Digital Commons®

DC Harvesting Tool: Step-by-Step Guide

Contents

Introduction

This guide describes how to use the Digital Commons Harvesting Tool to facilitate the addition of faculty publications from other sources into the repository.

The DC Harvesting Tool integrates data from ORCID, PubMed, Scopus, and Pure—with more sources planned for inclusion—and it features Sherpa-Romeo integration for simplified permissions checking.

Data retrieved for relevant faculty works is automatically mapped to Digital Commons schemas when you export your search results. The result is a prepopulated spreadsheet that's easy to ingest via batch import to a faculty publications series—or any DC publication structure where you want to showcase faculty work.

For an introductory overview of the tool and basic workflow, see Digital Commons Harvesting Tool: Automatically Populating the IR with Faculty Records.

If you have questions or need assistance with any of the steps in this guide: Please contact Consulting Services at dc-support@elsevier.com or 510-665-1200, option 2, weekdays 6:30 a.m.–7:30 p.m. North America Pacific Time.

Available Sources and Integrations

The DC Harvesting Tool integrates APIs from the following harvesting sources: Scopus, Pure, ORCID, and PubMed. In addition, integration with Sherpa-Romeo provides the option to include journal permissions information when harvesting from any source.

Some sources are available to Digital Commons subscribers by default and others require a subscription or additional setup, as noted below.

PubMed and **ORCID** harvesting are accessible to all users of the DC Harvesting Tool. It is not required that your institution be an ORCID institutional member.

Scopus data is available to Scopus subscribers with a one-time setup or as a Scopus add-on module for the DC Harvesting Tool. Scopus indexes over 25K publications including journals, conference proceedings, books, book series, and trade publications, from over 5000 different publishers.

Pure integration gives you the ability to showcase existing records without requiring you or faculty to resubmit to DC. To access Pure data within the DC Harvesting Tool, your institution needs to be a Pure subscriber, and you will need to provide us with an API key for a one-time setup. See Pure API key generation steps for instructions.

Sherpa-Romeo integration works with every content source in the Harvesting Tool and is available to all users as an option when exporting results. It provides reference information about the journal for each work as well as the specific details of the publisher's policies for posting different article versions in repositories.

Preparing to Use the Harvesting Tool

Before using the DC Harvesting Tool, it's helpful to review your strategy for organizing faculty publications in the IR and to choose an optimal harvesting workflow.

Setting up publication structures for faculty works:

The two most common strategies for managing faculty publications in Digital Commons are:

A. Upload to separate faculty publications series for each department/school/faculty OR
B. Upload to a single "bucket" faculty publications series; in addition, create departmental series and use the Collection Tool to set up auto-collect filters based on metadata criteria

Harvesting strategies for different approaches:

If you upload to separate faculty publications series for each department (option A above):

- Export from the Harvesting Tool to each separate faculty publications series and use the batch import tool in each series to add the works.
- Another approach is to generate one big spreadsheet for a single publication and then separate it out manually for the different series, adjusting for any differences in metadata fields between them.
- Option A may work best for institutions with decentralized series administrators or where there is a preference for department-specific labels in article-level URLs (these inherit from the series where the article is uploaded).

If you upload to a bucket series, and then auto-collect out to departmental series (option B above):

- Export from the Harvesting Tool to the bucket series and use its batch import tool to add the works to that single series. Then use the Collection Tool to auto-collect works to their respective departmental series using metadata criteria. See The Collection Tool guide for more about using auto-collect filters.
- Option B may work best for institutions with centralized IR administration.

With either of the above options, you can elect to use export sheets from the Harvesting Tool as a worklist and manually use the submission form in those series where you want to add faculty publications.

Checking metadata fields:

To take full advantage of the Harvesting Tool's automatic mapping between source metadata and Digital Commons metadata fields, you may want to review the list of mappings for each source included in this guide.

You may also want to review what fields you have enabled, and made required, for your publications/series. Any source data that you'll want to capture will need a corresponding default or custom metadata field in the target Digital Commons publication structure.

If the fields are present in the DC structure before using the Harvesting Tool (as is the case with most default fields), the generated batch import spreadsheet will be able to map data to those fields.

