

Google Scholar Indexing for Open Journal Systems (OJS)

Monica Westin, Google Scholar

Overview

1. How the Google Scholar indexing system works
2. Common OJS indexing problems and suggested fixes
3. Best practices for OJS journal indexing
4. Google Scholar indexing guidelines and resources for OJS
5. Questions

How the Google Scholar indexing system works

- Google Scholar crawls the entire web looking for scholarly publications: articles, books, reports, theses, conference proceedings, preprints ...
- The indexing system identifies scholarly content, determines each item's bibliographic metadata, and groups all versions of an item together with this metadata in search results



The screenshot shows the Google Scholar interface. At the top, the search bar contains the query "Plantas hospedeiras de cochonilhas-farinhentas (Hemiptera: Pseudococcida)". Below the search bar, the results are displayed. The first result is an article titled "Plantas hospedeiras de cochonilhas-farinhentas (Hemiptera: Pseudococcidae) em cultivos de videira" by FSC Lopes, JV de Oliveira, et al., published in 2019. The abstract indicates that the study identifies 37 species of plants that serve as hosts for scale insects in vineyards in the São Francisco region. The interface includes filters for time range (Any time, Since 2019, Since 2018, Since 2015, Custom range...), sorting options (Sort by relevance, Sort by date), and checkboxes for "include patents" and "include citations". A note at the bottom states "Showing the best result for this search. See all results".

Google Scholar

"Plantas hospedeiras de cochonilhas-farinhentas (Hemiptera: Pseudococcida)

Articles

Any time
Since 2019
Since 2018
Since 2015
Custom range...

Sort by relevance
Sort by date

include patents
 include citations

Plantas hospedeiras de cochonilhas-farinhentas (Hemiptera: Pseudococcidae) em cultivos de videira [PDF] ufg.br

FSC Lopes, JV de Oliveira... - ... Research in the ..., 2019 - revistas.ufg.br

Resumo As cochonilhas-farinhentas (Hemiptera: Pseudococcidae) são consideradas pragas de importância econômica em cultivos de videira. São altamente polífagas, alimentando-se de plantas hospedeiras de 250 famílias. Objetivou-se identificar plantas hospedeiras de cochonilhas-farinhentas em cultivos de videira, na região do Vale do São Francisco. Foram amostradas plantas daninhas, ornamentais, quebra-ventos, frutíferas e nativas do bioma Caatinga. Foram identificadas 37 espécies de plantas hospedeiras de ...

☆ 📄 Related articles 🔗

Showing the best result for this search. [See all results](#)

What Scholar needs for indexing

- URLs for all articles
- Bibliographic information in the form of machine-readable metadata tags (“metatags”)

Bibliographic metatags tell the Scholar indexing system what the metadata for an article is: title, author, publication date, etc.



“Citation_pdf_url” metatag tells the indexing system which file to associate with this metadata



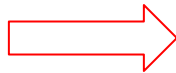
```
<meta name="citation_title" content="The testis isoform of the phosphorylase kinase catalytic subunit (PhK-T) plays a critical role in regulation of glycogen mobilization in developing lung">
<meta name="citation_author" content="Liu, Li">
<meta name="citation_author" content="Rannels, Stephen R.">
<meta name="citation_author" content="Falconieri, Mary">
<meta name="citation_author" content="Phillips, Karen S.">
<meta name="citation_author" content="Wolpert, Ellen B.">
<meta name="citation_author" content="Weaver, Timothy E.">
<meta name="citation_publication_date" content="1996/05/17">
<meta name="citation_journal_title" content="Journal of Biological Chemistry">
<meta name="citation_volume" content="271">
<meta name="citation_issue" content="20">
<meta name="citation_firstpage" content="11761">
<meta name="citation_lastpage" content="11766">
<meta name="citation_pdf_url" content="http://www.example.com/content/271/20/11761.full.pdf">
```

