



# **Le projet AMPERE 2.0**

## **Exploitation et annotation d'un corpus de documents, de natures et contenus hétérogènes**

**Christine Blondel, Aude Coustumer, Philippe Pons,  
Marco Segala  
(Centre Alexandre Koyré, Paris)**

ITEM, 2 décembre 2014

# Cadre institutionnel et pérennité d'un projet numérique

## ➤ Notre cadre

- Une UMR CNRS sans ingénieur informaticien
- Un financement ANR à court terme
- Des CDD limités par la loi Sauvadet

## ➤ Pérennité

- Archivage des données (images, fichiers XML) et leurs métadonnées sur la grille du TGIR Huma-num > pérennité assurée
- La maintenance de la plateforme et de ses fonctionnalités > hautement problématique

Condition nécessaire :

une infrastructure numérique stable et de taille suffisante.

# @. Ampère et l'histoire de l'électricité

La plateforme actuelle V1 (2009) [www.ampere.cnrs.fr](http://www.ampere.cnrs.fr) présente

- Le corpus Ampère
  - publications
  - correspondance
  - archives personnelles (fonds Ampère, Académie des sciences)
- Un parcours multimedia sur l'histoire de l'électricité  
(avec vidéos de répliques d'expériences historiques)

# Le site actuel (V1)



Ampère et l'histoire de l'électricité est une plateforme web évolutive, s'adressant aux chercheurs, aux enseignants, aux étudiants, et au grand public curieux. Pour un éclairage rapide sur Ampère voir le diaporama [Une vie en image](#), et pour un aperçu sur les charmes de l'électricité au 18<sup>e</sup> siècle voir la vidéo [L'électricité en spectacle au 18<sup>e</sup> siècle](#).

Derniers ajouts : traductions anglaises [Faraday, Ampere, and the mystery of continuous rotations](#)  
[Did Ampere invent the galvanometer, the telegraph, the electromagnet, the electric motor ... ?](#)  
[Some reflections on a "founding" instrument](#)

Recherche Google dans le site :

## Ampère

La première partie du site est consacrée à André-Marie Ampère (1775-1836) qui fut qualifié par le physicien anglais Maxwell de "Newton de l'électricité". Mais les intérêts intellectuels d'Ampère furent multiples. Ils se lisent dans ses [Publications](#). Sa vie personnelle et sa vie de savant sont éclairées par sa [Correspondance](#) dont une nouvelle édition est donnée ici.

Les [Manuscrits](#) d'Ampère, conservés dans ses archives à l'Académie des sciences, ont été numérisés (53 400 pages) et sont ouverts à la transcription. Une sélection commentée de documents issus de ces archives offre un [Aperçu](#) sur leur richesse.



[Une vie en images](#)



[Publications d'Ampère](#)



[Correspondance](#)



[Manuscrits](#)



[Documents et études sur Ampère](#)

## Histoire de l'électricité et du magnétisme

La deuxième partie du site élargit le propos à l'histoire de l'électricité et du magnétisme. Le [Parcours historique](#) propose une cinquantaine de dossiers multimedia comportant des [vidéos](#) avec reproductions d'expériences historiques. Ce Parcours s'appuie sur des [sources originales](#) présentées dans la partie [Sources primaires](#) (articles de l'*Encyclopédie* de Diderot et d'Alembert, textes de Dufay, Volta, Maxwell, etc. ).

Le [Laboratoire historique](#) offre quelques exemples de discussion et de répétition d'expériences du passé qui posent encore aujourd'hui des questions aux historiens et aux physiciens.



[Un parcours historique...  
de la boussole à la Fée électricité](#)



[Laboratoire historique](#)



[Sources et bibliographie](#)



[Chronologies](#)



[Autres ressources](#)

Ballanche, Pierre-Simon à Ampère, André-Marie | 13 floréal, an XIII [3 mai 1805]

A Monsieur Ampère, répétiteur d'analyse à l'École Polytechnique, place du Corps législatif, à Paris.

