

Digital Humanities Institute – Beirut
American University of Beirut
2-6 March 2015

INTERDISCIPLINARY CHALLENGE USING TEI FOR BUILDING MULTILINGUAL DIGITAL CORPORA A MAGHREB DH CASE STUDY

Mokhtar Ben Henda



The general scope

- **Creating a DH community of practice in the Maghreb region**
 - ❖ Treat multilingual and interdisciplinary resources
 - ❖ Acquire new research and editing habits with SH digital resources
- **Using TEI as technical solution**
 - ❖ For its collegiate dynamics, open to non-Western linguistic diversity
 - ❖ For its accuracy to maintain, enrich and document open guidelines on diversity
 - ❖ For its ability to ensure the interoperability of all produced resources



► Delimiters

Multilingualism as a core question

Hybridization of
oral and written languages
in the Maghreb



► Languages hybridization

- Use of classical or literary (scholar) Arabic in formal contexts :
 - ❖ Administration, politics, media, education, religion...
 - ❖ Dialects in daily life (trade, traditions, ...)
 - ❖ "Mother" tongue, "native" language ?
- Use of Arabic dialects:
 - ❖ Morocco : Haḍarī (Fez, Rabat), Bedui (Marrakech), 'Arūbī (west)
 - ❖ Algeria: wahrani (West), Ksantini (East), Dziri (Center)
 - ❖ Tunisia: Beldi (Tunis), Sahli (Center), Bedui (South)
- Ibn Žaldūn, al-Muqaddima, p.984 : (1377) «... Aboriginal Language [Algerian] has taken over the Arabic language. The result is a mixed language, in which Berber dominates and in which North African dialects are farthest from Classical Arabic ».

Writing hybridization



Turkish and Arabic manuscript dictionary. 1201 AH [1786/87 AD].



A diary page of November 1, 1895 with Ottoman-Turkish, Armenian, Hebrew, Latin, Arabic scripts

Writing hybridization

Notary records in Arabic and French (1342 Hg/1923 AD)

Strong claims of Berber as a national language



Berber languages

- It is estimated that there are between 14 and 25 million speakers of Berber languages

| | | |
|--------------------------|------------------|---------|
| <u>Tachelhit</u> | 3 million | Morocco |
| <u>Tamazight</u> | 3 million | Morocco |
| <u>Kabyle</u> | 2.5 to 6 million | Algeria |
| <u> Tarifit</u> | 1.7 million | Morocco |
| <u>Tachawit</u> | 1.4 million | Algeria |
| <u>Tamajaq</u> | 640,000 | Niger |
| <u>Tamasheq</u> (Tuareg) | 281,000 | Mali |
| <u>Tamajeq</u> | 250,000 | Niger |
| <u>Nafusi</u> | 167,000 | Libya |

Tamasheq has five varieties.

Tachelhit has a dialect called Susiua.

Tamazight has two major varieties: Central Altas and Southern Oran.

Tarifit has at least two dialects, one of which may be a separate language

Kabyle has two dialects: Greater and Lesser Kabyle.

Berber writing systems

- Berber languages are written with Roman, Arabic and Tifinagh alphabets

Sample text in Tamazight

Transliteration

Transliteration
Imdan, akkaen ma lian ttalien d lilejjen msawan di lhwerma d yizerfan-ghur sen tamsakwit d laqué u yessefk ad-till tegmatt gar asen.

All human beings are born free and equal in dignity and rights; they are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.



The female name "Saidq" written in tifinagh

POÉSIES POPULAIRES DE LA KABYLIE.
jetèrent dans les broussailles comme des sangliers; — toutes
descendirent la nuit dans les champs.

امّة يَقْهِمُ الْسَّقْرَ لِلْأَمْفَوْرَةِ فَتَبَيَّنَ

امله يطلع السمر لن اموري محبي
يسعى إلة تغراش تبت ألغان ورقبي
أحلاه خذمه أسلفة سلامة لفحة قلب

*Amelak ! ja Fot me en Sossoover ! — — la! emus amezzour d' el k'ensi !
iem is inmou'a l'droch — ibboui te, ther'ab over thelli.*

— la dame aux bandes et au henné ! — son nom était connu de toutes les

— l'ennemi l'a enlevé, elle a disparu. — La veuve Beni-Sliman, — ô mes larmes, coulez à torrent
 عَنْ بَعْدِ مُهَاجَرَةِ قَوْمٍ وَتَرَكَهُمْ شَفَقٌ نَّافِذٌ
 يَنْهَا مُهَاجِرَةً سَيِّئَتِ الْأَنْفُسُ أَنْجَارَةً
 وَزَانَتِهَا حَمْدًا لِلْمُنْكَرِ أَنْجَعَتِهَا

If this idiom d' est arch a — one the other's kind v'of munkri.

*me neko astta setim — el ha'el en ad i rebbi !
our neudi h'ad t'hara'ibd — qd' ini : Allah l'a rebbi !*

³ Voir la note à, page 126.
⁴ Soixante réaux font 150 francs, le réal vaut à fr. 50 cts. Les Ait-Iraten, les plus acharnés à la lutte, ont été les plus immobiles.

⁵ Les mots *Allak*, *la rébi* « Dieu, mon maître » sont une exclamation de pitie et de sympathie qu'il l'on adresse à une personne qui vient d'exprimer un malheur.

mis à la lutte, ont été les plus imposés. qui vient d'éprouver un malheur.

