projet HD-Muren<sup>119</sup> au sein d'IDEFI-CréaTIC (Université Paris 8, s. d.), un programme en formations

---

<sup>119</sup>HD-Muren : « Humanités Digitales multiculturelles en mode *crowdsourcing* via la recherche et l'enseignement »

innovantes qui lui sert de cadre institutionnel<sup>120</sup> et conditionne certains de ses objectifs. En effet, l'objectif de ce programme cadre est de mettre en œuvre de nouvelles méthodes pédagogiques associant enseignement et recherche notamment dans des formations liées à l'usage des dispositifs et technologies numériques pour l'enseignement, la création et le travail collaboratif. Dans ce cadre institutionnel structurant, HD-Muren vise à proposer la Text Encoding Initiative (TEI) comme cadre logiciel princeps du travail sur corpus entrepris mais aussi comme moteur de transformation d'une pédagogie active liant enseignement et recherche en Lettres, Sciences Humaines et Arts. Nous faisons aussi l'hypothèse qu'à l'ère du numérique, les patrimoines culturels, les ressources pédagogiques, les exercices des apprenants, les travaux divers des chercheurs ou encore les œuvres des créateurs, constituent de fait un même continuum potentiel à condition que l'ensemble de ces corpus soient balisés et structurés de façon standardisée, interopérable et normalisée. Ces corpus pourront ainsi s'associer de façon cumulative en utilisant les matériaux réalisés comme exercices pédagogiques en leur donnant une valeur scientifique ajoutée et en faisant circuler des ensembles structurés de corpus numériques. Cette collégialité respectueuse de ces diversités est précisément ce que permet la TEI. Nous pensons ainsi contribuer à la mutation en cours d'une nouvelle technoculture du numérique et de sa meilleure prise en compte et appropriation par les différents acteurs dans le secteur des lettres et des arts.

Les Humanités digitales et la TEI, des moteurs de mutation informationnelle de l'ère numérique, comme l'avaient été les Humanistes face à l'imprimé

Notre démarche nous paraît légitime parce qu'elle nous semble s'inscrire dans une logique de similarité entre la mutation humaniste de la Renaissance et les Humanités numériques aujourd'hui. Pour nous (et bien d'autres chercheurs), l'Humanisme numérique n'est pas seulement un type de contenu, ni même seulement une posture morale et philosophique (la médiation numérique de la croyance en l'Homme). C'est aussi une méthode : permettre une circulation mondiale et interopérable de corpus de documents parce qu'ils sont XMLisés, normalisés, structurés et balisés selon des schémas TEI. Cette détermination de méthode est explicitement inscrite dans les « Principes de Poughkeepsie » qui sont très largement repris dans le « Manifeste des Digital humanities » (Dacos, 2010).

## Historique de la TEI

Dans les années 80, un certain nombre de chercheurs, des grandes institutions impliquées dans la recherche et la gestion du domaine des humanités, mais aussi de la linguistique computationnelle se sont interrogés sur la meilleure façon de faire évoluer leurs habitus de chercheurs en fonction des nouvelles opportunités offertes par l'émergence du document structuré (SGML) et d'un réseau Internet

---

<sup>120</sup> IDEFI-CréaTIC est coordonné par l'Université Paris 8 avec d'autres partenaires. Pour ce qui est du « partenariat- patrimoine » soulignons celui des Archives nationales [de France].

pas encore World Wide Web. Cette volonté de faire évoluer leurs pratiques académiques et professionnelles correspondait aux besoins de certaines grandes institutions nationales et internationales ou de grandes fondations visant à préparer les bibliothèques, les musées, les grandes bases de gestion patrimoniales et les instances de recherches en littérature, en SHS et en Arts aux enjeux et opportunités du numérique en réseaux collaboratifs en devenir : inscrire des habits de chercheurs dans des schémas TEI qui n'étaient pas encore des schémas XML. Néanmoins, il faut souligner d'abord que le monde anglophone pouvait aborder plus précocement ces enjeux d'évolution dans la mesure où (surtout il y a 30 ans) l'informatique était presque exclusivement en anglais et que l'ingénierie linguistique et documentaire avait pris une énorme avance dans, et pour cette langue. Notons ensuite que la vieille Europe qui avait fondé ses institutions patrimoniales, académiques et de recherches bien avant l'Amérique du Nord avait globalement moins de facilité à envisager leur indispensable mutation que de l'autre côté de l'Atlantique où nombre de musées, de bibliothèques, d'universités sont des fondations privées évidemment plus réactives.

