



UNIVERSITAT DE  
BARCELONA

Facultat d'Informació i Mitjans Audiovisuals

<http://www.ub.edu/biblio>

Grau d'Informació i Documentació

Curs 2020-2021

## 2.2 Metadades per a Google Scholar

**Assignatura:** Metadades especialitzades

**Tema:** 2. Metadades per als continguts web...

**Professor:** Miquel Centelles



<http://creativecommons.org/licenses/by-nc-sa/3.0/es>

# Sumari

- Característiques i rellevància de Google Scholar.
- Beneficis per a col·leccions digitals.
- Esquemes de metadades per a Google Scholar.
- Per què Google Scholar posterga Dublin Core?
- Procés de creació de metadades per a Google Scholar amb Dspace.

# Presentació de Google Scholar

- Subproducte de Google iniciat el 2004 en fase beta.
- Especialitzat en **cercar i identificar** material bibliogràfic publicat, de caràcter científic i acadèmic.
- Google Scholar és diferent a Google pel que fa a **orientació i cobertura, índex generat i *crawlers*** aplicats.
- Proporciona:
  - Referències bibliogràfiques.
  - Accés a texts complets.
  - Cites.

# Abast de Google Scholar

- Google Scholar únicament indexa **documents acadèmics amb resum** allotjats en la **Web acadèmica...**
- Cobreix:
  - Revistes científiques i llibres.
  - Comunicacions i ponències a congressos.
  - Informes científicotècnics.
  - Tesis, tesines o memòries de grau.
  - Treballs dipositats en repositoris.

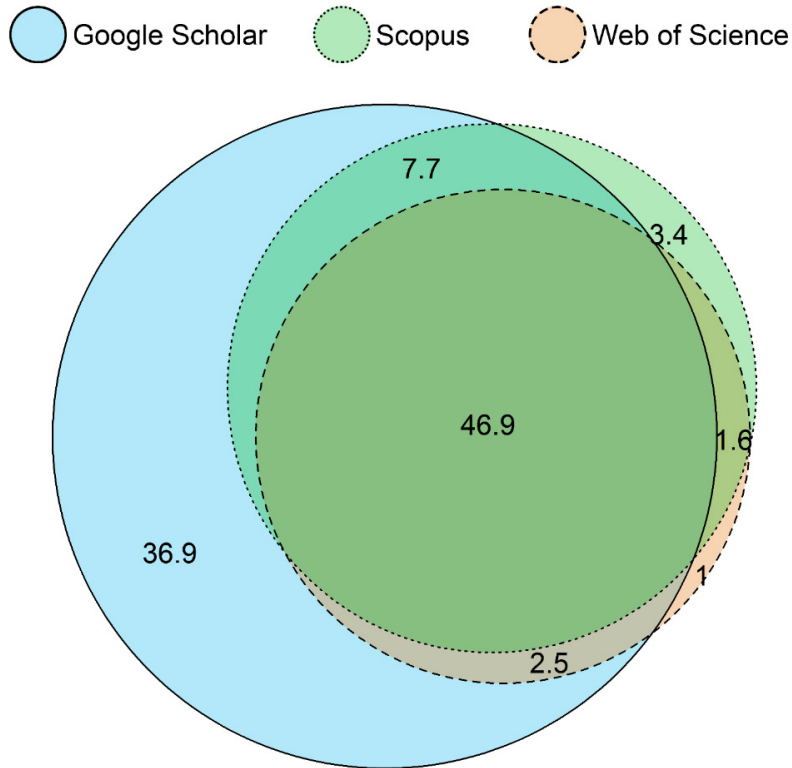
# Abast de Google Scholar (2)

- No cobreix:
  - Ressenyes de llibres i editorials de diaris i revistes
  - Llibres de text i monografies
  - Diaris i revistes comercials
- Però sí:
  - Patents
  - Disposicions legals

# Elements *<meta>* a Google Scholar

- Per què hem d'estar indexats en Google Scholar?
  - Mida superior de Google Scholar
  - Cobertura superior de Google Scholar
  - Gran facilitat d'ús
  - Diversitat i qualitat de les funcionalitats proporcionades
- Beneficis directes:
  - Grau d'ús de les col·leccions digitals i els seus fons.
  - Grau de visibilitat de les col·leccions digitals.

