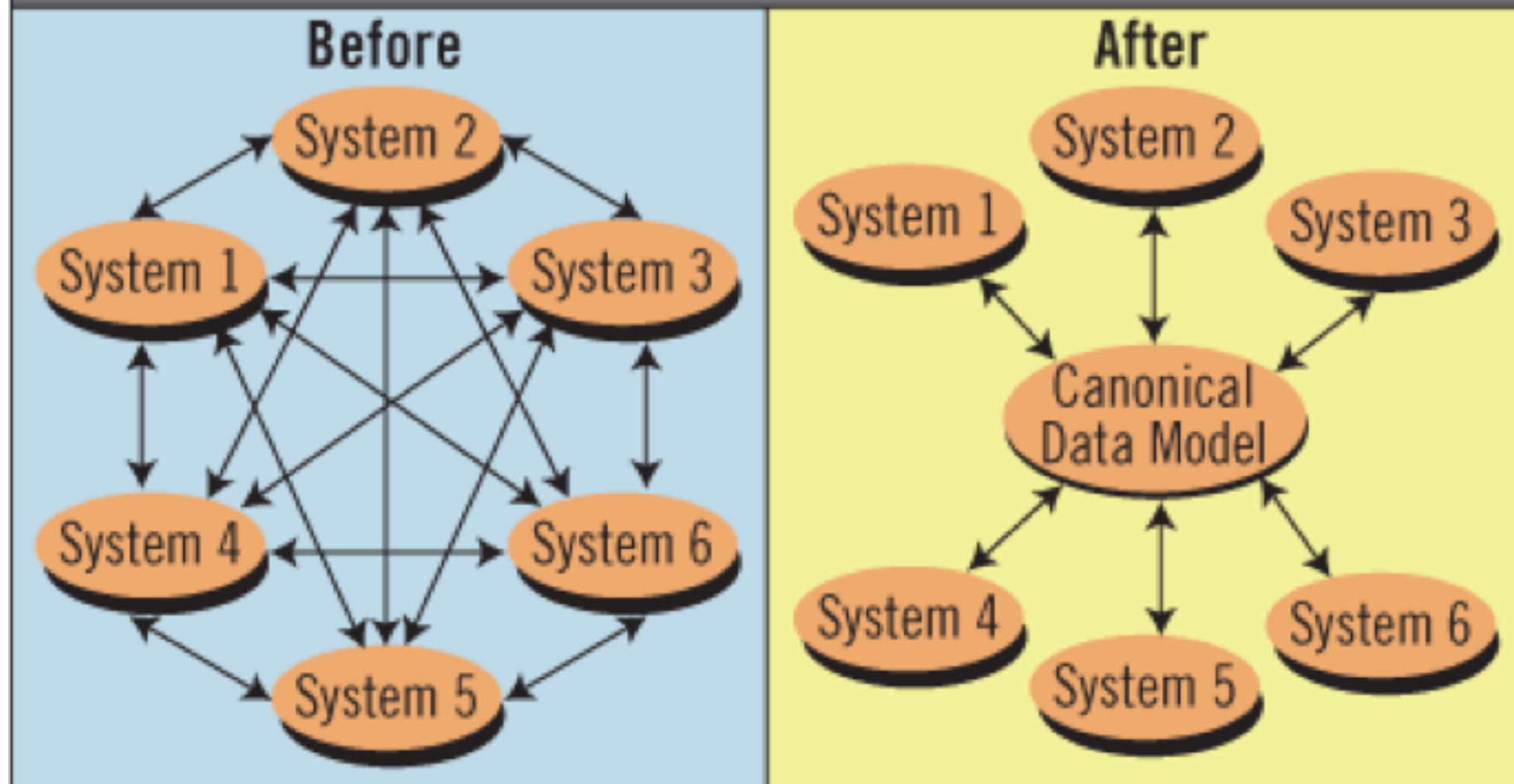


# TEI

Jean-Baptiste Camps & Simon Gabay

Objectif

**Figure 1: Simplified Example of the Use of a Canonical Data Model Before and After**



# Présentation de la TEI

# Présentation

- Fondation
  - Rencontre euro-américaine
  - Trois sociétés savantes
    - Association for Computers and the Humanities
    - Association for Computational Linguistics
    - Association for Literary and Linguistic Computing.
- Principes de Poughkeepsie (NY)