Associés à des experts de l'édition techno-documentaire en SGML, des chercheurs en littérature, des bibliothéconomes et des muséographes se sont réunis au Vassar College en 1987. À l'issue de leur rencontre, ils ont édicté les « Principes de Poughkeepsie » à l'origine de la fondation de la Text Encoding Initiative, TEI (Ide & Véronis, 1996). Notons que la TEI est ainsi soutenue par l'Association for Computers and the Humanities, l'Association for Computational Linguistics et l'Association for Literary and Linguistic Computing et que le projet a été financé par le National Endowment for the Humanities américain, la DG XIII de la CEE, la fondation Andrew W. Mellon et le Social Science and Humanities Research Council du Canada. Les « Principes de Poughkeepsie » se sont peu à peu précisés à travers une série de documents de travail et les « TEI Guidelines » ont été publiés en mai 1994. Ces « Recommandations de la TEI », partiellement disponibles en français sont continûment refondues et rééditées (TEI Consortium, 2015).

Dans ce sillage s'est agrégé un important mouvement dit des « Humanités numériques ». Ainsi, les principes d'interopérabilité normative des corpus, d'indépendance des solutions techniques, mais aussi de travail coopératif en réseaux au sein d'une communauté collégiale solidaire respectueuse des spécificités disciplinaires de tous, sont les paradigmes fondateurs qui constituent une véritable mutation dépassant de loin la seule détermination de numériser sans autre valeur ajoutée<sup>121</sup>. Les principes fondamentaux d'échange solidaire entre les chercheurs et d'interopérabilité d'un balisage respectueux des spécificités disciplinaires mais aussi de l'expression encore plus spécifique d'hypothèses ou

---

<sup>121</sup> De ce point de vue, nous sommes dans le même rapport de prise en compte du potentiel innovant que lorsque les premiers incunables reproduisaient en multiple à la presse des pseudo-manuscrits alors que quelques décennies plus tard c'est toute la typographie, toute la sémantique d'une nouvelle mise en page codifiée (normalisation des notes en bas de page, de la place du titre en première page, etc.) qui devaient se mettre en place.

de résultats de tel ou tel chercheur ou groupe de chercheurs deviennent d'années en années de moins en moins incontournables.

Il est trivial de dire que nous sommes confrontés à une mutation techno-culturelle similaire à celle que dut faire la Renaissance, période au cours de laquelle, l'imprimerie<sup>122</sup>, mais aussi la perspective pour l'image induisirent une transformation globale des habitus savants, de l'organisation des institutions académiques et des modes de publication ; cela généra notamment une conscience scientifique des arts et de sa mise en collection<sup>123</sup>. Précisons que la mutation est seulement « similaire », et non « semblable » dans la mesure où la technoculture dans laquelle nous vivons est aussi traversée par le mélange des cultures, des langues, des modes de médiation, que les progrès du Web sémantique, de la traductique, de l'hypermédia nous permettent déjà de dépasser. Comme à la Renaissance, il nous faudra attendre que nos institutions éducatives, patrimoniales, scientifiques se transforment en profondeur et qu'elles établissent entre elles de nouveaux équilibres, une nouvelle écologie informationnelle plus perfectionnée, plus adaptée et devenant donc plus durable. Mais cette mutation des institutions et des habitus des spécialistes face au numérique ne pourra qu'accompagner une appropriation mutante du fait numérique en large partie induite par la génération des digital natives exactement comme il a fallu un peu plus qu'une génération pour qu'apparaisse une culture mutante de l'imprimé. En cela, ce constat correspond bien aux thèses d'un historien des techniques comme Bertrand Gille (Gille, 1978) qui montre qu'à chaque mutation des paradigmes d'un ensemble (ou plutôt d'un environnement) technique doit correspondre non seulement une transformation des objets techniques mais aussi une transformation de la technoculture dans son ensemble.