If not all fields that you wish to capture are present in the DC publication structure, your consultant can add them for you. Another approach is to select the option in the Harvesting Tool to include additional unmapped metadata. This provides an opportunity to work out metadata mappings for additional fields manually, as described in the section, "Working with Metadata in the Export Spreadsheet."

If you have any questions or would like recommendations on how the Harvesting Tool can help with your institution's specific needs, please contact your consultant for more information.

Prepping IDs for Pure harvesting:

If you or your team also manage the Pure instance at your institution, you may find it helpful to prep tables of the Author IDs and Organizational Unit IDs from Pure to help facilitate searching in the DC Harvesting Tool.

DC teams who don't work with Pure may want to ask HR, the research office, or another department that works with Pure for a list of Author IDs and/or Organizational Unit IDs as tables to facilitate searching.

Access the Harvesting Tool

The DC Harvesting Tool is accessible via your My Account page in Digital Commons. Click the **Harvesting Tool** link under Site Administration Tools to access the tool.
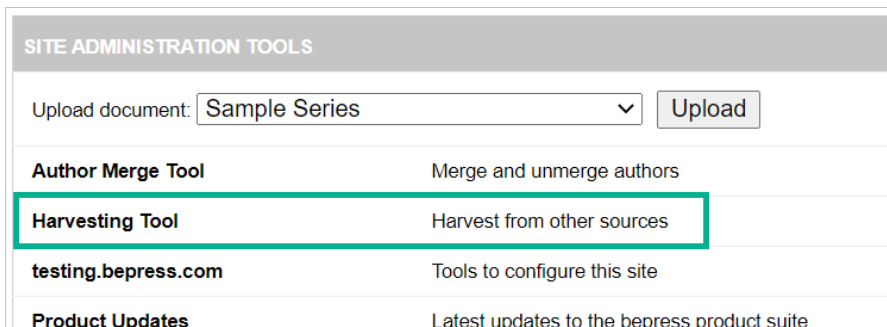
You need to be a DC site-level administrator with the "Can harvest content from 3rd parties" permission in order to see the **Harvesting Tool** link.

If you have the ability to modify other administrators' permissions (with the "Create administrators" permission), you can assign the Harvesting Tool permission to team members.

Refer to the Managing Administrator Permissions guide for more about permissions, and please contact Consulting Services if you have any questions.

Search Source Data

The first step in the Harvesting Tool is to perform a search, which accesses source data for relevant records by a particular author and/or affiliation. Selecting a source on the Harvest Search screen will display search fields unique to that source.

All search inputs use the "AND" Boolean operator by default unless otherwise specified.

Scopus

Make sure Scopus is selected as the source to start searching records in the Scopus database.

Author search: To search for an author's works, enter the first name/initial and last name in the relevant fields. Alternatively, you may enter last name plus affiliation; or Scopus author ID or ORCID ID to perform narrower author searches.

If you need to locate Scopus IDs for your authors, Scopus.com has a free search layer where authors can be identified, and their Scopus ID retrieved.

Affiliation search: Enter an affiliation only, without an author, if you want to perform a broader search of faculty records across your entire institution.

An affiliation-only search will open a pop-up to help choose the right organizations from the Scopus affiliation database. In the pop-up, choose the most relevant institution name, plus any previous names or affiliated organizations.

Publication Dates: You may use the publication date fields (by year) to add search parameters for finding backfiles or new works when searching by author and/or affiliation.

Pure

Select Pure as the source to start searching your institution's Pure records. The Pure option will show up if access has been enabled as described in the Pure API key generation steps.

Author search: Enter last name and/or first name to search by author. Alternatively, you may search using a Pure Author ID or ORCID ID.

Organizational Unit ID search: You can enter a Pure Organizational Unit ID to perform a broader search of the whole department or center that the unit ID corresponds to. Does not combine with author search.

Creation Dates: You can use the creation date fields to limit results to a specific date range.

Records with "Validated" workflow step: Selecting this option will return only results that are fully published and approved (with the "Validated" status in Pure), and will omit any items that are still in progress.

Records of Content Type: This option allows you to filter results by one or more Pure content types. Click the "Select type" button to open a modal showing the second and third levels of your three-level content type taxonomy from Pure.

ORCID

Select ORCID as the source to start searching data from ORCID profiles.