View source code from article landing page to view metatags

Right click or keyboard command to "View Page Source," depending on your browser



Search HTML source for "citation_" to view metatags



→ ↻ 🏠 ⓘ view-source:https://www.revistas.ufg.br/ateliê/article/view/54065

```
<meta name="citation_journal_title" content="Ateliê Geográfico" />
<meta name="citation_journal_abbrev" content="1" />
<meta name="citation_issn" content="1982-1956" />
<meta name="citation_author" content="Elaiz Aparecida Mensch Buffon" />
```

Metatags should match the published PDF:

→ ↺ ↻ ↶ (🌐) view-source:https://www.revistas.ufg.br/ateliê/article/view/54065

```
<meta name="citation_journal_title" content="Ateliê Geográfico"/>
<meta name="citation_journal_abbrev" content="1"/>
<meta name="citation_issn" content="1982-1956"/>
<meta name="citation_author" content="Elaiç Aparecida Mensch Buffon"/>
<meta name="citation_author_institution" content="Universidade Federal do Paraná"/>
<meta name="citation_author" content="Francisco de Assis Mendonça"/>
<meta name="citation_author_institution" content="Universidade Federal do Paraná"/>
<meta name="citation_title" content="Risco socioambiental à leptospirose humana no Aglomerado Urbano Metro
<meta name="citation_date" content="2019/04/20"/>
<meta name="citation_volume" content="13"/>
<meta name="citation_issue" content="1"/>
<meta name="citation_firstpage" content="44"/>
<meta name="citation_lastpage" content="61"/>
<meta name="citation_doi" content="10.5216/ag.v13i1.54065"/>
<meta name="citation_abstract_html_url" content="https://www.revistas.ufg.br/ateliê/article/view/54065"/>
<meta name="citation_language" content="pt"/>
<meta name="citation_pdf_url" content="https://www.revistas.ufg.br/ateliê/article/download/54065/33043"/>
```

- Title, author, and publication dates match
- Bibliographic metatags written in (only) language/script of the article
 - e.g. citation_title gives article title in Portuguese

Risco socioambiental à leptospirose humana no Aglomerado Urbano Metropolitano de Curitiba, Paraná-Brasil: uma proposta metodológica apoiada em SIG

Social-Environmental risk to human leptospirosis in Urban Agglomeration Metropolitan of Curitiba, Paraná-Brazil: a methodological proposal supported by GIS

Risque socio-environnemental pour la leptospirose humaine l'Agglomération Urbaine Métropolitaine de Curitiba, Paraná - Brésil: une proposition méthodologique basée sur le SIG

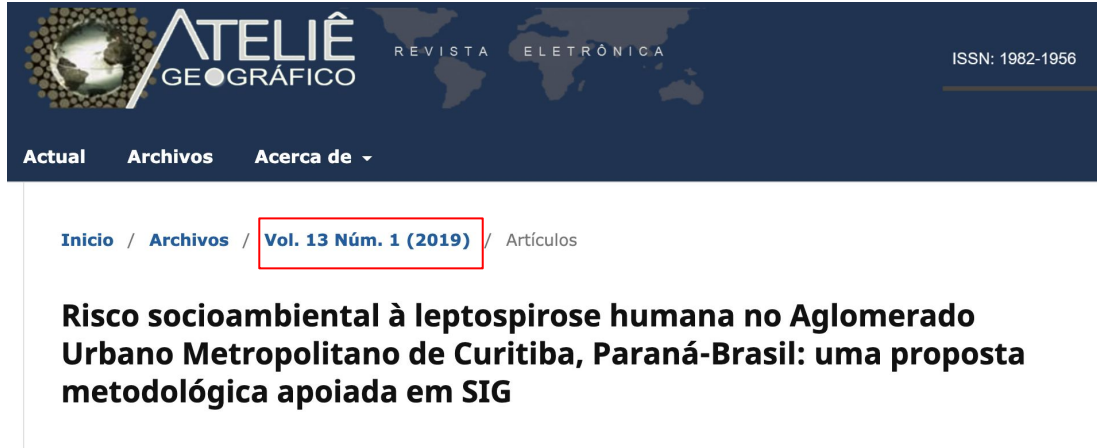
Elaiç Aparecida Mensch Buffon
Universidade Federal do Paraná
embuffon@gmail.com

