



EMERITES.LORRAINE

Association des Professeurs et Chercheurs Emérites de Lorraine



Enjeux de la maîtrise numérique de la connaissance

Introduction aux ateliers des émérites !

<https://wicri-demo.istex.fr/wicri-chanson-roland.fr/index.php/PaulMeyer2021>

Bonjour, ici Jacques Ducloy, 55 ans dans les sciences numériques pour l'ingénieur

- ▶ **1967** ENSEM, → Fac de Sciences
 - Analyse numérique (Jean Legras), compilation (Jean-Claude Derniame),
- ▶ **1971** IUCAL (Jean Legras puis Claude Pair)
 - CRPG, Cristallographie, Médecine... et notamment **TLF**, sans oublier Cyclades
- ▶ **1980** CNRS/SPI + Agence de l'Informatique (CNET, INRIA...)
 - Promotion des recherches en Génie logiciel, Intelligence Artificielle
 - Chaîne éditoriale et base documentaire sous Transpac
- ▶ **1988 ... 2007**, équipe de direction de l'INIST
- ▶ **1992 ... 2000** : Loria (Autoroutes de l'Information),
 - Digital libraries : INRIA → DELOS (Europe) → DCMI (Dublin Core)
- ▶ **2007** : DRRT Lorraine
 - Lancement de Wicri (DRRT Lorraine ... ISTEEX)

Le numérique pour la connaissance

- ▶ Quelles connaissances ?
 - Scientifiques, technologiques, culturelles, patrimoniales
- ▶ Quels enjeux ?
 - Stratégiques : compétitivité internationale // Santé
 - Culturels et patrimoniaux,
 - Personnels : comprendre le monde du numérique
- ▶ Quels moyens pour les émérites ?
 - Deux wikis (public, privé)
 - Pour s'entraîner, et peut-être pour développer
 - Le réseau Wicri
 - Accompagnement (ateliers, analyses de corpus)

Un wikis pour les émérites

- ▶ Même moteur que Wikipédia (MediaWiki)
- ▶ Contributions nominatives **réservées aux inscrits**
- ▶ Sur un plan technique 2 wikis :
 - Privé (documents internes ou bac à sable)
 - Public : initialisé avec les publications des émérites confinés
- ▶ Première visite :
 - Nouveau site Wicri : <https://wicri-demo.istex.fr/>
 - Wikis émérites : <https://wicri-demo.istex.fr/emerites.lorraine/fr>
 - Les [émérites confinés \(2021\)](#) // [un article](#)
 - L'onglet [Communautés](#)



Accès contributeur au réseau Wicri

- ▶ Réflexion sur une alternative à Wikipédia.
 - Des wikis pour des communautés culturelles ou scientifiques :
 - Thématiques,
 - Sciences et génie de l'environnement (eau, bois, sols urbains...)
 - Santé, musique
 - Géographiques
 - Lorraine, Maroc, Canada
 - Institutions, communautés
 - Colloques CIDE, H2PTM
 - Sujets,
 - Chanson de Roland
- ▶ Espace pour expérimentations



Ateliers



- ▶ Formule : 9h30 → 12 h, par exemple à la B.U.L.
 - 9h30 : petits rendez-vous
 - 10 h : conférence-démonstration // cours
 - 11 h : travaux personnels avec assistance (TP ou ???)
- ▶ Travaux
 - Découverte : Wikipédia, Wicri, (ISTEX, PubMed, Gallica...)
 - Formation : écriture hypertexte, analyse de corpus, xml
 - Développements :
 - Collections émérites, et pourquoi pas ALS ?
 - Groupe mémoire
 - Nouvelles applications
 - Prestations : analyse de corpus ISTEX ou PubMed...

Plan

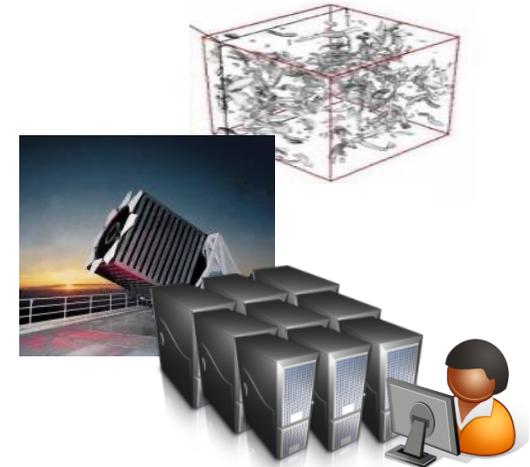
- ▶ Introduction aux ateliers (fait)
- ▶ Changements de paradigme dans la science
 - et dans la connaissance scientifique
- ▶ Enjeux au plan international, national... et lorrain
 - Enjeux culturels, patrimoniaux et citoyens
 - encyclopédies, dictionnaires, patrimoine numérique
 - Enjeux stratégiques (connaître l'essentiel de la recherche)
 - Intelligence économique par les corpus bibliographiques (INIST, PubMed)
- ▶ Aperçu de séances découvertes
 - Analyse des corpus à partir des wikis
 - Ecrire la musique
 - Humanités numériques avec la Chanson de Roland

2000 : vision E-Science : quatrième paradigme sciences physiques et expérimentales

- ▶ Thousand years ago – **Experimental Science**
 - Description of natural phenomena
 - ▶ Last few hundred years – **Theoretical Science**
 - Newton's Laws, Maxwell's Equations...
 - ▶ Last few decades – **Computational Science**
 - Simulation of complex phenomena
 - ▶ Today – **eScience or Data-centric Science**
 - Unify theory, experiment, and simulation
 - Using data exploration and data mining
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
 - Scientists over-whelmed with data
 - Computer Science and IT companies have technologies that will help
- (With thanks to Jim Gray)



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



... Merci à Tony Hey (Microsoft Research)

Changements de paradigmes pour les documents

PAGE



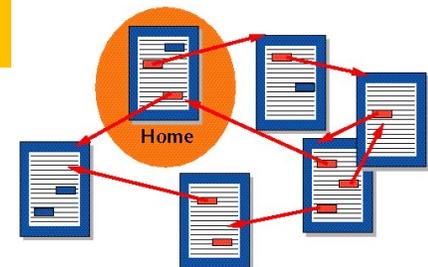
- ▶ 7000 av. J.-C. : jetons
- ▶ 3000 av. J.-C. : tablettes, papyrus, papyrus en rouleau

CODEX



- ▶ 300 av. J.C. : Codex manuscrit
- ▶ 1455 : Livre imprimé ; 1750 : Encyclopédie
- ▶ 1956 : Fieldata ; 1973 (Xerox Alto) ; 1984 : MacWrite ;
- ▶ 1992 : PDF

HYPERTEXTE



- ▶ 1963 : Hypertext ; 1987 : Hypercard ;
- ▶ 1989 : Html (le Web)
- ▶ 2001 : Wikis, Wikipédia (collectif et adaptatif)

Le document structuré

- ▶ Dans les années 80, on applique aux documents les mécanismes de structuration des programmes.

Application au Web (html)

Acte II, Scène 2

DON RODRIGUE À moi, Comte, deux mots.

LE COMTE Parle.

DON RODRIGUE Ôte-moi d'un doute.

Connais-tu bien Don Diègue ?

LE COMTE Oui.

DON RODRIGUE Parlons bas, écoute.

Sais-tu que ce vieillard fut la même vertu,

La vaillance et l'honneur de son temps ? Le sais-tu ?

```
<h1>Acte II, Scène 2</h1>  
<br/> <b>DON RODRIGUE</b> À moi Comte, deux mots.  
<br/> <b>LE COMTE</b>&nbsp;&nbsp;&nbsp;... &nbsp;&nbsp;&nbsp;Parle
```

La Text Encoding Initiative (TEI)

Corpus pour l'ordinateur, lisible par l'homme

```
<div type="Act" n="I"><head>Acte II</head>
  <div type="Scene" n="1"><head>Scène 2</head>
    <sp><speaker>Rodrigue</speaker>
      <l part="i">À moi, comte, deux mots.</l></sp>
    <sp><speaker>Comte</speaker><l part="m">Parle</l></sp>
    <sp><speaker>Rodrigue</speaker>
      <l part="f">Ôte-moi d'un doute</l></sp>
    <sp><speaker>Comte</speaker>
      <l part="i">Connais-tu bien Don Diègue ?</l></sp>
    <sp><speaker>Comte</speaker><l part="m">Oui</l></sp>
    <sp><speaker>Rodrigue</speaker>
      <l part="f">Parlons bas, écoute.</l>
      <l>Sais-tu que ce vieillard fut la même vertu,</l>
      <l>La vaillance et l'honneur de son temps ? Le sais-tu ?</l></sp>
    ...
  </div>
  ...
</div>
```

Acte II, Scène 2

DON RODRIGUE À moi, Comte, deux mots.

LE COMTE Parle.

DON RODRIGUE Ôte-moi d'un doute.

Connais-tu bien Don Diègue ?

LE COMTE Oui.

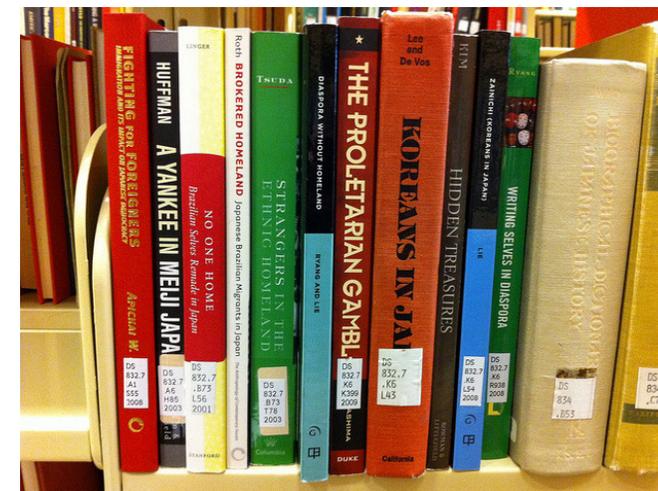
DON RODRIGUE Parlons bas, écoute.

