

TEI HAL import SWORD

Réseau Aramis – 30 mars 2017

TEI pour l'import HAL

Présentation

- Text encoding Initiative
- Basé sur XML, Langage de marquage qui permet d'échanger des données textuelles informatisées.
- Format standard pour l'import/export dans HAL
- Schéma TEI import = Schéma TEI export
 - <https://hal.archives-ouvertes.fr/documents/aofr.xsd>

Structure TEI

```

<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:hal="http://hal.archives-ouvertes.fr">
  <text>
    <body>
      <listBibl>
        <biblFull> <!-- Référence bibliographique structurée -->
          <titleStmt> <!-- Mention de titre (titre, personnes, responsables) --> </titleStmt>
          <editionStmt> <!-- Edition (Fichiers) --> </editionStmt>
          <publicationStmt> <!-- Publication (Licence) --> </publicationStmt>
          <seriesStmt> <!-- Collection --> </seriesStmt>
          <notesStmt> <!-- Notes, commentaires --> </notesStmt>
          <sourceDesc> <!-- Description bibliographique --> </sourceDesc>
          <profileDesc> <!-- Description des aspects non bibliographiques --> </profileDesc>
        </biblFull>
      </listBibl>
    </body>
    <back>
      <listOrg type="structures"> <!-- Définition de nouvelles structures --> </listOrg>
      <listOrg type="projects"> <!-- Définition de nouveaux projets (ANR, Eur) --> </listOrg>
    </back>
  </text>
</TEI>

```

Mention de titre

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:titleStmt

Financement

- Projet ANR

```
<!-- Accès au référentiel https://api.archives-ouvertes.fr/ref/anrProject?q=* &fl=* -->
<funder ref="#projanr-Identifiant du projet ANR dans AuréHAL"/>
<funder ref="#localProjanr-Identifiant du projet défini dans la partie back du fichier"/>
```

- Projet Européen

```
<!-- Accès au référentiel : http://api.archives-ouvertes.fr/ref/europeanProject?q=* &fl=* -->
<funder ref="#projeurop-Identifiant du projet européen dans AuréHAL"/>
<funder ref="#localProjeurop-Identifiant du projet défini dans la partie back du fichier"/>
```

- Autre financement

```
<funder>Financement</funder>
```

Section Edition

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:editionStmt

- Date d'écriture

```
<date type="whenWritten">Date d'écriture (ex: 2014-09-08)</date>
```

- Fichiers associés au dépôt

```
<!-- @titre : référentiel type de fichier http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:fileType -->
<!-- @subtype : référentiel origine du fichier http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:fileSource -->
<ref type="file" subtype="author" target="Lien vers le pdf"
n="1 (principal)">
  <desc>Description du fichier</desc>
  <date notBefore="Date d'embargo (2015-02-14)"/>
</ref>
```

Section Publication

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:publicationStmt

- Licence

```
<availability>
  <!-- @target : Référentiel des licences http://api.archives-
ouvertes.fr/ref/metadataList?q=metaName_s:licence -->
  <licence target="Licence"/>
</availability>
```

Collections

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:seriesStmt

- Tamponnage au moment du dépôt
 - Il faut être gestionnaire de collection et tamponner avec ses collections

```
<!-- @n : identifiant de la collection. Accès à la liste https://api.archives-ouvertes.fr/search/?q=\*&rows=0&facet=true&facet.field=collCode\_s -->  
<idno type="stamp" n="Identifiant de la collection"/>
```


