

SUR+ OA:

Configurazione del processo di import

Versione: 0.6 (draft) 25 agosto 2010

Autore: Andrea Bollini



1. Premessa

E' possibile configurare un mapping dei dati recuperati da banche dati esterne verso i metadati specifici dell'archivio OA, per ognuna delle fonti da cui viene richiesto di effettuare un import.

2. Tipologia dei documenti da importare

Alcune basi dati, come Crossref, espongono record di differenti tipologie (articoli, book, etc.) dandone indicazione in un metadato specifico. In questi casi, per ogni tipologia di materiale, è necessario indicare il form name corrispondente alla collezione di import in SUR+ OA. Se per una data tipologia non è indicato alcun form name o il form name non è valido, il sistema non consentirà l'import di record di tali tipologie. Se per una data tipologia non è presente una indicazione esplicita verrà utilizzata una collezione di default. Quest'ultima potrà anche non essere definita, per consentire di impostare il default a "tipologia non importabile".

3. Operazioni sui metadati

Prima dell'inserimento di un record in archivio vengono eseguite delle operazioni di arricchimento/trasformazione dei metadati che possono essere configurate sulle singole esigenze dell'archivio e forniscono punti di estensione per eventuali personalizzazioni aggiuntive.

In particolare esistono tre tipologie di operazioni/plugin:

- Converter: trasformazione di un singolo valore testuale in una forma equivalente (ad es., per la conversione di liste controllate)
- Enhance: aggiunta di nuovi metadati a partire dai metadati forniti dalla banca dati (ad es., costruzione delle stringhe allauthors e alleditors)
- Sanitize: trasformazione finale dei metadati prima dell'inserimento in archivio per rispondere a scelte di qualità dei dati (ad es., rimozione dei singoli autori in archivi SUR+ OA se non agganciati a SUR+ GA, appiattimento in un unico metadato dei subject, etc.)

4. Plugin

L'infrastruttura di import fornirà i seguenti plugin:

- Converter: trasforma un singolo valore testuale in una forma equivalente (ad es., per la conversione di liste controllate)
- Enhance: aggiunge nuovi metadati a partire dai metadati forniti dalla banca dati (ad es., costruzione delle stringhe allauthors e alleditors)
- Sanitize: trasforma i metadati prima dell'inserimento in archivio per rispondere a scelte di qualità dei dati (ad es., rimozione dei singoli autori in archivi SUR+ OA se non agganciati a SUR+ GA, appiattimento in un unico metadato dei subject, etc.)
- mapConverter[indicato in configurazione tramite il nome della mappa di conversione]: sostituisce un valore sulla base di un file di corrispondenza chiave valore. Se il valore di origine non è trovato

SUR+ OA: Configurazione del processo di import

nella mappa viene utilizzata la configurazione associata alla key “mapConverter.default”. I valori vuoti non vengono salvati

- converter removedot: elimina l'eventuale punto finale di una stringa, come ad esempio nei titoli provenienti da PubMed
- allnames: genera una stringa complessiva “Cognome, Iniziali; Cognome, Iniziali; ...”. Il metadato da cui vengono letti i singoli nomi può essere configurato (lo stesso plugin “lavora” quindi su author e editor)
- countnames: conta il numero di valori presenti per un dato metadato
- drop: elimina i metadati specificati
- allkeywords: concatena tutti i valori di un metadato specificato in una unica stringa della forma “valore1; valore2; ...; valoreN”. I metadati singoli vengono rimossi e viene salvata la stringa complessiva

5. Mapping dei metadati

- **CROSSREF**

Le tipologie di materiali vengono espresse come metadato itemType.

itemType	Valore singolo - Lista controllata	journal_title journal_issue journal_volume journal_article conference_title conference_series conference_paper book_title book_series book_content component dissertation report-paper_title report-paper_series report-paper_content standard_title standard_series standard_content
doi	Valore singolo – testo libero	
issn	Valore singolo – testo libero	
eissn	Valore singolo – testo libero	
isbn	Valore singolo – testo libero	Presente solo per le tipologie book* e conference*
seriesTitle	Valore singolo – testo libero	Presente solo per le tipologie book* e conference*
journalTitle	Valore singolo – testo libero	

SUR+ OA: Configurazione del processo di import

volumeTitle	Valore singolo – testo libero	Presente solo per le tipologie book* e conference*
authors	Valore ripetibile – Cognome, Nome	
editors	Valore ripetibile – Cognome, Nome	
translators	Valore ripetibile – Cognome, Nome	
chairs	Valore ripetibile – Cognome, Nome	
volume	Valore singolo – testo libero	
issue	Valore singolo – testo libero	
firstPage	Valore singolo – testo libero	
lastPage	Valore singolo – testo libero	
editionNumber	Valore singolo – testo libero	Presente solo per le tipologie book* e conference*
year	Valore singolo – testo libero	
publicationType	Valore singolo – Lista controllata	
articleTitle	Valore singolo – testo libero	