In the Author ID field, enter the ORCID iD number of the author for whom you'd like to find records.

To search multiple ORCID IDs at once, use the field labeled "Export results for multiple ORCID IDs" and enter one ID per line (press enter/return after each ID). Next, click the Export button.

When searching for multiple ORCID IDs, search results are skipped due to the way the ORCID API works. You'll go directly to the export step (described below), where you can select the "include additional unmapped metadata" option to add a spreadsheet column named "Origin ORCID Profile" noting the ORCID ID/profile each work came from.

PubMed

Select PubMed as the source to start searching data from the PubMed database.

You may search by any combination of author first/last names and initials, and/or by affiliation. Results can be filtered by publication date range (note: harvested works may have publication dates outside specified ranges, may appear out of order, or may differ from those found on the PubMed website due to technical limitations with the PubMed API).

For PubMed, affiliation search works like PubMed's advanced search: you can use AND, OR, or NOT Boolean operators with affiliations and use parenthetical expressions. Example: "Missouri State" NOT (Southeast OR Northwest). Put affiliation phrases in quotation marks as shown in the example.

Review Search Results

The search results in the Harvesting Tool include title, author(s), document type, and publication date. Scopus results also include Open Access labels where applicable (see below for details).

When searching for multiple ORCID IDs, as mentioned above, you will skip the search results step and go straight to the export step.

Use the **more authors** button to see additional authors if there are more than five authors.

Titles are hyperlinked to the publisher's site (wherever the data contains a DOI). Click the title of a record to verify the author or other publication details on the publisher page.

TIP: Result with the wrong author? Try a search combining author + affiliation (available with Scopus or PubMed) to check affiliation history. An author + affiliation search can be especially helpful when searching for common names. In general, it is recommended to over-specify search parameters whenever possible to obtain the most relevant results.

To refine your search results, click the **Modify Search** button.

Scopus Open Access labels

In Scopus results, an "Open Access" flag indicates if a record was originally published OA.

Additional open access labels indicate if a record is "Gold", "Hybrid Gold", "Green Final", "Green Accepted", or "Bronze" OA. These more granular OA labels, based on the Unpaywall open access database, help identify for which records it may be easier to acquire a permitted full-text copy to add to your IR. Definitions can be found by hovering over labels in the search results or by viewing the list here (select "Open Access for documents FAQs," then "Which Open Access filters are supported in Scopus?").

All OA labels will be passed through to the export spreadsheet as additional columns if the "Include additional unmapped metadata…" option is selected when exporting (see next section).

For content marked as open access, please note that the full text is not in Scopus. Additionally, open access status is separate from rights checking, so it's still a good idea to check the rights of the journal publisher.

Export Records to a Prepopulated DC Batch Spreadsheet

When you've finalized search results, click the **Export** button.

In the "Export options" pop-up, select a Digital Commons publication type. Then select the specific Digital Commons publication structure where you are intending to add the current set of results.

Check the box next to the Options that you would like to select, if any.

- **Include additional unmapped metadata from source in export:**
  Select this option if you wish to include additional unmapped metadata in your export (such as OA labels and funder information from Scopus). This option will require modifying, renaming, and/or deleting export spreadsheet columns prior to uploading to DC. See "Working with Metadata in the Export Spreadsheet" below for more information about the additional fields and how they map to Digital Commons fields.

- **Identify likely duplicates between result set and selected publication:**
  This option adds a flag to the spreadsheet that indicates where the results include likely duplicate records that are already in the target Digital Commons publication. Duplicates are detected using a machine-learning algorithm based on whether there are published items with the same DOI (if present), title, first author name, and publication/journal name. This may add time to the spreadsheet generation process, depending on the number of records. Only one duplicate check option may be selected at a time; this option will be grayed out if the below checkbox is selected.

- **Identify likely duplicates between result set and all publications of a given type:**
  This option adds a flag to the spreadsheet that indicates where the results include likely duplicate records across all publications matching the selected type—e.g., within all series in the IR, if you select series as the publication type. As with the previous option, duplicates are detected using a machine-learning algorithm based on DOI (if present), title, first author name, and publication/journal name. This may add significant time to the spreadsheet generation process, depending on the number of records. Only one duplicate check option may be selected at a time; this option will be grayed out if the above checkbox is selected.