Notes

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:notesStmt

```
<note type="commentary">Commentaire</note>
```

```
<note type="description">Description</note>
```

```
<!-- @n : audience; https://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:audience -->
```

```
<note type="audience" n="Audience"/>
```

```
<!-- @n : conférence invitée; http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:invitedCommunication -->
```

```
<note type="invited" n="Conférence invitée"/>
```

```
<!-- @n : vulgarisation; http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:popularLevel -->
```

```
<note type="popular" n="Vulgarisation"/>
```

```
<!-- @n : Comité de lecture ; https://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:peerReviewing -->
```

```
<note type="peer" n="Comité de lecture"/>
```

```
<!-- @n : Actes ; http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:proceedings -->
```

```
<note type="proceedings" n="Acte"/>
```

```
<!-- @n : Type de rapport ; http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:reportType -->
```

```
<note type="report" n="Type de rapport"/>
```

```
<!-- @n : Type d'image ; http://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:imageType -->
```

```
<note type="image" n="Type d'image"/>
```

```
<!-- @n : Niveau de cours ; https://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:lectureType -->
```

```
<note type="lecture" n="Niveau du cours"/>
```

Description Bibliographique

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:sourceDesc

```

<biblStruct> <!-- Référence bibliographique structurée -->
  <analytic><!-- Niveau analytique (titre, sous titre, auteurs) --></analytic>
  <monogr><!-- Niveau monographique --></monogr>
  <series><!-- Editeur collection, nom du cours --></series>
  <idno type="doi"><!-- Liste d'identifiants du document --></idno>
  <ref type="seeAlso"><!-- Complément sur le dépôt --></ref>
  <ref type="publisher"><!-- Lien vers le publisher --></ref>
  <relatedItem><!-- Ressources liées --></relatedItem>
</biblStruct>
  
```

Description de la source

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:sourceDesc/tei:biblStruct/tei:analytic

```

<?xml version="1.0" encoding="UTF-8"?>
<title xml:lang="en">Titre en anglais</title>
<title xml:lang="fr">Titre en français</title>
<title type="sub" xml:lang="fr">Sous-titre en français</title>
<title type="sub" xml:lang="de">Sous titre en allemand</title>
<!-- @role : fonction de l'auteur https://api.archives-ouvertes.fr/ref/metadataList?q=metaName_s:relator -->
<author role="aut">
  <persName>
    <forename type="first">Prénom</forename>
    <forename type="middle">Autre prénom</forename>
    <surname>Nom de famille</surname>
  </persName>
  <email>Adresse mail</email>
  <ptr type="url" target="URL du site web"/>
  <!-- liste des auteurs http://api.archives-ouvertes.fr/ref/author?q=*&fl=* -->
  <idno type="idhal">IdHal de l'auteur dans HAL</idno>
  <orgName ref="#struct-Identifiant d'une structure connue dans HAL"/>
  <!-- liste des structures dans AuréHAL http://api.archives-ouvertes.fr/ref/structure?q=*&fl=* -->
  <affiliation ref="#struct-Identifiant d'une structure connue dans HAL"/>
  <affiliation ref="#localStruct-Identifiant interne d'une nouvelle structure à définir dans la section back"/>
</author>
  
```

Ajouter un auteur (1)

- A partir d'un identifiant d'une forme auteur

```
<author role="crp">
  <idno type="halauthorid">49567</idno>
</author>
```

- A partir d'un idHAL

```
<author role="aut">
  <idno type="idhal" notation="string">laurentromary</idno>
  <!-- ou -->
  <idno type="idhal" notation="numeric">307</idno>
</author>
```

- A partir d'un identifiant externe

```
<author role="aut">
  <idno type="ORCID">0000-0002-0756-0508</idno>
</author>
```

Ajouter un auteur (2)

- Créer un nouvel auteur

```

<author role="aut">
  <persName>
    <forename type="first">Laurent</forename>
    <surname>Romary</surname>
  </persName>
  <email>laurent.romary@inria.fr</email>
  <affiliation ref="#localStruct-1"/>
</author>

<author role="aut">
  <idno type="halauthorid">98062</idno>
  <affiliation ref="#struct-3102"/>
</author>

<author role="aut">
  <idno type="idhal" notation="string">laurentromary</idno>
  <affiliation ref="#struct-3102"/>
</author>

