

IXe Congrès de l'Association des Francoromanistes allemands
LES « INTERFACES NUMÉRIQUES » DANS LA RECHERCHE AUJOURD'HUI



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

TEI et humanités digitales en milieu linguistique franco-maghrébin

PROJET HUDIMA



TEI


Mokhtar BEN HENDA
Henri HUDRISIER





9. Frankoromanistenkongress
**SCHNITTSTELLEN
 INTERFACES**
 24.-27.09.14
 Westfälische Wilhelms-Universität Münster
 DFG

Projet HumanitésDigitMaghreb


Humanités digitales






XML	TEI
<p>Concepts</p> 	<p>Syntax</p> <pre><element> <element attribute="value"> content </element></pre>
	<p>Language: vocabulary and grammar</p> <pre><p> <note type="foot"> </p></pre>

HumanitéDigitMaghreb



Inter discipl inarité
 SI ON TRAVAILLE EN COLLABORATION...

Interdisciplinarité



بلدية يسر
 ترحب بكم
 +ΣΟ×Σ +C+ΣΣ.Σ+
 .O+σ+Σ+Σ+I ΣO ÷O
 LA COMMUNE D'ISSER VOUS
 SOUHAITE LA BIENVENUE

Diversité Linguistique

HD : un point d'histoire

- 1 - Le *literacy and linguistic computing* (1960-1980)
- 2 - L'*humanities computing* (1980-1994)
- 3 - Les *digital humanities* (1994-...)
- « Nous sommes des experts de la maïeutique du texte, et c'est précisément ce qui définit la contribution des sciences humaines et sociales à l'élaboration du Web sémantique ».

Lou Burnard, « Du *literary and linguistic computing* aux *digital humanities* : retour sur 40 ans de relations entre sciences humaines et informatique », in Pierre Mounier (dir.), *Read/Write Book 2*, Marseille, OpenEdition Press (« Collection « Read/Write Book » »), 2012, p. 45-58



HD : au-delà de la textualité

- Déconstruction de la catégorie « texte » et prendre en compte les catégories parlées ou visuelles « Les Humanités dé-livrées » (*):

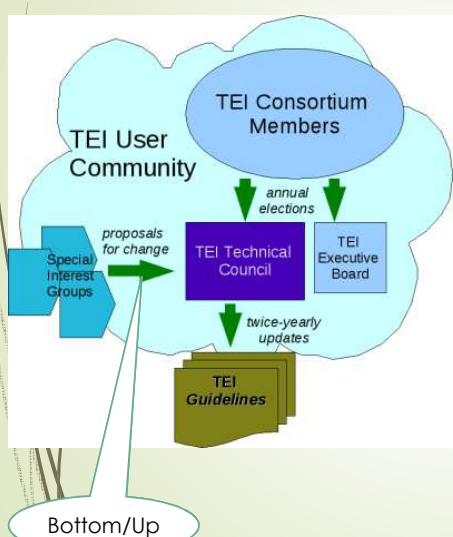
(*) Les « *humanités dé-livrées* » *Cultures parlées, visuelles et écrites, réinventées hors du livre* 1-2 octobre 2013, Amphimax 414, Université de Lausanne

- Développement d'une attention aux « *littératies plurielles* » (imprimé, informatique, audiovisuel)
- La TEI (Text Encoding Initiative) s'y adapte convenablement

TEI : Fondamentaux

- Faire des recommandations qui se basent sur des **consensus**
 - ❖ Concertation / collaboration / interopérabilité (Consortium TEI)
- Préférer les **solutions générales** à celles spécifiques à une discipline
 - ❖ Pratiques communautaires et inter-communautaires (Interdisciplinarité)
- En même temps permettant la **spécialisation** et l'**extension**
 - ❖ Appropriation / personnalisation / évolution

TEI : organisation



- ✓ 1987: Vassar College, Poughkeepsie
- ✓ 1990: P1 : SGML
- ✓ 1992: P2 : SGML
- ✓ 1994: P3 : SGML : 600 + elements
- ✓ 1995: TeiLite : SGML : 131 elements
- ✓ 1999: P3rev : SGML
- ✓ 2000: TEI Consortium
- ✓ 2001: P4 : XML comp
- ✓ 2001: TeixLite : XML
- ✓ 2005: P5 : XML
- ✓ 2007: P5 version 1
- ✓ 2010 : MEI

TEI : Modélisation (schémas)

► Spécialisation / harmonisation / interopérabilité

- ❖ La TEI est une architecture plastique, qui peut s'adapter aux nouveaux besoins et enjeux de la recherche en SHS
- ❖ Les aménagements de la TEI en schémas spécifiques :
 - **TEI Lite** : schéma pour la description de documents simples
 - **Bare TEI** : schéma basique avec un minimum d'éléments
 - **All** : schéma avec tous les modules inclus
 - **Corpus** : schéma pour encoder les corpus linguistiques
 - **MS** : schéma pour la description des manuscrits et les formes complexes
 - **Performance** : schéma pour la description d'oeuvres théâtrales et autres (audio, vidéo)
 - **Speech** : schéma pour la représentation du discours oral
 - **Verse** : schéma pour la description de poèmes
 - **Dictionaries** : schéma pour la description des dictionnaires
 - **MEI** : Music Encoding Initiative (2010)

Contexte maghrébin HumanitéDigitMaghreb

► Institutions / disciplines / chercheurs SHS

- ❖ Verticalité des structures institutionnelles
- ❖ Cloisonnement des disciplines
- ❖ Isolement des chercheurs

► Or :

- ❖ en matière de recherche littéraire, l'édition critique d'aujourd'hui est éminemment une œuvre non seulement collective mais aussi collaborative

Contexte maghrébin

Hybridation linguistique



- ❖ Ressources (orales et écrites) [simultanément] multilingues et multiécritures (Arabe classique & dialectes nat(reg)ionaux, Berbère & dialectes (kabyte), langues romanes
 - journaux, "imprimerie de ville", supports publicitaires, littérature populaire, contes, manuels pour l'apprentissage des langues, etc.
- ❖ Translittération entre latin, arabe et tfinagh
 - Sms, Réseaux sociaux, signalétique, manuscrits, annotations
- ❖ Nombreuses propriétés tonales, synonymie régionale et voyellation classique (écrit notarial) nécessitent un traitement spécial

