# **Google Scholar** Vs Institutional Repository

Tue Aug 27 2019, 2:30pm-3:30pm, Max Atria Garnet 219

#### Veerle kerstens



veerle.kerstens@kuleuven.be

#### Mehmet Celik

mehmet.celik@kuleuven.be

@celik.be

https://github.com/mehmetc





### Contents



- Introduction
- Issue: making our institutional repository discoverable in Google Scholar
- What we did
- Conclusion: experiences
- Possible plans for the future



# **KU LEUVEN**

LIBIS is a department of KU Leuven and supports central ICT applications for the management & discovery of information sources for libraries, archives, museums and other (heritage) collections (GLAM).

Division of Leuven Research and Development (LRD)



We provide consultancy on all Ex Libris products.

# •••• Institutional Repositories

- The Stakeholders
- Researchers: If publications are better findable -> more reads -> more citations -> more recognition -> more funding



- Institution:
  - More citations gives a better ranking
  - More visibility and credibility
- Collection managers: special collections









Institutional bibliography and repository of KU Leuven Association

KU LEUVEN ♠ LIRIAS		3	EN NL
Search Limo		Q	ADVANCED SEARCH
Welcome to the digital repository for KU Leuven Association research! Lirias is a document repository to archive the KU Leuven Association research output in a digital way. Lirias captures, stores, indexes, preserves, and distributes digital research material of the KU Leuven Association.	My Lirias Ku Leuven Association users can log in with their regular user-id and password on the "My Lirias" page.		
KU LEUVEN 2016 © KU Leuven Disclaimer Cookiebele	id Privacy Policy. Voig KU Leuv	ven via 🖪 🕻	) () in 🛗

- > 550.000 publications
- > 120.000 with files uploaded
- > 60.000 in Open Access
- Metadata managed in Elements
- Full text stored in DSpace
- Primo as end user interface





Repository of bachelor theses of colleges in Flanders

2



#### Zoekterm



#### Welkom op DoKS

DoKS maakt eindwerken van Vlaamse hogeschoolstudenten toegankelijk. Het eindwerk is het sluitstuk van de opleiding, Deze eindwerken bevatten dan ook waardevolle informatie over een breed spectrum van onderwerpen, van accountancy over media tot sociale zorg, en nog veel meer. Ze worden met toestemming van de student gratis ter beschikking gesteld van het brede publiek. Niet alleen studenten en docenten maar ook particulieren, bedrijven en overheidsinstellingen kunnen zo gebruik maken van de resultaten van het geleverd onderzoek. Dat leidt tot nieuwe samenwerkingsverbanden, eindwerkprojecten, aanwervingen, ...

De deelnemende hogescholen zijn :

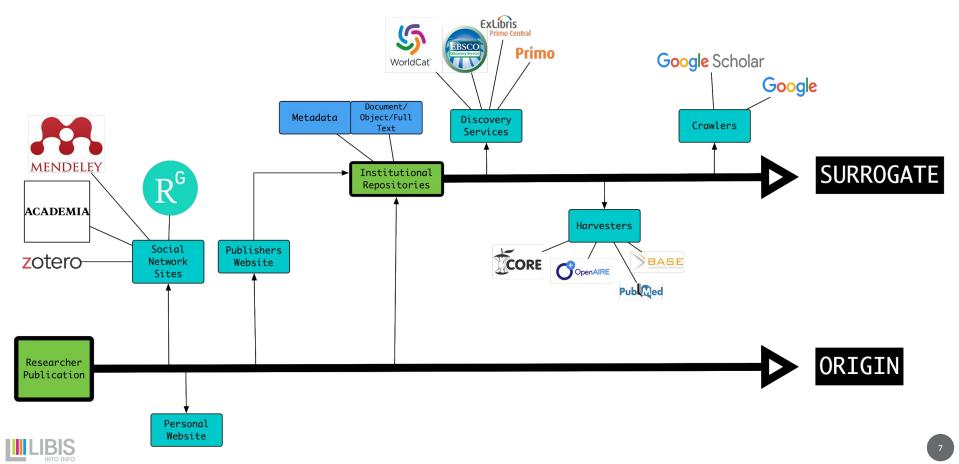
- Thomas More
- Vives
- UC Leuven-Limburg