```

Description monographique

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:sourceDesc/tei:biblStruct/tei:monogr

```

<idno type="nnt">Numéro National de Thèse</idno>
<idno type="isbn">ISBN</idno>
<idno type="eisbn">eISBN</idno>
<idno type="patentNumber">Numéro de brevet</idno>
<idno type="reportNumber">Numéro de rapport</idno>
<idno type="localRef">Référence interne</idno>
<!-- Référentiel AuréHAL des journaux https://api.archives-ouvertes.fr/ref/journal?q=* &fl=* -->
<idno type="halJournalId">Identifiant du journal dans HAL</idno>
<idno type="issn">ISSN</idno>
<idno type="eissn">eISSN</idno>
<title level="j">Nom du journal</title>
<title level="m">Titre de l'ouvrage</title>
<title level="m">Source</title>
<title level="u">Nom du cours</title>
  
```

- Ajout d'un journal à partir de son identifiant AuréHAL ou en le créant directement dans la TEI (issn, eissn, j)

Description monographique

```

<meeting>
  <title>Titre de la conférence</title>
  <date type="start">Date de début de la conférence</date>
  <date type="end">Date de fin de la conférence</date>
  <settlement>Ville</settlement>
  <country key="Code Pays"/>
</meeting>
<respStmt>
  <name>Organisateur de la conférence</name>
</respStmt>
<settlement>Ville</settlement>
<country key="Code Pays"/>
<editor>Editeur scientifique</editor>
<imprint>
  <publisher>Editeur commercial</publisher>
  <biblScope unit="serie">Série</biblScope>
  <biblScope unit="volume">Volume</biblScope>
  <biblScope unit="issue">Issue</biblScope>
  <biblScope unit="pp">Numéro de page</biblScope>
  <date type="datePub">Date de publication</date>
  <date type="dateDefended">Date de soutenance</date>
  <date type="dateEpub">Date de publication électronique</date>
</imprint>
<authority type="institution">Organisme de thèse, mémoire</authority>
<authority type="school">Ecole doctorale</authority>
<authority type="supervisor">Directeur de thèse</authority>
<authority type="jury">Jury</authority>

```

Identifiants, Liens, Ressources Liées

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:sourceDesc/tei:biblStruct/tei:idno | tei:ref | tei:relatedItem

- Identifiants externes

```
<!-- @type : doi, arxiv, bibcode, ird, pubmed, ads, pubmedcentral, irstea, sciencespo, oatao,
ensam, prodira -->
<idno type="Type d'identifiant">Identifiant</idno>
```

- Complément sur la ressource

```
<ref type="seeAlso">Lien vers la ressource complémentaire</ref>
```

- Ressources liées dans HAL

```
<!-- @type : liste accessible https://api.archives-
ouvertes.fr/ref/metadataList?q=metaName_s:relatedType-->
<relatedItem type="Type de relation" target="Lien vers le document" >
  Descriptif de la relation
</relatedItem>
```


Description non bibliographique

/tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:profileDesc

```

<profileDesc>
  <langUsage>
    <language ident="Code langue du texte"/>
  </langUsage>
  <textClass>
    <keywords scheme="author">
      <term xml:lang="Code langue du mot clé">Mot clé</term>
    </keywords>
    <classCode scheme="classification">Classification (pacs, msc, ...)</classCode>
    <classCode scheme="acm">Code ACM</classCode>
    <classCode scheme="mesh">Mot clé MESH</classCode>
    <classCode scheme="jel">Mot clé JEL</classCode>
    <!-- @n : disciplines HAL http://api.archives-ouvertes.fr/ref/domain?q=* -->
    <classCode scheme="halDomain" n="Code discipline HAL"/>
    <!-- @n : Type de document HAL http://api.archives-ouvertes.fr/ref/doctype?q=* -->
    <classCode scheme="halTypology" n="Type de dépôt"/>
  </textClass>
  <abstract xml:lang="Code langue du résumé">Résumé</abstract>
  <particDesc>
    <org type="consortium">Collaboration</org>
  </particDesc>
</profileDesc>
  
