



**HAL**  
open science

# Expérimentation de l'implémentation de l'API DTS par DoTS dans le cadre de JERIHNA et de BibINDEX

Jules Nuguet

► **To cite this version:**

Jules Nuguet. Expérimentation de l'implémentation de l'API DTS par DoTS dans le cadre de JERIHNA et de BibINDEX. 2024. hal-04750036

**HAL Id: hal-04750036**

**<https://hal.science/hal-04750036v1>**

Preprint submitted on 23 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open licence - etalab

# Expérimentation de l'implémentation de l'API DTS par DoTS dans le cadre de JERIHNA et de BiblINDEX

Jules Nuguet

02 avril 2024



Biblissima 

# Comment en est-on venu à s'intéresser à l'utilisation de DTS?

**Pour l'ANR Jerihna** : Étudier la réception des noms propres hébreux chez certains auteurs patristiques

**Pour BiblIndex** : Accéder aux textes patristiques contenant des citations bibliques recensées dans la base



# Historique de BibINDEX



🏠 / Citations

Nom latin

Sélectionner une bible de référence

Télécharger

⌵ Cliquer sur l'en-tête des colonnes pour classer les résultats

Référence bibl		Édition	Position	Type	Description	Auteur cité	Source
[B.] Gn 1:26	Dieu dit : Faisons l'homme à notre image, comme notre ressemblance, et qu'ils dominent sur les poissons de la mer, les oiseaux du ciel, les bestiaux, toutes les bêtes sauvages et toutes les bestioles qui rampent sur la terre.	22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 18 (p.76, L.26)			
[B.] Gn 1:26		22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 21 (p.108, L.11)			
[B.] Gn 1:28		22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.128, L.17)			
[B.] Gn 2:17	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.128, L.9)			
[B.] Gn 2:24	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.111, L.20)	➤		
[B.] Gn 3:1-6	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 10 (p.27, L.13)			
[B.] Gn 3:1-6	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.127, L.26)			
[B.] Gn 3:1	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.142, L.23)			
[B.] Gn 3:6	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.120, L.27)	➤		
[B.] Gn 3:6	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.121, L.1)	➤		
[B.] Gn 3:7	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 22 (p.129, L.5)			
[B.] Gn 3:14	JÉRÔME	Epistulae 1-22	HILBERG 1949 (CUF-JER_Ep1)	[EP] 7 (p.22, L.20)			

## Base de données actuelle

Type	Start	End
Book	✓	✓
Chapter	✓	✓
Paragraph	✓	✓
Page	✓	✓
Line	✓	✓
Type	value	
Edition	✓	
ID	✓	

### Attention :

Tous les champs ne sont pas remplis à chaque fois.

## Antiquité

- Perseids
- Alpheios

## Patristique

- Patristic Text Archive

### Attention :

Leurs implémentations n'utilisent pas les nouvelles recommandations liées aux éléments `<CiteStructure>` et `<CiteData>`.

Partager les données de façon similaire.

## Quels cas d'usage possibles pour le cluster 7 ?

- Partage de fichiers TEI lemmatisés
- Partage d'une édition TEI de dictionnaires
- Partage de relecture de citations
- Partage d'alignements de témoins multilingues

Normalisation du système de requêtes pour le partage de fragments de texte afin de simplifier leur citation



Rappel : L'API DTS, c'est 3 sorties principales (et 1 sortie secondaire) : <https://distributed-text-services.github.io/>

### **Collection**

Cette sortie est utile pour naviguer dans les collections. En effet, une collection comprend généralement des métadonnées, mais aussi un ou plusieurs membres qui représentent soit des sous-collections soit des ressources individuelles.

### **Navigation**

Cette sortie fournit des informations sur les structures internes d'une ressource, notamment comment ses différentes parties sont reliées entre elles et référencées. Elle décrit également les relations entre les divers segments textuels au sein du document.

### **Document**

Cette sortie sert exclusivement à obtenir le contenu d'un document particulier, contrairement aux métadonnées relatives au document et conservées dans les collections respectives.

L'usage de deux balises est recommandé afin de faire le pont entre l'API et les données :

*citeStructure* Déclare une structure et une méthode pour citer le document actuel. [lien](#)

- **@delim**
- **@match**
- **@use**

*citeData* Spécifie comment les informations peuvent être extraites des structures de citation. [lien](#)

- **@use**
- **@property**

# Collection et métadonnées

```
Corpus\  
  date\  
    Epistulae\  
      Epistulae-default1.xml  
      Epistulae-default2.xml  
      ...  
      Epistulae-lemma1.xml  
      Epistulae-lemma2.xml  
      ...  
  metadata\  
    dots_metadata_mapping.xml
```