La synergie collaborative de construction patrimoniale : de Diderot à Wikipédia, les nouvelles opportunités numériques hors des facilités de la « marchandisation publicitaire »

Nous pensons que pour être durable, la construction et la gestion des patrimoines et ressources numériques doivent indubitablement échapper à la seule responsabilité de leurs conservateurs. Ceci est d'autant plus nécessaire qu'au nom d'un réalisme économique, ceux-ci succombent souvent aux sirènes de la

---

<sup>122</sup> On sait qu'elle constitue une mutation profonde du savoir. Dès après son invention, émerge la conscience d'une « République des Arts, des lettres et des sciences » qui participe de cette évolution et qui prend des formes variées : institutionnelle (fondation des académies, création de nombreuses universités), éditoriale (transformation et surtout évolution des métiers du livre, invention de la typographie par mutation des habitus des copistes, révolution et on pourrait dire invention de l'auteur moderne), mutation aussi de la bibliothèque et invention en France par François 1<sup>er</sup> du Dépôt Légal avec l'ordonnance de Montpellier (1537). Marshall Mac Luhan, (McLuhan & Pàré, 1967), Elizabeth Eisenstein (Eisenstein, 1991) sont parmi les pionniers de la recherche moderne en la matière.

<sup>123</sup> On considère souvent David Teniers auteur du *Theatrum pictorium* (Téniers, 1658) comme un pionnier de la muséographie moderne.

« collaboration avec Google<sup>124</sup> », s'écartant ainsi de l'éthique des patrimoines comme « biens publics ». Pour pouvoir répondre à l'augmentation exponentielle des patrimoines numériques, il devient d'année en année plus indispensable de devoir partager cette création des patrimoines avec tous les utilisateurs, notamment des auteurs, des chercheurs, des enseignants et leurs élèves ou étudiants (voire aussi des séniors).

On a pu percevoir cette mutation dans les premières années du Web dans la mesure où le « Web pionnier » n'existait comme univers structurellement, référentiellement et sémantiquement cohérent et accessible qu'à la condition où les créateurs pionniers de pages HTML autodocumentaient, autostructuraient et éventuellement positionnaient les bons mots clés aux endroits adéquats. Cette culture pionnière de la construction collective du Web du début des années 90 a en grande partie disparu dans la mesure où les grands opérateurs ont réussi à marchandiser l'accès sémantique et référentiel en mettant un prix sur la fréquence des mots clés, le profil des utilisateurs et leur potentiel à consommer tels produits, tels services ou telles œuvres éditoriales. Néanmoins, l'existence de Wikipédia prouve, s'il en était besoin, que ces temps pionniers de la construction collective normalisée et cohérente sur le Web d'un patrimoine encyclopédique multilingue n'ont pas disparus et restent d'actualité. Wikipédia échappe à la marchandisation. Des milliers de pages se créent et s'améliorent continûment, même en employant très peu de gestionnaires permanents parce que les inventeurs de Wikipédia ont très vite proposé un schéma éditorial structuré permettant de documenter, de structurer de façon prévisible, de référencer, donc de contrôler automatiquement la construction de cette encyclopédie universelle et multilingue et d'en établir automatiquement les référencements croisés. De ce point de vue, Wikipédia constitue mutatis mutandis l'équivalent du travail d'accompagnement de la mutation accomplie par les Encyclopédistes des Lumières : à la fois théorique, c'est le but du « Prospectus » rédigé par Denis Diderot<sup>125</sup>, qui propose un schéma encyclopédique innovant, (cohérent et systématique) mais aussi pratique et éditorial en rassemblant des auteurs parmi les Philosophes, en contrôlant les articles, mais aussi en mettant en place un modèle de financement permettant de garantir leur indépendance rédactionnelle par la souscription (crowdfunding).