2020 RELEASE UNDER E.O. 14176

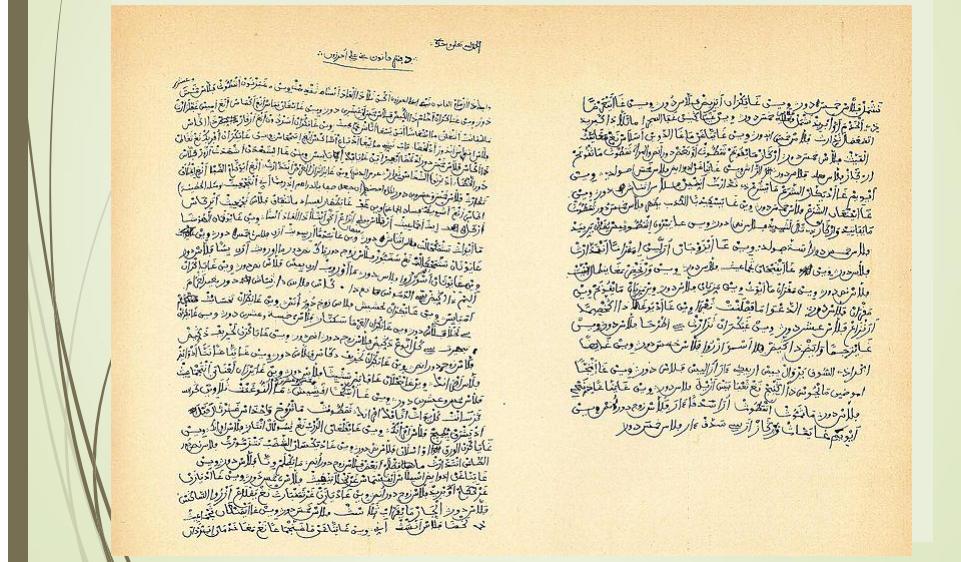
³ Les mots *Allah*, *ia rebbi* « Dieu, mon maître ! » sont une exclamation de pitié et de sympathie que l'on adresse à une personne qui vient d'éprouver un malheur.

qui vient d'éprouver un malheur.

Berber language poetry in Arabic script with its translation in French

Berber writing systems

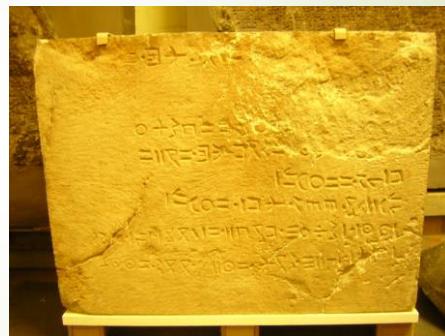
► Kabylian manuscript of the XVIII Century



Berber « Rosetta » stone

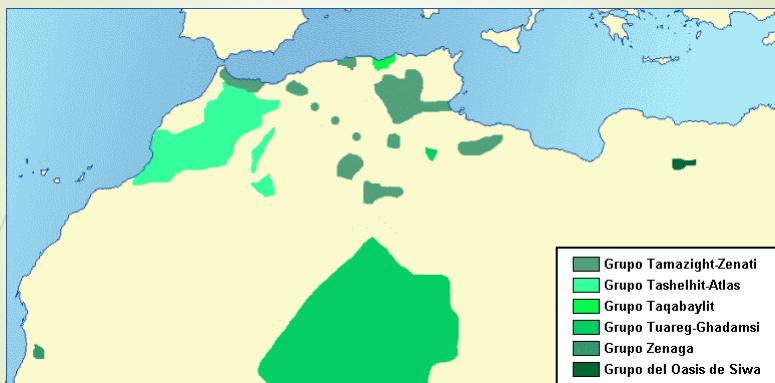
ΣΕΛΙΩΙ, οΚΚΩΙ Κο ΗΗΟι + + ΗΗΟι Λ ΣΗΗΗΕΣΩΙ ΕΘΩΠΩΙ ΛΣ
ΗΛΠΩΟΕΙ Λ ΣΣΜΩΟΕΙ - ΒΟ ΤΩΙ ΤΟΕΡΛΕΣΤ Λ ΗΟΛΕΗΙ :
ΣΣΟΟΗΕΚ οΛ-+ΣΗΣ +ΣΕΤ ΧΟΟ οθοι.

The Tifinagh alphabet is thought to have been derived from the Phoenician alphabet.



Berber « Rosetta stone » from Dugga (Tunisia)
Phoenician and Tifinagh scripts (British Museum)

Berber spreading in Maghreb



- 60% in Morocco, 50% in Algeria, 2% in Tunisia

Berber recognition as a national language

- Algeria: constitutional amendment (2002)

Chapitre I : De l'Algérie

Article 1er - L'Algérie est une République Démocratique et Populaire.
Elle est une et indivisible.

Art. 2 - L'Islam est la religion de l'Etat.

Art. 3 - L'Arabe est la langue nationale et officielle.

Art. 3 bis [Révision constitutionnelle du 10 avril 2002 adoptée suivant la procédure prévue à l'article 176 de la Constitution]- Tamazight est également langue nationale. L'Etat oeuvre à sa promotion et à son développement dans toutes ses variétés linguistiques en usage sur le territoire national.

Berber recognition as a national language

► Morocco « Constitutional amendment, July 2011)

Article 3

L'Islam est la religion de l'Etat, qui garantit à tous le libre exercice des cultes.

Article 4

L'emblème du Royaume est le drapeau rouge frappé en son centre d'une étoile verte à cinq branches.

La devise du Royaume est DIEU, LA PATRIE, LE ROI.

Article 5

L'arabe demeure la langue officielle de l'Etat.

L'Etat œuvre à la protection et au développement de la langue arabe, ainsi qu'à la promotion de son utilisation.

De même, l'amazighe constitue une langue officielle de l'Etat, en tant que patrimoine commun de tous les Marocains sans exception.