Voor meer informatie contacteer doks@thomasmore.be

- > 25.000 bachelor theses
- > 20.000 in Open Access
- Metadata managed in Alma
- PDF's stored in Rosetta
- Primo as end user interface



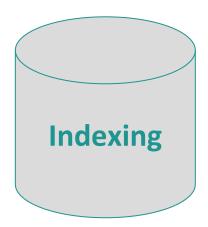
# •••• Publication Life Cycle

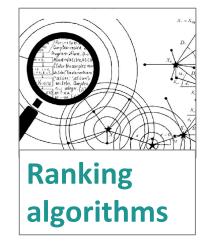


# •••• Google and Google Scholar: little in common...



Crawling













- Scope: scholarly publications (no books)
- No big documents (?)
- Crawls documents and bibliographic information
- Format: pdf or html
- Full text or at least abstracts needed (?)
- Updates records: 6-9 months to a year.



- Scope: "all"
- Crawls documents
- Many formats
- Updates records dependant on the importance of the site
- Follows any hyperlink







- Meta-tags: format Highwire Press, Eprints, BE Press or PRISM
- Grouping of results



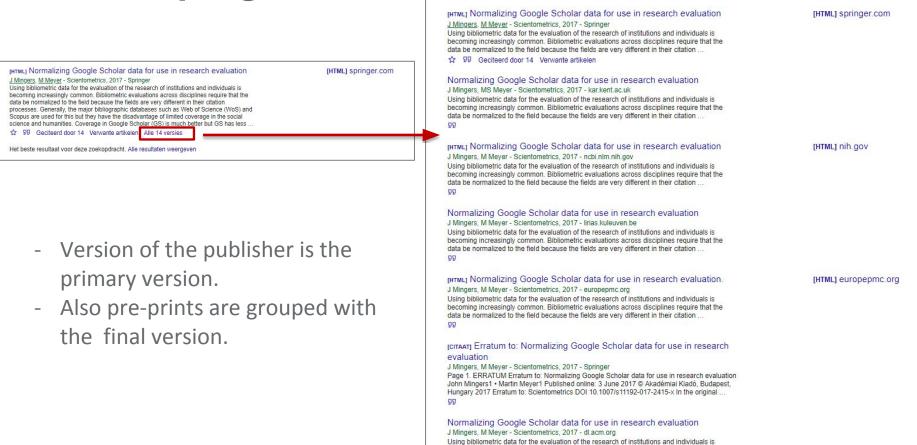
- Meta-tags: format schema.org
- Every url displayed separately in search results
   = > issue of duplicates



Indexing

# •••• Grouping of versions





increasingly common. Dibliometric avaluations coreas dissiplines require the

# •••• Other conditions

- Requirement: PDF in the same subdirectory/domain as metadata
- 1 PDF per document
- PDF's are text searchable (no scans)
- At least abstract is available (?)
- No pop-ups or requirements to log-in

Google Scholar Indexing Search Citations Inclusion Metrics Publishers About Libranes Inclusion Guidelines for Webmasters Overview Indexing Guidelines Content Google Scholar uses automated software, known as "parsers", to identify biblio papers, as well as references between the papers. Incorrect identification of bit Crawl references will lead to poor indexing of your site. Some documents may not be may be included with incorrect author names or titles, and some may rank lowe Indexina because their (incorrect) bibliographic data would not match (correct) reference papers. To avoid such problems, you need to provide bibliographic data and re-Troubleshooting automated "parser" software can process. 1. Preparing article URLs Questions

Place each article and each abstract in a separate HTML or PDF file. At this tim effectively index multiple abstracts on the same webpage or multiple papers in Likewise, we're unable to index different sections of the same paper in different have its own unique URL in order for it to be included in Google Scholar.