Contexte maghrébin

Hybridation linguistique

- ❖ Constat
 - Augmentation exponentielle du nombre de chercheurs intéressés par la collecte et l'étude des textes anciens classiques et des ressources orales et en dialectes locaux,
- ❖ Mesures potentielles ;
 - Produire des modèles de référence basées sur la TEI dans les langues locales
 - Proposer de nouvelles formes d'interfaces d'accès
 - Intégrer ces ressources dans le patrimoine culturel international

Contexte maghrébin

Solutions

- ❖ Création d'une communauté de pratique spécialisée dans le traitement des ressources spécifiques
- ❖ Acquisition de nouveaux habits de recherche sur les ressources numériques en SHS
- ❖ Démontrer que le futur de l'édition critique numérique (qui est le cœur de métier de la communauté TEI), est une question hautement stratégique qui doit être collectivement débattue par les chercheurs impliqués dans les Humanités digitales.

Ce qui nous intéresse dans la TEI, c'est précisément

- ❖ sa dynamique collégiale ouverte sur la diversité linguistique non-européenne
- ❖ sa rigueur pour maintenir, enrichir et documenter des directives ouvertes sur la diversité
- ❖ Sa capacité d'assurer l'interopérabilité de toutes les ressources produites

Montage du projet HumanitéDigitMaghreb



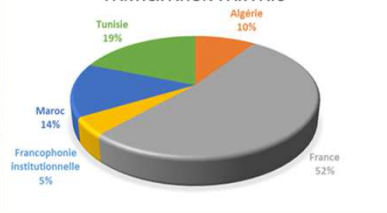
- ❑ Le projet est une recherche-action qui s'inscrit dans les orientations de recherche des Humanités Digitales
- ❑ Projet ISCC/CNRS mis en place en 2012 par une équipe de chercheurs de différentes institutions universitaires en France et au Maghreb et du Canada [des linguistes, chercheurs en littérature, culture, histoire, musique, arts,... tant en français qu'en arabe ou berbère]

Partenaires

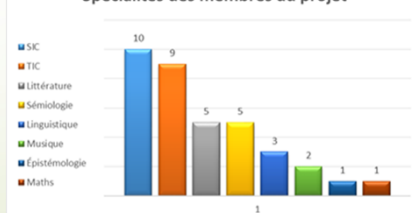


Institution	Origine
1. Agence Universitaire de la Francophonie	Francophonie
2. Chaire expérimentale internationale « Innovation, Transmission, Edition Numériques » (ITEN) à la Fondation Maison des Sciences de l'Homme (HH)	France
3. TGE-Adonis	France
4. Digital Visuel	France
5. École des Hautes Études en Sciences Sociales, Laboratoire d'Anthropologie Sociale (LAS), Paris	France
6. Faculté des Lettres et des Sciences Humaines, Université Ibn Zohr, Agadir	Maroc
7. Institut Francophone de l'Information et de la Communication, Tunis	Francophonie
8. Institut Supérieur de Documentation, Université La Manouba	Tunisie
9. Institut Supérieur de musique, université de Sousse	Tunisie
10. ISCC/CNRS	France
11. Laboratoire « Aménagement et enseignement de la langue amazighe » LAELA. Faculté de lettres et langues, Université Mouloud Mammeri de Tizi-Ouzou	Algérie
	Fax: 026411413
12. Laboratoire Paragraphe Paris 8 (EA349), MSH Paris Nord, Saint-Denis,	France
13. Laboratoire TACTIC EA 1738 Université Paris Ouest Nanterre La Défense,	France
14. MICA-E4432, Université Michel de Montaigne Bordeaux 3	France
15. TELECOM SudParis / Département ARTEMIS, Institut TELECOM, Paris	France
16. Université Mohamed Premier, faculté des Lettres, Centre Universitaire de Langues et Communication, Oujda	Maroc
17. Université Paris Ouest Nanterre La Défense	France
18. Université d'Evry, Labo Maths	France
19. LEDEN ¹ , Maison des Sciences de l'Homme Paris Nord	France