Une loi organique définit le processus de mise en œuvre du caractère officiel de cette langue, ainsi que les modalités de son intégration dans l'enseignement et dans les domaines prioritaires

IRCAM



French | Français | العربية | امازيغیت



- ▶ **Tunisia**
 - ❖ New constitution 2014: recognizing minorities without nominating them
 - ❖ NGOs acting for a constitutional amendment

[!\[\]\(89d1e09f668245d223896beda39443bd_img.jpg\)](#) [!\[\]\(5fff40472ad10d456ceb3775edf98340_img.jpg\)](#) [L'ESSENTIEL](#) [24H](#) [L'HEBDO](#) [LES +](#) [!\[\]\(207d79912909f888f5f071263296f6ac_img.jpg\) Di:TV](#) [DOSSIERS](#) [EMPLOI](#) [SPORT](#) [PAUSE CAFÉ](#)

Tunisie :

Une association dénonce l'exclusion des minorités dans la constitution

Par : Di avec TAP | 20 jan 2014 | 15 h 38 min | Mots-clés : citoyens, constitution, musulman, Politique, Presse, Société, Syndicat National Des Journalistes Tunisiens, Tunisie,

« La Constitution est la seule garantie qui protège les droits de tous les citoyens Tunisiens quelque soit leur appartenance ou diversité religieuse, ethnique ou culturelle » a indiqué, lundi, la présidente de l'Association Tunisienne de soutien aux minorités, Yamina Thabet.

How to subscribe to this challenge ?

- ▶ Need for a strategy
- ▶ Need for a working frame
- ▶ Need for competencies
- ▶ Need for data resources
- ▶ Need for ... incentives !

Incentives

▶ DH general context in France

❖ THATCamp Paris 2010 : HD Manifest

▶ ...

▶ III. Declaration

- ❖ 5. We, professionals of the digital humanities, are building a community of practice that is solidary, open, welcoming and freely accessible.
- ❖ 6. We are a community without borders. We are a **multilingual and multidisciplinary** community.
- ❖ 7. Our objectives are the advancement of knowledge, the improvement of research quality in our disciplines, the enrichment of knowledge and of collective patrimony, in the academic sphere and beyond it.
- ❖ 8. We call for the integration of digital culture in the definition of the general culture of the twenty-first century.



<http://tcp.hypotheses.org/411>

THATCAMP
PARIS

USER GENERATED UNCONFERENCE ON DIGITAL HUMANITIES
NON-CONFÉRENCE SUR LES DIGITAL HUMANITIES

Incentives

► France: 2014

Bac "Humanités numériques", enseignement de l'informatique... les 40 propositions du CNNum sur l'éducation

par Sylvain Arnulf - Publié le 03 octobre 2014, à 16h53

[Numérique](#), [Document](#), [Conseil national du numérique \(CNNum\)](#), [L'Usine Campus](#), [Formation](#)



► Bac "Humanités numériques", enseignement de l'informatique... les 40 propositions du CNNum sur l'éducation
► Védis Ronald - Filoir CC

Le CNNum a publié un rapport sur l'éducation et le numérique baptisé "Jules Ferry 3.0". Il contient 40 recommandations réparties en huit axes pour "bâtir une école créative et plus juste dans un monde numérique".

Comme beaucoup d'autres secteurs de la société, l'école entretient un rapport ambivalent avec le numérique, vu à la fois comme une chance de réinventer les enseignements, et comme une menace. Des start-up et initiatives privées (MOOC, nouvelles écoles du numérique comme Simplon et 42) disruptent le secteur. Le Conseil national du numérique a souhaité se saisir de cette question.

Le code informatique sera enseigné à l'école primaire dès la rentrée

par Elodie Valerey - Publié le 15 juillet 2014, à 11h56

[Vu sur le web](#), [Informatique](#), [Formation](#)



Dans une interview au Journal du Dimanche, le 13 juillet, le ministre de l'Education nationale Benoît Hamon estime que l'école ne peut plus "ignorer l'importance du numérique". Des élèves de primaire seront initiés au code informatique dès la rentrée de septembre, tandis que la discipline sera ajoutée aux programmes dans le secondaire, a-t-il annoncé.

Le ministre de l'Éducation nationale veut initier les petits Français au code informatique.

Incentives

► Research projects experience

❖ BNFB (Bibliothèque Numérique Franco-Berbère) : OIF/FFI Project - 2011 (100 000 €)

- Enrich and update, with automated procedures, records repositories that describe and make available documents in Berber languages (Timbuktu manuscripts)

❖ IDEFI-CréaTIC: Unesco Chair-ITEN (2012) : Innovative Training (5,2 million €)

❖ ISO standards:

- E-Learning vocabulary standards (SC36/WG1)
- E-Learning MLR (Metadata for Learning standard)
- E-learning MLO (Metadata for learning opportunities) (Afnor/GT1 mirror)



+01100+ +1100+ +010100+ +01001+



Incentives

► Research background

- ❖ Information and Communication Studies
- ❖ Multilingual data encoding standards
- ❖ Digital data processing



► Research affiliations

- ❖ Universities (Algeria, France, Morocco, Tunisia)
- ❖ Research Laboratories
- ❖ Professional associations
- ❖ International Organizations



Key incentive

► Support from international TEI referees

- ❖ Laurent ROMARY : director of Dariah-EU & member of the TEI council
- ❖ Lu BURNARD: Oxford Univ., godfather of TEI Guidelines & consultant TGIR Huma-Num, France

 **DARIAH-EU** | Digital Research Infrastructure for the Arts and Humanities



Laurent ROMARY



 **Huma-Num**
la TGIR des humanités numériques



Lu BURNARD



[CONTEXT]

The roadmap

- ▶ Building partnerships
- ▶ Fund raising
- ▶ Building work environment
- ▶ Work organization
- ▶ Implementation, production and evaluation



ISCC
Institut des sciences
de la communication
CNRS/PARIS SORBONNE/UPMC



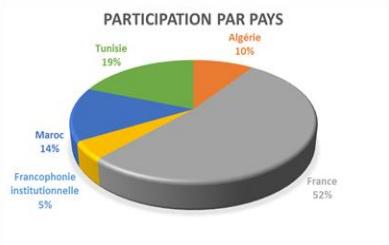
Partnerships

► Individuals and institutions

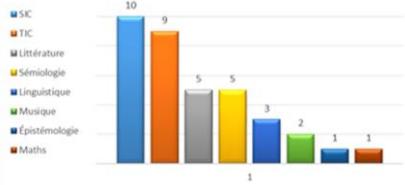


ΛΕΗΞ
Bibliothèque Numérique
Français-Berbère

| Institution | Origine |
|---|---------------|
| 1. Agence Universitaire de la Francophonie | Francophone |
| 2. Chaire expérimentale internationale « Innovation, Transmission, Édition Numériques » (ITEN) à la Fondation Maison des Sciences de l'Homme (HSH) | France |
| 3. TGE-Adonis | France |
| 4. Digital Visuel | France |
| 5. Ecole des Hautes Études en Sciences Sociales, Laboratoire d'Anthropologie Sociale (EHESS), 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 | Francophone |
| 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 » (ALEA), Faculté de lettres et langues, Université Mouloud Mammeri de Tizi Ouzou | 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'Évry, Labo Maths | France |
| 19. LEDEN ¹ , Maison des Sciences de l'Homme Paris Nord | France |