Sais-tu que ce vieillard fut la même vertu,
La vaillance et l'honneur de son temps ? Le sais-tu ?

Dans une bibliothèque numérique, on peut trouver tous les passages où parle Rodrigue...

On peut aussi représenter des [signalements bibliographiques](#)

Changements de paradigmes dans les bibliothèques



Une bibliothèque => Une URL
vers des hypertextes

Un wiki = une encyclopédie
Ou un wiki = une bibliothèque

Wicri = des encyclopédies
+ des bibliothèques
Avec des documents hypertextes

{BnF Gallica

ISTEX

L'excellence documentaire pour tous



Free Choral Music



Bibliothèques numériques

- ▶ 1970 : Progiciels spécialisés sur les métadonnées
 - Exemple : système de gestion de bibliothèques, moteur Mistral sur IRIS 80
- ▶ 1980 : SGBD relationnels
 - Très bien adapté aux applications transactionnelles fermées
 - Mais pas forcément aux connaissances potentiellement infinies
- ▶ 1990 : Entrepôt et archives ouvertes
 - 1989 : Web 1.0,
 - 1991 : Los Alamos National Laboratory → ArXiv
 - 2000 : Hal...
- ▶ 1995 : Bibliothèques (entrepôt) en Xml :
 - Erudit, PubMed Central,
 - Content Management Systems (CMS) Open Edition..
- ▶ 2000 : Web 2.0, Rédaction et **conception collective** (wikis)
 - Changement de paradigme bibliothéconomique, éditorial, rédactionnel

Wiki, modèle informatique

- ▶ Logiciel facile à utiliser par des amateurs
 - Ou complexe : des millions de pages
 - par des centaines de milliers d'amateurs
- ▶ Un wiki = un ensemble de pages
 - Chaque page a un nom unique géré par le contributeur
 - Une page contient du texte et des liens
 - Une page peut contenir une définition,
 - ou un livre...
 - Espaces de nommages
 - Catégories, modèles...
 - Sous-pages

[Charlemagne](#)

[Paul Meyer \(philologue\)](#)

[Utilisateur:Jacques Ducloy](#)

[Romania \(revue\)](#)

[Romania \(1933\) Ewert](#)

[Catégorie: Philologue français](#)

[Chanson de Roland/Manuscrit d'Oxford/Feuillet 20](#)

Wiki, gestion des contributions

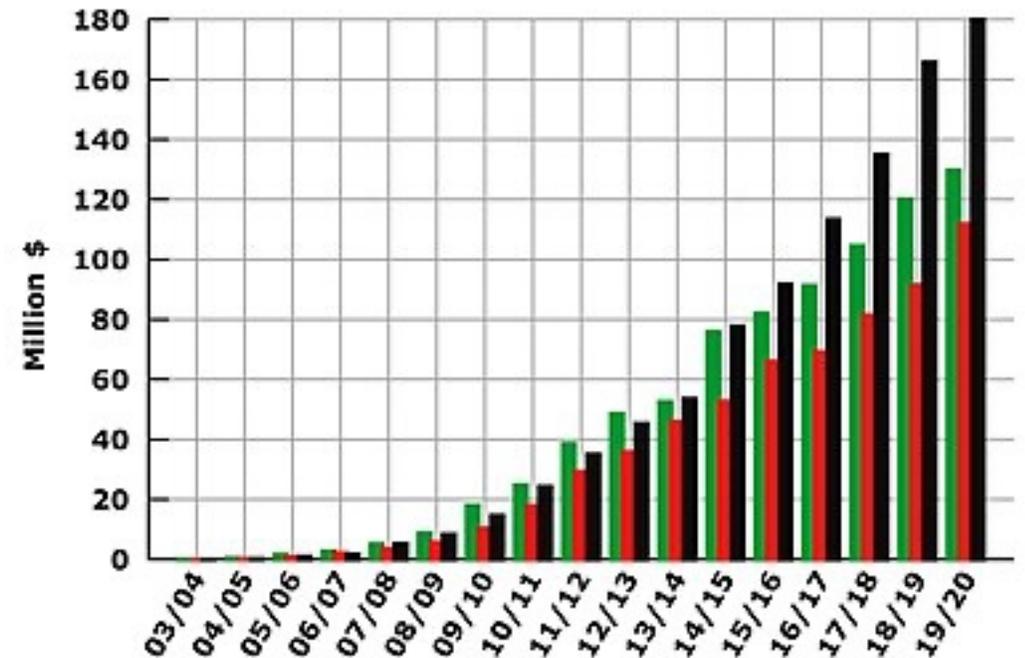
- ▶ Chaque wiki définit ses propres règles
 - Wikipédia : tout le monde peut contribuer (anonymat)
 - Wicri (public) : Tout contributeur est inscrit et visible
 - Vioramil, Préfet : accès limité aux inscrits
- ▶ Outils de suivi
 - [Trace des contributions](#) / liste de suivi
- ▶ Espaces de nommage spécialisés
 - Pages utilisateur
 - Pages de discussion

Wikipédia

- ▶ Géré par la Wikimedia Foundation
 - Siège : San Francisco,
 - effectif 409 : personnes
 - *Non profit organization*
 - C. A. 2020 : 124.468.145 \$
- ▶ Contributions libres
 - Inscriptions recommandées
 - Généralement amateurs motivés
 - avec pseudonymes
 - Outils efficaces pour
 - les vandalismes de base
 - Mais stratégies possibles
 - pour groupes de pression

Wikimedia Foundation Support and Revenue,
Expenses and Net Assets at Year End

(Net Assets do not include funds in Wikimedia Endowment and Tides Advocacy.
Expenses 2015–2020 include \$25 million paid to Wikimedia Endowment.)



MediaWiki, le moteur de Wikipédia

- ▶ Logiciel libre (PHP/MySQL)
 - Assez simple à installer pour faire un essai,
 - En réalité :
 - support et assistance professionnelle, (exemple INIST / UL)
 - léger mais de haut niveau
- ▶ Point fort, programmation par contributions
 - Une page peut contenir le code d'un modèle (ex. infobox)
 - Un amateur initié [peut appeler un modèle](#)
 - Un amateur bien formé [peut développer un modèle](#)
 - Un groupe amateur avec un expert est totalement autonome
- ▶ Extensions multiples
 - Musique, Sémantique, mathématiques...

| Paul Meyer | |
|---|--|
|  | |
| Paul Meyer en 1898. | |
| Fonctions | |
| Directeur École des chartes | |
| 1882-1916 | |
| ◀ Jules Quicherat | Maurice Prou ▶ |
| Président | |
| Société de l'École des chartes | |
| 1880-1881 | |
| ◀ Siméon Luce | |
| Biographie | |
| Naissance | 17 janvier 1840 ✎ Paris ✎ |

Galaxie Wikimedia Foundation



Ouvrages



FR: 2.388.429
11.728.290
EN : 6.436.960
54.946.563



Wiktionnaire
Le dictionnaire libre

FR : 4 298 952
EN : 6,912,797

Entrepôts



80.001.464
images
multimédia

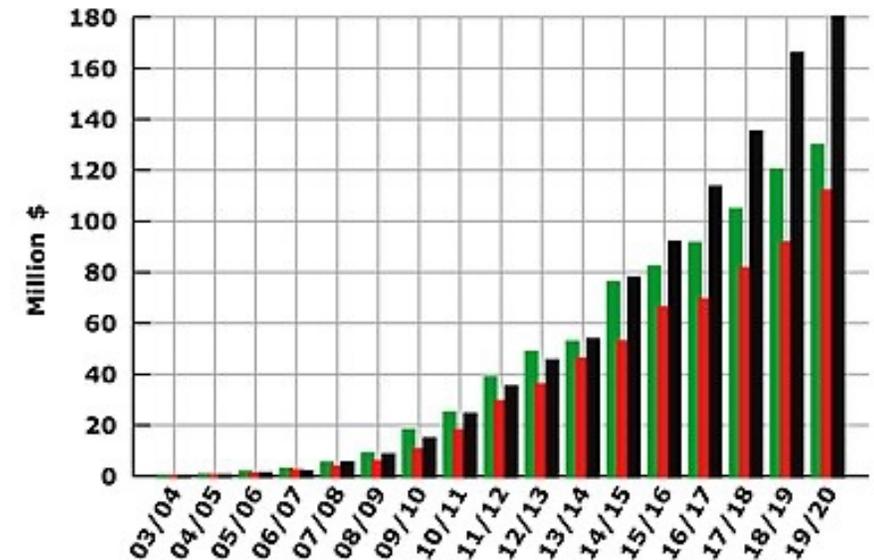
Bibliothèques



333.000
Livres en français

Wikimedia Foundation Support and Revenue, Expenses and Net Assets at Year End

(Net Assets do not include funds in Wikimedia Endowment and Tides Advocacy. Expenses 2015–2020 include \$25 million paid to Wikimedia Endowment.)



Sites MediaWiki -> réseau Wicri

► Dizaines de milliers



Free Choral Music



PROTEOPEDIA
— LIFE IN 3D —

► Réseau Wicri



Observatoire
Encyclopédique
Pilotage type CPER



Publications
Suite de
colloques



Encyclopédie + ...
+ bibliothèque



Encyclopédie
+ serveurs



Hypertexte +++

Liens sémantiques (Semantic MediaWiki)



sur un livret de
[[A pour auteur de livret::Lorenzo da Ponte]].