```

Disciplines et types de dépôt dépendent du portail

Métadonnées spécificités portails

- /tei:TEI/tei:text/tei:body/tei:listBibl/tei:biblFull/tei:editionStmt/tei:edition/tei:fs/tei:f

```

<editionStmt>
  <edition>
    <fs>
      <f name="metadonnée libre">
        <string>Valeur</string>
      </f>
    </fs>
    <fs>
      <f name="metadonnée liste prédéfinie">
        <numeric>Valeur</numeric>
      </f>
    </fs>
  </edition>
</editionStmt>
  
```

Annexes

/tei:TEI/tei:text/tei:back

- Définition des structures de recherches et projets (ANR, européens) non connus du référentiel HAL

```

<back>
  <listOrg type="structures">
    <org type="researchteam" xml:id="localStruct-Identifiant de la
structure">...</org>
    <org type="institution" xml:id="localStruct-Identifiant de la structure">...</org>
  </listOrg>
  <listOrg type="projects">
    <org type="anrProject" xml:id="localProjanr-Identifiant du projet">...</org>
    <org type="europeanProject" xml:id="localProjeurope-Identifiant du
projet">...</org>
  </listOrg>
</back>

```

Structure de recherche

/tei:TEI/tei:text/tei:back/tei:listOrg[@type = "structure"]

```

<!-- @type : researchteam, department, laboratory, institution -->
<org type="laboratory" xml:id="localStruct-1">
  <orgName>Centre pour la Communication Scientifique Directe</orgName>
  <orgName type="acronym">CCSD</orgName>
  <desc>
    <address>
      <addrLine>25 bd Pierre de Coubertin, 69100 Villeurbanne</addrLine>
      <country key="FR"/>
    </address>
    <ref type="url">http://ccsd.cnrs.fr</ref>
  </desc>
  <listRelation>
    <relation name="UMS3668" active="#struct-100000"/>
    <relation active="#struct-100001"/>
    <relation active="#localStruct-2"/>
  </listRelation>
</org>
<org type="institution" xml:id="localStruct-2">
  <orgName>Université de Lyon</orgName>
  <orgName type="acronym">PRES de Lyon</orgName>
  <desc>
    <address>
      <country key="FR"/>
    </address>
    <ref type="url">http://www.univ-lyon.fr</ref>
  </desc>
</org>
  
```

Projets

/tei:TEI/tei:text/tei:back/tei:listOrg[@type = "project"]

- Projet ANR

```

<org type="anrProject" xml:id="localProjanr-Identifiant du projet européen">
  <idno type="anr">Numéro du projet</idno>
  <orgName>Nom</orgName>
  <orgName type="program">Programme</orgName>
  <desc>Description</desc>
  <date>Année</date>
</org>
  
