

# Orfeo repository CSV Import: Instructies Rollout versie 19-03-2015

|   |   |
|---|---|
| CSV template                              | 2 |
| Formattering van de template              | 2 |
| Template headers                          | 2 |
| Aanlevering full-text                     | 4 |
| Bijlage A: Frascati lijst van onderwerpen | 7 |
| Bijlage B: Publicatie types               | 8 |

De bulk import van de onderzoeksitems in Orfeo zal gebeuren via een CSV ('Comma Separated Value') bestand, zowel voor de initiële import van de voorbije 5 jaar als voor toekomstige incrementele imports.

Dit bestand vormt de leidraad bij de CSV template die gebruikt zal worden bij import in Orfeo.

In dit bestand wordt gefocust op het technische aspect van de import. Voor meer informatie over het metadatamodel van Orfeo, zie het Vademecum.

## CSV template

De CSV template bestaat uit twee soorten lijnen. Op de eerste lijn worden de 'headers' gedefinieerd die de metadatavelden aangeven waarin de data van de onderzoeksitems opgeslagen zullen worden. Op de tweede en al de volgende lijnen worden de metadatawaarden van de onderzoeksitems gegeven. Elke lijn representeert hierbij de metadata van één onderzoeksitem.

Een voorbeeld van de CSV template kan gevonden worden in een aparte file genaamd csv-template.csv. Deze template bestaat uit twee lijnen: één lijn met de headers en één lijn met een voorbeeld van een onderzoeksitem.

## Formattering van de template

De CSV template gebruikt de volgende structuur om de metadata te structureren:

- Rond elke waarde van metadata staan aanhalingstekens: "metadatawaarde"
- De metadatavelden worden gescheiden met een komma: "metadataveld 1","metadataveld 2"
- Om meerdere waarden in eenzelfde veld aan te geven wordt een dubbele pijp gebruikt: "metadatawaarde 1|| metadatawaarde 2"

Voorbeeld met twee onderzoeksitems:

Header A, Header B, Header C

"Item 1 metadatawaarde A1|| Item 1 metadatawaarde A2", "Item 1 metadatawaarde B", "Item 1 metadatawaarde C"

"Item 2 metadatawaarde A", "Item 2 metadatawaarde B", "Item 2 metadatawaarde C1||Item 2 metadatawaarde C1"

Merk op: Het eerste onderzoeksitem op lijn 2 heeft 2 metadatawaarden in het veld 'Header A'. Het onderzoeksitem op lijn 3 heeft 2 metadatawaarden in het veld 'Header C'.

## Template headers

In de CSV template header worden de metadatavelden gedefinieerd. In deze header moeten de velden gedefinieerd worden zoals deze in Orfeo opgeslagen worden. De naam van de headers verschilt hierdoor van de labels die in het vademecum gebruikt worden en in de interface van Orfeo. In de onderstaande tabel wordt een overzicht gegeven van de mapping tussen de headers en de Orfeo metadatavelden. De velden die verplicht zijn in Orfeo staan in vet.

| Template header              | Orfeo veld     | Opmerkingen          |
|------------------------------|----------------|----------------------|
| <b>dc.title</b>              | <b>Title</b>   |                      |
| <b>dc.contributor.author</b> | <b>Creator</b> | Achternaam, voornaam |

| Template header             | Orfeo veld                       | Opmerkingen   |
|-----------------------------|----------------------------------|---|
| dc.contributor              | Contributor                      | Achternaam, voornaam  |
| dc.contributor.editor       | Editor                           | Achternaam, voornaam  |
| <b>dc.subject.frascati</b>  | <b>Subject (Frascati)</b>        | Lijst van Frascati onderwerpen (zie bijlage A)  |
| dc.subject.free             | Subject (Vrije keuze)            |   |
| <b>dc.description</b>       | <b>Description</b>               |   |
| dc.publisher                | Publisher                        |   |
| <b>dc.date</b>              | <b>Date</b>                      | JJJJ  |
| <b>dc.type</b>              | <b>Type</b>                      | Lijst van publicatietypes (zie bijlage B)   |
| <b>Orfeo.peerreviewed</b>   | <b>Peer-Review</b>               | Mogelijke waarden: <ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> <li>• <b>Not pertinent (default wanneer niet aangeleverd)</b></li> </ul> |
| dc.identifier.isbn          | Identifier: ISBN                 |   |
| dc.identifier.issn          | Identifier: ISSN                 |   |
| dc.identifier.publisherlink | Identifier: Publisherlink        |   |
| dc.identifier.doi           | Identifier: DOI                  |   |
| dc.identifier.isi           | Identifier: ISI                  |   |
| dc.identifier.pubmed        | Identifier: Pubmed               |   |
| dc.identifier.url           | Identifier: URL                  |   |
| dc.identifier.scopus        | Identifier: Scopus               |   |
| dc.identifier.worldcat      | Identifier: Worldcat             |   |
| dc.identifier.limo          | Identifier: Limo                 |   |
| dc.identifier.depositnumber | Identifier: Legal deposit number |   |
| dc.identifier.rmca          | Interne identifier: RMCA         |   |
| dc.identifier.belspo        | Interne identifier: Belspo       |   |
| dc.source.title             | Source: Title                    |   |
| dc.source.volume            | Source: Volume                   |   |
| dc.source.issue             | Source: Issue                    |   |