# Collection et métadonnées

```
<mapping>
  <dc:title scope="collection" format="tsv"
    ↪ source="./default_collections_titles.tsv" resourceId="id"
    ↪ value="title"/>
    <!-- ... -->
  <custom:sent_Name scope="document"
    ↪ xpath="//correspAction[@type='sent']/persName/text()[1]"/>
  <custom:sent_bibliissimaID scope="document"
    ↪ xpath="//correspAction[@type='sent']/persName/idno/text()"/>
  <custom:sent_role scope="document"
    ↪ xpath="//correspAction[@type='sent']/persName/roleName"/>
    <!-- ... (pareil pour le destinataire) -->
    <!-- XPath plus complexe pas limité à la v.1 -->
  <custom:contains_name scope="document" xpath="for $item in
    ↪ distinct-values(//name) order by $item return $item"/>
</mapping>
```

# Structure du texte

```
<refsDecl>
  <citeStructure unit="book"
    match="//body/div[@type='book']" use="@n">
    <citeStructure unit="chapter" match="div"
      use="position()">
    </citeStructure>
  </citeStructure>-
</refsDecl>
```

# Fichier lemmatisé

```
<refsDecl>
  <citeStructure unit="chapter"
    ↪ match="//body/div[@type='chapter']" use="@xml:id">
    <citeStructure unit="section"
      ↪ match="child::div[@type='section']" use="@xml:id">
        <citeStructure match="//w" use="@n" unit="word">
          <citeData use="@lemma" property="lemma"/>
          <citeData use="@pos" property="pos"/>
          <citeData use="@msd" property="msd"/>
          <citeData use="child::text()" property="text"/>
        </citeStructure>
        <citeStructure match="//name" use="@n" unit="name">
          <citeData property="type" use="@type"/>
        </citeStructure>
      </citeStructure>
    </citeStructure>
  </refsDecl>
```

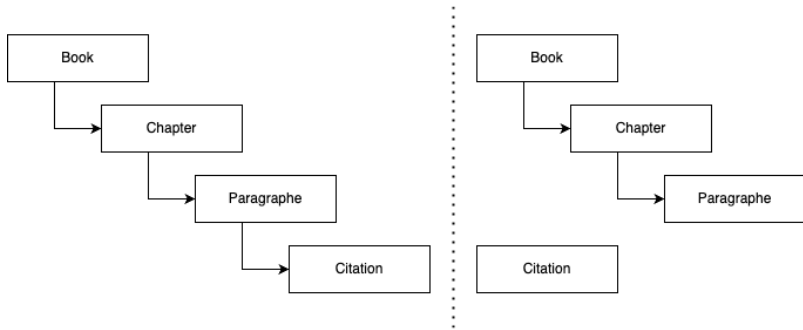
[api/dts/navigation?id=Epistulae-lemma1.xml&down=3](http://api/dts/navigation?id=Epistulae-lemma1.xml&down=3)

# Références externes

```
<citeStructure match="descendant::seg[@type='scriptural']"  
↳ use="@xml:id">  
  <citeData  
    ↳ use="concat('https://www.biblindex.org/fr/quotations/',@n)"  
    ↳ property="quotation"/>  
</citeStructure>
```

```
<seg type="scriptural" xml:id="cit2" n="1640205">quis sapiens et  
↳ intelletget ista, intellegens et cognoscet ea ? </seg>
```

[api/dts/navigation/?id=hos2.xml&down=3](https://www.biblindex.org/fr/quotations/?id=hos2.xml&down=3)



**Note :** À quel niveau doit-on placer une citation ? Au début d'un paragraphe ou encapsulée dans un autre paragraphe ?



# Références internes

```
Corpus/  
  data/  
    manuscripts/  
      C.xml  
      K.xml  
      Lag.xml  
      ...  
      Edition_PP.xml
```

```

<refsDecl>
  <citeStructure match="//entryFree" unit="entry" use="@xml:id">
    <citeData use="@xml:lang" property="lang" />
    <citeData use="if (@corresp) then
      ↪ (concat('api/dts/document?id=',tokenize(base-uri(),'/')[la
      ↪ st()],'.xml&amp;ref=',substring-after(@corresp,'#')))"
      ↪ property="est-une-traduction-de"/>
    <citeStructure unit="app"
      ↪ match="descendant::app[ancestor::entryFree]" use="position()"
      ↪ >
      <citeStructure match="descendant::rdg" use="position()">
        <citeData use=" let $x := ancestor::entryFree/@xml:id
          ↪ return (for $item in tokenize(@wit,' ') return
          ↪ concat('api/dts/document?id=',replace($ite
          ↪ m,'#',' '),'.xml&amp;ref=',$x)) "
          ↪ property="witness"/>
        </citeStructure>
      </citeStructure>
    </citeStructure>
  </refsDecl>

```

/api/dts/navigation/?id=edition\_PP.xml&down=1  
 /api/dts/navigation/?id=edition\_PP.xml&down=3

**Note** : Des choix doivent être faits quant au regroupement ou à la séparation des traductions et des témoins dans des fichiers distincts.

# Citation entre 2 milestones



- L'utilisation de Xpath uniquement semble insuffisante.