Semantic MediaWiki :
construction sémantique
Dans une approche RDF

A pour auteur de livret



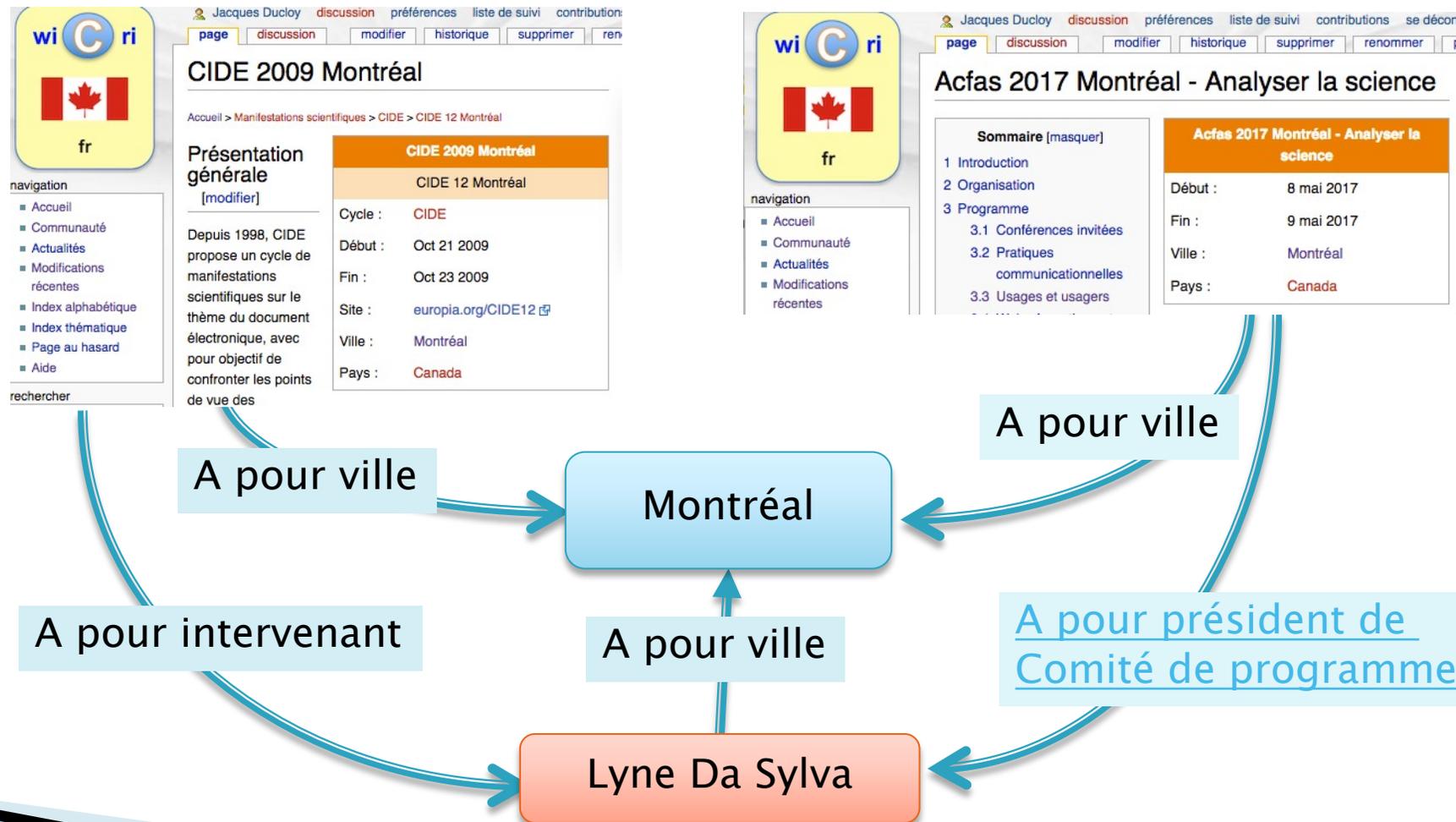
Liste des opéras ayant un livret de Lorenzo da Ponte

Sur ce wiki (par génération automatique)

- Axur, re d'Ormus (Antonio Salieri)
- Così fan tutte (Wolfgang Amadeus Mozart)

```
{{#ask: [[a pour auteur de livret::{{PAGENAME}}]]  
| format=ul | ?A pour compositeur=compositeur :
```

Wikis sémantiques et systèmes de recherche



Wicri réflexion alternative et orthogonale

- ▶ Galaxie Wikipédia,
 - Des services classiques
 - encyclopédie, dictionnaire, bibliothèque
 - Gigantisme pluri disciplinaire
 - Mécanismes de régulation souvent obscurs
- ▶ Wicri
 - Un réseau de sites
 - thématiques, géographiques ou institutionnels
 - Où divers services sont implantés
 - Comités scientifiques visibles
 - modération, orientation...



La tour de Babel



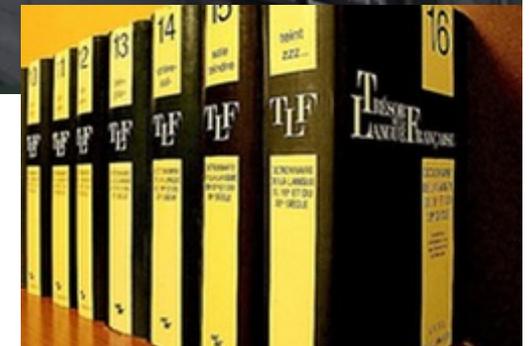
L'école d'Athènes
Socrate, Platon...

En Lorraine, la vision Wicri n'est pas un long fleuve tranquille

- ▶ 2007 : Mission Ticri, portée par la DRRT en lien avec Nancy Université
 - Blocage INPL (incompatibilité avec Dir. Inf. + périmètre institutionnel)
- ▶ 2008 : Expérimentation Loria sur le CPER, bien perçue par PRST SGE
- ▶ 2010 : Projet Wicri SGE, porté par le PRST SGE et NIT
 - Blocage Loria (Divergences technico scientifiques + périmètre institutionnel)
 - Migration site INPL (Nancy Université, NIT) – très bonne croissance
 - Blocage Université de Lorraine
 - profondes divergences scientifiques et politiques
 - leader proche de la retraite
 - périmètre institutionnel ([Euréka de la Région lorraine](#) et HAL)
- ▶ 2014 : Projet LorExplor (un retraité), porté par ISTEEX, soutenu par DIST/CNRS
 - Moyens : un retraité + 0.5 ETP IE
 - Bon bilan technico-scientifique => machine virtuelle à l'INIST
 - Difficultés suite à l'arrêt d'ISTEX et mutation de la DIST CNRS
- ▶ 2020 : Projet scientifique soutenu par Paris 8 / Paragraphe

Trésor de la langue française (dictionnaire)

- ▶ 1960 création de l'ARTLF (Recteur Paul Imbs)
 - Un Trésor (= corpus) numérique pour la recherche en linguistique
 - Valorisable par un pari : un dictionnaire
- ▶ 1963 : Acquisition du Gamma 60
- ▶ 1969 : 600 ouvrages saisis
- ▶ 1971 : Tome 1 du dictionnaire
- ▶ 1972 : Passage au Cii 10070
- ▶ 1982 : ARTLF avec Chicago
- ▶ 1994 : dernier tome

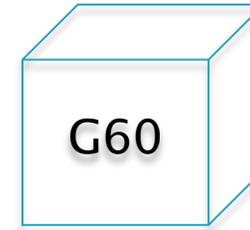


Outils de production, fiches, groupes binaires

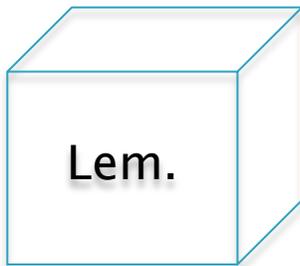
- ▶ Le Gamma 60 produit des concordances
 - parfait pour les mots de faible fréquence.
 - Exemple : *dodécaphonique* (4)
 - Les concordances sont éditées sur « listing »
 - Dans les autres cas : les groupes binaires
 - Co occurrences sur 5 termes sémantiques avant ou après un exemple
 - *Le corbeau fait son nid en haut des arbres.*
 - Le système propose (sur un listing)
 - un ensemble de mots associés avec des listes d'exemples
 - Multiples algorithmes de classement sélection
 - en fonction des fréquences des termes

[...] *pour chaque mot*
[...] *pour chaque occurrence*
un contexte de trois lignes,
le mot étudié figurant
obligatoirement
dans la ligne du milieu

Groupes binaires (pour la mémoire)

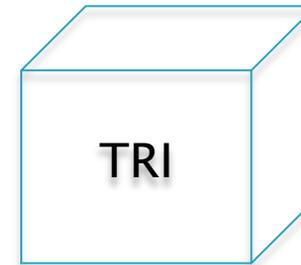


Texte corrigé



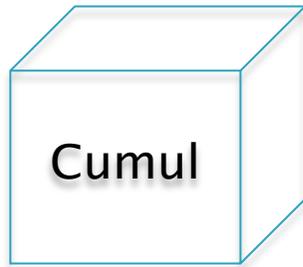
texte

mots



Mots triés

Groupes binaires (suite)

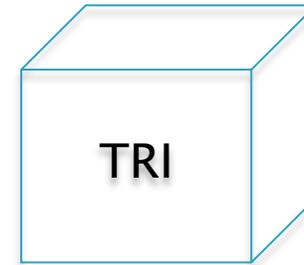


Mots triés

Abaca
Abaca
Abat
....

Couples
Fréquence - mot

2 abaca
5 abat
...