Lyon, 13 floréal, an XIII [3 mai 1805]

[...] Ce que vous me dites au sujet de vos succès en métaphysique me désole. Je vois avec peine qu'à 30 ans vous entriez dans une nouvelle carrière. On ne va pas loin quand on change tous les jours de route. Songez bien qu'il n'y a que de très grands succès qui puissent justifier votre abandon des mathématiques, où ceux que vous avez déjà eus présagent ceux que vous devez attendre. Mais je sais que vous ne pouvez mettre de frein à votre cerveau. Cette imagination ne fera-t-elle point quelque tort à vos sentiments religieux ? Prenez bien garde, mon cher et très cher ami, vous êtes sur la pointe d'un précipice [sic] ; pour peu que la tête vous tourne, je ne sais pas ce qui va arriver. Je ne puis m'empêcher d'être inquiet [...] Ce qui sera toujours étonnant, c'est que deux hommes aussi essentiellement différents l'un de l'autre que nous le sommes se conviennent cependant si bien ; car jamais il n'y aura personne qui puisse tenir dans mon coeur la place que vous y tenez. Mon cher ami, j'évite de me replier sur moi-même ; je voudrais que vous puissiez faire la même chose de votre côté ; mais je sens que cela ne vous est pas possible à cause du genre de vos études et de vos occupations habituelles ; c'est en cela que votre position est pire que la mienne. Vous m'avez conseillé de me marier ; j'en avais moi-même le désir ; mais je tremble devant l'avenir qui s'ouvrira devant moi lorsque je me marierai. J'aurai des moments de solitude effrayants parce que, si l'on ne dit pas tout à la personne avec qui on doit passer tous les instants de sa vie, on ne peut trouver le bonheur. Ensuite les enfants et tout le reste. Dans l'état de garçon, il est facile de dévorer son existence ; mais, dans l'état d'époux et de père, c'est bien différent [...]

SIMON

✚ **Source de l'édition électronique de la lettre :**

DE LAUNAY (Louis). *Correspondance du Grand Ampère*. tome I. Paris : Gauthier-Villars, 1936. p. 280-281

Voici le fac-similé de l'imprimé :



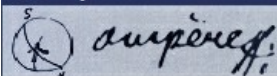
Voici le fac-similé du manuscrit :



✚ **Autre source de la lettre :** original manuscrit

Paris, Archives de l'Académie des sciences, fonds Ampère, carton XXIV, chemise 332





## Correspondance d'Ampère

[Fermer](#)

— 280 —

**141. Ampère à M. Clerc, professeur de mathématiques et de philosophie  
au Collège royal de Lyon, à Lyon (Rhône) (1).**

[1805].

Mon cher ami, j'ai beaucoup de reproches à me faire de ne vous avoir pas écrit plus tôt. Mais j'attendais d'avoir quelque chose à vous mander sur le second article de votre lettre, sur celui qui est relatif à la nouvelle organisation de l'Université; car, à l'égard du premier, sur la manière dont vous seriez payé, on m'a répondu que tout avait été réglé conformément à ce que vous pouviez désirer, et qu'on avait répondu à votre réclamation. On m'a dit aussi que le retard des paiements des honoraires des professeurs était également forcé pour tous par le manque de fonds, qu'on ne pouvait, à ce sujet, faire exception à l'égard de personne, mais que vous recevriez l'arriéré avec les autres à mesure qu'on pourrait payer.

Pour ce qui est de la nouvelle organisation de l'Université, j'ai eu beau faire, je n'ai pu me procurer aucune sorte de renseignements. J'en suis aussi inquiet que vous et, pour dire la vérité, j'ai bien plus à craindre. L'important pour vous est que cette organisation vous trouve en fonction, chargé du plus de travail possible. Dans un pareil moment, c'est une bonne fortune d'être chargé de trois cours. Je ne vois aucune possibilité que vous perdissiez rien aux modifications que peut éprouver l'instruction publique. Je voudrais avoir la même certitude pour moi-même.

M. Lacroix n'a point encore publié son troisième volume, je ne sais pas quand il paraîtra.

M. Beauregard m'a écrit pour que j'appuyasse une demande de retraite qu'il a adressée à la Commission de l'Instruction publique. Je crois qu'une pareille demande est tout ce qu'on peut faire de plus impolitique dans ce moment-ci. Tout ce que j'ai pu faire a été de parler à M. de Mussy de l'intérêt que je porte à M. Beauregard. Comme j'ai été longtemps retenu et que je le suis encore dans ma chambre par un rhume le plus violent que j'aie eu de ma vie, je n'ai pu m'informer de ce qui est résulté d'une démarche qui me paraît si inconsidérée. Comme je lui suis très attaché, je vous prie de me rappeler à son souvenir, et de m'excuser auprès de lui de ce que je n'ai pu lui répondre; le temps m'a manqué absolument pour le faire, ma table est couverte de lettres auxquelles il m'a été impossible de répondre.

Adieu, mon cher ami, je vous aime et vous embrasse de tout mon cœur.

A. AMPÈRE.

(1) Coll. L. De Launay. Quatre pages 19 × 25. Adresse sur la dernière.