- **Include Sherpa-Romeo journal permission information in export:**
  Selecting this option adds multiple columns to the spreadsheet with detailed permission checking metadata from Sherpa-Romeo. Results are included for a work if there is a match in the Sherpa-Romeo API with the journal ISSN, e-ISSN, or journal name. This may add significant time to the spreadsheet generation process, depending on the number of records.

Click **Export** in the "Export options" pop-up to request the generation of a prepopulated spreadsheet with metadata that matches your criteria.

The "Export Status" window will show the export spreadsheet name, export status, request date, # of records, DC publication, and duplication check choice (Y/N). Each DC publication link goes to that publication's batch upload page in Digital Commons.

Once an export completes, the status of "In Progress" will change to "Success" and the export spreadsheet will become available for download. **Click the .xls file name to download the generated spreadsheet.**

If an export is taking a bit longer to process, you can close the pop-up and continue using the Harvesting Tool. Click the Export Status link in the top right of the search and search results pages at any time to bring the Export Status pop-up back.

Working with Metadata in the Export Spreadsheet

When you export your search results, source data is automatically mapped to the metadata fields in the selected DC publication. The resulting Excel spreadsheet allows you to sort and modify the metadata as much as needed before proceeding to the batch import step.

If you exported without selecting any Options checkboxes:

- Default fields appear first, then custom fields, then authors. The DC metadata field names appear in the column headings, as shown in the figure below.
- You only need to fill in any further required fields that have not already been populated; otherwise the spreadsheet should be ready to batch import following the steps under "Batch Import Harvested Records to the Target Publication."
- If needed, you can see which fields are required by checking the submission form in the target publication or checking with your consultant.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	title	keywords	abstract	ocument_typ	doi	fpage	iss	isbn	issn	lpage	ublication_da	rce_publicat	volume	uthor1_fnam	uthor1_lnam	hor1_
2	GLL398, an	Breast cance	© 2020, Sprir	article	10.1007/s10:	359	2		01676806	368	2020-04-01	Breast Cance	180	Shanchun	Guo	Xavie
3	C-Reactive P	adult-onset,	© Copyright	article	10.3389/fcell.2020.00018						2020-02-20	Frontiers in C	8	Qiling	Li	Xi'an
4	New zampan	Anticancer a	© 2020 by th	article	10.3390/molecules250203	2					2020-01-15	Molecules	25	Guanglin	Chen	Califc
5	Metabolism a	Belinostat, H	© 2019 by th	article	10.3390/ph12040180		4				2019-12-01	Pharmaceutic	12	Changde	Zhang	Xavie
6	Targeting of I	breast cance	© Copyright	article	10.1089/ars.	1983	17		15230864	1998	2019-06-10	Antioxidants	30	Ke	Wu	Charl
7	Nitrogen-cont	Antiproliferati	© 2019 Elsev	article	10.1016/j.bio	227			00452068	239	2019-06-01	Bioorganic Cl	87	Pravien	Rajaram	Califc

If you chose one of the "Identify likely duplicates" options:

- Two extra columns will appear in the spreadsheet after the title field, with boundary columns on either side for clear identification.
- The first column will show a "Likely Duplicate" flag for a record if a duplicate is detected.
- The second column will list the URL(s) where the duplicate is located. Multiple URLs will appear separated by commas.
- Check the likely duplicate records (as needed) and remove any rows from the spreadsheet that you would not like to include in the batch import.
- Once you have modified the spreadsheet, be sure to delete the duplicate flag and boundary columns before attempting the batch import step.