Couples triés

...
5 abat
...
2 abaca

ETC durée 1 mois

Diffusion numérique, puis... arrêt

- ▶ 1972 : transfert sur IRIS 80 // Mais sans les concepteurs Bull
 - Simple transcription
- ▶ 1980 : Jacques Dendien rejoint le TLF
 - Stella : moteur de recherche grammatical dans les textes (Frantext)
 - Conception et maquettes du TLFi
- ▶ 2002 ATILF
 - 2004 : TLFi diffusé en CD-ROM En et sur Internet
 - Ortolang, CNRTL : des services pour les chercheurs

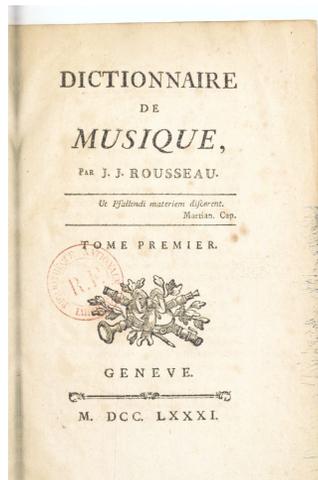
Mais pas de reprise de l'édition du dictionnaire (arrêtée en 1994)

Le TLFi sur Wicri

- ▶ Reprises et extensions d'article
- ▶ Exemples sur Wicri/Santé
 - [Respiratoire](#)
 - [Édentement](#)
 - [Mithridatiser](#)

Réédition numérique hypertexte Dictionnaire de musique (J.-J. Rousseau)

Romilly

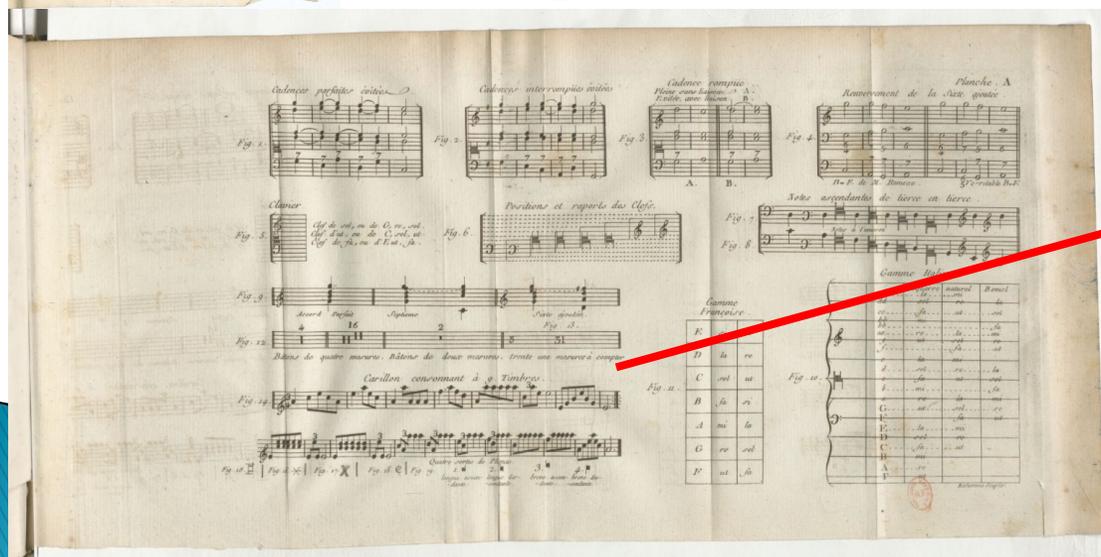


Dans le tome 1

pays des vaines spéculations.
CARRILLON. Sorte d'Air fait pour être exécuté par plusieurs Cloches accordées à différens Tons. Comme on fait plutôt le Carrillon pour les Cloches que les Cloches pour le Carrillon, l'on n'y fait entrer qu'autant de Sons divers qu'il y a de Cloches. Il faut observer de plus, que tous leurs Sons ayant quelque permanence, chacun de ceux qu'on frappe doit faire Harmonie avec celui

Lien

Dans le tome 2



Le lecteur « non solfégiste » peut écouter la musique

Dictionnaires hypertextualisés

Andante

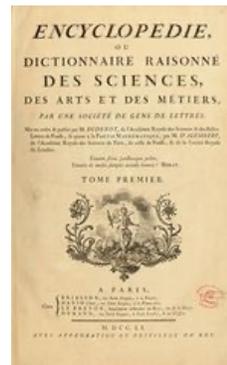
Brossard
1703



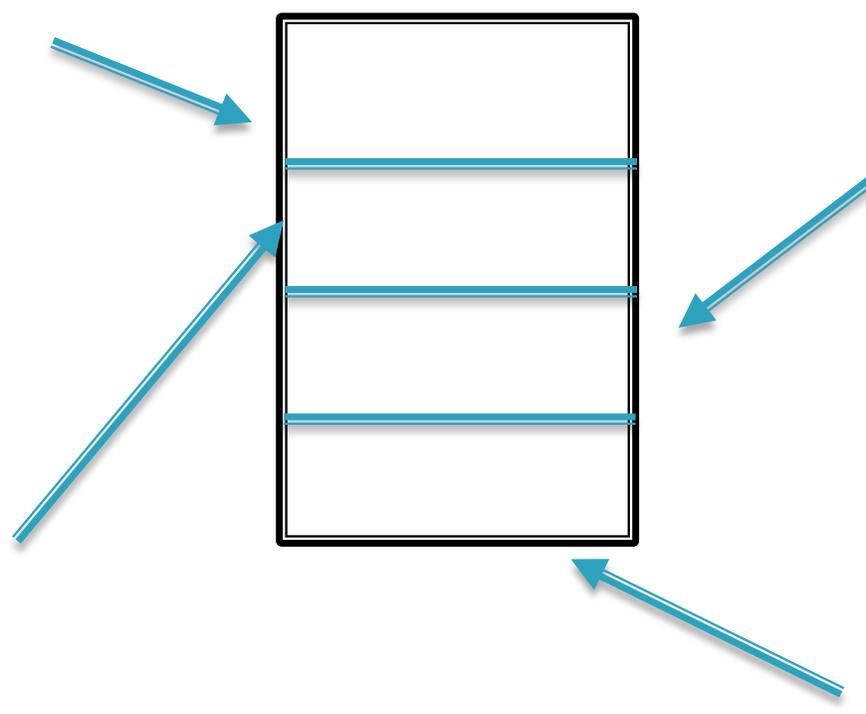
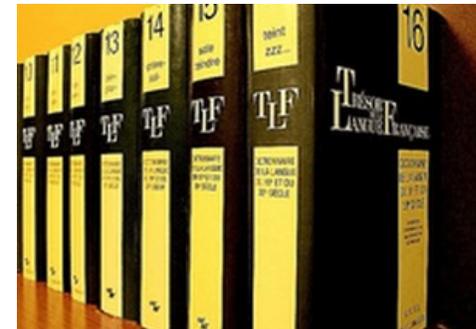
Jean-Jacques Rousseau
1871



Encyclopédie
1851



TLF
1970



Du culturel au stratégique

- ▶ Coût d'un dictionnaire de référence
 - TLF : 80 personnes dans un protocole très lourd
 - Mode wiki : 60 personnes / 50 + réseau
- ▶ Coût d'une encyclopédie
 - 3 permanents par spécialité (300) – mais pas à temps plein !!!
 - Réseau francophone
 - les études de l'existant pour les thèses encadrées par directeur de recherche
- ▶ Stratégie : détection de l'essentiel de la recherche
 - => Information scientifique et technique

Aux USA, les grands acteurs des années 80

- ▶ En santé, la NLM
 - 1836 : *Library of the Surgeon General's Office*
 - 1879 : Index Medicus → MEDLINE
 - 1981 250.000 refs par an (résumé d'auteur / thésaurus puissant MESH)
 - Automatisé sur un ordinateur IBM 370
- ▶ La Library of Congress crée les formats MARC en 1968
 - Culture réseau : chaque bibliothèque possède son catalogue
- ▶ OCLC : *Online Computer Library Center*
 - 1967 : *Ohio College Library Center* (Dublin) avec le format MARC
 - 1971 : catalogue collectif avec l'autre université de Dublin
 - 1978 : changement de statut (ouverture internationale)
 - 2020 : réseau de 15.000 bibliothèques (dont l'UL)
 - 1990 : Centre de recherche fédératif (1 à 3 membres de l'OCLC par projet)
 - **Le catalogue collectif complète les solutions locales sans les remplacer !!!**

Intelligence économique avec le CNRS

- ▶ 1946 : Bulletins signalétiques
 - Les sujets « intéressants » font l'objet d'un résumé en français...

Musique.

Généralités. Problèmes. Musicologie.

19-23-4618. ALBERSHEIM (G.). **Mind and matter in music.** (L'esprit et la matière dans la musique). *J. Aesth. Art It., U. S. A.* (1964), **22**, n° 3, 289-94. — L'espace et le mouvement sont les deux concepts fondamentaux qui permettent de comprendre la structure du monde tonal. Dans la musique tonale, l'auditeur recompose le morceau à mesure qu'il l'entend. Dans la musique moderne, cette coopération est abolie, ce qui révèle

pour QUANTZ « Versuch ». (De nouveaux exemples musicaux pour le « V. » de Q.). *J. amer. musical. Soc.* (1964), **17**, n° 2, 157-69. — Et, détaillée de l'interprétation des appoggiatures telle qu'elle est expliquée dans le *Versuch einer anweisung die Flöte traversiere zu spielen* (1752).

19-23-4630. ZUBILLAGA (L.). **De la música actual en latino-america : Juan Carlos Paz.** (De la musique contemporaine en Amérique latine). *Rev. nocion. Cull., Venez.* (1963), **25**, n° 160, 109-15.