Ballanche, Pierre-Simon à Ampère, André-Marie | 13 floréal, an XIII [3 mai 1805]

A Monsieur Ampère, répétiteur d'analyse à l'École Polytechnique, place du Corps législatif, à Paris.

Lyon, 13 floréal, an XIII [3 mai 1805]

[...] Ce que vous me dites au sujet de vos succès en métaphysique me désole. Je vois avec peine qu'à 30 ans vous entriez dans une nouvelle carrière. On ne va pas loin quand on change tous les jours de route. Songez bien qu'il n'y a que de très grands succès qui puissent justifier votre abandon des mathématiques, où ceux que vous avez déjà eus présagent ceux que vous devez attendre. Mais je sais que vous ne pouvez mettre de frein à votre cerveau. Cette imagination ne fera-t-elle point quelque tort à vos sentiments religieux ? Prenez bien garde, mon cher et très cher ami, vous êtes sur la pointe d'un précipice [sic] ; pour peu que la tête vous tourne, je ne sais pas ce qui va arriver. Je ne puis m'empêcher d'être inquiet [...] Ce qui sera toujours étonnant, c'est que deux hommes aussi essentiellement différents l'un de l'autre que nous le sommes se conviennent cependant si bien ; car jamais il n'y aura personne qui puisse tenir dans mon coeur la place que vous y tenez. Mon cher ami, j'évite de me replier sur moi-même ; je voudrais que vous puissiez faire la même chose de votre côté ; mais je sens que cela ne vous est pas possible à cause du genre de vos études et de vos occupations habituelles ; c'est en cela que votre position est pire que la mienne. Vous m'avez conseillé de me marier ; j'en avais moi-même le désir ; mais je tremble devant l'avenir qui s'ouvrira devant moi lorsque je me marierai. J'aurai des moments de solitude effrayants parce que, si l'on ne dit pas tout à la personne avec qui on doit passer tous les instants de sa vie, on ne peut trouver le bonheur. Ensuite les enfants et tout le reste. Dans l'état de garçon, il est facile de dévorer son existence ; mais, dans l'état d'époux et de père, c'est bien différent [...]

SIMON

✚ Source de l'édition électronique de la lettre :

DE LAUNAY (Louis). *Correspondance du Grand Ampère*. tome I. Paris : Gauthier-Villars, 1936. p. 280-281

Voir le fac-similé de l'imprimé :



Voir le fac-similé du manuscrit :



✚ Autre source de la lettre : original manuscrit

Paris, Archives de l'Académie des sciences, fonds Ampère, carton XXIV, chemise 332



# Transcription de manuscrits (V1)

Accueil

Défilement vertical

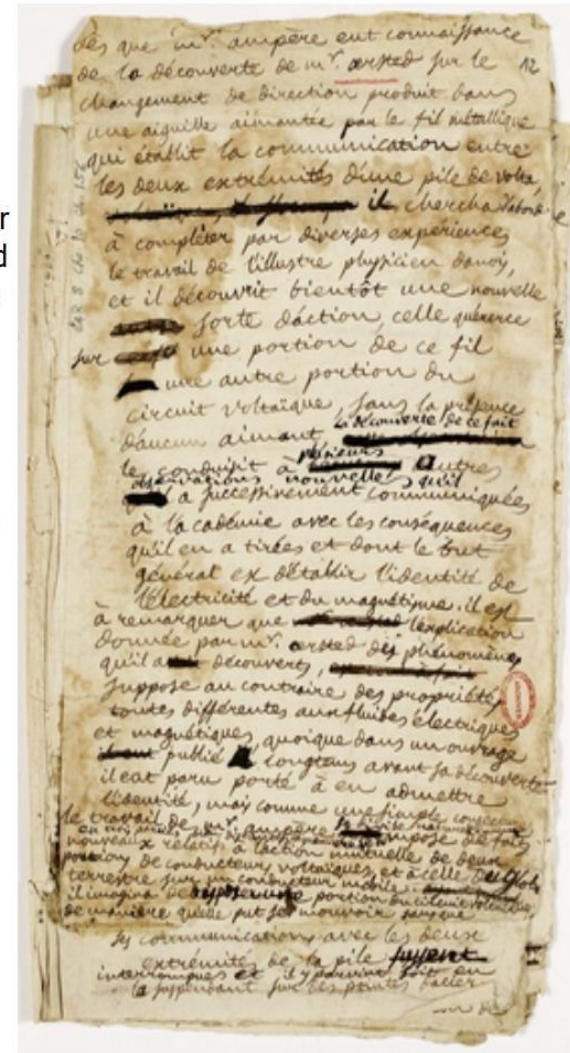
Manuscrits > André-Marie AMPÈRE, Chemise 156 [carton 8], 1775-1836.