PARTICIPATION PAR PAYS



Spécialités des membres du projet



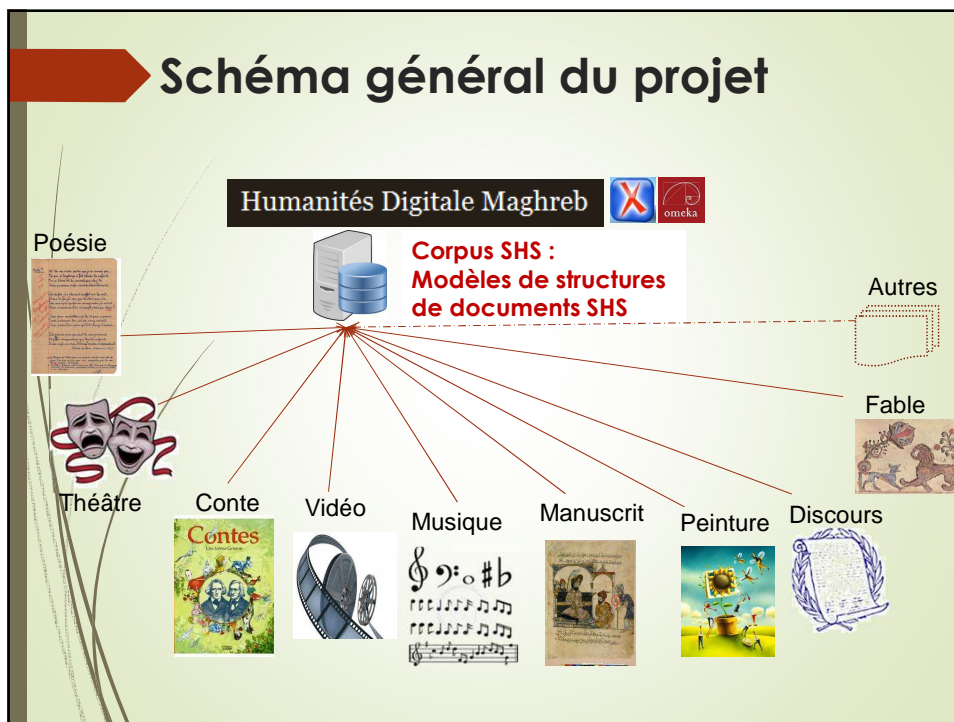
Objectifs généraux

- S'inscrire dans la dynamique des Humanités digitales comme définie dans les « Principes de Poughkeepsie » (résolutions finales de la conférence de Vassar College, New York, en novembre 1987) :
 - ❖ « 12. Nous nous engageons dans l'**édification d'une compétence collective** s'appuyant sur un vocabulaire commun, compétence collective qui procède du travail de l'ensemble des acteurs. Cette compétence collective a vocation à devenir un **bien commun**. Elle constitue une opportunité scientifique, mais aussi une opportunité d'insertion professionnelle, dans tous les secteurs.
 - ❖ 13. Nous souhaitons participer à la définition et à la diffusion de **bonnes pratiques**, correspondant à des besoins disciplinaires et **transdisciplinaires** identifiés, qui soient évolutives et issues d'un débat puis d'un consensus au sein des communautés concernées. L'ouverture fondamentale des digital humanities assure néanmoins une approche pragmatique des protocoles et des visions, qui maintient le droit à la coexistence de méthodes différentes et concurrentes, au profit de l'enrichissement de la réflexion et des pratiques.
 - ❖ 14. Nous appelons à la construction de **cyberinfrastructures** évolutives répondant à des besoins réels. Ces cyberinfrastructures se construiront de façon itérative, s'appuyant sur le constat de **méthodes et d'approches** qui font leurs preuves au sein des communautés de recherche. »

Objectifs concrets (Recherche-action)


- Définir des modèles de structure de corpus numériques en SHS fondés sur des typologies de documents (genres littéraires) proposés par les partenaires du projet
- Produire un environnement technologique de production/exploitation de corpus SHS normalisé par la TEI
- Mettre en ligne un corpus de démonstration de ressources mutualisées en SHS

Schéma général du projet



Une méthode de travail

- Les **SPÉCIALISTES SHS** identifient la structure logique d'un type de document dans leurs spécialités respectives (roman, poésie, théâtre, conte, manuscrit, peinture, musique, ...)



Une méthode de travail

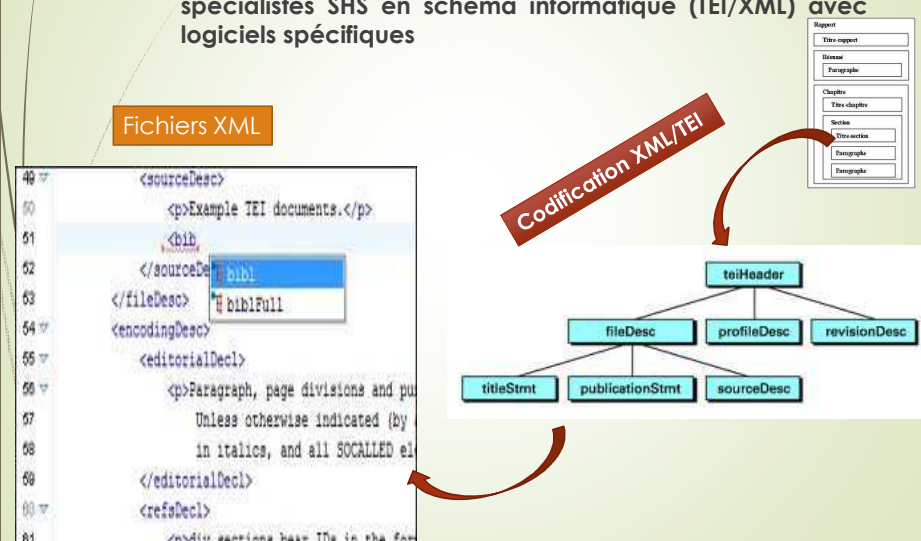
- Les **TECHNICIENS** convertissent le modèle défini par les spécialistes SHS en schéma informatique (TEI/XML) avec logiciels spécifiques

Fichiers XML

```

40 <sourceDesc>
50 <p>Example TEI documents.</p>
51 <bibl>
52 </sourceDesc>
53 </fileDesc>
54 <encodingDesc>
55 <editorialDecl>
56 <p>Paragraph, page divisions and pu
57 Unless otherwise indicated (by
58 in italics, and all SOCALLED el
59 </editorialDecl>
60 <refsDecl>
61 <p>div sections bear IDs in the for
    
```

Codification XML/TEI



Une méthode de travail

- Les **DOCUMENTALISTES** font le référencement des corpus par métadonnées et constituent les BD

CMS / BD

Référencement/indexation

The image shows a workflow for digital humanities. On the left, a screenshot of the 'Humanités Digitale Maghreb' CMS interface displays the 'Add an Item' form. The form includes fields for 'Title', 'Subject', and 'Description', each with an 'Add Input' button and a 'Use HTML' checkbox. A red arrow points from the 'Add an Item' button to a Notepad window on the right. The Notepad window shows XML code for a TEI header, including elements like <fileDesc>, <encodingDesc>, <profileDesc>, and <revisionDesc>. A green box labeled 'Référencement/indexation' is positioned above the Notepad window, and another red arrow points from it to the XML code.

Outils (interfaces)

- Oxygen (éditeur de schémas)

The image shows the Oxygen XML Editor website. The page title is 'Download <oXygen/> XML Editor, Author or Developer'. Below the title, there are three download options, each with a logo and a 'Download' button:

- <oXygen/> XML EDITOR
- <oXygen/> XML DEVELOPER
- <oXygen/> XML AUTHOR

The website header includes navigation links for Products, Downloads, Register, Shop, Resources, Support, and Company. A search bar is also visible.

Outils (interfaces)

► Oxygen : logiciel de production de corpus numérique en TEI

The diagram illustrates the Oxygen workflow. At the top, Documentum CMS handles Virtual documents, XML applications, and Media files. These are stored in Oxygen, which then publishes them to Desktop XML Authoring. The final output can be generated in various formats: PDF (Adobe), WebHelp, HTML, Eclipse Help, or CHM Help.