# Mida superior de Google Scholar



- Percentatge de citacions úniques i superposades a Google Scholar, Scopus i Web of Science (n = 2.448.055 cites de totes les àrees temàtiques).
- Font: Martín-Martín, A., Orduna-Malea, E., Thelwall, M., i López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of informetrics*, 12(4), 1160-1177.

# Cobertura **superior** de Google Scholar

	% GS (all cit.)	% WoS (all cit.)	% Scopus (all cit.)	% WoS cit. in GS	% Scopus cit. in GS	% WoS cit. in Scopus
Overall	94	52	60	95	92	93
Humanities, Literature & Arts	93	27	36	88	84	83
Social Sciences	94	35	43	93	89	89
Business, Economics & Management	96	28	35	93	92	89
Engineering & Computer Science	93	52	63	94	90	94
Physics & Mathematics	96	59	64	97	94	94
Health & Medical Sciences	94	54	62	95	91	93
Life Sciences & Earth Sciences	95	62	67	96	93	95
Chemical & Material Sciences	94	73	77	95	94	96

- Percentatge de cites a Google Scholar, Web of Science i Scopus, en relació al conjunt de cites i a les cites d'altres bases de dades.
- Font: Martín-Martín, A., Orduna-Malea, E., Thelwall, M., i López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of informetrics*, 12(4), 1160-1177.



# Gran facilitat d'ús

- És un recurs d'informació valorat com:
  - simple
  - senzill
  - ràpid
  - multilingüe
  - universal
  - gratuït

# Múltiples funcionalitats proporcionades

- Agrupació de resultats
- Ordenació per rellevància / data
- Extracció de cites
- Generació de citacions
- Perfil d'autors, bibliometria, etc.

# Beneficis per a col·leccions digitals

- En general, els dipòsits digitals (DD) que segueixen directrius mostren ràtios d'indexació superiors:
  - 88-98% dels DD que segueixen les directrius
  - 38-48% dels DD que no les segueixen
- La ràtio d'indexació a Google Scholar depén, en gran mesura, de la correcta aplicació de metadades en els recursos publicats en els dipòsits digitals.

# Esquemes de metadades per a Google Scholar

- Tot i que Google Scholar dóna suport a Dublin Core, resulten més eficaços els següents esquemes de metadades:
  - **Highwire Press tags (citation\_\*)**
  - BE Press tags (bepress\_\*)
  - Eprints tags (eprints.\*)
  - PRISM tags (prism.\*)
- Dublin Core (dc.\*/dcterms.\*) ha de ser la darrera opció!

# Esquemes de metadades per a Google Scholar: Highwire Press

- Esquema de metadades de la plataforma de publicació de HighWire Press.
- Opció destacada en els exemples de la documentació de Google Scholar.
- Característiques generals:
  - Noms dels elements iniciats amb prefix *citation\_*
  - Hi ha elements de metadades específics per a cada tipus de document (llibre, article, tesi...)

# Esquemes de metadades per a Google Scholar: Highwire Press (2)

- Elements mínims obligatoris:
  - Títol propi del document: *citation\_title*
  - Autor/s del document: *citation\_author*
  - Data de publicació: *citation\_publication\_date*

# Per què Google Scholar posterga Dublin Core? (1)

- Dublin Core resulta limitat per a descriure parts components de publicacions en sèrie.
  - No conté elements adients per a descriure **cites bibliogràfiques**:
    - Títol de la revista, número de volum i exemplar, i interval de pàgines s'introdueix generalment en un sol camp, com *dc.relation* o *dc.source*.
    - Les directrius per a la codificació d'informació de cites bibliogràfica utilitzant una qualificació de la propietat *dc.identifier* (anomenada "bibliographicCitation") segueix sent una única propietat.