|< < Image 26 > >|

| Adresse de citation : [http://www.ampere.cnrs.fr/ms-koyre\\_ampere-156-26-1.5.html](http://www.ampere.cnrs.fr/ms-koyre_ampere-156-26-1.5.html)

Dès que M. Ampère eut connaissance de la découverte de M. Ørsted sur le changement de direction produit dans une aiguille aimantée par le fil métallique qui établit la communication entre les deux extrémités d'une pile de Volta, il chercha d'abord à compléter par diverses expériences le travail de l'illustre physicien danois, et il découvrit bientôt une nouvelle sorte d'action, celle qu'exerce sur une portion de ce fil une autre portion du circuit voltaïque, sans la présence d'aucun aimant. La découverte de ce fait le conduisit à plusieurs autres observations nouvelles qu'il a successivement communiquées à l'académie avec les conséquences qu'il en a tirées et dont le but général est d'établir l'identité de l'électricité et du magnétisme. Il est à remarquer que l'explication donnée par M. Ørsted des phénomènes qu'il a découverts suppose au contraire des propriétés toutes différentes aux fluides électriques et magnétiques, quoique dans un ouvrage, publié longtemps avant sa découverte, il eut paru porté à en admettre l'identité, mais comme une simple conjecture.

Le travail de M. Ampère se divise naturellement en trois parties bien distinctes, la première se compose de faits nouveaux relatifs à l'action mutuelle de deux portions de conducteurs voltaïques et à celle du globe terrestre sur un conducteur mobile. Il imagina de disposer une portion de circuit voltaïque de manière qu'elle put se mouvoir sans que ses communications avec les deux extrémités de la pile fussent interrompues et il y parvint, soit en la suspendant sur des pointes d'acier,





# Objectifs scientifiques et techniques de la V2

## Exploitation et enrichissement du corpus

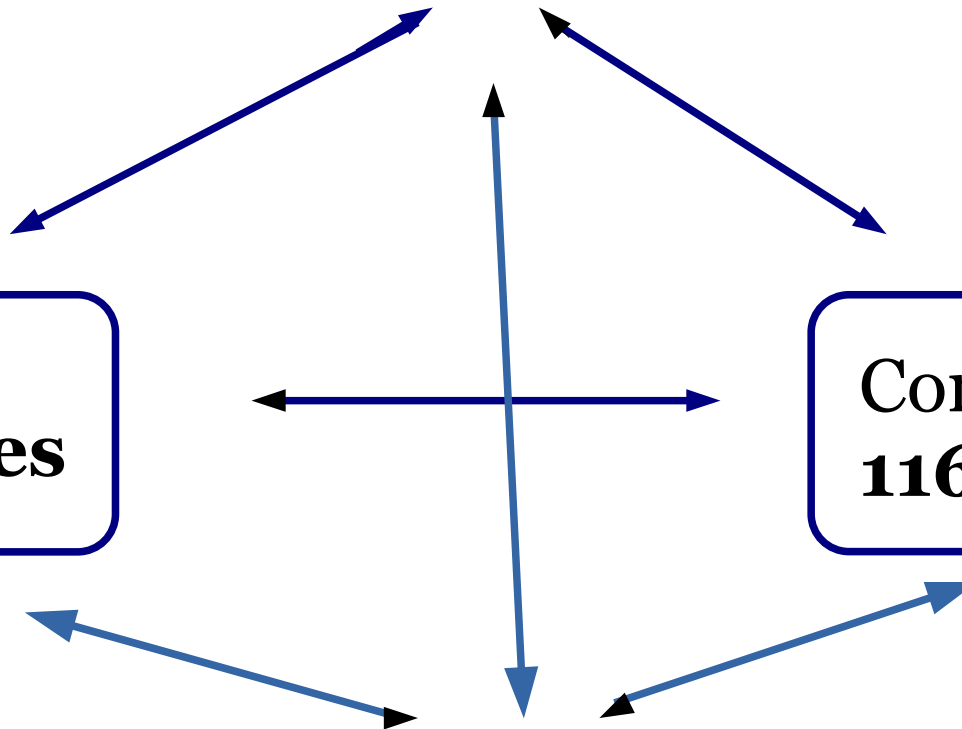
- Tous les domaines de la connaissance : mathématiques, astronomie, physique, sciences naturelles, philosophie, linguistique, poésie...
- Des documents très variés : imprimés, lettres, manuscrits, notes, brouillons, calculs, parfois sur une même page !
- Grande quantité de textes et d'images