Outils (interfaces)

TEI Roma: generating customizations for the TEI

<http://www.tei-c.org/Roma/>

TEI Roma is a tool for working with TEI customizations. A TEI customization is a document from which you can generate a schema defining which elements and attributes from the TEI system you want to use, along with customized HTML or PDF documentation of it. The schema generated can be expressed in any of DTD, RELAXNG W3C Schema or Schematron languages.

You can make or modify your TEI customization in several different ways:

- Build up: create a new customization by adding elements and modules to the smallest recommended schema
- Reduce: create a new customization by removing elements and modules from the largest possible schema
- Create a new customization starting from a template
- Use or modify an existing TEI-defined customization
- Upload a customization

Community-maintained customizations can be downloaded

The screenshot shows the TEI Roma interface. A dropdown menu is open, listing templates: TEI Absolutely Bare, TEI for Linguistic Corpora, TEI for Manuscript Description, TEI with Drama, TEI for Speech Representation, TEI for authoring ODD, TEI with SVS, TEI with MathML, TEI with Xinclude, and TEI with W3C internationalisation elements. Callout boxes link these options to their descriptions: 'Adapter à partir d'un schéma de base' (TEI Absolutely Bare), 'Adapter à partir d'un schéma exhaustif' (TEI for Linguistic Corpora), 'Adapter à partir d'un schéma spécifique' (TEI for Manuscript Description), 'Modifier une adaptation existante' (TEI with Drama, TEI for Speech Representation, TEI for authoring ODD), and 'Utiliser une adaptation existante' (TEI with SVS, TEI with MathML, TEI with Xinclude, TEI with W3C internationalisation elements).

Outils (interfaces)

TEI Roma: generating validators for the TEI

Set your parameters

[New](#)
[Customize](#)
[Language](#)
[Modules](#)
[Add Elements](#)
[Change Classes](#)
[Schema](#)
[Documentation](#)
[Save Customization](#)

Set your parameters

Title

Filename

Prefix for TEI pattern names in schema

Language

- English
- Deutsch
- Français
- Russian
- Svenska
- 日本語
- 中文

Author name

Description

Outils (interfaces)

TEI Roma: generating validators for the TEI

Add Element

[New](#)
[Customize](#)
[Language](#)
[Modules](#)
[Add Elements](#)
[Change Classes](#)
[Schema](#)
[Documentation](#)
[Save Customization](#)
[Sanity Checker](#)

[go back to list](#)

Defining a new element:

Name

Namespace

Description

Model classes

<input type="checkbox"/> model.addrPart	<input type="checkbox"/> model.addressLike
<input type="checkbox"/> model.applicationLike	<input type="checkbox"/> model.availabilityPart
<input type="checkbox"/> model.biblLike	<input type="checkbox"/> model.biblPart
<input type="checkbox"/> model.castItemPart	<input type="checkbox"/> model.catDescPart
<input type="checkbox"/> model.certLike	<input type="checkbox"/> model.choicePart
<input type="checkbox"/> model.common	<input type="checkbox"/> model.contentPart
<input type="checkbox"/> model.dateLike	<input type="checkbox"/> model.descLike
<input type="checkbox"/> model.dimLike	<input type="checkbox"/> model.div1Like
<input type="checkbox"/> model.div2Like	<input type="checkbox"/> model.div3Like
<input type="checkbox"/> model.div4Like	<input type="checkbox"/> model.div5Like

www.tei-c.org/Roma/startroma.php?mode=listAddedElements

Outils (interfaces)

- OxGarage : pour convertir des formats de documents numériques de et vers la TEI ;

OxGarage Conversion

Select the format into which you want to convert your document

Convert from: 2



Documents

- Compiled TEI ODD Document
- DocBook Document
- Microsoft Word Document (.doc)
- Microsoft Word Document (.docx)
- ODD Document
- Open Office Text Document (.odt)
- OpenOffice 1.0 Text Document (.sxx)
- Plain Text (.txt)
- Rich Text Format (.rtf)
- TEI P4 XML Document
- TEI P5 XML Document
- TEI Tite XML Document
- WordPerfect Document (.wpd)
- xHTML Document

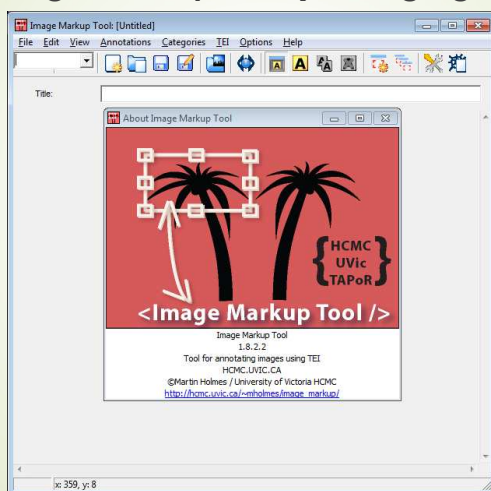
Convert to: 2

- Comma-Separated Values (.csv)
- ePub Document
- LaTeX Document
- Microsoft Excel Document (.xls)
- Microsoft Word Document (.docx)
- National Library of Medicine (NLM) DTD 3.0
- Open Office Spreadsheet (.ods)
- Open Office Text Document (.odt)
- OpenOffice 1.0 Spreadsheet (.sxc)
- OpenOffice 1.0 Text Document (.sxx)
- PDF Document
- Plain Text (.txt)
- RDF XML
- Rich Text Format (.rtf)
- Tab-Separated Values (.tsv)
- TEI P5 XML Document