Francisco de Assis Mendonça
Universidade Federal do Paraná
chico@ufpr.br

Resumo

O conhecimento das áreas de risco socioambiental à leptospirose humana pode auxiliar no processo de tomada de decisão por gestores. A quantificação e o mapeamento desse risco por meio de um índice sintético possibilitam uma análise integrada da problemática. Nesse sentido, a presente pesquisa tem como objetivo propor e avaliar uma metodologia para o mapeamento do risco socioambiental à leptospirose humana por meio de um índice sintético que considera as vulnerabilidades e a exposição à doença. A proposta foi aplicada no Aglomerado Urbano Metropolitano de Curitiba, visando apresentar uma análise no nível de desagregação intraurbano e intrametropolitano, considerando um cenário de 2007 a 2013. Foram utilizadas técnicas estatísticas associada a cartografia de síntese (análise multicritério, cálculo álgebra de mapas, cálculo de incidência e matriz de cruzamentos), que culminaram nos seguintes resultados: 1) a metodologia se mostrou eficaz para sintetizar um grande número de dados, que possibilitou reduzir fenômenos complexos de inter-relações em mensagens simples e, 2) os maiores índices de risco socioambiental estão localizados tanto em regiões periféricas,

Publication date metatag should match date of formal publication for the issue



The screenshot shows the header of the ATELIÊ GEOMÁTICO website. The header is dark blue with the logo on the left, the text 'REVISTA ELETRÔNICA' in the center, and the ISSN '1982-1956' on the right. Below the header is a navigation bar with 'Actual', 'Archivos', and 'Acerca de' (with a dropdown arrow). Below the navigation bar is a breadcrumb trail: 'Inicio / Archivos / Vol. 13 Núm. 1 (2019) / Artículos'. The article title is 'Risco socioambiental à leptospirose humana no Aglomerado Urbano Metropolitano de Curitiba, Paraná-Brasil: uma proposta metodológica apoiada em SIG'.

→ ↻ 🏠 ⓘ view-source:https://www.revistas.ufg.br/atelie/article/view/54065

```
<meta name="citation_date" content="2019/04/20"/>  
<meta name="citation_volume" content="13"/>  
<meta name="citation_issue" content="1"/>
```

OJS indexing errors

- Commonly: Incorrect or missing bibliographic metatags
 - Indexing system forced to skip sites with widespread errors
- Site outages that occur while the indexing system is looking for publications
- Crawler issues, including blocking the Googlebot crawler, slow responsiveness or errors to crawlers, or limiting the crawl speed.

Problem: Incorrect publication dates in citation_date metatag

- Known issue for OJS versions 3.0 and 2.4
- Usually upload date of article is included as publication date, e.g.
- `<meta name="citation_date" content="2019"/>`

versus issue date and date listed in PDF:



A produção de espaço urbano e os fatores do processo de expansão do mercado imobiliário em Mossoró-RN

The production of urban space and the factors of the expansion process of the real estate market in Mossoró-RN

La producción del espacio urbano y los factores del proceso de expansión del mercado inmobiliario en Mossoró-RN

Ademir Araújo da Costa
Universidade Federal do Rio Grande do Norte
ademir@ufrnet.br

Eduardo Alexandre do Nascimento
Universidade Federal do Rio Grande do Norte
eanascimentoa@gmail.com