19-23-4631. CHAILLEY (J.). **L'axiome de Strawinsky.** *J. Panthel. Et. Incl. Abbe* 1963, n° 4, 407-20. — La notation de

- ▶ 1971 création des bases Pascal Francis (Nathalie Dusoulier)
 - 400.000 références par an normalisées ISO 2709
 - Accessible sur 2 serveurs via le réseau Cyclades (Iria + CNET + Cii)
 - Avec le logiciel Mistral (navigation par thésaurus) réalisé par la Cii

Du CDST/CDSH à l'INIST



- ▶ 1978 : Nathalie Dusoulier à l'ONU
 - Genève (local), New-York, Genève (réseau ONU)
 - Présidente pour la normalisation CCF (Lc, NLM...)
- ▶ 1980 : situations du CDST et CSSH vécues comme difficile
 - Création de Télésystèmes Questel et fin de la CII
 - La compétence technologique CDST CDSH est externalisée
- ▶ 1989 retour N. Dusoulier au CNRS : INIST avec Goéry Delacôte
 - transfert CDST + CDSH à Nancy
 - Objectif : reprise de la maîtrise technologique
 - Informatisation de la bibliothèque (27 km de rayonnage)
 - fourniture de documents
 - Des photocopies aux documents numériques
 - Ticket modérateur
 - Filiale de commercialisation (modèle British Library)

Le Département Recherche et Produits nouveaux

- ▶ Objectif : Création d'un centre d'expertise/excellence dans la compétence numérique
- ▶ Renforcer les universités de Lorraine dans cet axe
- ▶ Modèle OCLC : 28 ITA pour encadrer 15 projets
- ▶ Objectif Indépendance technologique (bases de données)
 - Bases départementales thématiques (en vue d'indexation assistée)
 - Investissement SGML/XML (5 ans avant la Library Of Congress)
- ▶ Nouveaux produits : infométrie
 - Boite à outil SGML pour analyse de corpus

Une cascade de décisions hasardeuses

- ▶ 1991 départ de Goéry Delacôte dans un contexte difficile
 - Le marché FDP n'est pas tout à fait au rendez-vous
 - Conflits sur l'indépendance technologique (bases de données, serveurs)
 - Conflits sur les choix informatiques (confusion gestion / connaissance)
- ▶ Mission Eisenmann => Groupe INIST (à privatiser)
 - Remplacement des nouveaux cadres (dont Nathalie Dusoulier)
 - par des cadres venant du privé (sans compétences en recherche ou IST)
 - Suppression de la R&D
 - Oracle+ protocole qualité
 - Retour à une chaîne de production centralisée
 - Coopérations de plus en plus difficiles
 - Refus de vente pour la fouille de données sur Pascal
 - Pour garantir un monopôle des services sur Pascal
- ▶ 1998 : suite à l'échec de cette stratégie
 - Reprise du contrôle du CNRS Mais
 - La filiale est conservée
 - option indexation automatique (sans expérimentation préalable)
 - Les ingénieurs documentalistes sont invités à partir...

Pendant ce temps aux USA : PubMed

- ▶ Pas de FDP => les éditeurs alimentent PubMed
- ▶ 2000 : Accès devenant gratuit sur Internet
 - => position hégémonique
- ▶ Un nancéien (Olivier Bodenreider) dirige l'indexation assistée
- ▶ Fouille de donnée encouragée
- ▶ Projets collaboratifs (UMLS)
- ▶ 2020 : 900.000 articles par an
 - Chaque article est vu / indexé par 3 humains....
 - Un immense réseau de veille technologique

Des voies pour l'avenir

▶ Comparaisons

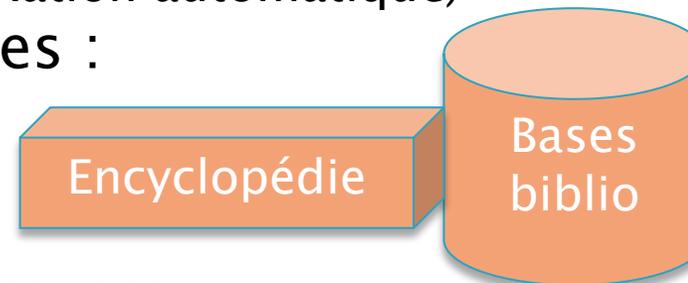
- USA : Les institutions
 - gèrent leurs bibliothèques et échangent des métadonnées
 - Conservent leur expertise en technologie numérique
 - Font confiance à l'expertise humaine (modération, indexation assistée)
- France : Les institutions
 - Délèguent vers des services centralisés (exemple HAL)
 - perdent leur maîtrise (et leur expertise)
 - Ne font pas confiance (contrôle a priori, indexation automatique)

▶ Quels moyens pour redémarrer les bases :

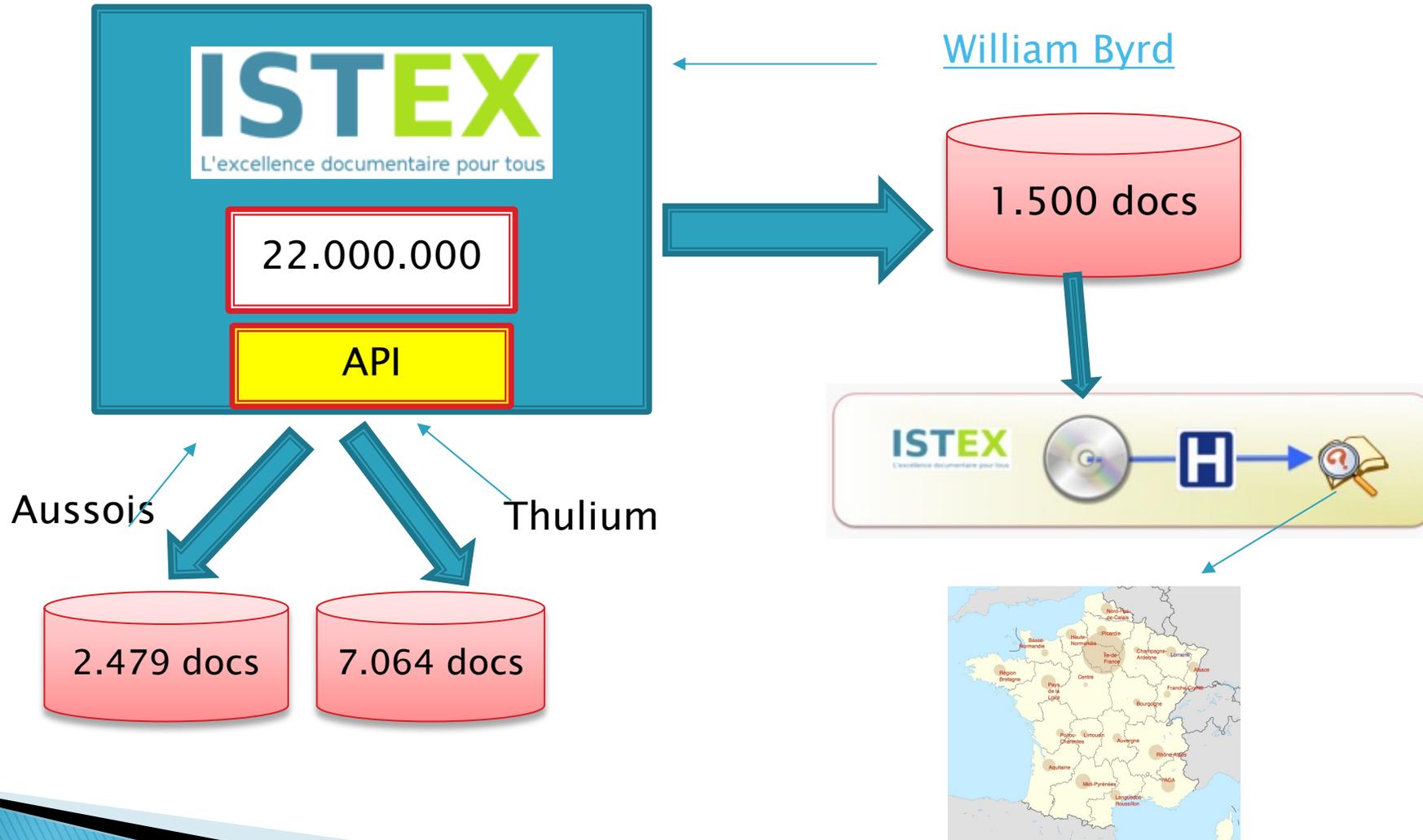
- les mêmes que l'encyclopédie
- pilote les sélections

▶ Quelle expertise

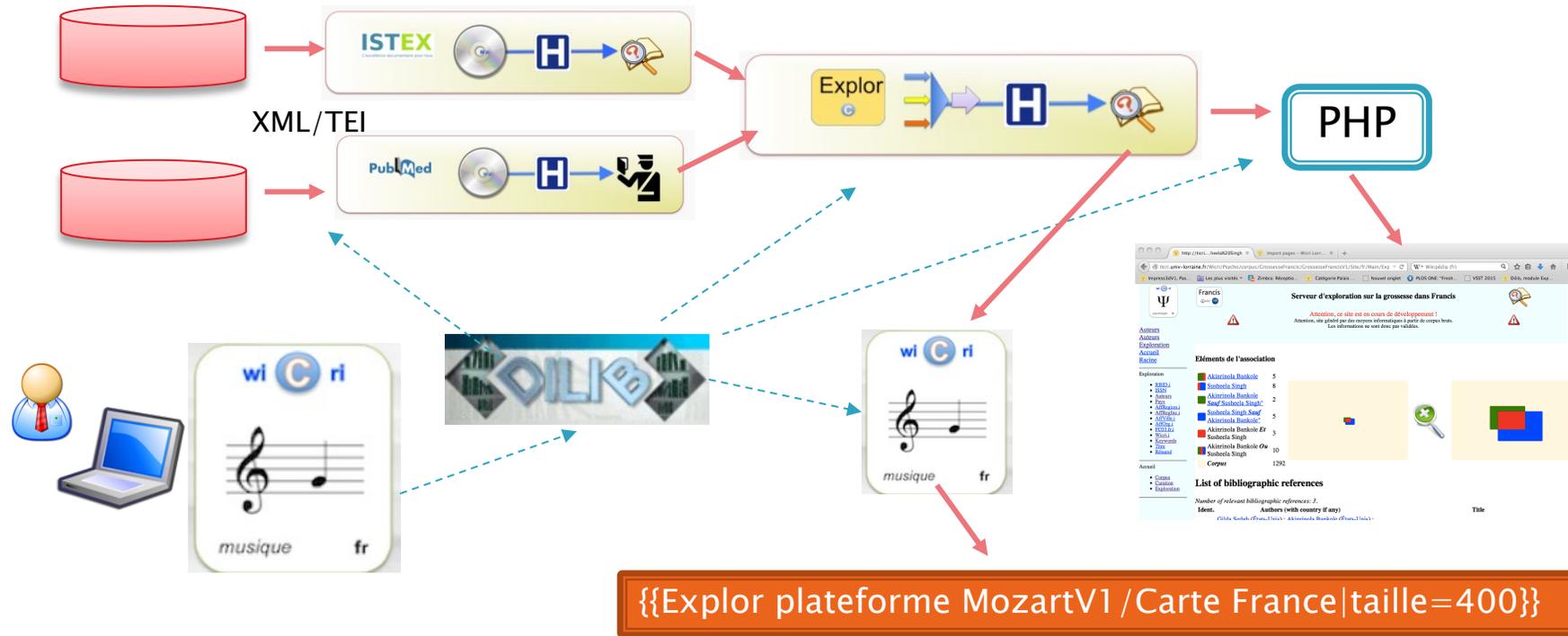
- édition numérique hypertexte de la connaissance
- En maîtrisant l'exploration de corpus