#### 2. Configuring the meta-tags

If you're using repository or journal management software, such as Eprints, DS or OJS, please configure it to export bibliographic data in HTML "<meta>" tags. supports Highwire Press tags (e.g., citation\_title), Eprints tags (e.g., eprints.title bepress\_citation\_title), and PRISM tags (e.g., prism.title). Use Dublin Core tags last resort - they work poorly for journal papers because Dublin Core doesn't ha for journal title, volume, issue, and page numbers. To check that these tags are abstracts and view their HTML source.

- A The title tag, e.g., citation\_title or DC.title, must contain the title of the pap title of the journal or a book in which the paper was published, or for the na repository. This tag is required for inclusion in Google Scholar.
- B. The author tag, e.g., citation\_author or DC.creator, must contain the author actual authors) of the paper. Don't use it for the author of the website or for than authors, e.g., thesis advisors. Author names can be listed either as "\$ "John Smith". Put each author name in a separate tag and omit all affiliation certifications, etc., from this field. At least one author tag is required for ince Scholar.
- C. The publication date tag, e.g., citation\_publication\_date or DC.issued, mus publication, i.e., the date that would normally be cited in references to this papers. Don't use it for the date of entry into the repository that should go citation\_online\_date instead. Provide full dates in the "2010/5/12" format it alone otherwise. This tag is required for inclusion in Google Scholar.

D. For journal and conference papers, provide the remaining bibliographic cit



1.1.1

## •••• Crawling & Indexing : monitoring





- No systematic checks possible
- Only via samples
- Contact person at Google Scholar

# Google

 Google Search Console to monitor process of crawling & indexing

🗴 III S	Q Een Uik, in https://imo.lkis.bo/ impectances	Google Search Console	=
	https://lmo.lbis.be/primo-explore/fs/ldiaplay/vid+00K5&amp.docid+32L885,ALMA_0551172922430001471&amp.fromSiaemap=1	• \wd.sidil.omi%ceqtri	n
GOOGLE-INDEX LIVE TEST	URL-inspectie	Oversicht	
	Detent op: 50 mai 2015, 00.20	Presides UR_incectie	
	URL is backfikkase rood coople An do can fill word gradienteeting repetitienteet do convertie. Non Acta anothen exergingenen in Coople backlearabilitation multi all reference automotion. Status: Informatike	· ·	otel D
	OGTESTE PADRA BECAREN	<ul> <li>Materia</li> <li>Mobiele braikbaarheid</li> </ul>	
	Beschlasanheld UBL kan worden gehodewend	ikgingapoblemen en v	
	Committance                Ø          Mobile Indiffusation          >	Links Instellingen	



### •••• Relevance ranking



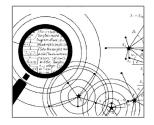
- Number of citations +++ (1)
- Search term in title ++
- Author name, Journal title++
- Search terms in document +
- Reputation of author, Journal+
- Date +-
- Number of times the search term appears in the document –
- Synonyms in document -

# Google

- Search terms in document ++
- Links to the document ++
- Number of times the search term appears in the document ++
- Synonyms in document+
- Quality of the links to the document +
- Date +



(1) Rovira, Cristòfol; Guerrero-Solé, Frederic; Codina, Lluís (2018). "Received citations as a main SEO factor of Google Scholar results ranking". El profesional de la información, v. 27, n. 3, pp. 559-569. https://doi.org/10.3145/epi.2018.may.09



# •••• How to make our publications discoverable in Google Scholar ?

Facilitate crawling of Primo?

- Required meta tags currently not available. (schema.org tags are not supported)
- Requirement of Google Scholar to have full text and metadata in the same domain

=> We developed our own solution

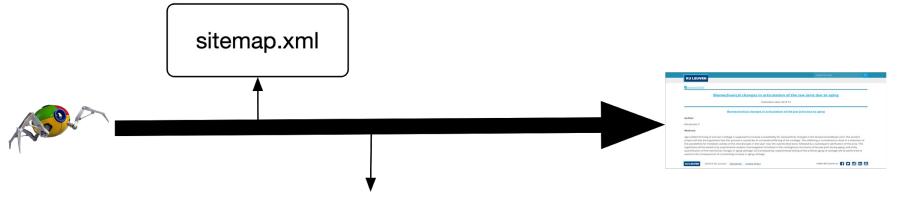


# •••• The institutional repository





### •••• Crawlers visiting







# ••••• Landing page

	Search in Lirias	Q
KU LEUVEN		

Download PDF

#### **Biomechanical changes in articulation of the jaw joint due to aging**

Publication date: 2018-10

Biomechanical changes in articulation of the jaw joint due to aging

Author:

Mirahmadi, F

#### Abstract:

Age-related thinning of articular cartilage is supposed to increase susceptibility for osteoarthritic changes in the temporomandibular joint. The present project will test the hypothesis that this process is caused by an increased stiffening of the cartilage. This stiffening is considered to result in a reduction of the possibilities for metabolic activity of the chondrocytes in the layer near the subchondral bone, followed by a subsequent calcification of this area. This hypothesis will be tested (i) by experimental analysis of propagation of solutes in the cartilageness structures of the jaw joint during aging, and (ii) by quantification of the mechanical changes in aging cartilage. (iii) Consequently, experimental testing of the artificial aging of cartilage will be performed to examine the consequences of crosslinking increase in aging cartilage.



Follow KU Leuven on 🗗 🗹 🚺 🛗

## •••• Primo to Landing page

PRIMO	<pre>rdddata&gt; caulats&gt;Hirahmdi caulats&gt;Hirahmdi caulats&gt;Caulats&gt;Caulats&gt; caulats&gt;Caulats&gt;Caulats&gt; cau&gt;Hirahmadi, P v cdita&gt; Biomechanical changes in articulation of the jaw joint due to aging <!--/dtita--> Biomechanical changes in articulation of the jaw joint due to aging  Caulats&gt; Cau</br></pre>
PR	nge-related thinning of articular cartilage is supposed to increase susceptibility for ostecarthritic changes in the temporomandibular joint. The present project will test the hypothesis that this process is caused by an increased stiffening of the cartilage. This stiffening is considered to result in a reduction of the possibilities for metabolic activity of the chondrocytes in the layer near the subchondral bone, followed by a subsequent calcitication of this hypothesis will be tested (i) by experimental analysis of propagation of solutes in the cartilaginous structures of

#### KU LEUVEN

Download PDF

#### **Biomechanical changes in articulation of the jaw joint due to aging**

Publication date: 2018-10

#### Biomechanical changes in articulation of the jaw joint due to aging

Author:

Mirahmadi, F

#### Abstract:

Age-related thinning of articular cartilage is supposed to increase susceptibility for osteoarthritic changes in the temporomandibular joint. The present project will test the hypothesis that this process is caused by an increased stiffening of the cartilage. This stiffening is considered to result in a reduction of the possibilities for metabolic activity of the chondrocytes in the layer near the subchondral bone, followed by a subsequent calcification of this area. This hypothesis will be tested (i) by experimental analysis of propagation of solutes in the cartilagence structures of the jaw joint during aging, and (ii) by quantification of the mechanical changes in aging cartilage. (iii) Consequently, experimental testing of the artificial aging of cartilage will be performed to examine the consequences of crosslinking increase in aging cartilage.

**KU LEUVEN** 

2019 © KU Leuven Disclaimer Cookie Policy

Follow KU Leuven on 🗗 🔽 💽 in 🛗

<meta name="citation publication date" content="2018-10" /> <meta name="citation\_author" content="Mirahmadi, F" /> <meta name="citation author institution" content="KU Leuven" /> <meta name="citation\_title" content="Biomechanical changes in articulation of the jaw joint due to aging" /> <meta property="og:description" content="Age-related thinning of articular cartilage is supposed to increase susceptibility for osteoarthritic changes in the temporomandibular joint. The present project will test the hypothesis that this process is caused by an increased stiffening of the cartilage. This stiffening is considered to result in a reduction of the possibilities for metabolic activity of the chondrocytes in the layer near the subchondral bone, followed by a subsequent calcification of this area. This hypothesis will be tested (i) by experimental analysis of propagation of solutes in the cartilaginous structures of the jaw joint during aging, and (ii) by guantification of the mechanical changes in aging cartilage. (iii) Consequently, experimental testing of the artificial aging of cartilage will be performed to examine the consequences of crosslinking increase in aging cartilage." /> <meta name="l:external" content="false" /> <meta name="citation pdf url" content="https://lirias.kuleuven.be/retrieve/517354" /> <meta name="citation language" content="eng" /> <meta name="citation\_public\_url" content="https://lirias.kuleuven.be/1652315" />