Spécialités des membres du projet



Fund raising

► OIF/FFI (2010-2013)

- ❖ 100 000 €
- ❖ BNFB project



► ISCC/CNRS (2012-2013)

- ❖ 10 000 €
- ❖ HumanitéDigitMaghreb



► LEDEN, MSH Paris-Nord (2014-2015)

- ❖ 3000 €
- ❖ TEI/MEI Euromed



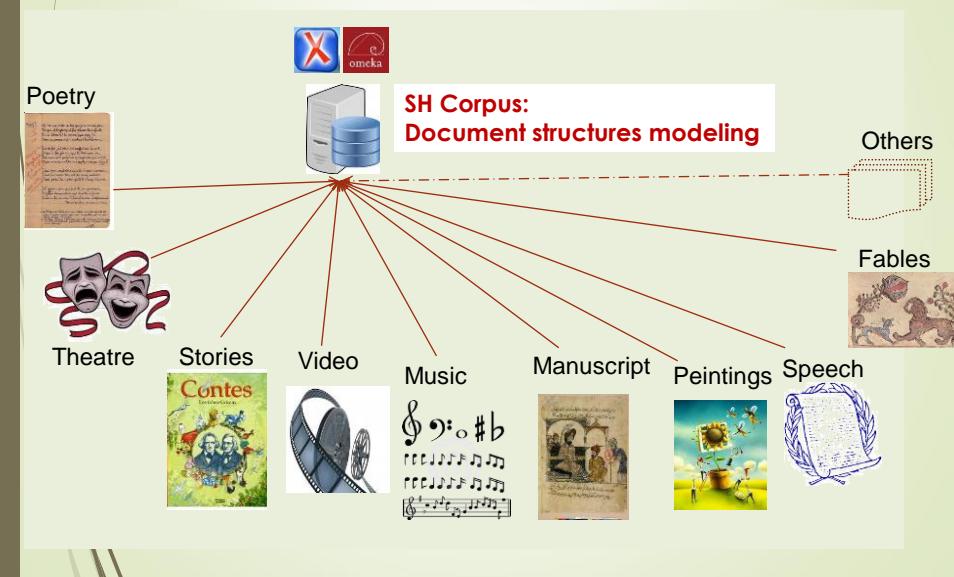
Maison des Sciences de l'Homme
Paris Nord

► Chaire Unesco-ITEN (provisional)

- ❖ HD-MUREN



Work organization (structuring)



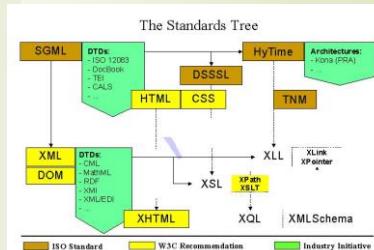
Work environment

► Preferring

- ❖ A data description system ?
 - Document centered system
 - Document structuring

► to

- ❖ A Data base management system
 - Flat scanning
 - Referencing metadata



Work environment

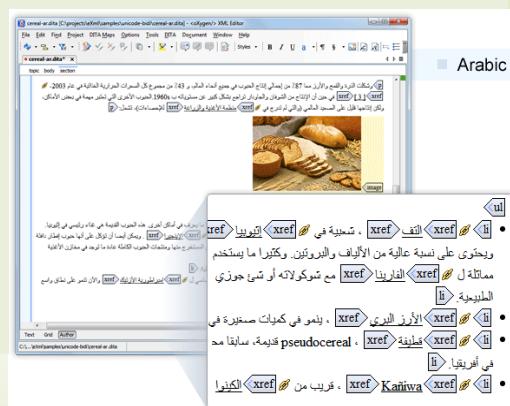
► oXygen



Work environment

► oXygen

- ❖ oXygen XML Editor supports a large number of text encodings.
- ❖ Bidirectional Text Support in Text Editing Mode (Unicode BIDI Algorithm)
- ❖ Editing right-to-left XML documents
- ❖ Ability to create and edit XML content in Arabic



Work environment

 **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 Aucun fichier sélectionné
- Use or modify an existing TEI-defined customization
- Upload a customization Aucun fichier sélectionné

Community-maintained customizations can be downloaded

[Start](#)

Work environment

 **OxGarage Conversion**

Select the format into which you want to convert your document

Convert from: [?](#)

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 (.sxw)
- Plain Text (.txt)
- Rich Text Format (.rtf)
- TEI P4 XML Document
- TEI P5 XML Document
- TEI Lite XML Document
- WordPerfect Document (.wpd)
- xHTML Document

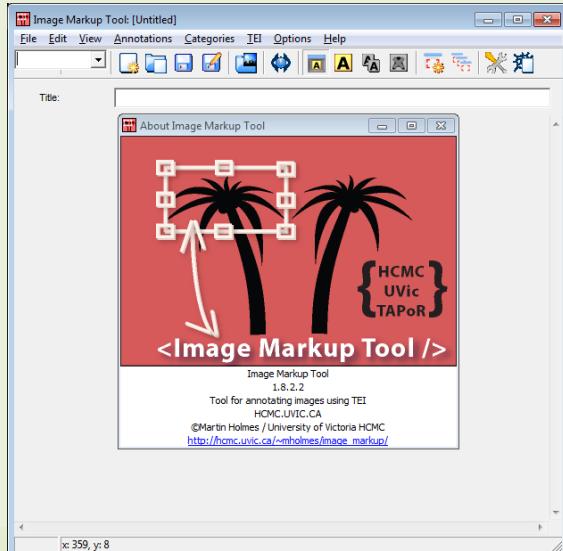
Convert to: [?](#)