Exploration de corpus avec ISTEEX



ISTEX – Serveur – génération multibase



| Pays | Région | Villes |
|---|--|--|
| 1. France (67) 🔗 | 1. Californie (11) 🔗 | 1. Paris (9) 🔗 |
| 2. États-Unis (31) 🔗 | 2. Île-de-France (9) 🔗 | 2. Marseille (5) 🔗 |
| 3. Royaume-Uni (14) 🔗 | 3. Occitanie (région administrative) (7) 🔗 | 3. Montpellier (4) 🔗 |
| 4. Allemagne (14) 🔗 | 4. Massachusetts (6) 🔗 | 4. Londres (4) 🔗 |
| 5. Canada (11) 🔗 | 5. Angleterre (6) 🔗 | 5. Grenoble (4) 🔗 |
| 6. Italie (10) 🔗 | 6. État de New York (5) 🔗 | 6. Berlin (4) 🔗 |
| 7. Espagne (8) 🔗 | 7. Maryland (5) 🔗 | 7. Toulouse (3) 🔗 |
| 8. Suisse (6) 🔗 | 8. Caroline du Nord (5) 🔗 | 8. Prague (3) 🔗 |
| 9. Australie (6) 🔗 | 9. Arizona (5) 🔗 | 9. Montréal (3) 🔗 |
| 10. Pays-Bas (5) 🔗 | 10. Washington (État) (4) 🔗 | 10. Zurich (2) 🔗 |
| Mots-clés anglais | Mots des titres | ISSN/revue |
| 1. Astrophysics (3) 🔗 | 1. data (10) 🔗 | 1. SPIE proceedings series (6) 🔗 |
| 2. State of the art (2) 🔗 | 2. analysis (7) 🔗 | 2. 1932-6203 (5) 🔗 |
| 3. Software package (2) 🔗 | 3. software (6) 🔗 | 3. Lecture Notes in Computer Science (4) 🔗 |
| 4. Real time (2) 🔗 | 4. microbial (6) 🔗 | 4. Eos Trans. AGU (3) 🔗 |
| 5. Québec (2) 🔗 | 5. marine (5) 🔗 | 5. 2034-9250 (3) 🔗 |
| 6. Perspective (2) 🔗 | 6. genome (5) 🔗 | 6. 1091-6490 (3) 🔗 |
| 7. Open source software (2) 🔗 | 7. distributed (5) 🔗 | 7. 0096-9941 (3) 🔗 |
| 8. Measurement sensor (2) 🔗 | 8. genomic (4) 🔗 | 8. 0027-8424 (3) 🔗 |
| 9. Library network (2) 🔗 | 9. control (4) 🔗 | 9. 2047-217X (2) 🔗 |
| 10. Information policy (2) 🔗 | 10. web (3) 🔗 | 10. 1545-7885 (2) 🔗 |



Santé : Serveurs à génération rapide (5')

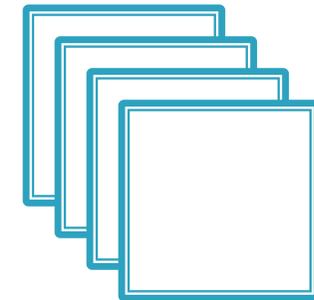
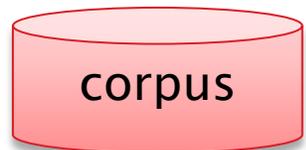
PubMed : 33 millions de citations indexées



NlmPubMedExplorCorpus -q 'influenza AND canada' -s 1000



| PubMed | MeSH | MeSH |
|---------------------|--------------------------|---------------------|
| 1. France (1) [p] | 1. California (1) [p] | 1. Paris (1) [p] |
| 2. Canada (1) [p] | 2. New York (1) [p] | 2. Montreal (1) [p] |
| 3. Quebec (1) [p] | 3. Ontario (1) [p] | 3. Montreal (1) [p] |
| 4. Ontario (1) [p] | 4. Massachusetts (1) [p] | 4. Quebec (1) [p] |
| 5. Quebec (1) [p] | 5. Ontario (1) [p] | 5. Quebec (1) [p] |
| 6. Ontario (1) [p] | 6. Ontario (1) [p] | 6. Quebec (1) [p] |
| 7. Ontario (1) [p] | 7. Ontario (1) [p] | 7. Quebec (1) [p] |
| 8. Ontario (1) [p] | 8. Ontario (1) [p] | 8. Quebec (1) [p] |
| 9. Ontario (1) [p] | 9. Ontario (1) [p] | 9. Quebec (1) [p] |
| 10. Ontario (1) [p] | 10. Ontario (1) [p] | 10. Quebec (1) [p] |



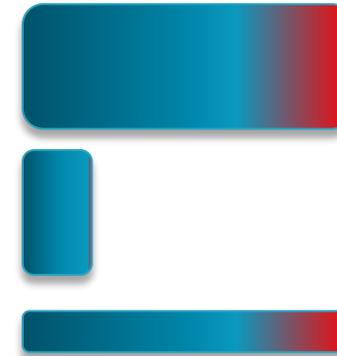
Pages paramètres



Base Xml

Corpus : méfiance / curation

- ▶ Exemple : Mozart
 - 15.000 documents (Musique + médecine)
 - Quelques problèmes de type « avenue Mozart »
 - Plus sérieux :
 - Musique : peu de signalement d'affiliations
 - Médecine : forte politique d'affiliations
 - Les statistiques se focalisent sur la médecine...
- ▶ Exemple : Parkinson en France
 - Parkinson : 90.000 documents
 - Extrait de 4000 documents :
 - peu de bruit
 - Parkinson en France :
 - beaucoup de bruit.
- ▶ **Quelle formation donner à un bibliothécaire pour accompagner un chercheur dans une démarche de curation?**



Règles de curation des données



| | | |
|-----------------|--------------------|---|
| Myriam Chimènes | Myriam Chimènes | affiliation : Institut de recherche en musicologie |
| Denis Herlin | Denis Herlin | affiliation : Institut de recherche en musicologie ; affiliation : Université de Tours |
| Paul Henry Lang | 0027-4631:P. H. L. | affiliation : Université Columbia |
| Edward Lowinsky | Edward E. Lowinsky | affiliation @from=1961 : Université de Chicago |



| | | |
|---------------------|---------------------|--|
| Université Columbia | Columbia University | country : États-Unis ; region @type=state : État de New York ; settlement @type=city : New York |
| Université Cornell | Cornell University | country : États-Unis ; region @type=state : État de New York ; settlement @type=city : Ithaca (New York) |

Exploration, Filtrage



- ▶ Quelles sont les œuvres de Mozart les plus citées dans un corpus ?
 - Idée générale : utiliser le catalogue Köchel
 - Résultat : Sonate KV. 448

```
HfdCat Data/Main/Exploration/biblio.hfd \
| SxmlFindText -r "[K][Vv]*[ \.]*[0-9][0-9]* » \
| SxmlSelect -p @5 -p @1 | sort | IndexBuildRec
```

- ▶ Quelles sont les applications de « *dance therapy* » avec une dimension artistique ?
 - Recherche de présence de chorégraphes (nom-prénom) en utilisant un filtre créé pour les noms binomiaux

Ecrire la musique



<score>

```
{  
  \time 2/4  
  \clef bass  
  c4 c g g a a g2  
}
```

Les commandes sont précédées d'un \

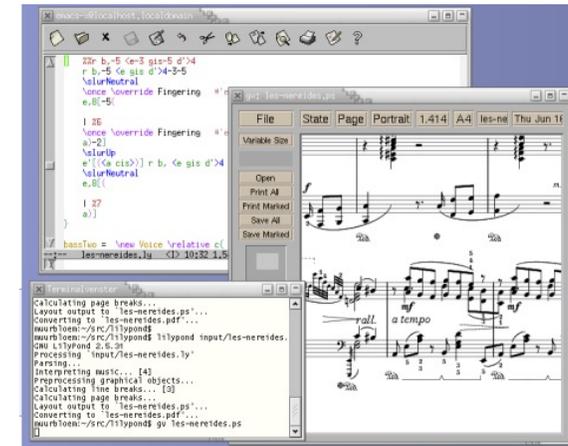
Les lettres représentent les notes

Les chiffres annoncent la durée

</score>



a = la
b = si
c = do
d = ré
...



Apprendre le solfège avec



- ▶ Exemple avec une stagiaire
- ▶ Sur une semaine, 4 à 5 séances de 10 minutes à 1 h

Il court il court le furet

J'ai du bon tabac



Alignement du texte sur la partition

Ecriture de la ligne musicale

Études pour choristes amateurs curieux..

Lassus, musica ficta, contrafacta

Impression pour musicien professionnel

Altérations implicites

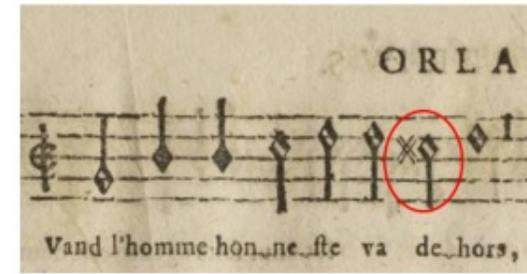


S V P E R I V S.

