

Dépôt Orfeo

Import CSV : Mode d'emploi

Version "Déploiement"

19-03-2015

| | |
|---|---|
| Modèle CSV | 2 |
| Formatage du modèle | 2 |
| Entêtes | 2 |
| Import de texte intégral | 4 |
| Annexe A : Liste des sujets de Frascati | 7 |
| Annexe B: Types de publication | 8 |

L'import par lots des publications de recherche dans Orfeo se fait via un fichier CSV ("Comma Separated Value", valeurs séparées par une virgule), tant pour l'import initial des notices des cinq dernières années que pour les futurs imports.

Ce document constitue le mode d'emploi pour la constitution d'un modèle de fichier CSV qui servira à l'import de notices dans Orfeo.

Dans ce document, l'accent sera mis sur les aspects techniques de cet import/ Pour plus d'information sur le modèle de métadonnées, veuillez vous référer au Vademecum.

Modèle CSV

Le fichier CSV doit être composé de plusieurs lignes. La première ligne contient les entêtes, qui définissent les champs de métadonnées dans lesquels doivent être enregistrées les valeurs du fichier. La seconde ligne, et toutes les suivantes, se retrouvent les valeurs pour ces différents champs. Chaque ligne représente les métadonnées d'une notice.

Un exemple d'un modèle de fichier CSV peut être trouvé dans un fichier séparé, "csv-template.csv". Ce fichier se compose de deux lignes : une ligne contenant les entêtes, et un exemple de notice à importer.

Formatage du modèle

Le modèle CSV répond à la structure suivante :

- Chaque valeur est encadrée par des guillemets : "Valeur de la métadonnée"
- Les champs de métadonnées sont séparés par des virgules : "métadonnée 1", "métadonnée 2"
- Pour attribuer plusieurs valeurs à un même champ, une double barre verticale peut être utilisée : "métadonnée 1||métadonnée 2"

Exemple avec deux notices de recherche

Entête A, Entête B, Entête C

"Notice 1 métadonnée A1|| Notice 1 métadonnée A2", "Notice 1 métadonnée B", "Notice 1 métadonnée C"

"Notice 2 métadonnée A", "Notice 2 métadonnée B", "Notice 2 métadonnée C1|| Notice 2 métadonnée C1"

Attention : La première notice (seconde ligne) contient deux valeurs pour le champ "Entête A". La seconde notice (troisième ligne) contient deux valeurs pour la métadonnée "Entête C".

Entêtes

Les champs de métadonnées sont définis dans l'entête, à savoir la première ligne du fichier CSV. Ces champs doivent être utilisés conformément à leur définition dans Orfeo. Attention, le nom des champs diffère du libellé qui leur est attribué dans l'interface d'Orfeo. Dans le fichier CSV, le nom des champs, et non leur libellé, doit être renseigné. La table ci-dessous donne un aperçu du mapping entre le nom du champ et son libellé. Les champs obligatoires dans Orfeo sont affichés en caractères gras.

| Nom | Libellé | Remarque |
|------------------------------|-----------------------------------|--|
| dc.title | Title | |
| dc.contributor.author | Creator | Nom, prénom |
| dc.contributor | Contributor | Nom, prénom |
| dc.contributor.editor | Editor | Nom, prénom |
| dc.subject.frascati | Subject (Frascati) | Sujet de la liste de Frascati (voir annexe A) |
| dc.subject.free | Subject (Vrije keuze) | |
| dc.description | Description | |
| dc.publisher | Publisher | |
| dc.date | Date | AAAA |
| dc.type | Type | Liste des types de publication (voir annexe B) |
| Orfeo.peerreviewed | Peer-Review | Valeurs possibles : <ul style="list-style-type: none"> • Yes • No • Not pertinent (valeur par défaut si laissé vide) |
| dc.identifieur.isbn | Identifieur: ISBN | |
| dc.identifieur.issn | Identifieur: ISSN | |
| dc.identifieur.publisherlink | Identifieur: Publisherlink | |
| dc.identifieur.doi | Identifieur: DOI | |
| dc.identifieur.isi | Identifieur: ISI | |
| dc.identifieur.pubmed | Identifieur: Pubmed | |
| dc.identifieur.url | Identifieur: URL | |
| dc.identifieur.scopus | Identifieur: Scopus | |
| dc.identifieur.worldcat | Identifieur: Worldcat | |
| dc.identifieur.limo | Identifieur: Limo | |
| dc.identifieur.depositnumber | Identifieur: Legal deposit number | |
| dc.identifieur.rmca | Interne identifieur: RMCA | |
| dc.identifieur.belspo | Interne identifieur: Belspo | |
| dc.source.title | Source: Title | |

| Nom | Libellé | Remarque |
|----------------------------|---------------------|--|
| dc.source.volume | Source: Volume | |
| dc.source.issue | Source: Issue | |
| dc.source.page | Source: Page | |
| dc.source.editor | Source: Editor | |
| dc.relation.ispartofseries | Source: Series | |
| dc.identifier.citation | Citation | |
| dc.language | Language | Valeurs possibles : <ul style="list-style-type: none"> • nld • eng • fra • por • spa • ita • deu • N/A |
| dc.relation.project | Relation | |
| dc.audience | Audience | Valeurs possibles : <ul style="list-style-type: none"> • Scientific • Educational • General Public • Policy-Oriented |
| dc.coverage.spatial | Coverage (spatial) | |
| dc.coverage.temporal | Coverage (temporal) | |
| file | | Ce champ est utilisé pour l'ajout de fichiers de texte intégral. Plus d'information à ce sujet sous la rubrique "Import de texte intégral". |

L'ordre des champs dans l'entête est libre. Il est par contre indispensable de respecter le même ordre pour toutes les notices dans les lignes suivantes.