Publications d'Ampère  
**321 refs (158 princeps)**

Archives  
**53417 images**

Correspondance  
**1161 lettres**

Bibliographie des études sur Ampère  
**291 refs**



# Objectifs scientifiques et techniques de la V2

- Exploitation et enrichissement du corpus
- Edition du corpus avec des transcriptions différentes pour différents publics
  - > encodage TEI
- Annotation du corpus en ligne
  - > logiciel d'annotation
- Mise en relation des documents (hyperliens, indexation, annotation)
  - > **nouvelle plateforme**

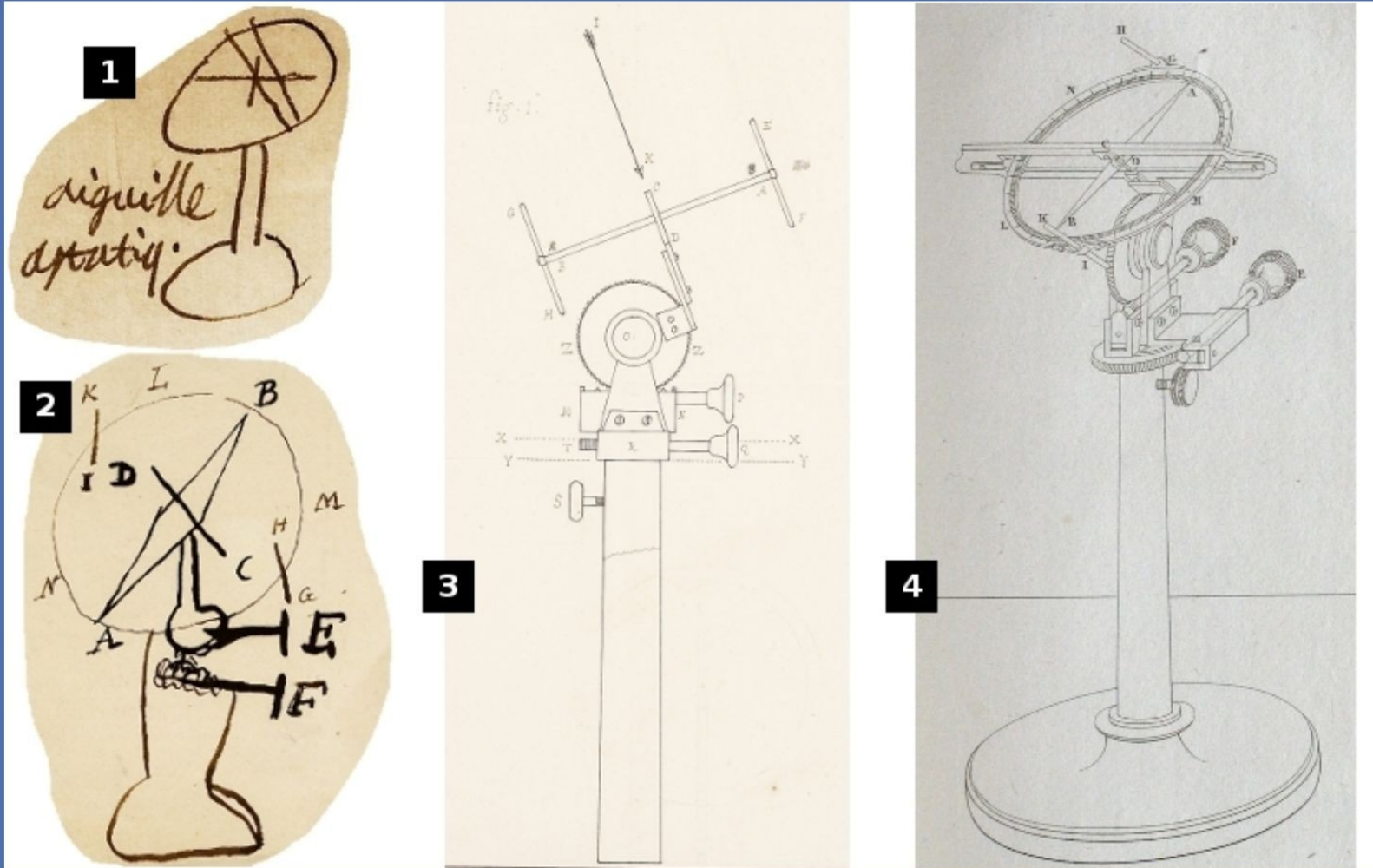


# Et pour la recherche en histoire des sciences ?

- Donner un accès simultané à des documents de statuts et localisations distincts.
- Répondre à des questions de recherche auxquelles on ne peut pas répondre avec une édition papier
- Susciter de nouvelles questions de