	A	B	C	D	E	
1	title	=> BEGIN DUPLICATE FLAG SECTION	Likely Duplicate in Selected Publication	Duplicate Location(s)	<= END DUPLICATE FLAG SECTION	(M(
2	The Effect		-	-		
3	Using plac		-	-		
4	SMART PI		-	-		
5	TARGETIN		-	-		
6	Introductio		-	-		
7	Evaluating		-	-		
8	Measuring		-	-		
9	The effect		-	-		
10	A Special		-	-		
11	Growing s		Likely Duplicate	https://hsrc.himmelfarb.gwu.e		
12	Labor-dem		-	-		
13	How Effec		-	-		
14	The social		-	-		
15	Measuring		-	-		
16	The kalam		-	-		
17	Supplemer		Likely Duplicate	https://hsrc.himmelfarb.gwu.e		
18	Simulating		-	-		

If you checked the "Include additional unmapped metadata" option:

- The unmapped fields will appear in the spreadsheet with boundary columns on either side for clear identification.
- If you want to include any of the unmapped fields in the import, the field will need to exist in the Digital Commons publication.
  - If there's an existing DC field you want to use, you can replace the Scopus, Pure, PubMed, or ORCID field name in the column heading with the Digital Commons back-end field name (make sure to delete that DC column from elsewhere in the spreadsheet). Conversely, you can cut the harvested contents in the source column and paste them into the column with the Digital Commons back-end field name.
  - If you'd like to create a new custom field to hold metadata from an unmapped field, contact your consultant to request the new field. Then add that field's back-end name to the column heading once you have confirmation it's been created.
- After modifying the spreadsheet, be sure to delete the boundary columns and any other unedited unmapped field columns before attempting the batch import step.

TIP: You can see back-end field names in a DC publication structure by doing a batch export in that publication. The field names will appear in the DC batch export spreadsheet column headings. See the Batch Upload, Export, and Revise guide for more info.

If you chose the "Include Sherpa-Romeo permission information" option:

- Multiple columns will appear in the spreadsheet after the title field (or after the duplicate flag section, if present). Boundary columns appear on either side for clear identification.
- The first several columns include the Sherpa-Romeo URI, last modified date, and the journal URI on the publisher site (to double-check the most current policy, if desired).
- Columns then appear for all relevant Sherpa-Romeo pathways and accompanying details. Pathways are not included that have OA fees or which have limited locations not applicable to IRs. The OA fee column will always be empty as a result.
- An empty row means none of the three criteria fields (ISSN, e-ISSN, or journal name) matched a Sherpa-Romeo result.
- Once you have finished permissions checking, be sure to delete all Sherpa-Romeo and boundary columns before attempting the batch import step.

Adding full text files to the spreadsheet:

If your IR requires that a full text copy of an article be obtained before it is loaded onto the site, you may be able to take care of several steps with the Sherpa-Romeo option described above. If needed, you can also use the spreadsheet as a working document while doing additional work "offline." This work may include further permissions checking, reaching out to faculty, or putting the article on an appropriate server or storage platform for DC batch import.

Once you are ready, add the file URL to the **fulltext_url** column in the spreadsheet. Full-text import links from Pure will be automatically included on the export spreadsheet, mapped to the fulltext_url field, if the file exists in the Pure record.

See the Batch Upload, Export, and Revise guide or ask your consultant for more information about where to store files for batch import.

Batch Import Harvested Records to the Target Publication

The final harvesting step is to import the records in the spreadsheet to the selected DC publication, using that publication's batch import tool. For a direct route, you may use the publication link in the Export Status window (click "Export Status" in the Harvesting Tool to reopen the window if needed and view exports from the last 72 hours).

Detailed information about batch importing to Digital Commons publications is available in the Batch Upload, Export, and Revise guide.

You can skip to step 3 in the batch import process, since you will already have a filled in spreadsheet generated by the Harvesting Tool.
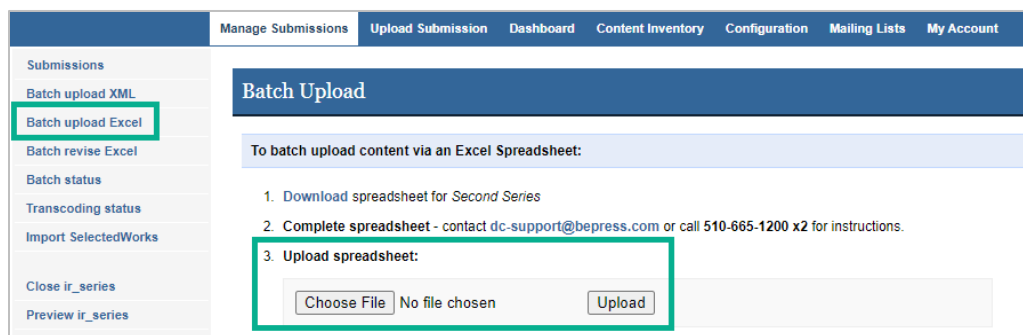
Metadata and Document Type Mappings

The below fields from each source map automatically if the corresponding DC metadata fields are present in your DC publication structure. If the DC field isn't present in the publication structure, the corresponding source metadata is only exported if you select "include additional unmapped metadata" in the export pop-up. Contact your consultant if you wish to enable any of these fields or add custom fields to contain any of your exported content.