Exemple :

dc.title, dc.contributor.authors
"Titre", "Smith, John|Smith, Jane"

aura le même effet que :

dc.contributor.authors, dc.title
"Smith, John|Smith, Jane", "Titre"

Import de texte intégral

Le texte intégral, ainsi que toutes les informations y relatives, doit être importé par l'intermédiaire du champ "file". Ce champ doit présenter la structure suivante :

<chemin_vers_le_fichier>::version:<version>::accessrights:<permissions>,

où les données entre les symboles “<” et “>” doivent être complétées.

Chemin vers le fichier

Il existe deux façons d’importer des documents de texte intégral dans Orfeo :

1. via des URL pointant vers un emplacement publiquement accessible auquel le texte intégral peut être téléchargé
2. via un fichier ZIP qui contient tous les fichiers de texte intégral

1. Import via une URL

Lorsque l’import se fait directement via une URL, l’URL renseignée dans le champ doit pointer directement vers le fichier, et non vers une page via laquelle le fichier peut être généré en cliquant sur un lien ou un bouton.

2. Import via un fichier ZIP

Lorsque l’import se fait via un fichier d’archive ZIP, la colonne “file” doit être complétée avec l’emplacement du fichier de texte intégral dans la structure du fichier ZIP, à la place de l’URL.

Exemple :

Voici un exemple de structure pour un fichier ZIP :

- Mainfolder
 - Subfolder_1
 - Full-Text_1
 - Subfolder_2
 - Subfolder_3
 - Full-Text_2
 - Full-Text_3
 - Full-Text_4

La valeur du champ <chemin_vers_le_fichier> pour chaque fichier sera, dans ce cas :

- Full-Text_1: Mainfolder/Subfolder_1/Full-Text_1
- Full-Text_2: Mainfolder/Subfolder_2/Subfolder_3/Full-Text_2
- Full-Text_3: Mainfolder/Full-Text_3
- Full-Text_4: Mainfolder/Full-Text_4

Afin de s’assurer que le fichier correct est importé, il est indispensable que le chemin vers ce fichier soit unique au sein de l’archive ZIP.

Syntaxe du champ “file”

La table ci-dessous détaille la façon dont ce champ doit être renseigné.

| Élément du champ | Contenu | Exemple |
|-----------------------------|---|--|
| <chemin_vers_le_fichier> | URL ou chemin vers le fichier. L'information doit pointer directement vers le fichier. | <u>URL:</u> https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1 <u>Archive Zip :</u> Mainfolder/Subfolder_1/Full-Text_1 |
| version:<version> | Valeurs possibles : <ul style="list-style-type: none"> • Published (valeur par défaut lorsque laissé vide) • Pre-print • Post-print | version:Published |
| accessrights:<permissions > | Valeurs possibles : <ul style="list-style-type: none"> • public access (valeur par défaut lorsque laissé vide) • restricted access • under embargo: AAAA-MM-JJ | accessrights:public access |

Tant les paramètres "version" que "accessrights" ont une valeur par défaut. Ils n'est donc pas obligatoire de les renseigner. Le séparateur entre les paramètres est "::" (double double-point).

Exemple de champs "file" contenant une URL :

- Tous les paramètres sont renseignés :
"https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1::version:Published::accessrights:public access"
- Seule l'URL est renseignée, les paramètres "version" et "accessrights" sont laissés par défaut :
"https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1"

Lorsque seule l'URL est renseignée, le fichier sera donc automatiquement publié en libre accès, et la version qualifiée de "Published".

Exemples de champs "file" se référant à une archive ZIP :

- Tous les paramètres sont renseignés :
"Mainfolder/Subfolder_1/Full-Text_1::version:Published::accessrights:public access"
- Seul le chemin dans l'archive est renseigné, les paramètres "version" et "accessrights" sont laissés par défaut :
"Mainfolder/Subfolder_1/Full-Text_1"

A nouveau, dans ce dernier cas, le fichier sera automatiquement publié en libre accès, et la version qualifiée de "Published".

Annexe A : Liste des sujets de Frascati

Les sujets doivent être sélectionnés parmi cette liste. Le(s) sujet(s) sélectionné(s) peu(ven)t appartenir tant au premier qu'au second niveau. Lorsque possible, il est conseillé d'être le plus spécifique possible, et donc de privilégier la sélection de sujets du second niveau.

- **Natural sciences**
 - Mathematics
 - Computer and information sciences
 - Physical sciences
 - Chemical sciences
 - Earth and related Environmental sciences
 - Biological sciences
- **Engineering and technology**
 - Civil engineering
 - Electrical engineering, Electronic engineering, Information engineering
 - Mechanical engineering
 - Chemical engineering
 - Materials engineering
 - Medical engineering
 - Environmental engineering
 - Environmental biotechnology
 - Industrial biotechnology
 - Nano-technology
- **Medical and Health sciences**
 - Basic medicine
 - Clinical medicine
 - Health sciences
 - Medical biotechnology
- **Agricultural sciences**
 - Agriculture, Forestry and Fisheries
 - Animal and Dairy science
 - Veterinary science
 - Agricultural biotechnology
- **Social sciences**
 - Psychology
 - Economics and Business
 - Educational sciences
 - Sociology
 - Law
 - Political science
 - Social and economic geography
 - Media and communications
- **Humanities**
 - History and Archaeology
 - Languages and Literature
 - Philosophy, Ethics and Religion
 - Arts

Annexe B: Types de publication

- Article
- Book
- Book chapter
- Catalog
- Conference
- Report
- Bachelorthesis
- Masterthesis
- Doctoral thesis
- Cartographic document
- Review
- Working paper
- Lecture