Vand mon mary vient de dehors, Ma



0:00 / 0:00



ORLA

Vand l'homme hon.ne.ste va de.hors,



0:00 / 0:00

Pour amateurs

Altérations explicites



Quand mon mary vient de dehors,
Ma rente est d'estre battue :
Il prend la cuillier du pot
À la teste il me la rue.

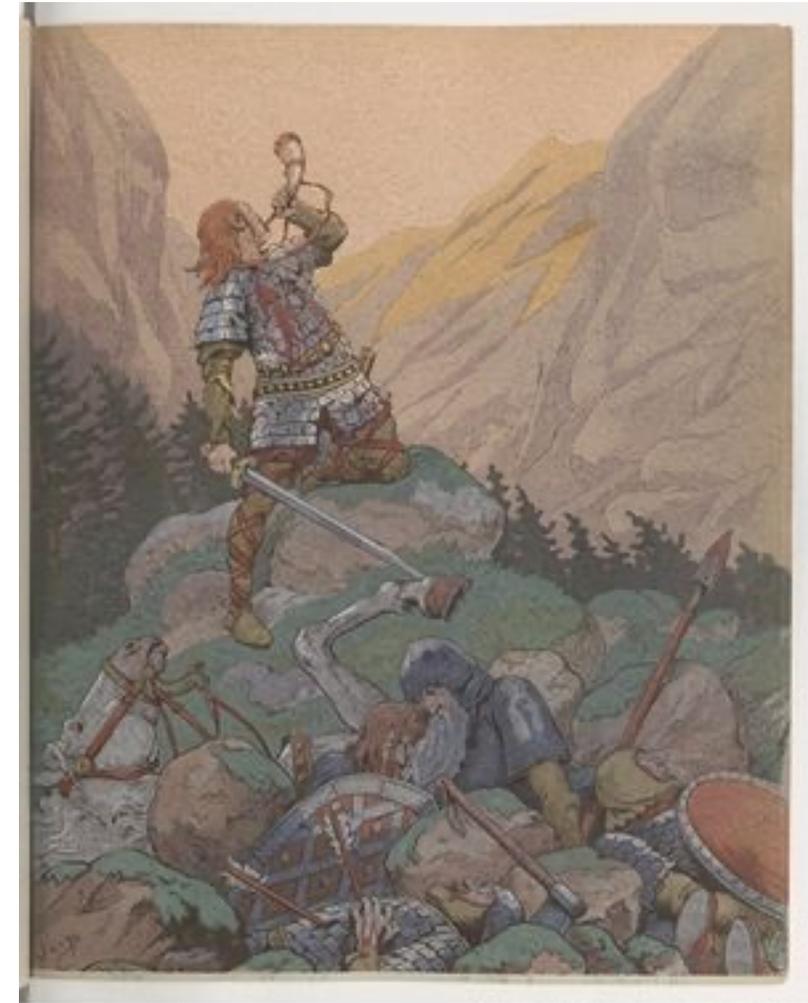
J'ay grand peur qu'il ne me tue.
C'est un faux vilain, jaloux
C'est un vilain, rioteux, grommeleux.
Je suis jeune et il est vieux.

Quand l'homme honneste va dehors
Sa femme n'est par la rue
Ainsi à la besogne alors
Menagère elle se rue

S'il revient elle le salue
Et lui fait accueil gracieux
Il n'est un vilain, rioteux, grommeleux.
Si elle jeune et s'il est vieux.

La Chanson de Roland, une histoire, une épopée

- ▶ Le 14 août 778, Charlemagne de retour d'Espagne perd l'arrière garde de son armée au col de Roncevaux
 - Ainsi est mort Roland
- ▶ La légende introduit la trahison de Ganelon
 - L'attitude héroïque de Roland
 - Qui attend de mourir pour sonner de l'olifant
 - Qui brise le rocher avec Durandal son épée

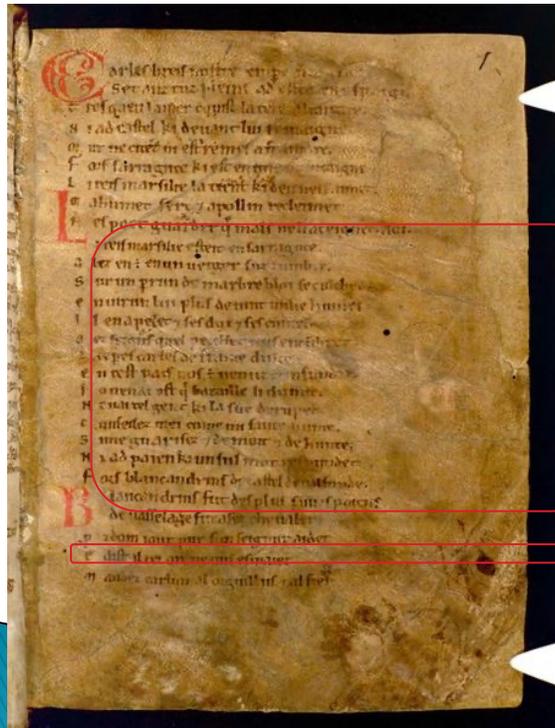


Source gallica.bnf.fr / Bibliothèque nationale de France

La chanson de Roland, des manuscrits

- ▶ Point de départ : inconnu, nombreuses hypothèses,
 - nombreux manuscrits

Manuscrit d'Oxford (XIIe siècle ?)



72 feuillets
Recto-verso

28?
Couplets
(Laises)

4202 vers

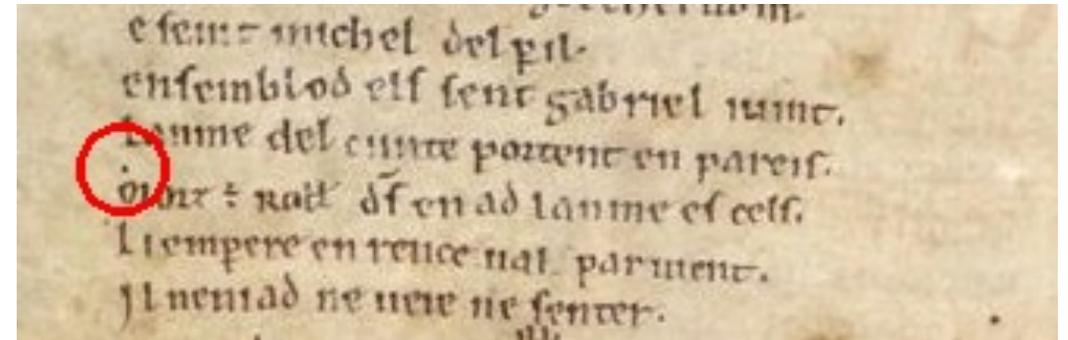
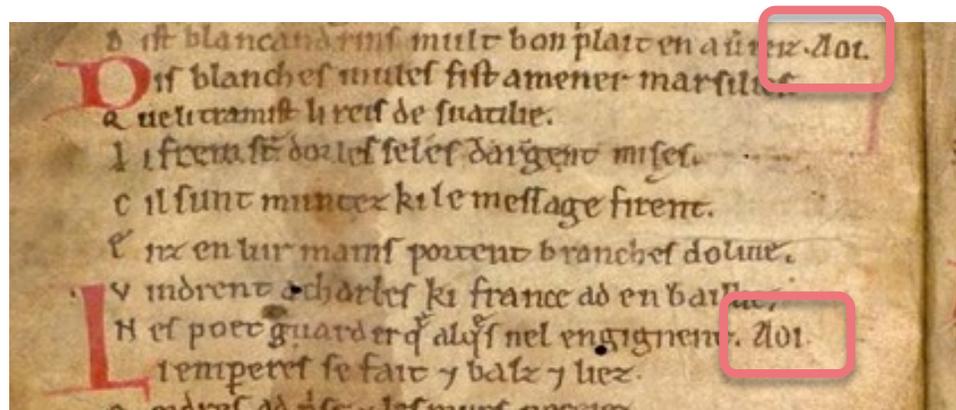
Manuscrit de Venise
XIIIe siècle
6000 vers
419 couplets



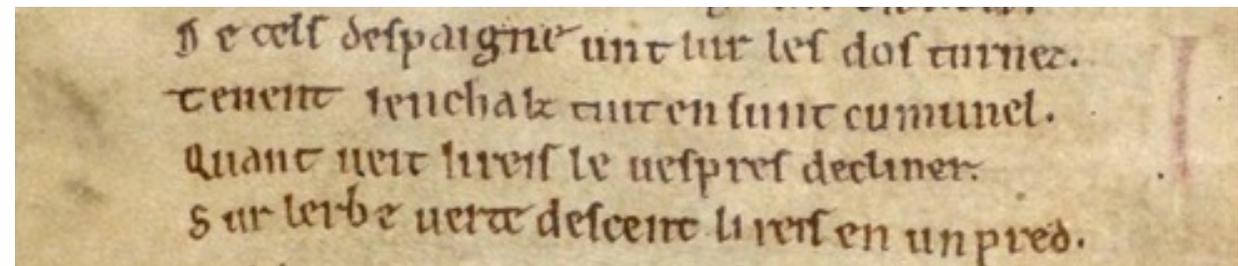
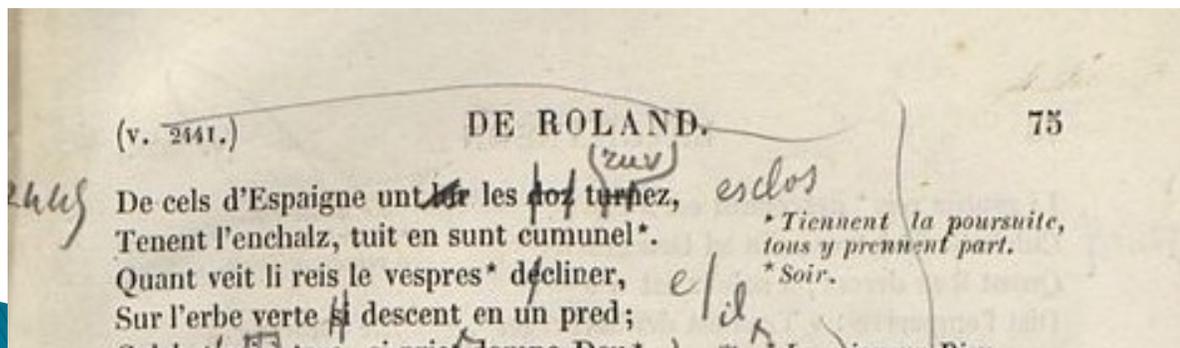
Manuscrit de Paris
XIIIe siècle