Each list of document types below shows how source document types map to DC document types. If a DC publication has custom document types, you can manually edit the mappings in the export spreadsheet. Please let your consultant know if you'd like to add any custom document types for

this purpose. If you'd like to see what the original document type was for a source, that is available in the spreadsheet when using the "Include additional unmapped metadata" export option.

Scopus to DC metadata field mappings:

Scopus Metadata	DC Metadata Field
Article Title	title
Publication Title	source_publication
ISSN	issn
E-ISSN	eissn
ISBN	isbn
Issue	issnum
Volume	volnum
Page Range	fpage, lpage
Publication Date	publication_date
DOI	doi
Abstract	abstract
Keywords	keywords
Document type	document_type
Author	authorX_fname, authorX_lname
Author Affiliation	authorX_institution
Pubmed ID	pubmedid
Funding Number	grant
Funding Sponsor	fundref
Article Number	articlenum
Scopus ID (for work)	identifier, external_article_id
URL prefix + DOI	source_fulltext_URL

Scopus to DC document type mappings:

Scopus Document Type	DC Document Type
Article-ar	article (default)
Abstract Report-ab	article (default)
Book-bk	series default
Book Chapter-ch	series default
Business Article-bz	article (default)
Conference Paper-cp	conference
Conference Review-cr	conference
Data Paper-dp	article (default)
Editorial-ed	editorial
Erratum-er	article (default)
Letter-le	letter
Multimedia-mm	series default
Note-no	article (default)
Press Release-pr	news

Report-rp	article (default)
Retracted-tb	article (default)
Review-re	article (default)
Short Survey-sh	article (default)

Pure to DC metadata field mappings:

Pure Metadata	DC Metadata Field
items/abstract/text/value	abstract
items/articleNumber	articlenum
items/personAssociations/0/name/firstName	author1_fname
items/personAssociations/0/externalOrganisations/0/name/text/0/value	author1_institution
items/personAssociations/0/name/lastName	author1_lname
items/bibliographicalNote/text/0/value	comments
items/type/term/text/0/value	document_type
items/electronicVersions/0/doi; items/electronicVersions/1/doi	doi
items/publicationSeries/0/electronicIssn	eissn
items/info/additionalExternalIds/0/value + (items/info/additionalExternalIds/1/idSource)	external_article_id
items/pages	fpage
items/electronicVersions/0/file/fileURL; items/electronicVersions/1/file/fileURL	fulltext_url
items/isbns/0	isbn
items/journalAssociation/issn/value	issn
items/journalNumber	issnum
items/keywordGroups/0/keywordContainers/0/freeKeywords/0/freeKeywords/0	keywords
items/pages	lpage
items/publicationStatuses/0/publicationDate/month; items/publicationStatuses/0/publicationDate/day; items/publicationStatuses/0/publicationDate/year	publication_date
items/info/additionalExternalIds/0/value (ONLY if items/info/additionalExternalIds/1/idSource = "PubMed")	pubmedid
items/electronicVersions/0/doi; items/electronicVersions/1/doi	source_fulltext_URL
items/pureId	identifier

Pure to DC document type mappings:

Pure template (2nd level)	Pure sub-type (3rd level)	DC Document Type
Book/Report	Anthology	book
Book/Report	Book	book
Book/Report	Commissioned report	default
Book/Report	Other report	default