Resumo

As ações das forças políticas e do capital, em território mossoroense, têm engendrado a reunião de um conjunto de condições favoráveis ao expressivo desenvolvimento do mercado imobiliário na cidade. Tal desenvolvimento se evidencia pela significativa quantidade de investimentos aplicados na produção de condomínios residenciais fechados, destinados às classes, média e alta, e na construção de conjuntos habitacionais "populares" dentro do Programa Minha Casa Minha Vida (PMCMV). *Diante dos impactos da dinâmica do setor imobiliário sobre a economia e sobre a reconfiguração do espaço urbano local, é preciso indagar: quais fatores têm impulsionado o mercado de imóveis residenciais em Mossoró a partir dos primeiros anos de 2000 até o presente momento?* Partindo dessa questão, define-se como objetivo central deste trabalho a análise das principais condições e fatores responsáveis pelo processo de expansão do mercado imobiliário em Mossoró a partir dos primeiros anos de 2000.

Palavras-chave: Mossoró-RN; Mercado Imobiliário; Espaço Urbano; PMCMV

Fix: Incorrect publication dates in citation_publication_date metatag

- Patches for supported versions of OJS include:
 - OJS 3.0.1 and 3.0.2:
<https://github.com/pkp/ojs/commit/9105b7ebc397f3647d500d62d30654b56a4f1e60.diff>
 - OJS 3.0.0:
<https://github.com/asmecher/ojs/commit/c46a9d7a0873ba21ab7fec1cd1453cae215f9a3f.diff>
 - OJS 2.4.x:
<https://github.com/pkp/ojs/commit/6cf0078cff0556231696cf9979377381d18d721d.diff>
- More information here: <https://github.com/pkp/pkp-lib/issues/2739>

Problem: Metatags no longer included after OJS upgrade

- Known issue for journals upgrading from OJS 2.x to OJS 3.0.1, 3.0.2, 3.1.0, and 3.1.1.
- In these cases, the Google Scholar plugin was not automatically enabled
- To test, view the source code for a few articles in each journal that upgraded.
 - If there is no citation_title tag, your site is affected

Fix: Metatags no longer included after OJS upgrade

- **Fix:** re-enable Google Scholar plugin manually via admin dashboard for OJS instances with a small number of journals
- **Fix:** upgrade to OJS 3.1.2
- **Fix:** use an SQL command for large OJS instances with many journals:

https://github.com/pkp/ojs/blob/stable-3_1_2/dbscripts/xml/upgrade/3.1.2/update.xml#L41..L42

Website Settings

[Appearance](#) [Information](#) [Archiving](#) [Languages](#) [Plugins](#) [Announcements](#) [Navigation Menus](#)[Installed Plugins](#) [Plugin Gallery](#)

Plugins

Name	Description	Enabled
Metadata Plugins (3)		
MODS 3.4 meta-data	Contributes MODS 3.4 schemas and application adapters.	<input checked="" type="checkbox"/>
Dublin Core 1.1 meta-data	Contributes Dublin Core version 1.1 schemas and application adapters.	<input checked="" type="checkbox"/>
OpenURL 1.0 meta-data	Contributes OpenURL 1.0 schemas and application adapters.	<input checked="" type="checkbox"/>
Authorization Plugins (0)		

No Items

Block Plugins (7)

Language Toggle Block

Subscription Block

"Developed By" Block

"Hosted By PKP | PS" Block

Information Block

"Make a Submission" Block

Browse Block

Gateway Plugins (1)

Resolver Plugin

Generic Plugins (2)