# "Génétique des images"

... l'aiguille "astatique" montre qu'une boussole qui n'est plus soumise au magnétisme terrestre s'oriente perpendiculairement au courant :



[www.ampere.cnrs.fr/ms-koyre\\_ampere-206-281-1.5.html](http://www.ampere.cnrs.fr/ms-koyre_ampere-206-281-1.5.html) suspendue à un fil est attirée par un courant :

Versions successives d'un nouvel instrument

# Qu'apporte l'encodage ?

- structure le document pour permettre l'interaction homme-machine
- attache des métadonnées au document pour permettre le moissonnage
- permet une exploitation du texte par des recherches simples ou multicritères



# Qu'est-ce que la TEI ?

- "The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form."  
(<http://www.tei-c.org>)
- La TEI est un langage standard, international, promu par les institutions (W3C, ANR, Huma-Num...)
- La TEI est flexible et s'adapte à (quasiment) tout type de texte

L'encodage en TEI permet d'attacher de l'information à toute partie du texte (mot, phrase, paragraphe,...) :

- information physique/logique : chapitres, pages, paragraphes,...
- information éditoriale : citations, notes, ratures, orthographe ancienne...
- information sémantique : noms, dates, institutions, notions, titres d'ouvrages...

→ l'encodage implique un travail d'analyse du texte.

# Encodage : des choix à faire

Une partie de l'encodage est systématique, une autre partie est ouverte.

Il faut préciser :

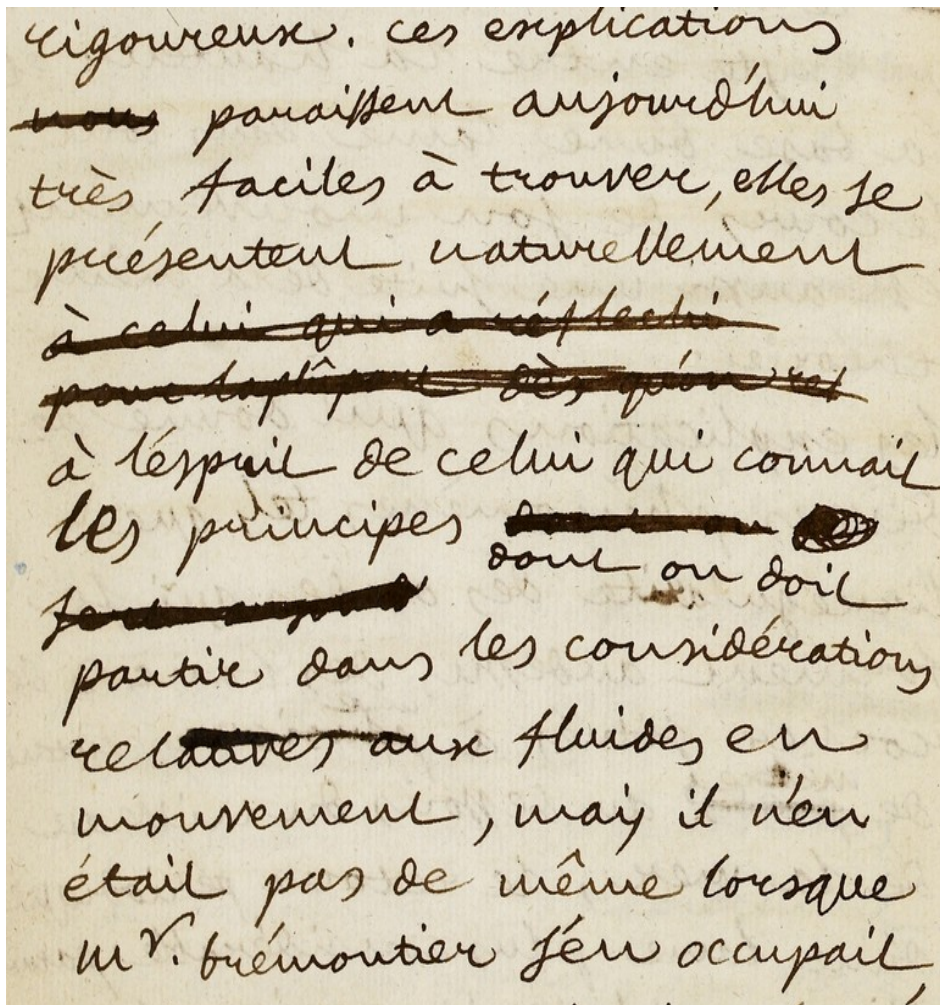
- la structuration du document (ex : formule de politesse dans une lettre)
- le niveau d'encodage (orthographe ancienne, encodage sémantique)
- Viser l'homogénéité de l'encodage sur l'ensemble du corpus
- Le balisage d'un document prend du temps...