```

- Projet européen

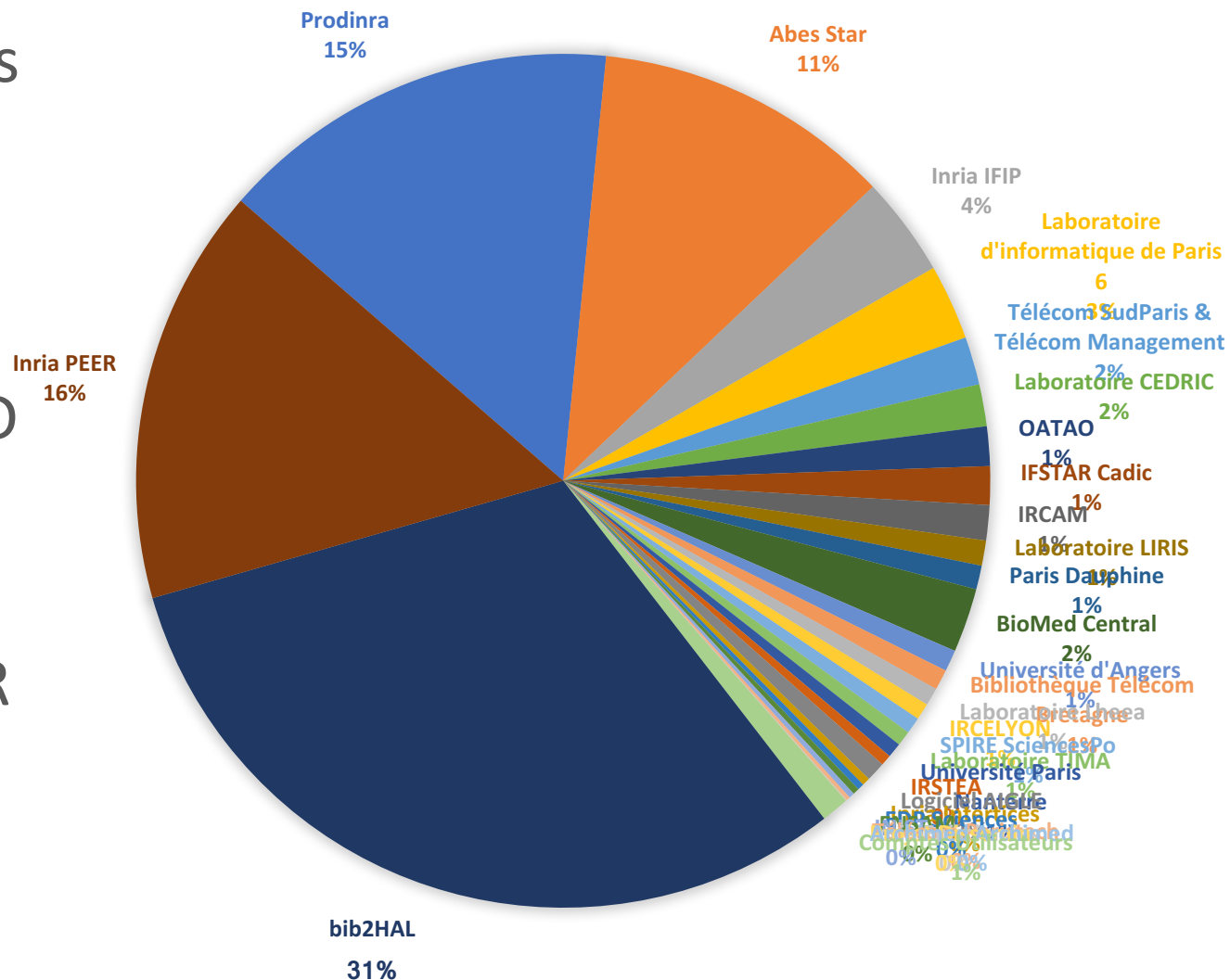
```

<org type="europeanProject" xml:id="localProjeurope-Identifiant du projet européen">
  <idno type="program">Programme</idno>
  <idno type="number">Numéro de projet</idno>
  <idno type="call">Code</idno>
  <orgName>Nom</orgName>
  <desc>Description</desc>
  <date type="start">Date de début</date>
  <date type="end">Date de fin</date>
</org>
  
```

IMPORT **SWORD**

SWORD dans HAL (mars 2017)

- 131 300 dépôts effectués en SWORD (dont 53 612 avec le texte intégral)
- Service SWORD mis en place dans le cadre du projet européen PEER
- 460 comptes utilisateurs



Simple Web-service Offering Repository Deposit

- Basé sur *Atom Publishing Protocol* (**APP**)
- Utilisé par d'autres AO & logiciels
 - arXiv, Dspace, Eprints, ...
- API REST HTTP
- Verbes HTTP :
 - **GET** : statut d'une ressource
 - **POST** : nouveau dépôt
 - **PUT** : modification des métadonnées / nouvelle version
 - **DELETE** : suppression d'une ressource

<http://swordapp.org/>

Entêtes HTTP (1)