### **Pourquoi le projet HD-Muren ?**

HD-Muren hérite, comme bien d'autres projets, du long cheminement de tous ses partenaires. On peut ainsi souligner l'implication souvent ancienne de nombre d'entre eux dans des projets et études destinés à accompagner les mutations de patrimoines écrits ou multimédia. Par exemple dès 1990 « l'étude d'impact des NTIC pour la Très Grande Bibliothèque » (nom provisoire de l'actuelle BnF) dont le paradigme clé était le PLAO (Poste de « Lecture savante » Assisté par Ordinateur) (Hudrisier, 1999). Selon les recommandations de l'étude, la BnF serait devenue

---

<sup>124</sup> On ne peut que regretter les nombreux contrats de collaboration entre par exemple Gallica et Google, mais aussi nombre de grands musées ou bibliothèques partout dans le monde.

<sup>125</sup> Pour une édition moderne on peut lire par exemple (Sibertin-Blanc, 1957)



ainsi dès 1993 un pôle partenaire TEI, mais de nombreux facteurs institutionnels ont bloqué cette opportunité. Cependant, ceux qui ont conçu et fait exister Gallica, sont ceux-là mêmes qui ont mis en œuvre pour le CNRS le projet TGE Adonis, devenu aujourd'hui HumaNum<sup>126</sup>, qu'on peut considérer comme le point de départ en France (mais aussi en Europe) de la TEI à une large échelle. L'un de nos auteurs travaillant à l'INRIA à Nancy appartenait à des équipes directement impliquées dans la recherche appliquée à la linguistique computationnelle et à la mise en œuvre de très grands corpus de textes<sup>127</sup>. Il est devenu très tôt un des responsables de la TEI en France. D'autres encore étaient de par leur histoire fortement impliqués dans les échanges entre communautés linguistiques et culturelles. Ce point est fondamental dans la mesure où le potentiel de traductique associé aux besoins d'échanges interlinguistiques et de localisations interculturelles ont considérablement augmenté avec la mondialisation numérique.

La facilité avec laquelle, certains membres de notre équipe savent circuler entre plusieurs cultures a aussi son importance pour faire exister une masse critique et une proportion suffisante de « corpus humanistes » notamment en langue arabe mais aussi dans des langues berbères<sup>128</sup>. Il serait en effet très opportun que soit mise en œuvre la numérisation systématique des très nombreux textes qui constituent l'Islam des Lumières et la Renaissance arabe qui précède de beaucoup la Renaissance européenne. Ainsi, pourront se mettre en œuvre des traductions systématiques de ces textes ce qui permettra de proposer des concordances d'équivalence thématiques entre l'Islam des Lumières et les Lumières européennes. Il serait important aussi de montrer que nombre des textes de l'Antiquité grecque ou romaine nous sont parvenus grâce à des traductions arabes qui ont permis la transmission beaucoup plus proche de l'original de ces textes alors que le Moyen-âge européen les transformaient de recopie manuscrite en recopie manuscrite à travers le seul prisme de son dogmatisme chrétien. Cette visibilité comparée des grands textes du Monde arabe constituerait un rempart culturel beaucoup plus efficace face aux risques terroristes que bien d'autres réponses sécuritaires proposées. Il faudrait aussi, symétriquement, que soient proposés des corpus de textes des Poètes libertins (émergence de l'agnosticisme et de l'athéisme), d'écrits des Lumières, des débats philosophiques argumentant la légitimité morale d'une loi non obligatoirement divine, des textes historiques de la séparation de l'Eglise et de l'Etat en 1905, etc. Nous avons l'ambition avec HD-Muren d'expérimenter la faisabilité de ces enjeux.