▶ Citation Style Language

▶ Web Feed Plugin

TinyMCE Plugin

PDFJS PDF Viewer

▶ Custom Block Manager

HTML Article Galley

eLife Lens Article Viewer

Google Scholar Indexing Plugin

Generic Plugins (32)

eLife Lens Article Viewer	This plugin provides rendering support for JATS XML galleys using eLife Lens.	<input checked="" type="checkbox"/>
TinyMCE Plugin	This plugin enables WYSIWYG editing of textareas using the TinyMCE content editor.	<input checked="" type="checkbox"/>
PDFJS PDF Viewer	This plugin uses the pdf.js PDF viewer to embed PDFs on the article and issue galley view pages.	<input checked="" type="checkbox"/>
Dublin Core Indexing Plugin	This plugin embeds Dublin Core meta tags in article views for indexing purposes.	<input checked="" type="checkbox"/>
▶ Web Feed Plugin	This plugin produces RSS/Atom web syndication feeds for the current issue.	<input checked="" type="checkbox"/>
HTML Article Galley	This plugin provides rendering support for HTML Article Galleys.	<input checked="" type="checkbox"/>
Google Scholar Indexing Plugin	This plugin enables indexing of published content in Google Scholar.	<input checked="" type="checkbox"/>

Fix: Select the “Google Scholar Indexing Plugin” checkbox

Problem: Language mix up for multilingual metatags

- Combining different languages/scripts in metatags results in mixed, duplicate bibliographic information that prevents the item from being indexed properly in Scholar search results — and is confusing for users
- ex: including the translated version of the title in title metatags, e.g.
 - `<meta name="citation_title" content="War and Peace == Война и мир" />`
- ex: listing authors in native script/language of home institution when it is not the language in which the article was written, e.g.
 - `<meta name="citation_author" content="Толстой, Лев Николаевич" />`
 - `<meta name="citation_author" content="Tolstoy, Lev Nikolayevich" />`

Problem: Using English by default for all metatags

- Using English in metatags, e.g. in the title metatags, when the article is in a different language results in citations to articles being missed and articles not being ranked as they should be

ex: <meta name="citation_title" content=

"Socio-environmental risk to human leptospirosis in the Metropolitan Urban Parcel of Curitiba, Paraná-Brazil: a methodological proposal based on GIS" />

Risco socioambiental à leptospirose humana no Aglomerado Urbano Metropolitano de Curitiba, Paraná-Brazil: uma proposta metodológica apoiada em SIG

Social-Environmental risk to human leptospirosis in Urban Agglomeration Metropolitan of Curitiba, Paraná-Brazil: a methodological proposal supported by GIS

Risque socio-environnemental pour la leptospirose humaine l'Agglomération Urbaine Métropolitaine de Curitiba, Paraná - Brésil: une proposition méthodologique basée sur le SIG

Elaiz Aparecida Mensch Buffon
Universidade Federal do Paraná
eambuffon@gmail.com

Francisco de Assis Mendonça
Universidade Federal do Paraná
chico@ufpr.br

Resumo

O conhecimento das áreas de risco socioambiental à leptospirose humana pode auxiliar no processo de tomada de decisão por gestores. A quantificação e o mapeamento desse risco por meio de um índice sintético possibilitam uma análise integrada da problemática. Nesse sentido, a presente pesquisa tem como objetivo propor e avaliar uma metodologia para o mapeamento do risco socioambiental à leptospirose humana por meio de um índice sintético que considera as vulnerabilidades e a exposição à doença. A proposta foi aplicada no Aglomerado Urbano Metropolitano de Curitiba, visando apresentar uma análise no nível de desagregação intraurbano e intrametropolitano, considerando um cenário de 2007 a 2013. Foram utilizadas técnicas estatísticas associada a cartografia de síntese (análise multicritério, cálculo álgebra de mapas, cálculo de incidência e matriz de cruzamentos), que culminaram nos seguintes resultados: 1) a metodologia se mostrou eficaz para sintetizar um grande número de dados, que possibilitou reduzir fenômenos complexos de inter-relações em mensagens simples e, 2) os maiores índices de risco socioambiental estão localizados tanto em regiões periféricas,

Fix: Always use language of article for metadata

- Use only the language/script of the published PDF for all metatags. Don't duplicate metatag information in multiple scripts.