<http://www.tei-c.org/oxgarage/>

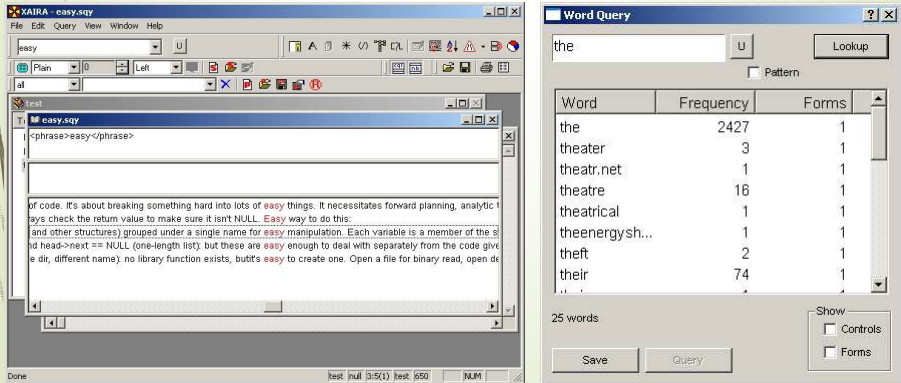
Outils (interfaces)

- Image Markup Tool (balisage graphique TEI)



Outils (interfaces)

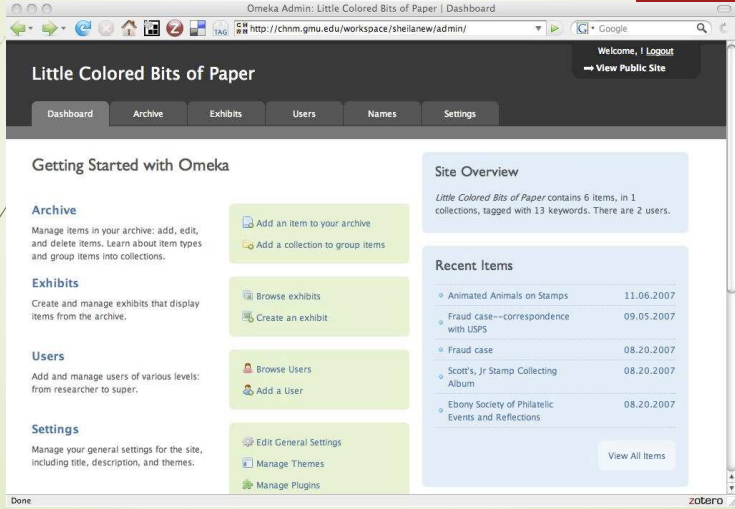
- Xaira (moteur de recherche xml)**
 - Explorer le langage d'un corpus :
 - chercher des mots, des motifs, des phrases ou des balises XML
 - analyser la façon dont ils sont utilisés.



Word	Frequency	Forms
the	2427	1
theater	3	1
theatr.net	1	1
theatre	16	1
theatrical	1	1
theenergysh...	1	1
theft	2	1
their	74	1

Outils (interfaces)

- OMEKA (Aggrégateur/CMS)**



Omeka Admin: Little Colored Bits of Paper | Dashboard

Little Colored Bits of Paper

Getting Started with Omeka

- Archive**
 Manage items in your archive: add, edit, and delete items. Learn about item types and group items into collections.
 - Add an item to your archive
 - Add a collection to group items
- Exhibits**
 Create and manage exhibits that display items from the archive.
 - Browse exhibits
 - Create an exhibit
- Users**
 Add and manage users of various levels: from researcher to super.
 - Browse Users
 - Add a User
- Settings**
 Manage your general settings for the site, including title, description, and themes.
 - Edit General Settings
 - Manage Themes
 - Manage Plugins

Site Overview

Little Colored Bits of Paper contains 6 items, in 1 collections, tagged with 13 keywords. There are 2 users.

Recent Items

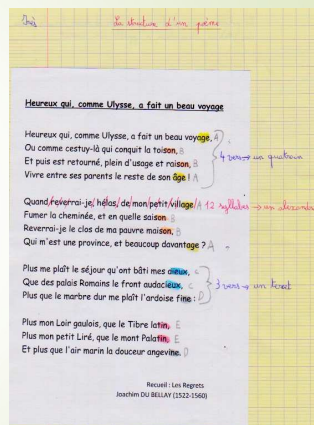
- Animated Animals on Stamps 11.06.2007
- Fraud case—correspondence with USPS 09.05.2007
- Fraud case 08.20.2007
- Scott's, Jr Stamp Collecting Album 08.20.2007
- Ebony Society of Philatelic Events and Reflections 08.20.2007

View All Items

PRODUCTION : Structuration/encodage d'un poème (TEIVerse)

► Identifier dans un poème

- ❖ La mesure des vers
- ❖ Les différents types de vers
- ❖ Les groupes de vers (couplets, tercets, quatrains)
- ❖ La strophe
- ❖ La forme de la strophe
- ❖ La rime
- ❖ L'enjambement
- ❖ Le rejet et le contre-rejet
- ❖ ...



Le choix de granularité peut varier entre grands segments et éléments plus petits

PRODUCTION : Structuration/encodage d'un poème (TEIVerse)

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <Title>Heureux qui comme
      Ulysse</Title>
      <respStmt>
        <resp>Auteur</resp>
        <Name>Joachim Du
        Belay</Name>
      </respStmt>
    </titleStmt>
    <sourceDesc>
      <Bibl>Recueil Les regrets
      </Bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>...
```

```
<text>
  <body>
    <head>Heureux qui, comme Ulysse, a fait un beau
    voyage </head>
    <lg>
      <l>Heureux qui, comme Ulysse, a fait un beau voyage
      </l>
      <l>Ou comme cestuy-là qui conquit la toison </l>
      <l>Et puis est retourné, plein d'usage et raison </l>
      <l>Vivre entre ses parents le reste de son âge ! </l>
    </lg>
    <lg>
      <l>Quand reverrai-je, hélas, de mon petit village </l>
      <l>Fumer la cheminée, et en quelle saison </l>
      <l>Reverrai-je le clos de ma pauvre maison </l>
      <l>Qui m'est une province, et beaucoup davantage ?
      </l>
    </lg>
    <lg>
      <l>Plus me plaît le séjour qu'ont bâti mes aïeux </l>
      <l>Que des palais Romains le front audacieux </l>
      <l>Plus que le marbre dur me plaît l'ardoise fine : </l>
    </lg>
    <lg>
      <l>Plus mon Loir gaulois, que le Tibre latin </l>
      <l>Plus mon petit Liré, que le mont Palatin </l>
      <l>Et plus que l'air marin la douceur angevine </l>
    </lg>
  </body>
</text>
```


PRODUCTION : Structuration/encodage d'une pièce de théâtre (TEIDrama)

► Identifier dans une pièce de théâtre :

- ❖ L'interprétation et la mise en scène
- ❖ La parole sur scène
- ❖ Les didascalies
- ❖ Les types d'exposition
- ❖ Le dénouement
- ❖ Le prologue
- ❖ ...

► Interventions et indications scéniques