---

<sup>126</sup> Le TGE Adonis (Très Grand Equipement du CNRS, chargé des apports numériques à la recherche en sciences humaines et sociales) est maintenant fusionné dans Huma-Num, TGIR des humanités numériques (HumaNum, s. d.).

<sup>127</sup> Frantice, Le TLF (Trésor de la Langue Française) et un certain nombre de bases de données proposées par l'ATILF

<sup>128</sup> Nombre d'entre nous ont été les partenaires fondateurs de la BNB (Bibliothèque Numérique Berbère, projet soutenu par la Région Ile de France), de la BNFB (Bibliothèque Numérique Franco-Berbère, projet soutenu par l'Organisation Internationale de la Francophonie, Fonds Francophone des Inforoutes) et de HumanitésDigitMaghreb (un projet soutenu par le CNRS-ISCC), ce dernier projet ayant réalisé des ateliers TEI sur des corpus arabes et berbères.

Notre projet s'appuie sur trois premiers constats :

La transformation du fait numérique, de son économie, de sa prise en compte professionnelle et institutionnelle est grandement conditionnée par son appropriation de masse (digital natives) et par l'évolution d'une informatique surtout locale au réseau mondial.

La synergie enseignement/recherche/patrimoine qui de notre point de vue avait été la règle depuis l'invention de la science (Antiquité puis Renaissance) et jusqu'au milieu du XX<sup>ème</sup> siècle avait perdu beaucoup de son importance après la Deuxième guerre mondiale, mais redevient d'actualité grâce aux réseaux numériques.

Constat directement lié au point précédent, le « patrimonial numérique », mais aussi des « ressources pédagogiques en ligne » sont deux domaines dont on ne peut continuer à garantir la croissance exponentielle actuelle, sans recourir à des participations externes à leurs pôles métiers respectifs : jouer les synergies entre enseignement, recherche et gestion ou conservation des patrimoines. Si nous néglignons trop longtemps d'activer ces potentiels de symbiose entre ces trois pôles, nous augmenterions indubitablement les risques de marchandisation Google de ces biens communs patrimoniaux. Face à ces risques, le recours à une production participative des utilisateurs (crowdsourcing) est indispensable, les deux candidats-partenaires potentiels étant d'une part les seniors et d'autre part les élèves, les étudiants et leurs professeurs. Dans le cas de notre projet, il nous paraît parfaitement légitime que les élèves et les étudiants (mais aussi les enseignants) qui sont les premiers bénéficiaires mais aussi utilisateurs de la recherche en Lettres et sciences humaines soient incités à devenir contributeurs. Sur l'autre versant, les chercheurs en Lettres, Arts et SHS<sup>129</sup>, mais aussi les gestionnaires de patrimoines ont eux aussi grand intérêt à partager en synergie la construction de leurs corpus. Par contre, cette synergie doit s'accompagner de règles éthiques : les élèves, les étudiants de première année et les professeurs doivent pouvoir bénéficier de nombreuses rencontres avec les chercheurs et les professionnels du patrimoine pour se sentir ainsi acteurs à part entière des progrès scientifiques et de l'augmentation des patrimoines. Ces contacts doivent susciter des vocations et permettre de contribuer à la plus-value scientifique ou patrimoniale. Les élèves et les professeurs utilisateurs doivent ressentir à quel point ces corpus sont des biens communs précieux, difficiles et dispendieux à mettre en œuvre et constater par l'expérience participative que les musées et les bibliothèques numériques peuvent aussi croître de par leurs contributions en nature. Il nous apparaît aussi que cette conscience contributive peut devenir le gage de leur future appétence pour utiliser ces patrimoines numériques. A l'inverse, les chercheurs en littérature par exemple, qui pourront apprécier d'être déchargés de nombreuses tâches de numérisation et du balisage référentiel et structurel de leurs corpus d'étude pourront constater combien il peut leur être utile de bénéficier du travail de plusieurs classes pilotes, encadrées par leurs professeurs, qui auront pu baliser systématiquement et

---

<sup>129</sup> Qui sont très faiblement subventionnés par rapport aux sciences exactes et expérimentales.

exhaustivement sur tout l'ensemble de l'œuvre d'un auteur des figures de styles, des types de versification, des hypothèses d'interprétation sémantiques des textes ; ou même avoir travaillé sur commande explicite du chercheur.