17

Book/Report	Scholarly edition	book
Chapter in Book/Report/ Conference proceeding	Chapter	default
Chapter in Book/Report/ Conference proceeding	Conference contribution	default
Chapter in Book/Report/ Conference proceeding	Entry for encyclopedia/dictionary	default
Chapter in Book/Report/ Conference proceeding	Foreword/postscript	default
Chapter in Book/Report/ Conference proceeding	Other chapter contribution	default
Contribution to conference	Abstract	conference
Contribution to conference	Other	default
Contribution to conference	Paper	conference
Contribution to conference	Poster	conference
Contribution to journal	Article	article
Contribution to journal	Book/Film/Article review	default
Contribution to journal	Comment/debate	response
Contribution to journal	Conference article	conference

ORCID to DC metadata field mappings:

ORCID Metadata	DC Metadata Field	Notes
short-description	abstract	
For searched-for author: /person/email/email For non-searched-for author(s): /contributors/contributor/0/contributor-email	author1_email	If ORCID profile being harvested from has a public email address, that email address goes into the "authorX_email" field in spreadsheet
For searched-for author: /person/name/given-names/value For non-searched-for author(s): contributors/contributor/0/credit-name	author1_fname	Mapped from (in order of priority): contributor credit name; bibtex citation; profile given and family names
For searched-for author: profile.activities-summary/employments/affiliation-group/summaries/0/organization/name For non-searched-for author(s): nil	author1_institution	searched-for "organization" comes from profile
For searched-for author: /personal-details/family-name For non-searched-for author(s): contributors/contributor/0/credit-name	author1_lname	Mapped from (in order of priority): contributor credit name; bibtex citation; profile given and family names
type	document_type	
external-ids/external-id/0/external-id-value	doi	(if associated external-id-type = doi)

[external-ids/external-id/0/external-id-value + (external-ids/external-id/0/external-id-type Description)];...[external-ids/external-id/n/external-id-value + (external-ids/external-id/n/external-id-type Description)]	identifier	(if associated external-id-type is NOT "doi", "ISBN", "pmid", or "grant_number")
external-ids/external-id/0/external-id-value	grant	(if associated external-id-type = grant_number)
external-ids/external-id/0/external-id-value	isbn	(if associated external-id-type = isbn)
language-code	language	
publication-date/year/value; publication-date/month/value; publication-date/day/value	publication_date	3 ORCID fields concatenated into single DC field
external-ids/external-id/0/external-id-value	pubmedid	(if associated external-id-type = pmid)
doi/value	source_fulltext_URL	
journal-title/value	source_publication	
title/title/value	title	
volume (in bibtex citation)	volnum	If bibtex citation present
number (in bibtex citation)	issnum	If bibtex citation present
pages (in bibtex citation)	fpage	If bibtex citation present
pages (in bibtex citation)	lpage	If bibtex citation present

ORCID to DC document type mappings:

ORCID Document Type	DC Document Type
annotation	response
artistic-performance	default
book	book
book-chapter	default
book-review	bookreview
conference-abstract	conference
conference-paper	conference
conference-poster	conference
data-management-plan	default
data-set	default
dictionary-entry	default
disclosure	default
dissertation	dissertation
dissertation-thesis	default
edited-book	book
encyclopedia-entry	default
invention	default
journal-article	article
journal-issue	article

lecture-speech	default	
license	default	
magazine-article	article	
manual	default	
newsletter-article	news	
newspaper-article	news	
online-resource	default	
other	default	
patent	default	
physical-object	default	
preprint	default	
registered-copyright	default	
report	default	
research-technique	default	
research-tool	default	
software	default	
spin-off-company	default	
standards-and-policy	default	
supervised-student-publication	article	
technical-standard	default	
test	default	
trademark	default	
translation	default	
website	default	
working-paper	default	

PubMed to DC metadata field mappings:

PubMed Metadata	DC Metadata Field	Notes
PubmedArticle.MedlineCitation.Article.Abstract.AbstractText[0].#text	abstract	
PubmedArticle.MedlineCitation.Article.ArticleTitle	title	
PubmedArticle.MedlineCitation.Article.AuthorList.Author[0].AffiliationInfo.Affiliation	author1_institution	Middle initial is imported with the first name into first name field.
PubmedArticle.MedlineCitation.Article.AuthorList.Author[0].ForeName	author1_fname	Splits on space, taking the first word
PubmedArticle.MedlineCitation.Article.AuthorList.Author[0].ForeName	author1_mname	Splits on space, taking the second word
PubmedArticle.MedlineCitation.Article.AuthorList.Author[0].LastName	author1_lname	
PubmedArticle.MedlineCitation.Article.ELocationID	doi	Can appear in MedlineCitation.Article.ElocationID or