# Per què Google Scholar posterga Dublin Core? (2)

- Dublin Core resulta limitat per a parts components de publicacions en sèrie.
  - No facilita la distinció entre diferents documents acadèmics:
    - Pre-print vs. Article de revista
    - Capítol de llibre vs. Llibre
    - Document de treball vs. Comunicació vs. Tesi

**Aquestes condicions dificulta l'anàlisi i indexació precises de les dades en els seus components bibliogràfics individuals.**



# Procés de creació de metadades per a Google Scholar amb Dspace

- Fases del procés:
  1. Preparació de les URL dels documents en DSpace
  2. Configuració del dispositiu de mapatge a DSpace
  3. Marcatge de les referències bibliogràfiques
  4. Introducció de les metadades Dublin Core
  5. Generació de les metadades de Google Scholar
  6. Revisió de metadades de Google Scholar
- Font d'informació de requeriments i recomanacions:  
[Inclusion Guidelines for Webmasters](#)

# 1. Preparació de les URL dels documents en DSpace

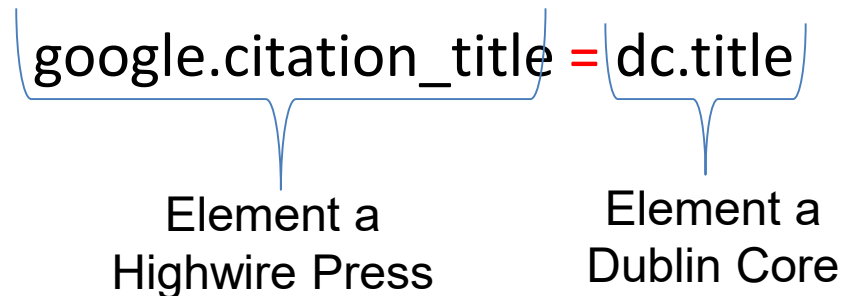
- Google Scholar indexa elements <meta> situats en la capçalera dels fitxers HTML.
  - El gestor de col·leccions digitals **Dspace** permet generar aquests elements a partir dels registres de metadades dels documents lliurats.
- Regla bàsica: cada document i cada abstract del document han d'estar situats en un fitxer PDF (indexat) o en un fitxer HTML **separat**.
  - Els gestor de col·leccions digitals **DSpace** facilita aquesta disposició.

## 2. Configuració del dispositiu de mapatge a Dspace (1)

- Dspace té un dispositiu de mapatge que connecta els elements de les metadades Dublin Core amb els elements de l'esquema Highwire Press en els elements <meta> d'HTML.
- El dispositiu de mapatge es basa en un fitxer de mapatge que defineix les correspondències entre l'esquema d'origen (Dublin Core) i l'esquema de destinació (Highwire Press).

## 2. Configuració del dispositiu de mapatge a Dspace (2)

- Fitxer de mapatge:
  - Es denomina “Google metadata field configuration” o google-metadata.properties
  - Està implementat en un format *.properties*, que es pot editar amb Notepad++ o similar.
  - Les correspondències s’expressen de forma molt simple:



# 2. Correspondència entre elements dels esquemes (1)

## Correspondències d'elements generals

HighwirePressMetadata	DublinCore
citation_author	dc.creator
citation_date/citation_publication_date	dc.date.issued
citation_title	dc.title
citation_publisher	dc.publisher
citation_keywords	dc.subject
citation_language	dc.language
citation_doi	dc.identifier.doi
citation_pmid	dc.identifier.pmid
citation_issn	dc.identifier.issn
citation_isbn	dc.identifier.isbn
citation_volume	dc.identifier.volume
citation_issue	dc.identifier.issue
citation_firstpage	dc.identifier.spage
citation_lastpage	dc.identifier.epage