- 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 (.sxw)
- 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/>

Work environment

► Image Markup Tool



Work environment

► Xaira (XML search engine)

- ❖ Explore a corpus language:
 - Search words, forms, phrases or XML tags
 - Analyze the way how they are used (frequency, stats, forms)

A screenshot of the Xaira XML search engine interface. On the left, there is a query editor window titled "XAIRA - easy.sqy" showing an XML query: <phrase>easy</phrase>. The main workspace contains a text area with code related to breaking down code into smaller components. On the right, there is a "Word Query" window with a search field containing "the". Below the search field is a table showing word frequency and forms:

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

25 words

Show Controls Forms

Save Query

Work environment

► Omeka CMS platform



A screenshot of the Omeka Admin Dashboard titled "Little Colored Bits of Paper". The dashboard has a dark header with tabs for "Dashboard", "Archive", "Exhibits", "Users", "Names", and "Settings". The main content area is divided into sections: "Getting Started with Omeka" (Archive, Exhibits, Users, Settings), "Site Overview" (listing 6 items, 13 keywords, and 2 users), and "Recent Items" (a list of items with dates). A "View All Items" button is at the bottom right of the recent items section.

Work plan (provisional)

- **Sharing tasks and responsibilities**
 - ❖ Describing roles
 - ❖ Training sessions
- **Resource analysis**
 - ❖ Structuring (segmentation)
 - ❖ TEI/XML Encoding
- **Collections organization**
 - ❖ Referencing (TeiHeader)
 - ❖ DC indexing (metadata harvesting)
- **Exploitation**
 - ❖ Content research
 - ❖ Communication

Work organization (sharing)

SH Specialists

humities www.fotosearch.com

Arts & SH Corpus

XML/TEI technicians

Info-Com. specialists

Code Snippet (XML/TEI)

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE document SYSTEM "http://www.w3.org/1999/xhtml">
<document>
  <!-- Header -->
  <!-- Body -->
</document>
    