La Chanson de Roland, problèmes et divergences



[Feuillet 43 verso](#)



La Chanson de Roland

Des transcriptions, des traductions (centaines)

Francisque Michel (1837)
Première transcription
(Manuscrit d'Oxford)

LA CHANSON
DE ROLAND
de Roncevaux

DE ED. MICHEL

PARIS CHEZ LA CHANSON DE ROLAND

PARIS CHEZ LA CHANSON DE ROLAND



PARIS
CHEZ LA CHANSON DE ROLAND

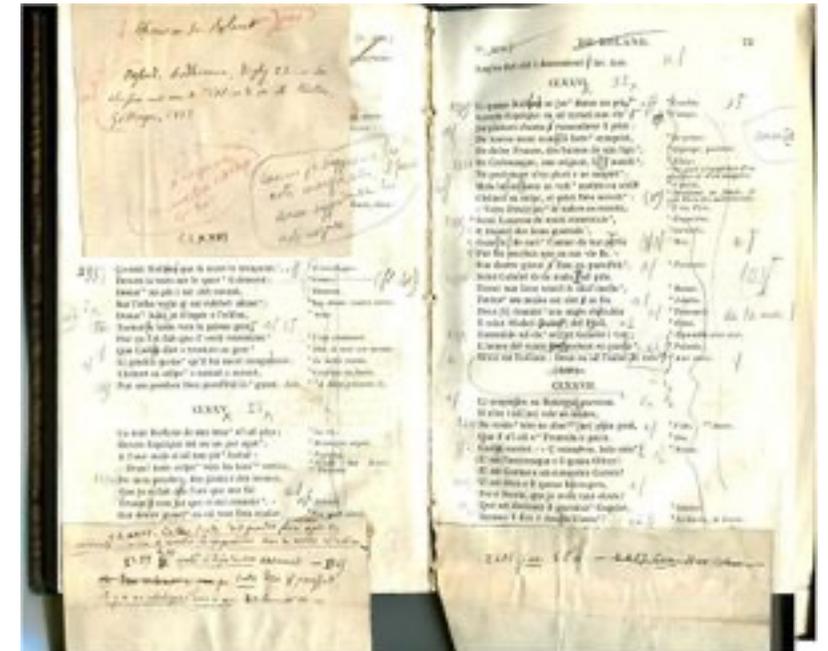
1837

12 1/2

[Léon Gautier](#)
[Edition populaire](#)
[1881](#)



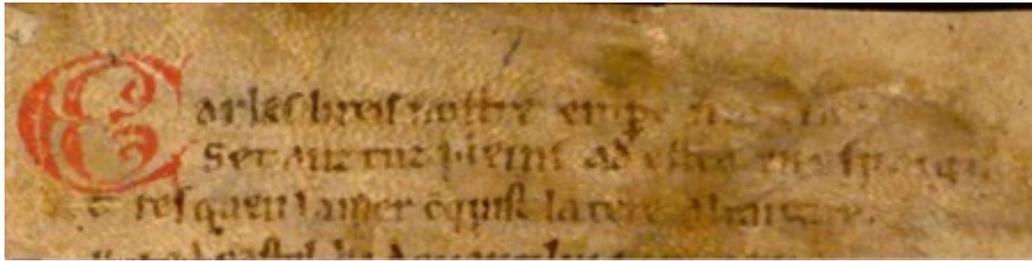
[Francisque Michel \(1869\)](#)
[Annoté par Paul Meyer](#)



La Chanson de Roland

Un oratorio de Gilles Mathieu

▶ Des vers en assonance



Carles li reis, nostre emperere magne*,
Set anz tuz pleins* ad ested en Espaigne,
'Tresqu'en* la mer cunquist la tere altaigne**;
N'i ad castel ki devant lui remaigne*,

▶ Une mise en musique

- En 10 mouvements
- Sur une sélection de 200 vers
- Pour chœurs SATB
- Cor, orchestre à cordes

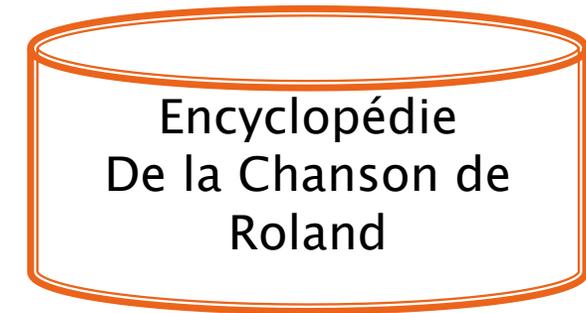
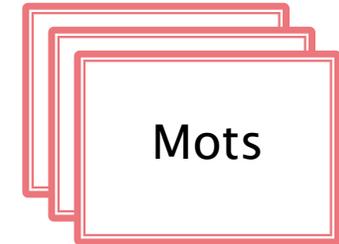
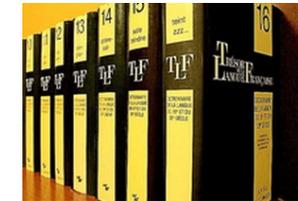
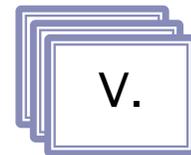
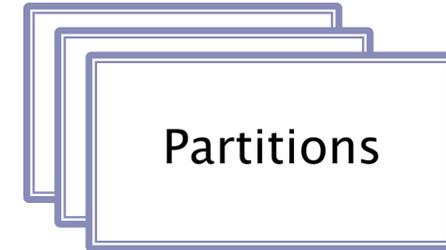
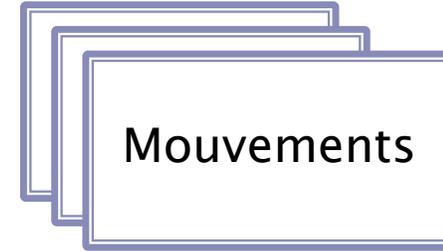
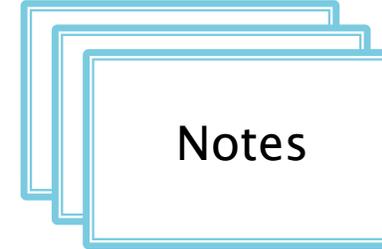
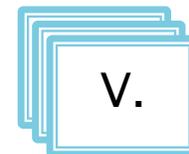
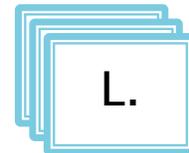
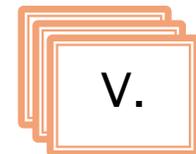
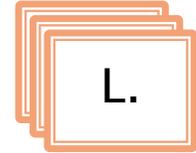
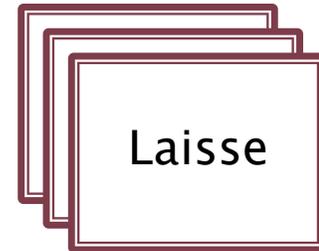
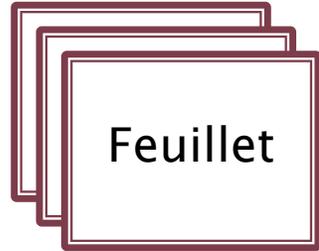
f Ca - rles li - re - is nostr' em - pe - rer mag - nes Set anz tuz pleins ad e - stet en

f Ca - rles li - re - is nostre ma - gnes Set anz tuz pleins ad e - stet en

f Car - les - li reis nostr' em - pe - rer mag - nes Set anz tuz pleins ad e - stet en

f Car - les li reis nostr' em - pe - rer mag - nes Set anz tuz pleins ad e - stet en

La Chanson de Roland un gigantesque hypertexte



Autour de Paul Meyer

Navigation: Accueil, Communauté, Actualités, Modifications récentes, Index alphabétique, Index thématique, Page au hasard, Aide

Rechercher dans Wiki: Continuer, Rechercher

Outils: Pages liées

A pour auteur

Alfred Ewert

Décès : **Oxford**
22 Octobre 1969

Université : **Université d'Oxford**

Est dans la revue

Navigation: Accueil, Communauté, Actualités, Modifications récentes, Index alphabétique, Index thématique, Page au hasard, Aide

Rechercher dans Wiki: Continuer, Rechercher

Outils: Pages liées

A pour créateur

Paul Meyer (philologue)

Naissance : **Paris**
18 janvier 1840

Décès : **Saint-Mandé**
9 septembre 1917

Paul Meyer est un savant français du

A pour créateur

Gaston Paris

Naissance : **Avenay-Val-d'Or**
9 août 1939

Autour du wiki Wicri/Chanson de Roland

- ▶ Navigation hypertexte dans les laisses, le feuillet les paragraphes des transcriptions, traduction, et...
 - Les systèmes de l'oratorio de Gilles Mathieu,
 - Avec un blog pour dialoguer avec le compositeur...
- ▶ A votre disposition pour des visites guidées...
- ▶ Projet musical (en imagination)
 - ACJ Lorraine/Grand Est + Wallonie + Luxembourg + Rhénanie
 - Introduction ouverture musicale Moyen Âge
 - Epilogue pacifique (style The Armed man Jenkins)
 - Site web trilingue Français Anglais Allemand (arabe ?)

Et maintenant

- ▶ Merci pour votre attention
- ▶ Allons-y !!!