Mais il faut aussi tenir compte des opportunités qui nous sont ouvertes (et quelquefois des rigidités ou des résistances auxquelles nous sommes confrontés). Ainsi :

1) Les Humanités numériques et la TEI sont à des niveaux de développement et d'appropriation très inégaux selon les pays, les disciplines et les institutions : notons la très faible participation des pays en développement et à quelques exceptions notables (Japonais, Coréens) la très faible présence des communautés linguistiques à écriture non latine, à l'exception des corpus en langues mortes.

2) On sait par ailleurs qu'on est passé de la traduction automatique pionnière utilisant des analyses de contexte restreint parce que « propriétaires » à une traductique utilisant comme contexte l'exhaustivité des ressources linguistiques structurées précisément selon des normes proposées par le Consortium TEI en lien avec l'ISO130.

3) Les médias analogiques traditionnels (textes imprimés, audiovisuels) rendaient difficile la démarche interdisciplinaire non exclusivement textuelle : étude du cinéma, de la photographie, des arts, de la musique, des corpus oraux, etc. mais aussi l'étude savante des manuscrits. La convergence des médias et l'unicité multimédia numérique peut devenir notamment en TEI (et en MEI/Music Encoding Initiative) le mode privilégié de la communication et de l'interopérabilité interdisciplinaire et cross-média.

4) Dès son apparition, l'imprimerie avait été perçue comme un moyen jusque-là inédit de parfaire les textes, notamment du fait du travail éditorial collectif<sup>131</sup>. Soulignons cependant que les éditeurs de la Renaissance, notamment quand ils avaient pour objectif de parfaire des classiques (grec ou latin) ou des textes scientifiques, ont jugé nécessaire de s'associer en tant « qu'éditeurs humanistes » et d'établir des règles typographiques (normalisation de la mise en page et de la typographie<sup>132</sup>). Ainsi, pour rendre possible l'accès à l'ensemble de la littérature

---

<sup>130</sup> Notamment tout le travail de l'ISO-SC4 (Normalisation des ressources linguistiques) précisément présidé par un de nos auteurs.

<sup>131</sup> « Le pouvoir que nous donne l'imprimerie d'améliorer, de corriger continuellement nos œuvres dans des éditions successives, écrivait David Hume [1711-1776] à son éditeur, me paraît le principal avantage de cet art. » : cité par (Eisenstein, 1991, p.100).

<sup>132</sup> Recherche de lisibilité des caractères (Antique, Italique), mais aussi des standards de présentation et des zones fonctionnelles du volume (systématisation du principe de l'unicité d'un volume pour une œuvre, place du titre, d'un avertissement servant de 4<sup>ème</sup> de couverture et de l'entité page (normalisation de la page de titre, des notes de bas de page ou de fin de volume).

grecque (du Vème Sc. av. JC jusqu'aux écrits byzantins contemporains de la Renaissance), il était indispensable de s'entendre sur une norme graphique de la forme des lettres. La création de la police dite Grec du Roi est emblématique des nombreux autres efforts collégiaux, tant techniques qu'intellectuels qui durent être déployés pour faire exister l'Humanisme. Beaucoup moins régaliennes qu'elles ne le sont aujourd'hui, les académies irriguaient en réseaux le monde intellectuel et artistique jusqu'au cœur des petites cités provinciales. Il nous paraît indispensable d'analyser les Humanités numériques en recherchant les similarités d'efforts techno-intellectuels que réalise notamment une communauté comme la TEI. Ainsi, la TEI manuscrit propose des modalités normalisées pour décrire, référencer et baliser différentes « mains » d'un manuscrit, noter le degré de certitude d'une attribution, gérer des paginations complexes et/ou incertaines, paralléliser une transcription,... ou encore décrire, mesurer la matérialité de l'objet manuscrit, sa date, son degré de conservation, etc. Le chapitre TEI manuscrit cristallise ainsi de façon coopérative et normalisée les habits de cette discipline. Il cherche de plus à proposer, (toujours en consensus avec les acteurs de ce collège disciplinaire), une « valeur ajoutée informatique et réseau » : tout ce que l'ordinateur pourra détecter, retrouver, comptabiliser, comparer de façon plus rapide que des humains. La même démonstration pourrait se faire pour TEI verse qui permet de formaliser, de référencer, de décrire sémantiquement et stylistiquement des corpus versifiés, TEI performance (ex TEI drama) pour des corpus de théâtre ou des scénarios, TEI speech pour des corpus oraux, TEI critical apparatus pour des appareils critiques, etc.