| Template header            | Orfeo veld          | Opmerkingen   |
|----------------------------|---------------------|---|
| dc.source.page             | Source: Page        |   |
| dc.source.editor           | Source: Editor      |   |
| dc.relation.ispartofseries | Source: Series      |   |
| dc.identifier.citation     | Citation            |   |
| <b>dc.language</b>         | <b>Language</b>     | Mogelijke waarden: <ul style="list-style-type: none"> <li>• nld</li> <li>• eng</li> <li>• fra</li> <li>• por</li> <li>• spa</li> <li>• ita</li> <li>• deu</li> <li>• N/A</li> </ul> |
| dc.relation.project        | Relation            |   |
| <b>dc.audience</b>         | <b>Audience</b>     | Mogelijke waarden: <ul style="list-style-type: none"> <li>• Scientific</li> <li>• Educational</li> <li>• General Public</li> <li>• Policy-Oriented</li> </ul>                       |
| dc.coverage.spatial        | Coverage (spatial)  |   |
| dc.coverage.temporal       | Coverage (temporal) |   |
| file                       |                     | Dit veld wordt gebruikt voor de aanlevering van de full-text. Meer informatie kan gevonden worden in de sectie "Aanlevering full-text".   |

De volgorde van de 'headers' in de CSV is vrij te kiezen. Het is hierbij belangrijk dat de metadata van de onderzoeksitems op de volgende lijnen deze volgorde respecteert.

Voorbeeld:

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

zal resulteren in hetzelfde onderzoeksitem als:

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

## Aanlevering full-text

De full-text en alle informatie omtrent deze full-text zal aangeleverd worden onder de header 'file'. Dit veld heeft de volgende syntax:

<locatie van full-text>::version:<version>::accessrights:<accessrights>,

waarbij de waarden ingevuld moeten worden in <...>.

## Locatie van de full-text

Er zijn twee manieren om full-texten aan te leveren in Orfeo:

1. Via urls naar een publiek toegankelijke locatie waar de full-text gedownload kan worden.
2. Via een zip file die alle full-texten bevat.

### 1. Aanlevering via url

Wanneer de files aangeleverd worden via de url, moet deze url rechtstreeks naar de file wijzen en bijvoorbeeld niet naar een pagina waar op een knop gedruwd moet worden om de file te downloaden.

### 2. Aanlevering via een zip file

Naast het gebruik van urls kunnen de full-texten die bij de onderzoeksitems horen, ook aangeleverd worden via een zip file. In dit geval moet er in de 'file' kolom verwezen worden naar de locatie van de full-text binnen de zip file, in plaats van naar de url.

Voorbeeld:

Hieronder wordt een voorbeeld gegeven van een mogelijke structuur van een zip file:

- Mainfolder
  - Subfolder\_1
    - Full-Text\_1
  - Subfolder\_2
    - Subfolder\_3
      - Full-Text\_2
  - Full-Text\_3
  - Full-Text\_4

De paden naar de verschillende full-texten, die aangegeven kunnen worden in de file kolom, zijn dan:

- 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

Om de correcte full-text te kunnen halen uit de zip file, dient elk pad naar een full-text uniek te zijn.

## Syntax van het file veld

In de onderstaande tabel staat verder uitgelegd hoe dit veld gebruikt moet worden.

| 'File' onderdelen           | Inhoud van <...>   | Voorbeeld  |
|-----------------------------|--|--|
| <locatie van full-text >    | De url of het pad naar de locatie van de full-text. Deze moet rechtstreeks naar de file verwijzen.   | <u>URL:</u><br><a href="https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1">https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1</a><br><br><u>Zip file:</u><br>Mainfolder/Subfolder_1/Full-Text_1 |
| version:<version>           | Mogelijke waarden: <ul style="list-style-type: none"> <li>• <b>Published (Default wanneer niet aangeleverd)</b></li> <li>• Pre-print</li> <li>• Post-print</li> </ul>                            | version:Published  |
| accessrights:<accessrights> | Mogelijke waarden: <ul style="list-style-type: none"> <li>• <b>public access (Default wanneer niet aangeleverd)</b></li> <li>• restricted access</li> <li>• under embargo: JJJJ-MM-DD</li> </ul> | accessrights:public access   |

Zowel 'version:<version>' als 'accessrights:<accessrights>' hebben een default waarde en hoeven dus niet aangeleverd te worden. De separator tussen de verschillende onderdelen van het 'file'-veld is '::'.

#### Voorbeeld van het 'file' veld waarbij urls gebruikt worden

- Alle velden zijn ingevuld:  
<https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1>::version:Published::accessrights:public access”
- Enkel de url is ingevuld, voor de versie en de access rights mogen de default waarden gebruikt worden:  
<https://openknowledge.worldbank.org/bitstream/handle/10986/20483/DB15-Full-Report.pdf?sequence=1>”

Enkel de url aanleveren onder de header 'file' zal dus resulteren in een full-text met public access rechten en version 'Published'.

#### Voorbeeld van het 'file' veld waarbij een zip file gebruikt wordt

- Alle velden zijn ingevuld:  
Mainfolder/Subfolder\_1/Full-Text\_1::version:Published::accessrights:public access”
- Enkel het pad naar de full-text is ingevuld, voor de versie en de access rights mogen de default waarden gebruikt worden:  
Mainfolder/Subfolder\_1/Full-Text\_1”

In dit geval zal de full-text 'public access' rechten en version 'Published' krijgen.

## Bijlage A: Frascati lijst van onderwerpen

De onderwerpen dienen gekozen te worden uit de volgende lijst. Onderwerpen kunnen zowel uit het tweede niveau als het eerste niveau gekozen worden. Indien mogelijk vragen we om het onderwerp zo specifiek mogelijk te kiezen.

- **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

## Bijlage B: Publicatie types

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