# Présentation

- Les recommandations visent à fournir un *format* standard pour favoriser l'échange de données dans les humanités.
- Les recommandations visent aussi à suggérer des *principes abstraits* pour l'encodage des données.
- Les recommandations doivent :
  - définir une syntaxe recommandée pour ce format,
  - définir un métalangage pour la description des structures d'encodage de données,
  - décrire ce format et ces structures, à la fois dans ce métalangage et en langage naturel.
- Les recommandations doivent proposer des ensembles de conventions d'encodage adaptés à plusieurs applications différentes.
- Notamment, les recommandations doivent inclure un ensemble minimal de conventions pour l'encodage de nouveaux données.
- Les recommandations seront rédigées par plusieurs commissions coordonnées par un comité d'organisation représentant les principaux organismes impliqués (financièrement ou pas). On distinguera :
  - la documentation (métadonnées) des données (*Committee for Text Documentation*)
  - la *représentation* des composants textuels (*Committee for Text Representation*)
  - l'analyse et l'*interprétation* du texte (*Committee for Text Analysis and Interpretation*)
  - la définition du métalangage, ainsi que la description de structures textuelles proposées ou existantes (*Metalanguage and Syntax Committee*)
- La compatibilité avec des standards existants sera maintenue autant que possible.
- Plusieurs grandes bibliothèques de données sont d'accord sur le principe de soutenir les recommandations dans leur fonction de format d'échange. Nous encourageons tous les financeurs à soutenir le développement d'outils pour faciliter cet échange.
- La conversion de données numériques existants vers ce nouveau format implique la traduction de leurs conventions dans la syntaxe du nouveau format. Aucune information supplémentaire ne sera exigée pour la conversion dans ce nouveau format.

# Présentation

- Fondation
- Repères chronologiques
  - 1990 : **TEI Proposal 1** (TEI P1), Guidelines for the Encoding and Interchange of Machine-Readable Texts, dir. Michael Sperberg-McQueen et Lou Burnard
  - 1992-1993 : TEI **P2**, phase d'expansion
  - 1994 : TEI **P3**, considérée comme la première version complète
  - 2000 : naissance du TEI Consortium
  - 2001-2004, TEI **P4** : introduction du XML (à côté de SGML)
  - 2007-... : TEI **P5** (v.1.0) : abandon de SGML, autre phase d'accroissement ; mises à jour bi-annuelles ; 26 juillet 2013, v. 2.5 ; v. 2.9.1, 15 octobre 2015.

# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
  - Bureau
- TEI Board of Directors
- 5 membres élus pour deux ans
- Les minutes des (régulières) rencontres sont publiées sur <http://www.tei-c.org/Board/>



# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
  - Bureau
  - Conseil technique
- TEI technical Council
- 11 membres élus pour deux ans
- La liste est disponible sur <http://www.tei-c.org/Activities/Council/index.xml>

# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
  - Bureau
  - Conseil technique
  - Membres institutionnels et individuels
- Actuellement 53 membres institutionnels dont la liste est publiée sur <http://members.tei-c.org/Institutions>
  - Infrastructures européennes (DARIAH) et nationales (humaNum)
  - Projets internationaux (Perseus)
  - Bibliothèques universitaires (UCLA...)
  - Maisons d'édition (University of Virginia Press)
  - Des chaires (Humboldt Chair of Digital Humanities)

# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
  - Bureau
  - Conseil technique
  - Membres institutionnels et individuels
  - Workgroups
- Travailler aux modifications et aux extensions des TEI guidelines.
- Migration Workgroup (arrêté en 2004)
- Metalanguage Workgroup (arrêté en 2005)
- Physical Bibliography Workgroup (créé en 2004)
- ... (<http://www.tei-c.org/Activities/Workgroups/>)

# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
  - Bureau
  - Conseil technique
  - Membres institutionnels et individuels
  - Workgroups
  - SIG
- Special Interest Groups (SIGs)
  - Structurés autour d'une liste de diffusion
  - Chercheurs aux intérêts communs
    - Bibliothèques
    - Manuscrits
    - TEI pour les linguistes
    - Outils
    - Correspondance

# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
  - Bureau
  - Conseil technique
  - Membres institutionnels et individuels
  - Workgroups
  - SIG
  - Communauté
- Nombreux projets
  - Bibliothèques virtuelles (e-codices, Fribourg)
  - Outils linguistiques (TXM, Lyon)
  - Corpus (Deutsches Textarchiv)
  - ... (<http://www.tei-c.org/Activities/Projects/>)

# Présentation

- Fondation
- Repères chronologiques
- Consortium de la TEI
- Activités
- Un site : <http://www.tei-c.org/>
- Mailinglist anglophone et francophone
- Un Github : <https://github.com/TEIC>
- Members meetings
  - 2016 : Vienne
  - 2017 : Victoria (British Columbia)
  - 2018 : Tokyo
  - 2019 : Graz

# Les Guidelines

# Guidelines

- Que sont les guidelines ?
  - Les *guidelines* définissent et documentent le balisage nécessaire à l'encodage d'un texte, et notamment les sources primaires à des fins de recherche et d'analyse. Les *guidelines* prennent la forme d'un schéma modulaire, extensible et documenté, qui est publié en *open source*.
  - *Grosso modo* c'est un cadre pour encoder un document en TEI.



# Guidelines

- Que sont les guidelines ?

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 3.3.0. Last updated on 31st January 2018, revision f4d8439

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



### Front Matter

#### Title

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- ⊕ iv. [About These Guidelines](#)
- ⊕ v. [A Gentle Introduction to XML](#)
- ⊕ vi. [Languages and Character Sets](#)

### Back Matter

- ⊕ Appendix A [Model Classes](#)
- ⊕ Appendix B [Attribute Classes](#)
- ⊕ Appendix C [Elements](#)
- ⊕ Appendix D [Attributes](#)
- ⊕ Appendix E [Datatypes and Other Macros](#)
- ⊕ Appendix F [Bibliography](#)
- ⊕ Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

### Text Body

- ⊕ 1 [The TEI Infrastructure](#)
- ⊕ 2 [The TEI Header](#)
- ⊕ 3 [Elements Available in All TEI Documents](#)
- ⊕ 4 [Default Text Structure](#)
- ⊕ 5 [Characters, Glyphs, and Writing Modes](#)
- ⊕ 6 [Verse](#)
- ⊕ 7 [Performance Texts](#)
- ⊕ 8 [Transcriptions of Speech](#)
- ⊕ 9 [Dictionaries](#)
- ⊕ 10 [Manuscript Description](#)
- ⊕ 11 [Representation of Primary Sources](#)
- ⊕ 12 [Critical Apparatus](#)
- ⊕ 13 [Names, Dates, People, and Places](#)
- ⊕ 14 [Tables, Formulæ, Graphics and Notated Music](#)
- ⊕ 15 [Language Corpora](#)
- ⊕ 16 [Linking, Segmentation, and Alignment](#)
- ⊕ 17 [Simple Analytic Mechanisms](#)
- ⊕ 18 [Feature Structures](#)
- ⊕ 19 [Graphs, Networks, and Trees](#)
- ⊕ 20 [Non-hierarchical Structures](#)
- ⊕ 21 [Certainty, Precision, and Responsibility](#)
- ⊕ 22 [Documentation Elements](#)
- ⊕ 23 [Using the TEI](#)

### TEI sourcecode

- [Getting and Using the TEI Sources.](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

# Guidelines

- Que sont les guidelines ?
- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html>
- Un exemple : `<div>`