```
<sp who="P">
  <speaker>Paul</speaker>
  <p>Il fait chaud
  aujourd'hui...</p>
</sp>
<stage>Paul allume une
cigarette</stage>
```

ACTE PREMIER

SCÈNE I

Elseneur. — Une plate-forme devant le château.

FRANCISCO montant la garde, BERNARDO vient à lui.

BERNARDO. — Qui va là ?

FRANCISCO. — Non, répondez vous-même. Arrêtez-vous et faites-vous reconnaître.

BERNARDO. — Vive le roi !

FRANCISCO. — Bernardo ?

BERNARDO. — En personne.

FRANCISCO. — Vous venez très-soigneusement à votre heure.

BERNARDO. — Minuit vient de sonner : va regagner ton lit, Francisco.

FRANCISCO. — Pour cette délivrance, mille grâces. Le froid est aigre, et j'ai le cœur saisi.

PRODUCTION : Structuration/encodage d'un manuscrit (TEIManuscript)

<surface> : une page, une stèle, tout objet avec une inscription

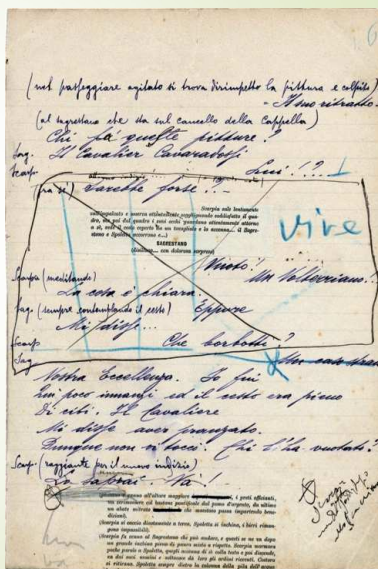
- ❖ La surface contient des zones et des lignes
- ❖ Elle a des coordonnées

<zone> : Une aire de la superficie définie de façon arbitraire à des fins éditoriaux. Les zones peuvent se superposer : la superposition est définie selon des coordonnées spatiales

- ❖ Peut contenir des <line>
- ❖ Dispose de coordonnées

<line> : une suite de texte identifiée de façon claire par l'éditeur

- ❖ Peu contenir du texte et des <zone>
- ❖ Ne dispose pas de coordonnées



PRODUCTION : Structuration/encodage d'un manuscrit (TEIManuscript)

1 April 2009

Samarina is a Greek island of water that comes from the natural springs of Stilos, in Crete.

Fed Birds in the park today.
Might write an article about the Thick-billed Warbler.

PRODUCTION : Structuration/encodage d'un manuscrit (TEIManuscript)

```

<surface
  ulx="0"
  uly="0"
  lrx="700"
  lry="1000">
  <!-- ... -->
</surface>
<zone
  ulx="93"
  uly="681"
  lrx="967"
  lry="1568">
  <graphic url=" gb.jpg"/>
</zone>
<sourceDoc>
  <surface ulx="0" uly="0" lrx="200" lry="300">
    <zone ulx="10" uly="43" lrx="185" lry="84"
      rotate="0">
      <zone>
        <line rend="right"> 1 April 2009</line>
      </zone>
      <line>Fed Birds in the park today.</line>
      <line>Might write an article about</line>
      <line>the Thick-billed Warbler.</line>
      </zone>
      <zone ulx="9" uly="20" lrx="70" lry="60"
        rotate="90">
        <line>Samarina is a Greek</line>
        <line>brand of water that</line>
        <line>comes from the natural</line>
        <line>springs of Stilos, in</line>
        <line>Crete</line>
      </zone>
    </surface>
  </sourceDoc>

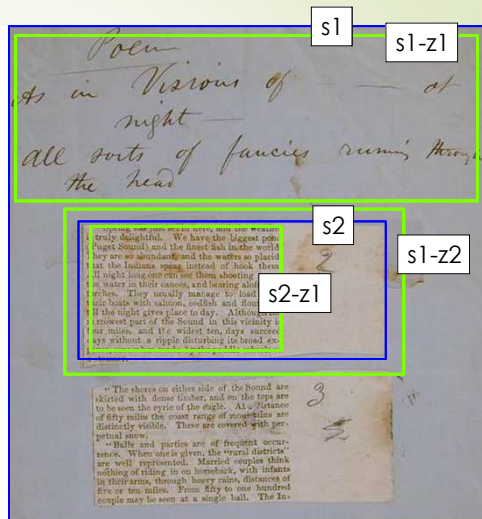
```

PRODUCTION : Structuration/encodage d'un manuscrit (TEIManuscript)

```

<surface xml:id="s1" ulx="0" uly="0"
  lrx="50" lry="50">
  <zone xml:id="s1-z1" ulx="1" uly="1"
    lrx="10" lry="10">
    <line>Poem</line>
    <!-- ... -->
    <line>the head</line>
  </zone>
  <zone xml:id="s1-z2" ulx="4" uly="4"
    lrx="20" lry="20">
    <surface xml:id="s2" ulx="0"
      uly="0" lrx="100" lry="100">
      <zone xml:id="s2-z1" ulx="10"
        uly="10" lrx="90" lry="95"> Spring has just
        set in here, and the weather [...] a
        steamer </zone>
      </surface>
    </zone>
  </surface>

```



PRODUCTION : Structuration/encodage d'un graphique (Graph)

Image Markup Tool