```

Work organization (structuring)

► **SHS SPECIALISTS** identify logical structure of documents according to the genres (Novel, poem, theatre, story, painting, music, ...)

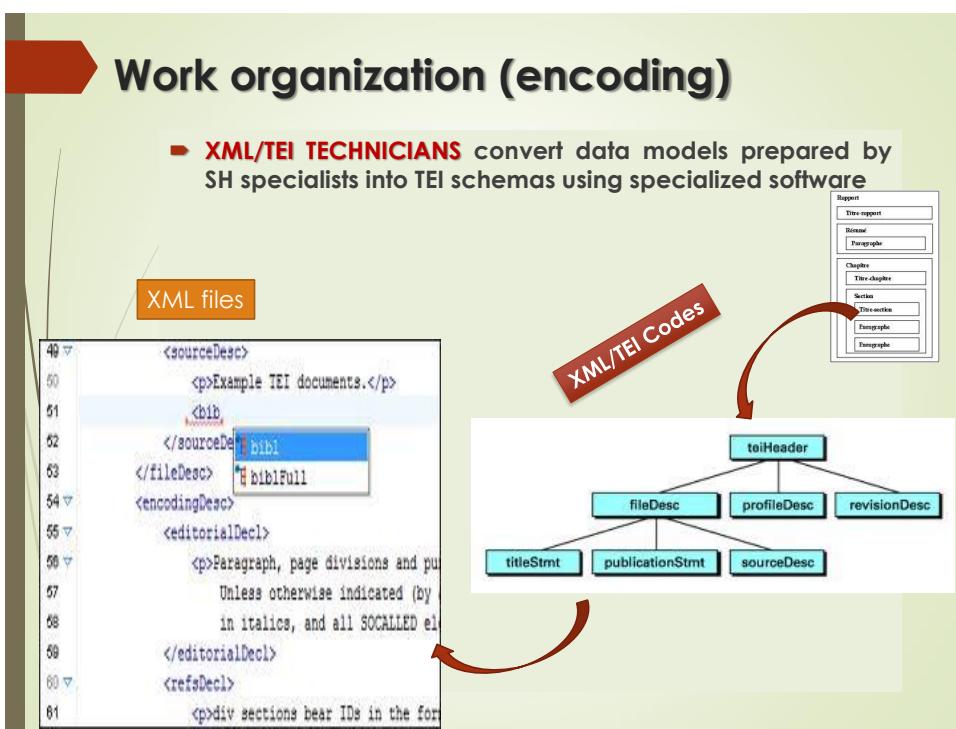
Literary Genres

- Fiction**: Historical Fiction - Science Fiction - Fantasy - Western - Fiction - Mystery
- Nonfiction**: Biography - Autobiography - Informational
- Folktales**
- Drama**
- Poetry**

Legend: Fablie - Fairy Tale - Myth - Folk Tale

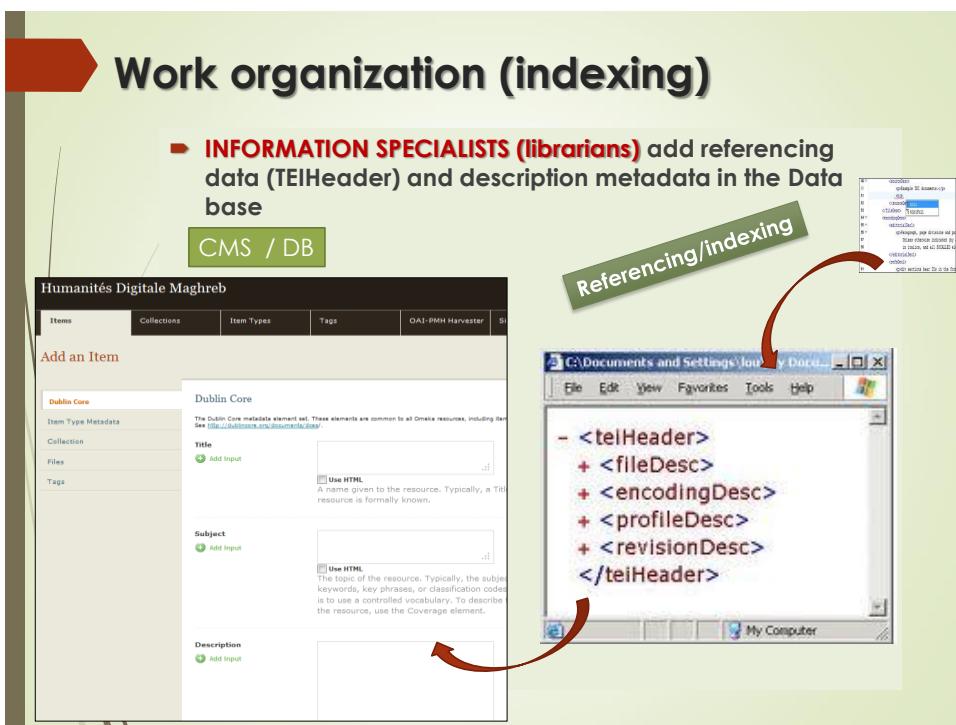
Work organization (encoding)

- XML/TEI TECHNICIANS convert data models prepared by SH specialists into TEI schemas using specialized software



Work organization (indexing)

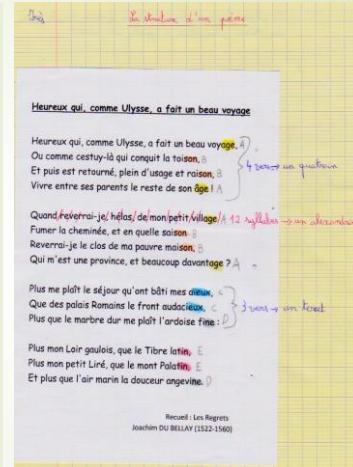
- INFORMATION SPECIALISTS (librarians) add referencing data (TEIHeader) and description metadata in the Data base



Some examples (TEIVerse)

► Identify within a poem

- ❖ Verse metric
- ❖ Different types of verse
- ❖ Different verse clusters (verses, quatrains)
- ❖ Patterns of stanzas
- ❖ Rhythm
- ❖ ...



The granularity may vary between large segments of text and very tiny character strings

Some examples (TEIVerse)

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

Some examples (TEIDrama)

Identify within a piece of theatre:

- ❖ Interpretation and production
- ❖ Speech on stage
- ❖ Stage direction ("didascalies")
- ❖ Types of exposure
- ❖ Dénouement
- ❖ 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.

Some examples (TEIManuscript)

<surface> : a page, a stela, all objects with inscription

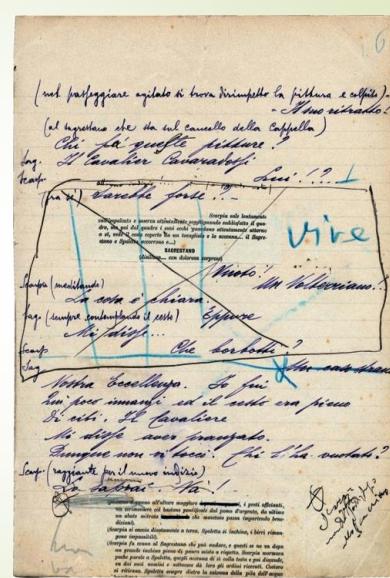
- ❖ Surface contains zones and lines
- ❖ It has coordinates

<zone>: An area of the surface determined arbitrarily for editorial needs. Areas can overlap: overlapping is defined according to spacial coordinates

- ❖ Can contain <line>
- ❖ Disposes of coordinates

<line> : a continuum of text clearly identified by editor

- ❖ Can contain text and <zone>
- ❖ Without coordinates



Some examples (TEIManuscript)

Samaria is a Greek
brand 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.

1 April 2009

```
<surface  
    ulx="0"  
    uly="0"  
    lrx="700"  
    lry="1000">  
    <!!-- ... -->  
</surface>
```

```
<zone  
    ulx="93"  
    uly="681"  
    lrx="967"  
    lry="1568">  
    <graphic url="gb.jpg"/>  
</zone>
```

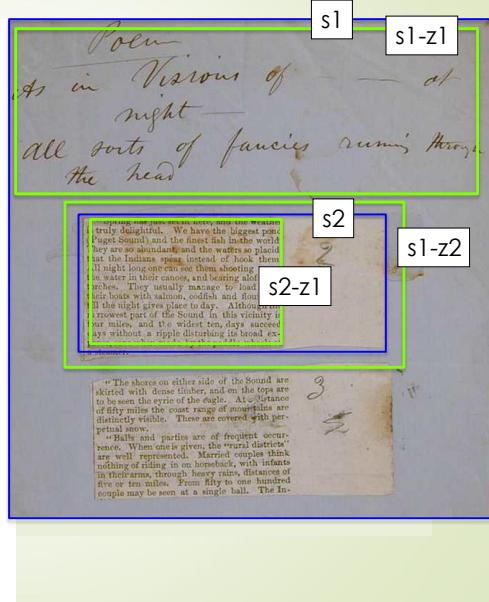
```
<sourceDoc>  
    <surface ulx="0" uly="0" lrx="200" lry="300">  
        <zone ulx="10" uly="43" lrx="185" lry="84"  
            rotate="0">  
            <line rend="right"> 1 April 2009</line>  
        </zone>  
        <line>Fed Birds in the park today.</line>  
        <line>Might write an article about</line>  
        <line>the Thick-billed Warbler.</line>  
    </surface>  
    <zone ulx="9" uly="20" lrx="70" lry="60"  
        rotate="90">  
        <line>Samaria 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>  
</sourceDoc>
```

Some examples (TEIManuscript)