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules

# Guidelines

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	<a href="#">17 Simple Analytic Mechanisms</a>
certainty	Certainty and Uncertainty	<a href="#">21 Certainty, Precision, and Responsibility</a>
core	Common Core	<a href="#">3 Elements Available in All TEI Documents</a>
corpus	Metadata for Language Corpora	<a href="#">15 Language Corpora</a>
dictionaries	Print Dictionaries	<a href="#">9 Dictionaries</a>
drama	Performance Texts	<a href="#">7 Performance Texts</a>
figures	Tables, Formulae, Figures	<a href="#">14 Tables, Formulae, Graphics and Notated Music</a>
gaiji	Character and Glyph Documentation	<a href="#">5 Characters, Glyphs, and Writing Modes</a>
header	Common Metadata	<a href="#">2 The TEI Header</a>
iso-fs	Feature Structures	<a href="#">18 Feature Structures</a>
linking	Linking, Segmentation, and Alignment	<a href="#">16 Linking, Segmentation, and Alignment</a>
msdescription	Manuscript Description	<a href="#">10 Manuscript Description</a>
namesdates	Names, Dates, People, and Places	<a href="#">13 Names, Dates, People, and Places</a>
nets	Graphs, Networks, and Trees	<a href="#">19 Graphs, Networks, and Trees</a>
spoken	Transcribed Speech	<a href="#">8 Transcriptions of Speech</a>
tagdocs	Documentation Elements	<a href="#">22 Documentation Elements</a>
tei	TEI Infrastructure	<a href="#">1 The TEI Infrastructure</a>
textcrit	Text Criticism	<a href="#">12 Critical Apparatus</a>
textstructure	Default Text Structure	<a href="#">4 Default Text Structure</a>
transcr	Transcription of Primary Sources	<a href="#">11 Representation of Primary Sources</a>
verse	Verse	<a href="#">6 Verse</a>

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
- Chaque module est documenté dans un chapitre des Guidelines
- ... définit un certain nombre de composantes du modèle (éléments, classes, macros)
- ... pointe vers des spécifications techniques
- ... correspond à une problématique particulière
- ... si certains sont obligatoires, d'autres sont propres à un type d'objet, une approche, une discipline



# Guidelines

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	<a href="#">17 Simple Analytic Mechanisms</a>
certainty	Certainty and Uncertainty	<a href="#">21 Certainty, Precision, and Responsibility</a>
core	Common Core	<a href="#">3 Elements Available in All TEI Documents</a>
corpus	Metadata for Language Corpora	<a href="#">15 Language Corpora</a>
dictionaries	Print Dictionaries	<a href="#">9 Dictionaries</a>
drama	Performance Texts	<a href="#">7 Performance Texts</a>
figures	Tables, Formulae, Figures	<a href="#">14 Tables, Formulae, Graphics and Notated Music</a>
gaiji	Character and Glyph Documentation	<a href="#">5 Characters, Glyphs, and Writing Modes</a>
header	Common Metadata	<a href="#">2 The TEI Header</a>
iso-fs	Feature Structures	<a href="#">18 Feature Structures</a>
linking	Linking, Segmentation, and Alignment	<a href="#">16 Linking, Segmentation, and Alignment</a>
msdescription	Manuscript Description	<a href="#">10 Manuscript Description</a>
namesdates	Names, Dates, People, and Places	<a href="#">13 Names, Dates, People, and Places</a>
nets	Graphs, Networks, and Trees	<a href="#">19 Graphs, Networks, and Trees</a>
spoken	Transcribed Speech	<a href="#">8 Transcriptions of Speech</a>
tagdocs	Documentation Elements	<a href="#">22 Documentation Elements</a>
tei	TEI Infrastructure	<a href="#">1 The TEI Infrastructure</a>
textcrit	Text Criticism	<a href="#">12 Critical Apparatus</a>
textstructure	Default Text Structure	<a href="#">4 Default Text Structure</a>
transcr	Transcription of Primary Sources	<a href="#">11 Representation of Primary Sources</a>
verse	Verse	<a href="#">6 Verse</a>

# Guidelines

analysis	Analyse simple et interprétations d'éléments textuels (phrase, mots...)
certainty	Niveaux de certitude, de précision et marquage de la responsabilité
core	Eléments utilisés dans tous les documents TEI
corpus	Métadonnées pour les corpus
dictionaries	Ressources lexicales
drama	Scénario, pièce de théâtre ou toute transcription de performance
figures	Images, graphiques, notation musicale dans du texte
gaiji	Caractères non-standards et glyphes
header	En-tête TEI (métadonnées du document)
iso-fs	Structures de traits
linking	Liens, segmentation, alignements
msdescription	Description de manuscrits
namesdates	Noms, dates, personnes, lieux
nets	Graphes réseaux et arbres
spoken	Transcription de documents oraux
tagdocs	Documentation des éléments XML pour les schémas et les DTD
tei	Définition des classes, macros et types de données
textcrit	Apparat critique
textstructure	Eléments de base pour structurer un texte de type livre
transcr	Transcription de sources primaires
verse	Documents versifiés

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
- Organisent les éléments du modèle.
- Les éléments héritent
  - Directement des propriétés de la classe à laquelle ils appartiennent
  - Indirectement des propriétés de la classe de cette classe (superclasse)
  - Ces propriétés sont des attributs ou un emplacement



# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
    - Classe d'attribut
- att. Même ensemble d'attributs
- Classe `att.naming` fournit des attributs communs (`@role`) aux éléments qui concernent des lieux (`<placeName>`), des personnes (`<persName>`), des organismes (`<orgName>`)...  
  