```

<teiHeader>
  <fileDesc>
    <titleStat><title>Titre de la photo</title></titleStat>
  </fileDesc>
  <publicationStat>
    <p></p></publicationStat>
  <sourceDesc>
    <p>593 x 344 4608 x 3456</p></sourceDesc></fileDesc>
  <encodingDesc></encodingDesc></teiHeader>

```


PRODUCTION : Codification/encodage d'une chanson (paroles) <TeiHeader> [TEI]

Espace de nom

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/sc
3 <?xml-stylesheet type="text/css" href="7-Taht-El-Yasmina.css"?>
4 <TEI xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5   xmlns="http://www.tei-c.org/ns/1.0"
6   xmlns:math="http://www.w3.org/1998/Math/MathML"
7   xmlns:xi="http://www.w3.org/2001/XInclude"
8   xmlns:svg="http://www.w3.org/2000/svg">
9   <teiHeader>
10     <fileDesc>
11       <titleStmt>
12         <title>تحت الياسمينه في الليل</title>
13         <author role="chanteur">الهادي الجويني</author>
14         <author role="compositeur">الهادي الجويني</author>
15         <author role="parolier">عزة العبيدة</author>
16       </titleStmt>
17       <publicationStmt>
18         <p></p>
19       </publicationStmt>
20       <sourceDesc>
21         <p></p>
22       </sourceDesc>
23     </fileDesc>
24   </teiHeader>

```

Métadonnées (TEI Header)

PRODUCTION : Codification/encodage d'une chanson (paroles) <text><body>

Stanza (Quatrin)

```

25 <text>
26 <body>
27 <head>تحت الياسمينه</head>
28 <lg type="stanza">
29 <l>تحت الياسمينه في الليل</l>
30 <l>نسمه والورد محاذيني</l>
31 <l>الأغصان عليا تسميل</l>
32 <l>تمسحلي في دمه عيني</l>
33 </lg>
34 <lg type="stanza">
35 <l>تحت الياسمينه اتكيت</l>
36 <l>عدلت السعود وغنيت</l>
37 <l>وناطر دمعي ويكيت</l>
38 <l>تفكرتك كيف كنت تجيني</l>
39 </lg>
40 <lg type="stanza">
41 <l>جنية مزينها النوار</l>
42 <l>فاحت من ريحت الأزهار</l>
43 <l>تفكرتك شعبلت النار</l>
44 <l>عملت لهليه في كنييني</l>
45 </lg>
46 <lg type="stanza">
47 <l>متوحش وحدي مختار</l>
48 <l>لاقمرة ولا حس أطيّار</l>
49 <l>كان النسمه ع الأشجار</l>
50 <l>توانس فيا وتواسيني</l>
51 </lg>
52 </body>

```

- تحت الياسمينه في الليل
- نسمه والورد محاذيني
- الأغصان عليا تسميل
- تمسحلي في دمه عيني
- تحت الياسمينه اتكيت
- عدلت السعود وغنيت
- وناطر دمعي ويكيت
- تفكرتك كيف كنت تجيني
- جنية مزينها النوار
- فاحت من ريحت الأزهار
- تفكرتك شعبلت النار
- عملت لهليه في قلبي
- متوحش وحدي مختار
- لاقمرة ولا حس أطيّار
- كان النسمه ع الأشجار
- توانس فيا وتواسيني

PRODUCTION : Codification/encodage d'une chanson (notes musicales) [MEI]



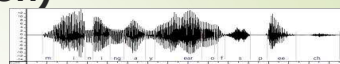
```

13 <music>
14 <body>
15 <mdiv type="chanson">
16 <score>
17 <scoreDef meter.count="3" meter.unit="4" key.sig="2f" key.mode="dorian"/>
18 <section n="mesito8">
19 <measure n="1">
20 <staff>
21 <layer>
22 <note pname="a" oct="4" dur="4" stem.dir="up" syl="نجد"/>
23 <note pname="b" oct="4" dur="4" stem.dir="down" syl="البا سمينة"/>
24 <note pname="b" oct="4" dur="4" stem.dir="up" syl="في الليل"/>
25 </layer>
26 </staff>
27 </measure>
28 <measure n="2">
29 <staff>
30 <layer>
31 <note pname="g" oct="4" dur="4" stem.dir="up" syl="تسمة"/>
32 <note pname="f" oct="4" dur="4" stem.dir="up" syl="والورد"/>
33 <note pname="g" oct="4" dur="4" stem.dir="up" syl="معا ذيني"/>
34 </layer>
35 </staff>
36 </measure>
37 </section>
38 </score>
39 </mdiv>
40 </body>
41 </music></mei>

```

PRODUCTION : Structuration/encodage d'un fichier son (TEISpeech)

- Un énoncé se définit comme une « séquence attribuée à un locuteur à un instant » : *flux temporel* « Timeline »
- Les énoncés sont regroupés dans des balises <div> (division)
- Les segments de discours composant l'énoncé sont balisés par <s> ou <seg>
- Opérer, au moyen de balises et d'attributs, une description très fine des différents phénomènes de communication oraux et non-oraux qui font partie du discours :
 - ❖ pauses, chevauchements de paroles, changements d'intonation, de voix ou de langue, expressions vocalisées (tousseur, rire, se moucher, grogner...), gestes, etc.



```

<u who="#locuteur" sync="#T234">
<seg type="interrupted">
<kinesic>
<desc>coughs</desc>
</kinesic>
<w>you</w>
<w>must</w>
</seg>
<seg type="declarative">
<w>you</w>
<w>should</w>
<w>let</w>
<pause dur="short"/>
<w>it</w>
<w>be</w>
</seg>
<seg type="emphatic">
<vocal>
<desc>laughs</desc>
</vocal>
<w>please</w>
</seg>
</u>

```


Organisation (Saisie / Référencement / Indexation / Recherche)

Objet Collection Typologie Meta Doc Exposition Batch TEI

HumanitésDigitMaghreb

Welcome, [okhtar](#) | [Log Out](#)

Items Collections Item Types Tags Simple Pages Exhibits Dropbox TEI Config OAI-PMH Harvester

Dashboard

✓ A new version of Omeka is available for download. [Upgrade to 2.2.2](#)

Getting Started with Omeka

Items

[Add a new item to your archive](#) Manage items in your archive: add, edit, and delete items.
[Browse your items](#)

Collections

[Add a collection to group items](#) Manage collections in your archive: add, edit, and delete collections.
[Browse your collections](#)

Users

[Browse Users](#) Manage users of various levels: from

Site Overview

HumanitésDigitMaghreb contains 6 items, in 10 collections, tagged with 15 keywords. There are 13 users.