```

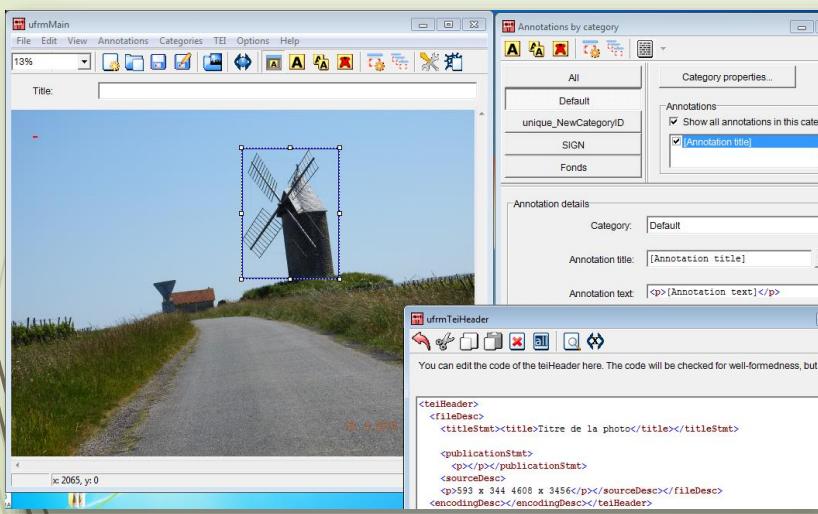
<surface xml:id="s1" ulx="0" uly="0"
lrx="50" lry="50">
<zone xml:id="s1-z1" ulx="1" uly="1"
lrx="10" lry="10">
<line>Poem</line>
<!-- ... -->
<line>the head</line>
</zone>
<zone xml:id="s1-z2" ulx="4" uly="4"
lrx="20" lry="20">
<surface xml:id="s2" uly="0"
uly="0" lrx="100" lry="100">
<zone xml:id="s2-z1"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

```



Some examples (Graph)

Image Markup Tool



Some examples (Graph)

```

<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>The Image Markup Logo</title>
    </titleStmt>
    <publicationStmt>
      <p></p>
    </publicationStmt>
    <sourceDesc>
      <p>377 x 259</p>
    </sourceDesc>
  </fileDesc>
  <encodingDesc>
  </encodingDesc>
</teiHeader>

```



Some examples (Graph)

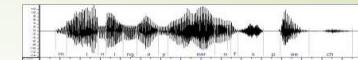
► Image Markup Tool

The screenshot shows the Image Markup Tool interface. On the left, a thumbnail list of images is visible. In the center, a historical document page is displayed with several red annotation zones highlighted. A detailed XML dialog box is open on the right, showing the XML code for the selected zone. The XML code includes terms like 'idx_indians', 'idx_alcohol', and 'idx_proclamation'. The dialog box also shows the corresponding text from the document and the XML attributes for each term.

| Name | xml:id | Attribute | Value |
|----------|----------------|-----------|------------------------------|
| #comment | | #comment | &#amp;#amp;#... .cm =====... |
| = div | | type... | |
| = head | | | |
| = opener | | | |
| = p | para_001 | xml:... | I have the hono... |
| = #text | | | |
| + index | | | |
| + hi | | rend... | |
| + #text | | | day of Septem... |
| + note | B00101 | n=+... | prohibiting the ... |
| + #text | | | |
| + index | idx_indians | xml:... | |
| + index | idx_alcohol | xml:... | |
| + index | idx_proclam... | xml:... | |
| = term | | | Proclamation |
| ... | | | |
| + index | | | |
| #text | | | Indians of Fras... |
| + hi | | rend... | |
| Cancel | | | |
| OK | | | |

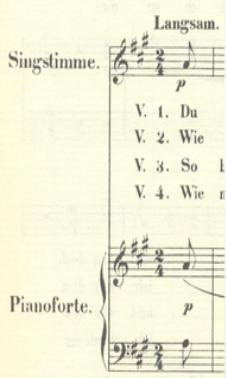
Some encoding examples (TEISpeech)

- ▶ Encoding speech according to a « Timeline »
- ▶ Possibility to describe different speech phenomena of oral communication and sounds during a speech
 - ❖ pauses, words overlaps, tonal changes, vocal expressions (coughs, laughs, sneezing, ...), gestures 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>
```

Some examples (MEI)



```
<music>
  <body>
    <mdiv>
      <score>
        <scoreDef meter.count="2" meter.unit="4" key.sig="3s">
          <staffGrp symbol="line">
            <staffDef n="1" label="Singstimme." lines="5" clef.shape="G" clef.line="2"/>
            <staffDef n="2" lines="5" clef.shape="G" clef.line="2"/>
            <staffDef n="3" lines="5" clef.shape="F" clef.line="2"/>
          </staffGrp>
          <staffGrp symbol="brace" label="Pianoforte.">
            <staffDef n="2" lines="5" clef.shape="G" clef.line="2"/>
            <staffDef n="3" lines="5" clef.shape="F" clef.line="2"/>
          </staffGrp>
        </scoreDef>
      </score>
    </mdiv>
  </body>
</music>
```

A tunisian song encoded in TEI <TeiHeader>

The diagram illustrates the mapping between TEI Header elements and their corresponding Arabic descriptive terms. The vertical axis on the left lists the TEI Header elements, and the horizontal axis shows the Arabic terms. Red arrows indicate the mapping from each element to its term.

| TEI Header Element | Arabic Description |
|-----------------------------|------------------------|
| <teiHeader> | تحت الباسمينة في التبل |
| <fileDesc> | |
| <titleStmt> | |
| <title> | نحوت الباسمينة |
| <author role="chanteur"> | الهادي الجوبني |
| <author role="compositeur"> | الهادي الجوبني |
| <author role="parolier"> | المقدمة عزة |
| <publicationStmt> | |
| <p></p> | |
| <sourceDesc> | |
| <p></p> | |
| </fileDesc> | |
| </teiHeader> | |