`<persName role="author">Victor Hugo</nom>`
- Classe `att.global` concerne tous les éléments: `@xml:id`, `@n`, `@xml:lang`, `@rend`, `@rendition`, `@xml:base`, `@xml:space`

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
    - Classe d'attribut
    - Classe du modèle
- `model`. Même emplacement dans le modèle
- Classes structurelles/hiéarchiques pour des éléments qui ont le même emplacement dans la structure hiérarchique d'un document TEI
  - Classe `model.divPart` pour tout ce qui peut être contenu dans une `<div>`
  - Classe `model.phrase` pour tout ce qui peut être contenu dans une phrase
- Classes sémantiques pour les éléments avec similarité de sens
  - Classe `model.nameLike` pour les divers types de noms
  - Classe `model.pLike` pour les éléments de types paragraphes

# Guidelines

- Que sont les guidelines ?
- Ce sont des contraintes de contenu
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros
    - *Standard content Models*
- Spécifier le contenu d'éléments
- PCDATA (de texte) ? Un ou des éléments ? Répétés combien de fois ? Dans quel ordre ?

# Guidelines

1. `macro.phraseSeq` (*phrase sequence*) : mélange de PCDATA et d'éléments de niveau phrase (0 ou plus et dans n'importe quel ordre)
2. `macro.paraContent` (*paragraph content*) contenu d'élément de type paragraphe
3. `macro.specialPara` (*'special' paragraph content*) contenu de paragraphes spéciaux, qui peuvent contenir soit du contenu structuré de type sous-éléments, soit du contenu type paragraphe
4. `macro.phraseSeq.limited` (*limited phrase sequence*) mélange de texte et de contenu de niveau phrase, excluant les éléments servant à la transcription de sources
5. `macro.limitedContent` (*paragraph content*) la même chose, mais au niveau paragraphe
6. `macro.xtext` (*extended text*) mélange de PCDATA et d'éléments gaiji

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros
    - *Standard content Models*
    - *Datatype Macros*
- Spécifier la "valeur" d'un `@attribut`
- Chaîne de caractères ? Nom XML ? Liste séparée par des espaces ?

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros
    - *Standard content Models*
    - *Datatype Macros* généralistes
- `data.word` : un mot
- `data.enumerated` : une valeur (un mot) parmi une liste fermée de valeurs possibles
- `data.text` : une chaîne de caractères
- `data.name` : un nom XML
- `data.numeric` : valeur numérique
- `data.count` : compteur

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros
    - *Standard content Models*
    - *Datatype Macros*  
généralistes, spécifiques à la TEI
- `data.certainty` : degré de certitude (`data.certainty = "high" | "medium" | "low" | "unknown"`)
- `data.probability` : probabilité
- `data.interval` : intervalle
- `data.percentage` : pourcentage



# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros
    - *Standard content Models*
    - *Datatype Macros*  
généralistes, spécifiques à la TEI, valeurs normalisées
- `data.duration.w3c` : durée exprimée selon la spécification W3C
- `data.duration.iso` : durée exprimée selon la spécification ISO 8601

# Guidelines

- Que sont les guidelines ?
- Un exemple : `<div>`
- TEI Infrastructure
  - Modules
  - Classes
  - Macros
    - *Standard content Models*
    - *Datatype Macros*  
généralistes, spécifiques à la TEI, valeurs normalisées, valeurs spécialisées
- `data.namespace` : espace de nom
- `data.pattern` : une expression régulière
- `data.pointer` : pointeur
- `data.version` : une version de la TEI ou d'Unicode
- `data.versionNumber` : un numéro de version
- `data.xpath` : une expression XPath