Production participative enseignement–recherche ayant un objectif d'optimisation économique et équitable.

La production participative (crowdsourcing) permet de répondre à l'enjeu stratégique de pouvoir disposer de ressources littéraires numériques sans faire obligatoirement appel aux acteurs marchands du Net. Cela permet aussi de réaliser ces ressources de façon systématique en mode image ET en mode texte de façon structurée et normalisée. Par exemple, Gallica met à disposition de nombreuses ressources, mais celles-ci sont souvent produites en collaboration avec Google et sont souvent uniquement en mode image.

L'hypothèse du projet consiste à utiliser des cohortes volontaires de lycéens ou d'étudiants en première ou deuxième année pour leur faire réaliser de façon participative (avec des contrôles mutuels en double et triple aveugle) des corpus numériques (scanner, reconnaissance des caractères et balisage TEI minimum). A partir de cette étape initiale, ces lycéens ou étudiants seront impliqués dans des projets d'enseignement-recherche en approfondissant le balisage littéraire savant : pour des poèmes par exemple, ils affineront la description métrique, l'analyse des figures de stylistiques etc. En associant en synergie un enseignant et un chercheur (ou un groupe de chercheurs), l'enseignant peut ensuite impliquer les lycéens ou les jeunes étudiants dans des processus plus complexes comme le repérage des variantes et des incertitudes d'interprétation dans un appareil

critique, la description savante de manuscrits et leur mise en parallèle avec leur transcription, la mise en parallèle de traductions...

Même si ce type de pédagogie ouverte ne correspond pas à tous les profils d'apprenants, il nous apparaît que les élèves et professeurs impliqués ont tout à gagner de ce type de projet entrant notamment dans les objectifs du nouveau baccalauréat des Humanités numériques proposés en France dès cette année 2015 (Conseil National du Numérique, 2014). Sa complexité fait aussi sa richesse, dans l'entrelacs entre des contenus intellectuels et des techniques de communication via un processus de redocumentarisation.

Pour conclure, il nous paraît important d'insister sur l'importance des objectifs de durabilité du processus de production qui est pour nous centrale dans le projet.

Comme le fait remarquer Michel Serres, « l'enfant apprend à ses parents ou à ses grands-parents, l'aptitude à la maîtrise des outils techniques et l'adulte n'a qu'à transmettre les fondamentaux de la civilisation ». En l'occurrence l'Humanisme que les jeunes générations sauront naturellement adapter et traduire en Humanités numériques. Il serait logique que les patrimoines littéraires numériques soient « naturellement aménagés » par les générations de *digital natives* afin que les décideurs, aménageurs ou enseignants qui ne le sont pas puissent interagir en synergie pour transmettre les données savantes, littéraires, philosophiques ou historiques.

La durabilité du modèle économique proposé, mais aussi sa dynamique interculturelle, interdisciplinaire, interprofessionnelle, intergénérationnelle, nous paraît importante à souligner. Cependant, de nombreuses dimensions devront être testées dans le projet : l'incidence sur la formation des maîtres, la complémentarité de différents élèves ou jeunes étudiants mettant en synergie des compétences diverses, tant matérielles comme scanner ou ocriser, qu'intellectuelles, la capacité des chercheurs à interagir et à équilibrer les profits et la dissémination de la recherche, la capacité des bibliothèques ou des bibliothécaires à coopérer avec de tels dispositifs et enfin, la capacité des communautés territoriales ou des services de l'Etat à utiliser un tel type de projet pour créer du lien social.