A tunisian song encoded in TEI <text>

تحت الباسمية في الليل
 سمة والورد محاذبي
 الأغصان عليا تتميل
 نمسحلي في دمعة عيني

تحت الباسمية اكتب
 عدل السعود وغشت
 وناظر مهني ونكت
 تفكرتك كيف كنت تجني

حسنة مزيناها النوار
 فاخت من ريح الزهار
 تفكرتك شعلت النار
 عملت لهليبة في قلبى

متوجه وحدى محثار
 لا فمها ولا حس أطمار
 كان النسمة ع الأشجار
 توانس فيها وتواسينى

```

<text>
  <body>
    <head> تحت الباسمية </head>
    <lg type="stanza">
      <l><1> تحت الباسمية في الليل</1>
      <l><1> سمة والورد محاذبي</1>
      <l><1> الأغصان عليا تتميل</1>
      <l><1> نمسحلي في دمعة عيني</1>
    </lg>
    <lg type="stanza">
      <l><1> تحت الباسمية اكتب</1>
      <l><1> عدل السعود وغشت</1>
      <l><1> وناظر مهني ونكت</1>
      <l><1> تفكرتك كيف كنت تجني</1>
    </lg>
    <lg type="stanza">
      <l><1> حسنة مزيناها النوار</1>
      <l><1> فاخت من ريح الزهار</1>
      <l><1> تفكرتك شعلت النار</1>
      <l><1> عملت لهليبة في قلبى</1>
    </lg>
    <lg type="stanza">
      <l><1> متوجه وحدى محثار</1>
      <l><1> لا فمها ولا حس أطمار</1>
      <l><1> كان النسمة ع الأشجار</1>
      <l><1> توانس فيها وتواسينى</1>
    </lg>
  </body>

```

► A tunisian song encoded in MEI <music>

A tunisian song encoded in MEI <music>

A Arabic/French fable encoded with Oxygen

A Arabic/French fable encoded with Oxygen

A TEI-based critical apparatus of an Arabic document

Critical diagnosis

► Facing many problems

- ❖ DH & TEI are very new concepts for academic community (resistance)
 - ❖ Vertical clustering of disciplines in the same university
 - ❖ Some cultural and linguistic sensitiveness observed between partners

- Need for more training and interdisciplinary research

- ❖ Nomadic training team
 - ❖ Group publishing

The stylistic notes by Rétif de la Bretonne: a hypertext before digital era
Henri HUDRISIER (1), Sofia BEZINA (2), Hichem ISMAIL (3), Rachid ZGHIBI (4), Sihem ZGHIDI (5), Loula Abd-elrazak (6), Laurent ROMARY (7), Mokhtar BEN HENDA (8), Arnaud LABORDERIE (9), Ghislaine AZEMARD (10)



General orientation

► A strategy of support and orientation

- ❖ French translation of the full version of "TEI P5 Guidelines" (1500 pages counting appendices including about 600 pages are translated to date)
- ❖ An Arabic TEI P5 version is also under attempt



Henri HUDRISIER, PAris8



Rachid ZGHIBI, Tunis

General orientation

► Further research projects

- ❖ 2014-2015: Project & MEI TEI-Euromed
(Encoding with MEI Maghreb
traditional music)



Sylvaine MARTIN,
LEDEN, MSH Paris Nord



The screenshot shows a news article from the ILES D IMESLI website. The article is titled "La MEI (Music Encoding Initiative) : un standard au service de la musique kabyle" and is attributed to "Sylvaine Martin DE GUISE". The website has a blue header with the title "ILES D IMESLI". Below the header, there are navigation links for "INFORMATIONS", "PAGE D'ACCUEIL", "À PROPOS", "SE CONNECTER", "S'INSCRIRE", and "RECHERCHER". The "INFORMATIONS" section contains links for "Pour les lecteurs", "Pour les auteurs", and "Pour les bibliothécaires". The "UTILISATEUR" section includes fields for "Nom", "Prénom", "Mot de passe", and a checkbox for "Mémoriser mon mot de passe". A "Se connecter" button is also present.

General orientation

► Further research projects

- ❖ 2012: The Unesco Chair-ITEN:
"Innovation, Transmission, Edition
Numérique"
- ❖ Bringing academic and
entrepreneurial partners together to
discuss the contemporary changes
education related to the effects of
digital innovation
- ❖ 2012: IDEFI-CréaTIC project
 - ❖ budget of 5.2 M€ on 8 years
 - ❖ 17 training programs
 - ❖ 24 workshops open to more than 500
students



Training

ARTS

- Virtual Imaging Arts and Technologies
- International Artistic Cooperation
- National Drama Academy Graduate Course
- Stage Direction and Dramaturgy
- Audiovisual Screenwriting

CULTURE & COMMUNICATION

- Analysis and Promotion of Digital Uses
- Digital Creation and Publishing
- Audiovisual Designer
- Information and Document Management
- Digital Issues and Technologies

LANGUAGE

...
OK



INITIATIVE D'EXCELLENCE
EN FORMATIONS INNOVANTES



INVESTISSEURS
D'AVENIR

DIGITAL WORK ENVIRONMENT

INTERNATIONAL STUDENT MOBILITY

COMMUNICATION

- Press Releases
- Contact
- Media Ressources
- Press Review

NEWSLETTER

Nom & Prénom:

Votre nom et prénom



HD-Muren

- ▶ Implement and test an environment where different stakeholders in the academic communities learn together about new computing technologies and their impact on the humanities
- ▶ DH & TEI at school?



PROJET Iten-Unesco

PROJET HD-MUREN

Des Humanités digitales multiculturelles en mode crowdsourcing via la recherche et l'enseignement

Proposition de projet coordonné par

Mokhtar Ben Henda et Henri Hudrisier

pour la chaire Paris 8, ITEN-Unesco

dans le cadre de IDEFI CratCic

General orientation

- More positioning within international communities of practice in the DH

- ❖ Humanistica (2014 membership)
- ❖ ThatCamps: St Malo, 2013 / Lyon 2014 / Beirut, 2015 / Paris 2015)



Thank you for your attention