		PubmedData.ArticleIdList
PubmedArticle.MedlineCitation.Article.ELocationID	source_fulltext_url	URL version of DOI
PubmedArticle.MedlineCitation.Article.Journal.ISSN.#text	eissn	If issn "type" field is "Electronic"
PubmedArticle.MedlineCitation.Article.Journal.ISSN.#text	issn	
PubmedArticle.MedlineCitation.Article.Journal.JournalIssue.Issue	issnum	
PubmedArticle.MedlineCitation.Article.Journal.JournalIssue.PubDate	publication_date	
PubmedArticle.MedlineCitation.Article.Journal.JournalIssue.Volume	volnum	
PubmedArticle.MedlineCitation.Article.Journal.Title	source_publication	
PubmedArticle.MedlineCitation.Article.Language	language	
PubmedArticle.MedlineCitation.Article.Pagination.MedlinePgn	fpage	Extrapolated from MedlinePgn page range (ex: '1461-1468')
PubmedArticle.MedlineCitation.Article.Pagination.MedlinePgn	lpage	Extrapolated from MedlinePgn page range (ex: '1461-1468')
PubmedArticle.MedlineCitation.Article.PublicationTypeList.PublicationType[0].#text	document_type	
PubmedArticle.MedlineCitation.KeywordList.Keyword	keywords	
PubmedArticle.MedlineCitation.MeshHeadingList.MeshHeading	subject_area	
PubmedArticle.MedlineCitation.PMID.#text	pubmedid	
PubmedArticleSet.PubmedArticle.PubmedData.ArticleIdList	identifier	

PubMed to DC document type mappings:

PubMed Document Type	DC Document Type
Adaptive Clinical Trial	default
Address	default
Autobiography	book
Bibliography	default
Biography	book
Case Reports	default
Classical Article	article
Clinical Conference	conference
Clinical Study	default
Clinical Trial	default
Clinical Trial Protocol	default
Clinical Trial, Phase I	default

Clinical Trial, Phase II	default
Clinical Trial, Phase III	default
Clinical Trial, Phase IV	default
Clinical Trial, Veterinary	default
Collected Works	book
Comment	response
Comparative Study	default
Congress	conference
Consensus Development Conference	conference
Consensus Development Conference, NIH	conference
Controlled Clinical Trial	default
Dataset	default
Dictionary	default
Directory	default
Duplicate Publication	default
Editorial	editorial
Electronic Supplementary Materials	default
English Abstract	default
Equivalence Trial	default
Evaluation Study	default
Expression of Concern	response
Festschrift	book
Government Publication	default
Guideline	default
Historical Article	article
Interactive Tutorial	default
Interview	default
Introductory Journal Article	article
Journal Article	article
Lecture	default
Legal Case	default
Legislation	default
Letter	letter
Meta-Analysis	default
Multicenter Study	default
News	news
Newspaper Article	news
Observational Study	default
Observational Study, Veterinary	default
Overall	conference
Patient Education Handout	default
Periodical Index	default
Personal Narrative	default
Portrait	default

Practice Guideline	default
Pragmatic Clinical Trial	default
Preprint	default
Publication Components	default
Publication Formats	default
Published Erratum	default
Randomized Controlled Trial	default
Research Support, American Recovery and Reinvestment Act	default
Research Support, N.I.H., Extramural	default
Research Support, N.I.H., Intramural	default
Research Support, Non-U.S. Gov't	default
Research Support, U.S. Gov't, Non-P.H.S.	default
Research Support, U.S. Gov't, P.H.S.	default
Retracted Publication	default
Retraction of Publication	default
Review	article
Scientific Integrity Review	default
Study Characteristics	default
Support of Research	default
Systematic Review	default
Technical Report	default
Twin Study	default
Validation Study	default
Video-Audio Media	default
Webcast	default

Do you have questions about the DC Harvesting Tool? Please contact Consulting Services at dc-support@elsevier.com or 510-665-1200, option 2, weekdays 6:30 a.m.–7:30 p.m. North America Pacific Time.