- **Content-Type :**
 - [text/xml](#) pour le dépôt d'un fichier XML
 - [application/zip](#) pour le dépôt d'une archive ZIP contenant un fichier XML + les fichiers associés
- **Content-Disposition :**
 - [attachment; filename=tei.xml](#) à utiliser pour indiquer le nom du fichier XML dans le ZIP
- **Packaging :**
 - <http://purl.org/net/sword-types/AOfr> : format AOfr - TEI HAL
 - <http://jats.nlm.nih.gov/publishing/tag-library/> : format JATS
- **Content-MD5 :**
 - [Signature MD5](#) : vérifie l'intégrité du contenu envoyé

Entêtes HTTP (2)

- On-Behalf-Of : Dépôt pour le compte d'un utilisateur HAL
 - UID ou LOGIN du compte HAL
 - IDHAL ou ORCID (*prochainement*)
- Export-To-Arxiv
 - true / false : indique si le dépôt doit être transféré sur ArXiv
- Export-To-PMC
 - true / false : indique si le dépôt doit être transféré sur Pubmed Central
- Hide-For-RePEc
 - true / false : permet de cacher le dépôt du réservoir accessible à RePEc dans l'archive HAL
- Hide-In-OAI
 - true / false : permet de cacher le dépôt du réservoir OAI-PMH et du Sitemap
- X-Allow-Completion :
 - idext : récupération des métadonnées à partir d'un identifiant externe DOI, arXivID, ...
 - grobid : récupération des métadonnées à partir du PDF soumis via l'outil GROBID
 - affiliation : recherche des affiliations des auteurs en se basant sur le référentiel AURÉHAL

Statut d'une ressource

- GET `api.archives-ouvertes.fr/sword/[identifiant]`

```
curl -X GET -u login:pwd https://api.archives-ouvertes.fr/sword/hal-00000001v2
```

- Réponse
 - HTTP/1.1 200 OK
 - statut : `accept` | `verify` | `update` | `delete`

```
<?xml version="1.0" encoding="utf-8"?>  
<document id="hal-00000001" version="2">  
  <status>accept</status>  
  <comment></comment>  
</document>
```

Suppression d'une ressource

- DELETE [api.archives-ouvertes.fr/sword/\[identifiant\]](https://api.archives-ouvertes.fr/sword/[identifiant])

```
curl -X DELETE -u login:pwd https://api.archives-ouvertes.fr/sword/hal-01039627
```

- Réponse
 - Suppression OK : [HTTP/1.1 204 No Content](#)
 - Suppression NOK : [Code Erreur](#)

Dépôt d'une ressource (1)

- POST `api.archives-ouvertes.fr/sword/[portail]`
 - Liste des portails accessible via l'API : <https://api.archives-ouvertes.fr/ref/instance>
- Format pivot basé sur la TEI
 - <https://hal.archives-ouvertes.fr/documents/aofr.xsd>
- Dépôt d'un fichier XML (`Content-Type:text/xml`)
 - Notice ou Texte intégral non intégré au dépôt
- Dépôt d'une archive ZIP (`Content-Type:application/zip`)
 - Texte intégral : fichier(s) intégré(s) au dépôt

Dépôt d'une ressource (2)

```
curl -X POST -u login:pwd https://api.archives-ouvertes.fr/sword/hal -H "X-Packaging:http://purl.org/net/sword-types/AOfr" -H "Content-Type:text/xml" -data-binary @tei.xml
```

- Réponse
 - Dépôt accepté (passage direct en ligne) :
[HTTP/1.1 202 Accepted](#)
 - Dépôt créé (en modération dans HAL) :
[HTTP/1.1 201 Created](#)
 - Problème lors du dépôt :
[Code Erreur](#)