Des choix à faire sur la partie du corpus à traiter et le niveau d'encodage, en fonction des besoins et des moyens (financiers et humains).



# Du papier à la page web (1)

## Fac-similé



rigoureux. ces explications  
~~vous~~ paraissent aujourd'hui  
très faciles à trouver, elles se  
présentent naturellement  
~~à celui qui a réfléchi~~  
~~pour l'aptitude des opérations~~  
à l'esprit de celui qui connaît  
les principes ~~de~~  
~~de~~ ~~de~~ ou doit  
partir dans les considérations  
relatives aux fluides en  
mouvement, mais il n'en  
était pas de même lorsque  
M. Brémontier s'en occupait,

## Transcription normalisée

rigoureux. Ces explications nous paraissent aujourd'hui très faciles à trouver, elles se présentent naturellement à l'esprit de celui qui connaît les principes dont on doit partir dans les considérations relatives aux fluides en mouvement, mais il n'en était pas de même lorsque M. Brémontier s'en occupait,

# Du papier à la page web (2)

## Transcription diplomatique

rigoureux. Ces explications  
~~nous~~ paraissent aujourd'hui  
très faciles à trouver, elles se  
présentent naturellement  
~~à celui qui a réfléchi~~  
~~pour la plupart dès qu'on rel~~  
à l'esprit de celui qui connaît  
les principes ~~dont on se sert~~ aujourd  
dont on doit

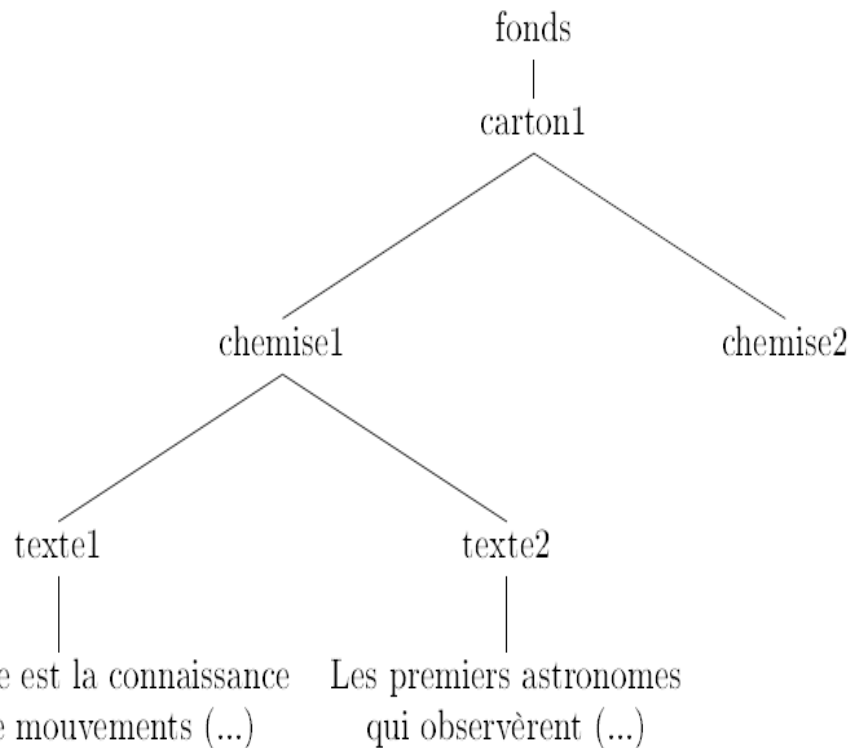
partir dans les considérations  
relatives aux fluides en  
mouvement, mais il n'en  
était pas de même lorsque  
M. Brémontier s'en occupait,

## Encodage en XML-TEI

```
rigoureux. Ces explications<lb/>
<del>nous</del> paraissent aujourd'hui<lb/>
très faciles à trouver, elles se<lb/>
présentent naturellement<lb/>
<del>à celui qui a réfléchi<lb/>
pour la plupart dès qu'on rel</del><lb/>
à l'esprit de celui qui connaît<lb/>
les principes <subst> <del>dont on se<lb/> sert aujourd</del>
  <add>dont on doit</add> </subst><lb/>
partir dans les considérations<lb/> relatives aux fluides en<lb/>
mouvement, mais il n'en<lb/> était pas de même lorsque<lb/>
M. <persName ref="#pers B786">Brémontier</persName> s'en occupait,<lb/>
```