- PUBMED

Le tipologie di materiali vengono mappati su una singola collezione (i.e. Articolo su rivista).

pubmedID	Valore singolo – testo libero	
doi	Valore singolo – testo libero	
issn	Valore singolo – testo libero	
eissn	Valore singolo – testo libero	
journalTitle	Valore singolo – testo libero	
title	Valore singolo – testo libero	
publicationModel	Valore singolo – Lista controllata	Print Print-Electronic Electronic Electronic-Print

SUR+ OA: Configurazione del processo di import

year	Valore singolo – testo libero (yyyy[-mm-dd])	
volume	Valore singolo – testo libero	
issue	Valore singolo – testo libero	
language	Valore singolo – testo libero (codice 3 lettere)	
type	Valore ripetibile – testo libero	Questi i valori emersi durante i test: Journal Article (caso frequente), Case Reports, Comparative Study, Evaluation Studies,
primaryKeywords	Valore singolo – testo libero	
secondaryKeywords	Valore singolo – testo libero	
primaryMeshHeadings	Valore singolo – testo libero (proveniente dal mesh)	
secondaryMeshHeadings	Valore ripetibile – testo libero (proveniente dal mesh)	
startPage	Valore singolo – testo libero	
endPage	Valore singolo – testo libero	
summary	Valore singolo – testo libero	
status	Valore singolo – Lista controllata	received accepted epublish ppublish revised aheadofprint retracted
authors	Valore ripetibile – Cognome, Nome	

6. Configurazione

-- CROSSREF

```
target_collection.default =  
target_collection.journal_article = Articolo su rivista  
target_collection.book_content = Contributo in libro  
target_collection.book_title = Monografia  
target_collection.unspecified = Monografia  
doi = dc.identifier.doi  
itemType =  
issn = dc.relation.issn  
eissn = dc.relation.eissn
```

SUR+ OA: Configurazione del processo di import

```
isbn = dc.relation.isbn
seriesTitle = dc.relation.ispartof
journalTitle = dc.relation.ispartof
volumeTitle = dc.relation.ispartof
authors = dc.contributor.author
editors = dc.contributor.editor
translators = dc.contributor.translator
chairs =
volume = dc.relation.volume
issue = dc.relation.issue
firstPage = dc.relation.firstpage
lastPage = dc.relation.lastpage
editionNumber =
year = dc.date.issued
publicationType =
articleTitle = dc.title
Articolo\ su\ rivista.seriesTitle =
Articolo\ su\ rivista.volumeTitle =
Articolo\ su\ rivista.tranlators =
Contributo\ in\ libro.tranlators =
Monografia.volumeTitle = dc.title
Monografia.tranlators =
Monografia.articleTitle =

enhance.dc.description.allauthors = allnames
enhance.dc.description.alleditors = allnames
enhance.dc.description.numberofauthors = countnames
sanitize.dc.contributor.author = drop
sanitize.dc.contributor.editor = drop

-- PUBMED

target_collection.default = Articolo su rivista
pubmedID = dc.identifier
doi = dc.identifier.doi
issn = dc.relation.issn
eissn = dc.relation.eissn
journalTitle = dc.relation.ispartof
title = dc.title
pubblicationModel =
year = dc.date.issued
volume = dc.relation.volume
issue = dc.relation.issue
language = dc.language.iso
type = dc.type.contribution
primaryKeywords = dc.subject
secondaryKeywords = dc.subject
primaryMeshHeadings = dc.subject
secondaryMeshHeadings = dc.subject
startPage = dc.relation.firstpage
endPage = dc.relation.lastpage
summary = dc.description.abstract
status = dc.type.publicationstatus (pubstatusPubmed)
authors = dc.contributor.author

enhance.dc.description.allauthors = allnames
enhance.dc.description.numberofauthors = countnames
sanitize.dc.subject = allkeywords
sanitize.dc.contributor.author = drop
```

SUR+ OA: Configurazione del processo di import

-- Configurazione Converter pubstatusPubmed

```
mapConverter.default = Sottoposto a rivista  
ppublish = Pubblicato  
epublish = Pubblicato
```

-- Configurazione Plugin Enhance countnames e allnames

```
AllNamesEnhanceMetadata.allauthors = dc.contributor.author  
AllNamesEnhanceMetadata.alleditors = dc.contributor.editor  
CountNamesEnhanceMetadata.numberofauthors = dc.contributor.author
```

7. Attivazione di un import

Per effettuare l'import, l'utente dovrà loggarsi in OA ed accedere alla propria pagina MyDSpace. Qui troverà un pulsante di import. Il sistema fornisce la possibilità di inserire uno o più id/doi contemporaneamente e un dropdown per la scelta del formato. Una volta attivato il processo, la/le pubblicazioni importate saranno visibili nell'elenco delle pubblicazioni dell'utente. Entrando in modifica del documento, i metadati importati saranno già compilati nelle varie pagine della submission. L'utente potrà quindi completare i metadati mancanti e terminare la submission.