Dépôt d'une ressource (3)

```

<?xml version="1.0" encoding="utf-8"?>
<entry xmlns="http://www.w3.org/2005/Atom"
xmlns:sword="http://purl.org/net/sword/terms/"
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:hal="http://hal.archives-ouvertes.fr/">
  <title>Accepted media deposit to HAL</title>
  <id>hal-01040864</id>
  <hal:password>9322one</hal:password>
  <hal:version>1</hal:version>
  <updated>2017-03-28T08:56:23+02:00</updated>
  <summary>A media deposit was stored in the HAL workspace</summary>
  <sword:treatment>stored in HAL workspace</sword:treatment>
  <sword:userAgent>HAL SWORD API Server</sword:userAgent>
  <source>
    <generator uri="https://api.archives-ouvertes.fr/sword"
version="1.0">hal@ccsd.cnrs.fr</generator>
  </source>
  <link rel="alternate" href="https://hal.archives-ouvertes.fr/hal-01040864"/>
</entry>

```

Modification des métadonnées

- PUT [api.archives-ouvertes.fr/sword/\[identifiant\]](https://api.archives-ouvertes.fr/sword/[identifiant])

```
curl -X PUT -u login:pwd https://api.archives-ouvertes.fr/sword/hal-01040864 -H "X-Packaging:http://purl.org/net/sword-types/AOfr" -H "Content-Type:text/xml" --data-binary @new_tei.xml
```

- Réponse
 - Modification OK : [HTTP/1.1 200 OK](#)
 - Modification NOK : [Code Erreur](#)

Dépôt d'une nouvelle version

- Idem modification des métadonnées + ajout du texte intégral
- PUT [api.archives-ouvertes.fr/sword/\[identifiant\]](https://api.archives-ouvertes.fr/sword/[identifiant])

```
curl -X PUT -u login:pwd https://api.archives-ouvertes.fr/sword/hal-01040864 -H "X-Packaging:http://purl.org/net/sword-types/AOfr" -H "Content-Type:text/xml" --data-binary @new_tei.xml
```

- Réponse
 - Dépôt OK : [HTTP/1.1 201 Created](#)
 - Dépôt NOK : [Code Erreur](#)

Gestion des erreurs (1)

- HTTP/1.1 4XX
- 406 Not Acceptable
 - Packaging non reconnu
 - Content-type non reconnu
 - Erreur dans chargement du fichier XML
- 412 Precondition Failed
 - Problème dans la vérification du MD5
- 403 Forbidden
 - Problème d'authentification
- 405 Method Not Allowed
 - Utilisation d'un verbe HTTP non accepté
- 413 Request Entity Too Large
 - Taille du fichier supérieur à la limite (200Mo)
- 400 Bad Request
 - Erreur d'enregistrement

Gestion des erreurs (2)

- Xpath : `/sword:error/sword:verboseDescription`

```
<?xml version="1.0" encoding="utf-8"?>
<sword:error xmlns:sword="http://purl.org/net/sword/error/" xmlns="http://www.w3.org/2005/Atom"
href="http://purl.org/net/sword/error/ErrorException">
  <title>ERROR</title>
  <updated>2017-03-28T10:13:50+02:00</updated>
  <author>
    <name>HAL SWORD API Server</name>
  </author>
  <source>
    <generator uri="https://api.archives-ouvertes.fr/sword"
version="1.0">hal@ccsd.cnrs.fr</generator>
  </source>
  <summary>Some parameters sent with the request were not understood</summary>
  <sword:treatment>processing failed</sword:treatment>
  <sword:verboseDescription>{"meta":{"abstract":{"isEmpty":"Vous devez remplir ce
champ"}}}</sword:verboseDescription>
  <link rel="alternate" href="https://api.archives-ouvertes.fr" type="text/html"/>
</sword:error>
```

Liens utiles

- Schéma XML-TEI
 - <https://hal.archives-ouvertes.fr/documents/aofr.xsd>
- Exemples fichiers TEI
 - <https://github.com/CCSDForge/HAL>
- Documentation API SWORD
 - <https://api.archives-ouvertes.fr/docs/sword>