# Le langage XML (eXtensible Markup Language)

## Arborescence



## Représentation XML

```
<fonds>
  <carton>
    <chemise>
      <text n="1">
        L'astronomie est la connaissance<lb/>
        de tant de mouvements (...)
      </text>
      <text n="2">
        Les premiers astronomes<lb/>
        qui observèrent (...)
      </text>
    </chemise>
  </carton>
</fonds>
```

# Chaîne de travail

- Définition de l'encodage
- Création du "schéma TEI" : liste des éléments à utiliser pour le corpus
- Rédaction de normes de transcriptions et d'un manuel d'encodage
- Récupération des données existantes (bases de données)
- Transcription et encodage
- Export du document TEI vers du HTML et du traitement de texte pour faciliter la relecture

# Mais...

- Nécessité d'adapter le nombre de balises utilisées (parmi les 500 existantes) à la nature du corpus et aux moyens disponibles.
- L'encodage n'est pas défini pour certaines catégories textuelles, la correspondance par exemple, mais la communauté y travaille.
- Pour certains types de textes, il existe d'autres standards XML que la TEI : inventaires d'archives (EAD), formules mathématiques (MathML) etc



# État des lieux à mi-parcours

- Publications : transcription et encodage par la société Azentis, contrôlé et affiné ensuite (fin prévue pour décembre 2014).
- Correspondance :
  - export en TEI des lettres publiées
  - rétablissement des coupures à partir des ms originaux
  - encodage sémantique
  - ajout de nouvelles lettres
- Archives : une centaine de chemises transcrites et encodées (env. 5000 pages)
- Index communs aux trois types de documents : noms de personnes, institutions scientifiques, noms de lieux, index thématique

# Version bêta de la nouvelle plateforme sous BaseX

BaseX (logiciel libre) est un système de gestion de base de données XML native, conçu pour la gestion de gros corpus.

ex : Crealscience

- Logiciel adopté par l'Atelier des Humanités Numériques (AHN) de l'ENS Lyon et l'ANR "Guides de Paris".
- Collaboration en vue de la création d'une nouvelle « brique » pour publier sur le web les textes encodés.

# L'Atelier des Humanités Numériques

## Projets AHN

- Chaîne Publication
- Kit Corpus AHN
- MutEC

## Projets Labos

- Bibliothèque Virtuelle Montesquieu
- Bibliparl
- Dictionnaire Montesquieu
- Hyperdonat
- HyperMachiavel
- HyperPrince
- Morand de Jouffrey
- Mosare

## Projets

### Projets de l'AHN

L'AHN est un lieu d'échange qui permet de repérer des problématiques méthodologiques et techniques communes à la plupart des projets relevant des humanités numériques, de discuter de solutions possibles et de déployer des environnements logiciels permettant de tester ces hypothèses.

Ressources

### Projets des laboratoires

Les ingénieur-es participant à l'AHN sont impliqué-es dans leurs laboratoires respectifs dans plusieurs projets de recherche relevant des humanités numériques (éditions critiques en ligne, corpus numériques, bases de données...) Les activités et les réflexions de l'AHN se nourrissent des acquis méthodologiques et techniques, mais aussi des difficultés rencontrées dans la réalisation de ces différents projets.

# Annotation

- Enrichir le texte avec les annotations des chercheurs
- Annoter en ligne les pages web sans intervenir sur les fichiers XML-TEI
- Etablir des relations entre les différents types des documents

# Annotation

- Auteur et date de l'annotation
- Sujet et type d'annotation
- Annotation privée ou partagée
- Annotation collaborative



# Annotation

- La mise en ligne du corpus pour stimuler la recherche partagée
- L'annotation est à la base de la recherche sur un corpus
- Visualisation des annotations et des relations établies par les différents chercheurs
- Le site devient la référence standard de la communauté

# Annotation



Version XIX<sup>e</sup> siècle



Version XXI<sup>e</sup> siècle

➤ **Pundit** (Net7 in Pisa)

possibilité d'annoter en ligne des pages web sans intervenir sur les fichiers XML-TEI.

- a semantic annotation tool for the web:

Linked data, Share & Reuse, Open Source

- Chosen by DM2E (Digitised Manuscripts to Europeana)