# 2. Correspondència entre elements dels esquemes (2)

Correspondències d'elements per a tipus específics de documents

HighwirePressMetadata	Tipus	DublinCore
citation_inbook_title	Capítols de llibre	dc.relation.ispartof
citation_conference_title	Congressos	dc.relation.ispartof
citation_technical_report_number	Informes	
citation_technical_report_institution	Informes	dc.publisher
identifiers.technical_report	Informes	dc.type:TechnicalReport
citation_patent_country	Patents	
citation_patent_number	Patents	
identifiers.patent	Patents	dc.type:Patent
citation_journal_title	Revistes	dc.relation.ispartof
citation_dissertation_name	Tesis	dc.title
citation_dissertation_institution	Tesis	dc.publisher
identifiers.dissertation	Tesis	dc.type:Thesis

## 2. Correspondència entre elements dels esquemes (3)

### Enllaç de fitxers i àabstracts

En aquest tipus de mapatge, cal aplicar components específics dels sistemes de gestió de col·leccions digitals, com DSpace

Highwire Press Metadata	Descripció	Dublin Core	DSpace
<code>citation_abstract_html_url</code>	Enllaç pàgina descriptiva (metadades)		\$handle
<code>citation_fulltext_html_url</code>	Enllaç text complet en html		
<code>citation_pdf_url</code>	Enllaç text complet	dc.identifier	\$simple-pdf

# 3. Marcatge de les referències bibliogràfiques

- Abans de lliurar un document a Dspace (PDF o HTML), cal preparar les referències bibliogràfiques que contenen.
  - La secció ha d'estar identificada amb una capçalera estàndard, com ara "Referències " o "Bibliografia", en una línia diferenciada.
  - Cada referència dins aquesta secció ha d'estar numerada:
    - Números "1. - 2. - 3." o "[ 1] - [ 2 ] - [3 ]" en el cas de PDF.
    - Elements <ol>, en el cas de documents HTML.
  - El text de cada referència ha de ser una cita bibliogràfica formal en un format d'ús comú.



# 4. Introducció de les metadades Dublin Core

- Realitzem un lliurament amb DSpace i incorporem les metadades de Dublin Core en el registre corresponent.

The screenshot shows the DSpace interface for adding metadata. At the top, there's a navigation bar with a logo on the left and a user profile 'Connectat com miquel.centelles@...' on the right. Below this is a menu with buttons: 'Descriu' (highlighted), 'Descriu', 'Carrega fitxer', 'Verifica', 'Llicència', and 'Complet'. The main heading is 'Envia: descriu el document' with a question mark icon. Below the heading is a paragraph of instructions: 'Si us plau, ompliu la informació que es demana a continuació sobre el document. En la majoria de navegadors, és possible utilitzar la tecla 'tabulador' per moure el cursor a la següent casella o botó. Per guardar haureu d'utilitzar el ratolí cada vegada.' Below this are several form sections: 1. 'Authors': A label 'Enter the names of the authors of this item.' followed by two input fields: 'Cognoms, e.g. Vidal' and 'Nom(s) + "Jr", e.g. Donald Jr', and a '+ Afegeix-ne més' button. 2. 'Title \*': A label 'Enter the main title of the item.' followed by a single input field. 3. 'Other Titles': A label 'If the item has any alternative titles, please enter them here.' followed by a single input field and a '+ Afegeix-ne més' button. 4. 'Date of Issue \*': A label 'Please give the date of previous publication or public distribution. You can leave out the day and/or month if they aren't applicable.' followed by three input fields: 'Mes:' with a dropdown menu (showing '(Sense mes)'), 'Dia:', and 'Any:'. 5. 'Publisher': A label 'Enter the name of the publisher of the previously issued instance of this item.' followed by a single input field. At the bottom of the form, the text 'metadades especialitzades' is visible.