ex: <meta name="citation_title" content=

"Risco socioambiental à leptospirose humana no Aglomerado Urbano Metropolitano de Curitiba, Paraná-Brasil: uma proposta metodológica apoiada em SIG" />

Risco socioambiental à leptospirose humana no Aglomerado Urbano Metropolitano de Curitiba, Paraná-Brasil: uma proposta metodológica apoiada em SIG

Social-Environmental risk to human leptospirosis in Urban Agglomeration Metropolitan of Curitiba, Paraná-Brazil: a methodological proposal supported by GIS

Risque socio-environnemental pour la leptospirose humaine l'Agglomération Urbaine Métropolitaine de Curitiba, Paraná - Brésil: une proposition méthodologique basée sur le SIG

Elaiz Aparecida Mensch Buffon
Universidade Federal do Paraná
eambuffon@gmail.com

Francisco de Assis Mendonça
Universidade Federal do Paraná
chico@ufpr.br

Resumo

O conhecimento das áreas de risco socioambiental à leptospirose humana pode auxiliar no processo de tomada de decisão por gestores. A quantificação e o mapeamento desse risco por meio de um índice sintético possibilitam uma análise integrada da problemática. Nesse sentido, a presente pesquisa tem como objetivo propor e avaliar uma metodologia para o mapeamento do risco socioambiental à leptospirose humana por meio de um índice sintético que considera as vulnerabilidades e a exposição à doença. A proposta foi aplicada no Aglomerado Urbano Metropolitano de Curitiba, visando apresentar uma análise no nível de desagregação intraurbano e intrametropolitano, considerando um cenário de 2007 a 2013. Foram utilizadas técnicas estatísticas associada a cartografia de síntese (análise multicritério, cálculo álgebra de mapas, cálculo de incidência e matriz de cruzamentos), que culminaram nos seguintes resultados: 1) a metodologia se mostrou eficaz para sintetizar um grande número de dados, que possibilitou reduzir fenômenos complexos de inter-relações em mensagens simples e, 2) os maiores índices de risco socioambiental estão localizados tanto em regiões periféricas,

Problem: Badly formatted author names

- Common examples include incorrect “first name, last name” format, incomplete author names, and errors in spelling and capitalization, e.g.
 - `<meta name="citation_author" content="Filipa, Pimenta" />`
 - `<meta name="citation_author" content="Leal" />`
 - `<meta name="citation_author" content="Joao maoco" />`

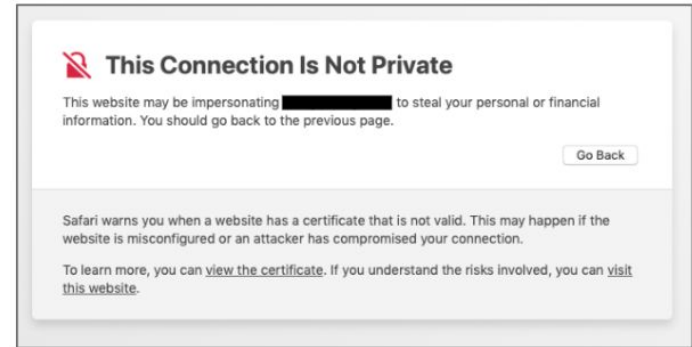
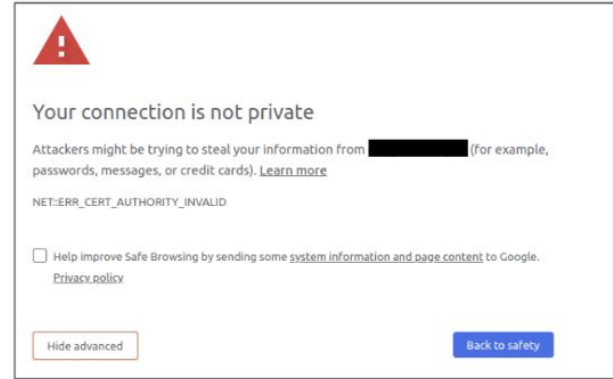
vs. author names listed in published PDF:

Filipa Pimenta¹, Isabel Leal^{1,2} & João Maroco^{1,2}

- **Fix:** List complete author names in citation_author tags as they are written in the published PDF. Use either “last name, first name” or “first name last name” format in metatags.