<meta name="citation pii" content="1652315" />

## **••••** Highwire Press Metadata

<meta name="citation publication date" content="2018-10" /> <meta name="citation author" content="Mirahmadi, F" /> <meta name="citation author institution" content="KU Leuven" /> <meta name="citation title" content="Biomechanical changes in articulation of the jaw joint due to aging" /> <meta property="oq:description" content="Age-related thinning of articular cartilage is supposed to increase susceptibility for osteoarthritic changes in the temporomandibular joint. The present project will test the hypothesis that this process is caused by an increased stiffening of the cartilage. This stiffening is considered to result in a reduction of the possibilities for metabolic activity of the chondrocytes in the layer near the subchondral bone, followed by a subsequent calcification of this area. This hypothesis will be tested (i) by experimental analysis of propagation of solutes in the cartilaginous structures of the jaw joint during aging, and (ii) by quantification of the mechanical changes in aging cartilage. (iii) Consequently, experimental testing of the artificial aging of cartilage will be performed to examine the consequences of crosslinking increase in aging cartilage." /> <meta name="l:external" content="false" /> <meta name=[citation pdf url" content="https://lirias.kuleuven.be/retrieve/517354" /> <meta name="citation language" content="eng" /> <meta name= citation public url " content="https://lirias.kuleuven.be/1652315" />

```
<meta name="citation_pii" content="1652315" />
```