# 5. Generació de les metadades de Google Scholar

- En completar el lliurament d'un document en una col·lecció de Dspace es dispara el dispositiu de mapatge.
- Les metadades s'incorporen en la pàgina del *handle*, o identificador únic de document.
  - Per exemple:  
<http://www.tabuladecimal.info:8080/jspui/handle/123456789/412>

# 6. Revisió de metadades de Google Scholar



## DSpace JSPUI

DSpace preserva i permet l'accés fàcil i obert a tot tipus de contingut digital, incloent text, imatges, imatges en moviment, mpegs i conjunts de dades

Més informació

Repositori de pràctiques de la Facultat de BID / Grau en ID / Metadades / Metadades

Si us plau utilitzeu sempre aquest identificador per citar o enllaçar aquest document:

<http://www.tabuladecimal.info:8080/jspui/handle/123456789/412>

Títol:	Las 10 mejores recomendaciones para cocinar
Autors:	Pittrón López, Anna
Matèria:	Cuina
Data de publicació:	27-de -2020
Publicat per:	Planeta
URI:	<a href="http://www.tabuladecimal.info:8080/jspui/handle/123456789/412">http://www.tabuladecimal.info:8080/jspui/handle/123456789/412</a>
ISBN:	012345678910
Apareix en les col·leccions:	Metadades

```
18 <!DOCTYPE html>
19 <html>
20 <head>
21 <title>Repositori de pràctiques de la Facultat de BID: Las 10 mejores recomendaciones para cocinar</title>
22 <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
23 <meta name="generator" content="DSpace 6.3" />
24 <meta name="viewport" content="width=device-width, initial-scale=1.0">
25 <link rel="shortcut icon" href="/jspui/favicon.ico" type="image/x-icon"/>
26 <link rel="stylesheet" href="/jspui/static/css/jquery-ui-1.10.3.custom/jquery-ui-1.10.3.custom.css" type="text/css" />
27 <link rel="stylesheet" href="/jspui/static/css/bootstrap/bootstrap.min.css" type="text/css" />
28 <link rel="stylesheet" href="/jspui/static/css/bootstrap/theme.min.css" type="text/css" />
29 <link rel="stylesheet" href="/jspui/static/css/bootstrap/dspace-theme.css" type="text/css" />
30 <link rel="search" type="application/opensearchdescription+xml" href="/jspui/open-search/description.xml" title="DSpace"/>
31
32 <link rel="schema:DC" href="http://purl.org/dc/terms/" />
33 <link rel="schema:DC" href="http://purl.org/dc/elements/1.1/" />
34 <meta name="DC.creator" content="Pittrón López, Anna" />
35 <meta name="DC.dateAccepted" content="2020-02-27T10:17:00Z" scheme="DCTERMS.W3CDTF" />
36 <meta name="DC.dateIssued" content="2020-02-27T10:17:00Z" scheme="DCTERMS.W3CDTF" />
37 <meta name="DC.identifier" content="012345678910" />
38 <meta name="DC.identifier" content="http://www.tabuladecimal.info:8080/jspui/handle/123456789/412" scheme="DCTERMS.URI" />
39 <meta name="DC.language" content="es" xml:lang="ca_ES" scheme="DCTERMS.RFC1766" />
40 <meta name="DC.publisher" content="Planeta" xml:lang="ca_ES" />
41 <meta name="DC.subject" content="Cuina" xml:lang="ca_ES" />
42 <meta name="DC.title" content="Las 10 mejores recomendaciones para cocinar" xml:lang="ca_ES" />
43 <meta name="DC.type" content="Book" xml:lang="ca_ES" />
44 <meta name="citation_keywords" content="Cuina; Book" />
45 <meta name="citation_isbn" content="012345678910" />
46 <meta name="citation_title" content="Las 10 mejores recomendaciones para cocinar" />
47 <meta name="citation_publisher" content="Planeta" />
48 <meta name="citation_language" content="es" />
49 <meta name="citation_author" content="Pittrón López, Anna" />
50 <meta name="citation_pdf_url" content="http://www.tabuladecimal.info:8080/jspui/bitstream/123456789/412/1/T2metadades_Exercici02_esquemes_codificacio_html_meta.html" />
51 <meta name="citation_date" content="2020-02-27" />
52 <meta name="citation_abstract_html_url" content="http://www.tabuladecimal.info:8080/jspui/handle/123456789/412" />
53
```

Fitxers d'aquest document:

Fitxer	Descripció	Dimensions	Format
T2Metadades_Exercici02_esquemes_codificacio_html_meta.html		4,24 KB	HTML

Mostrar/Obrir