Recent Items

- Fichier DOC [TEST]
- Fichier PDF [TEST]
- Fichier DOCX [TEST]
- Fichier PDF [TEST]
- Fichier MP3 [TEST] : chant des partisans

Recherche



<http://www.hudima.org/admin/>

Organisation (saisie)

OMEKA : agrégateur de corpus/collection/document



Digital Humanities Maghreb

Articles Collections Types d'éléments Balises OAI-PMH batteuse Pages simples

Parcourir les collections (5 au total) [Ajouter une collection](#)

✓ La collection "Musique inuite" a été ajoutée avec succès!

ID	Nom	Collectionneurs	Date d'ajout	Nombre total d'articles	Modifier?
1	Préliminaires	Pas de collectionneurs	24 août 2012	1	Éditer
2	Documents techniques	Pas de collectionneurs	24 août 2012	Voir	Éditer
4	Poésie arabe	Pas de collectionneurs	25 août 2012	Voir	Éditer
5	Contes populaires berbères	Pas de collectionneurs	25 août 2012	Voir	Éditer
6	Musique inuite	Pas de collectionneurs	26 août 2012	Voir	Éditer

Organisation (référencement)

► OMEKA : référencement par métadonnées (Dublin Core)



Digital Humanities Maghreb

Articles Collections Types d'éléments Balises OAI-PMH batteuse Pages simples

Ajouter un article Public: En vedette:

Dublin Core

Les métadonnées de type objet

- Collection
- Fichiers
- Balises

Dublin Core

L'élément de métadonnées Dublin Core réglé. Ces éléments sont communs à toutes les ressources Omeka, y compris des articles, des fichiers, des collections, des expositions et des entités. Voir <http://dublincore.org/documents/dces/>.

Titre

[Ajouter une entrée](#)

Utilisez HTML
Un nom donné à la ressource. Typiquement, le titre sera le nom par lequel la ressource est officiellement connue.

Sujet

[Ajouter une entrée](#)

Utilisez HTML
Le sujet de la ressource. Typiquement, le sujet sera représenté à l'aide des mots-clés, phrases clés ou de codes de classification. Il est recommandé.

Organisation (indexation)

► OMEKA : Moissonnage par protocole OAI-PMH



Digital Humanities Maghreb Bienvenue, Mokhtar

Articles Collections Types d'éléments Balises **OAI-PMH batteuse** Pages simples

OAI-PMH batteuse

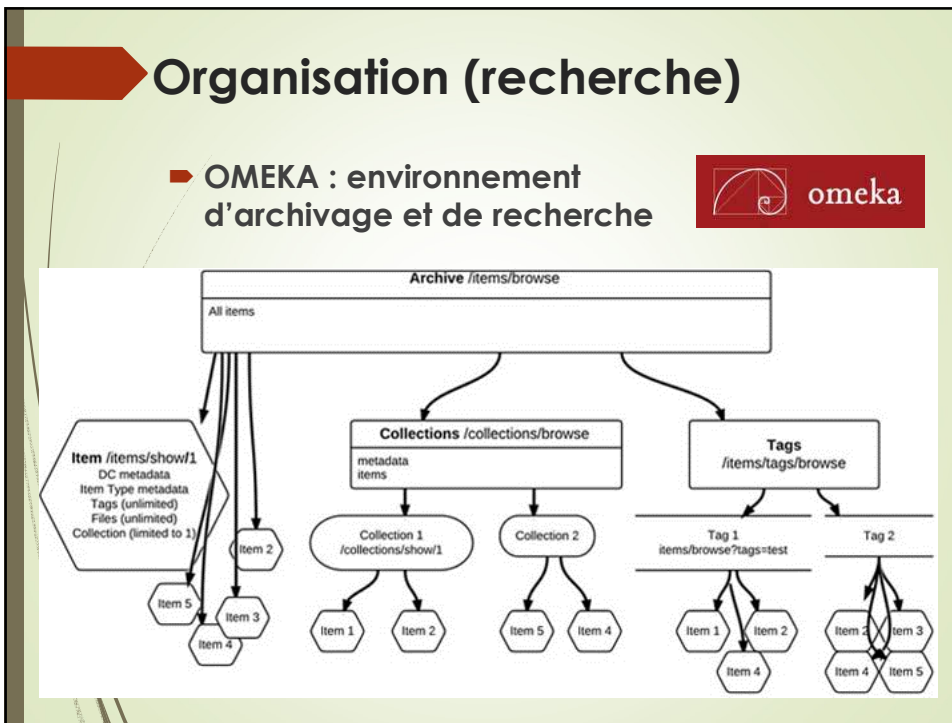
Base de l'URL

L'URL de base du fournisseur de données OAI-PMH.

[Voir Règle](#)

Récoltes

ID	URL de base	Préfixe métadonnées	Réglez Spec	Définir le nom	Statut
1	http://oai.datacite.org/oai	oai_dc			Erreur



Organisation (recherche)

▶ Recherche multicritère

HumanitésDigitMaghreb

Items	Collections	Item Types	Tags	Simple Pages	Exhibits
-------	-------------	------------	------	--------------	----------

Advanced Search

Search for Keywords

Narrow by Specific Fields contains

Add a Field

Search by a range of ID#s (example: 1-4, 156, 79)

Search By Collection

Search By Type

Perspectives

► Une stratégie d'appui et d'accompagnement

- ❖ Traduction en français de la version complète de la « TEI P5 Guidelines » (1500 pages en comptant les annexes dont environ 600 pages sont traduites à ce jour).
- ❖ Une version arabe est aussi une action liée à ce projet
- ❖ Un nouveau sujet linguistique est actuellement traité par TEI: l'arabe est la première écriture de droite à gauche traitée par la TEI

► Une dynamique de projets de recherche-action

- ❖ Projet MEI & TEI-Euromed
- ❖ Projet HumanEduc

► Une stratégie de positionnement élargie à d'autres communautés de pratiques dans les SHS

- ❖ Maghreb
- ❖ Francophonie plurielle