Problem: invalid HTTPS certificate

- When the https version of a site returns errors due to an invalid certificate, the indexing system is blocked from crawling the site and is forced to removed it from the Scholar index.
- **Fix:** To test, open an article in several different browsers (Chrome, Safari, etc). If you see warnings that the connection is not private, contact your SSL certificate provider to fix.



Problem: hacked OJS sites

- Large number of OJS sites have been hacked
 - Close to 150 hacked sites identified by indexing system
 - Hacked sites are used for commercial spamming
 - Hackers replace OJS code to return spammy pages to Google crawler
- Scholar indexing tries to identify hacked OJS sites
 - Hacked OJS sites are removed from the index

Fix: hacked OJS sites

- **Test:** fetch an article URL emulating the crawler

```
curl -A Googlebot URL_ON_YOUR_SITE
```

- Check a broad sample of randomly selected articles to see if they return a different page from what you see in the browser, or redirect to another site.
- Alert your hosting provider
- Alert the PKP community forum at <https://forum.pkp.sfu.ca>

Best practices for OJS journal indexing

- **Avoid customizing URLs**, which makes it more difficult to identify a journal site, and as a result takes longer to index
 - There is no need to register your site for it to be indexed. The Scholar crawler will automatically find the site.
- The standard OJS URL structure has worked well for a long time

Example of standard URL:

<http://seer.unipampa.edu.br/index.php/agropampa/article/view/23142>

Best practices for OJS journal indexing

- **Set up article-level HTTP 301 (permanent) redirects** whenever URLs for publications in the site change, e.g. when journal site moves or items are renumbered

- Example:

<http://ojs.statsbiblioteket.dk/index.php/her/article/view/25673>

now redirects to

<https://tidsskrift.dk/index.php/her/article/view/25673>

Best practices for OJS journal indexing

- **Avoid outages during migrations**
 - If you are migrating to a new site, keep the old site functional while developing the new version
 - When the new version is ready, change DNS lookup from the old server to the new one

Best practices for OJS journal indexing

- **Avoid repeated site outages**

- Outages result in dead pages in search results
- Repeated outages obstruct the Scholar crawler from finding your publications:

Date	06-20	06-21	06-22	06-23	06-24	06-25	06-26	06-27	06-28	06-29	06-30	07-01	07-02
Status	Green	Green	Red	Light Green	Green	Red	Light Green	Red	Red	--	Red	--	Red

Best practices for OJS journal indexing

- **Avoid extended site outages**

- Extended site outages obstruct the Scholar crawler from finding your publications
- Extended outages can also result in the site dropping out of the index entirely:

Date	06-20	06-21	06-22	06-23	06-24	06-25	06-26	06-27	06-28	06-29	06-30	07-01	07-02
Status	Green	Green	Green	Red	--	Red	Red	Red	Red	--	Red	--	Red

- **Don't keep your OJS site down** for extended periods of time, e.g. multiple days

Google Scholar indexing guidelines and resources for OJS

1. Google Scholar inclusion & troubleshooting guidelines

<https://scholar.google.com/intl/en/scholar/inclusion.html>

2. “Indexing Repositories: Pitfalls & Best Practices” presentation from 2015 Open Repositories conference (targeted to repositories, but has good general guidelines for both Scholar & web indexing)

<https://www.or2015.net/wp-content/uploads/2015/06/or-2015-anurag-google-scholar.pdf>

3. PKP Community Forum

<https://forum.pkp.sfu.ca>

Thank you for joining us!
Questions?

Contact me: mwestin@google.com