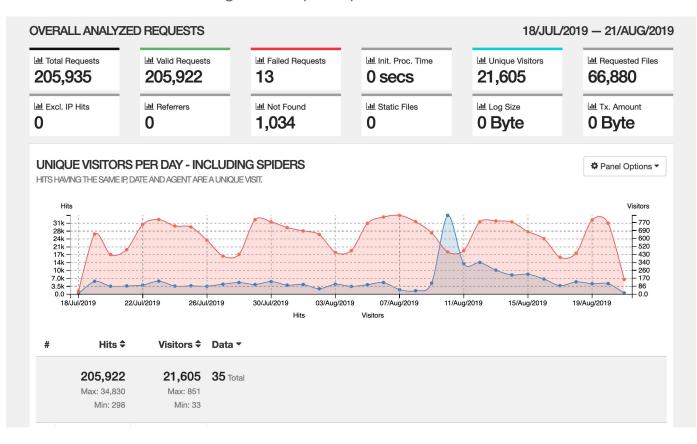
## •••• Bitstream proxy



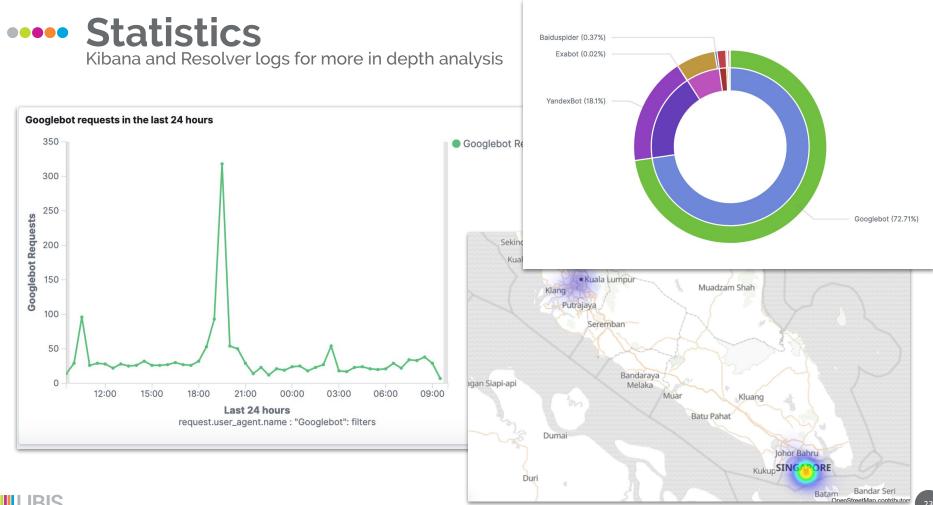




GoAccess on HTTP server logs for simple reports







INTO INFO

# •••• How to deal with records that have more than one file attached ?

Google Scholar can only link to one file

- Freely available and restricted version ?
   => only freely available version is exposed for crawling
- More than 1 freely available file ? scenario's:
  - Files are different parts of the work
  - Files are different versions of the work



# •••• Files are different parts of the work



PHD THESIS

Posttraumatic stress disorder and disasters in Peru : the role of personality and social support Gargurevich Liberti, Rafael: Conveleyn, Jozef (Supervisor) : Luyten, Patrick (Co supervisor) 2006-12

VIEW ONLINE	
DETAILS	View Online REPORT A PROBLEM
LINKS	A. caratula.pdf [freely available]🖸 >
SEND TO	Acknowledgments.pdf [freely available][2] > B. Initial summary.pdf [freely available][2] >
TAGS	C. Table of Contents.pdf [freely available]@ > D. Introduction.pdf [freely available]@ > E. Chapter 1.pdf [freely available]@ > F. Chapter 2.pdf [freely available]@ > G. Chapter3 section 1.pdf [freely available]@ > H. Chapter3 section 2.pdf [freely available]@ > I. Chapter 3 section 3.pdf [freely available]@ > J. Chapter3 section 4.pdf [freely available]@ > K. Chapter 4.pdf [freely available]@ > L. APPENDIX 123.pdf [freely available]@ > M. APPENDIX 4.pdf [freely available]@ >

Advise to researchers not to split up their work into parts.

 $\rightarrow$ 



## •••• Files are different versions of the work

Currently no possibility to indicate in the metadata which file is the preferred one

- if accepted and published version are available
   => published version is selected for crawling
- Otherwise: all files are presented for crawling
   => first occurrence is taken by Google Scholar





### Can happen if

- Failed to set embargo
- Metadata gets replaced

Possible consequences

- Lower ranking of repository data
- Metadata removal
- Link roth

Send **410**: "Gone" instead or better forward to the correct entry



## •••• Our experiences

- SEO for Google Scholar appeared to be a long lasting process
- What helped us to solve issues:
  - Systematic checks in GS
  - Monitoring of the harvesting
  - Contact person at Google Scholar
- Advice on best practices for researchers is important



# •••• Best practices for researchers

- Good metadata is important, always provide an abstract
- Quality of the uploaded files is important.
  - Don't upload image-scans of papers
  - Need of well formed citations section in the document
- Do not upload more than one freely available file per paper
- Open Access papers will get more exposure

[статом] Views and attitudes towards Evidence-Based Practice in a Dutch social work organization R Van der Zwet. D Beneken genaamd Kolder... - Journal of Evidence ... - lirias kuleuven be

Journal of Evidence-Based Social Work. ISSN: 1543-3714. Publisher: Taylor & Francis (Routledge). Views and attitudes towards Evidence-Based Practice in a Dutch social work organization. Author: Van der Zwet, R. Beneken genaamd Kolder, D ; Schalk, R ; Van ... 99 ≫

#### Record without an abstract

Understanding social innovation as an innovation process: Applying the innovation journey model

PRA Oeij, W van der Torre, F Vaas... - Journal of Business ..., 2019 - Iirias.kuleuven.be The innovation journey is a process model distinguishing between the initiation, developmental and implementation/termination period of innovations; it looks at drivers and barriers, like innovation managers, investors, setbacks, adaptation, infrastructure. We ... 99

#### Record with an abstract

# •••• Possible plans for the future

- Implement the landing pages for our the repository of Master Theses DokS
- Use the monitoring tool to report on the use of repository items (page views and full text downloads)
- Facilitate crawling of by Google (include schema.org tags in the landing pages)
- Landing page as default ?





# Questions?