## Bibliographie

Ben Henda, M. (2015). "Interdisciplinarity dynamics using TEI for building multilingual digital corpora: a Maghreb-based case study". Digital Humanities Institute-Beirut, American University of Beirut, 2-6 March 2015.

Conseil national du numérique (2014). Jules Ferry 3.0, Bâtir une école créative et juste dans un monde numérique | Conseil national du numérique (No. 3.0). Paris. Consulté à l'adresse <http://www.cnnumerique.fr/education-2/>

Dacos, M. (2010). Manifeste des Digital humanities [Billet]. Consulté à l'adresse: <http://tcp.hypotheses.org/318>

Eisenstein, E. I. (1991). *La révolution de l'imprimé. A l'aube de l'Europe moderne*. Paris: La Découverte.

Gille, B. (1978). *Histoire des techniques : Technique et civilisations, technique et sciences*. Gallimard.

Hudrisier, H. (1999). La « lecture assistée par ordinateur » et ses applications savantes ou pédagogiques dans un contexte interactif normalisé : la TEI (Text Encoding Initiative). *Passerelles (Revue de l'Université de Paris 8)*, (24), 57-64.

HumaNum. (s. d.). TGIR Huma-Num | Très Grande Infrastructure de Recherche dédiée aux humanités numériques. Consulté 30 mai 2015, à l'adresse <http://www.huma-num.fr/>

Ide, N., & Véronis, J. (1996). Présentation de la TEI. *Cahiers GUTenberg, TEI : Text Encoding Initiative*(24), 4-10.

Lehmans, A & Liquète, V. (2015) Conditions for a sustainable information transculture. Actes du colloque Media and information Futures, 8-9 mai 2014, Tampere (Finlande), à paraître.

Liquète, V. & Kovacs, S. (coord.) (2013) La lutte des classements. Introduction générale. *Hermès*, n°66, 2013, p. 9-15.

McLuhan, M., & Pàré, J. (1967). Marshall McLuhan. *La Galaxie Gutenberg : La genèse de l'homme typographique*. The Gutenberg galaxy. Traduit de l'anglais par Jean Paré. Editions HMH.

Romary, L. & Hudrisier, H. (2003). Le balisage normalisé des concepts et document en liaison avec les normes de l'EAD. Colloque Normes & standards pour l'apprentissage en ligne, Versailles, 19/03/03. Consulté 30 mai 2015, à l'adresse [http://www.initiatives.refer.org/Initiatives-2003/\\_notes/\\_notes/henri.htm](http://www.initiatives.refer.org/Initiatives-2003/_notes/_notes/henri.htm)

Sibertin-Blanc, C. (1957, janvier 1). Un nouvel exemplaire du prospectus de l'encyclopédie [Text]. Consulté 30 mai 2015, à l'adresse <http://bbf.enssib.fr/consulter/bbf-1957-01-0051-007>

TEI Consortium. (2015). Guidelines for Electronic Text Encoding and Interchange. (originally edited by Sperberg-McQueen, C.M. and Burnard, L., Version 2.8.0. avril 2015). Consulté 30 mai 2015, à l'adresse <http://www.tei-c.org/Guidelines/P5/>

Téniers, D. (1658). *Theatrum pictorium*. Consulté à l'adresse [http://data.bnf.fr/15575552/david\\_teniers\\_theatrum\\_pictorium/](http://data.bnf.fr/15575552/david_teniers_theatrum_pictorium/)

Université Paris 8. (s. d.). IDEFI-CréaTIC - Université Paris 8. Consulté 30 mai 2015, à l'adresse <http://www.univ-paris8.fr/IDEFI-CreaTIC>