



**Illinois Wesleyan University**  
**Digital Commons @ IWU**

---

Digital Commons Great Lakes User Group Annual Meeting

---

Aug 9th, 11:15 AM - 12:00 PM

## Let There Be Light! Indexing Materials from Digital Commons in Apache Solr via OAI-PMH

Graham Hukill  
Wayne State University, [graham.hukill@wayne.edu](mailto:graham.hukill@wayne.edu)

Follow this and additional works at: <https://digitalcommons.iwu.edu/dcglug>



Part of the [Library and Information Science Commons](#)

---

Hukill, Graham, "Let There Be Light! Indexing Materials from Digital Commons in Apache Solr via OAI-PMH" (2013). *Digital Commons Great Lakes User Group Annual Meeting*. 1. <https://digitalcommons.iwu.edu/dcglug/2013/posters/1>

This Event is protected by copyright and/or related rights. It has been brought to you by Digital Commons @ IWU with permission from the rights-holder(s). You are free to use this material in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This material has been accepted for inclusion by faculty at Illinois Wesleyan University. For more information, please contact [digitalcommons@iwu.edu](mailto:digitalcommons@iwu.edu).

©Copyright is owned by the author of this document.

# Let There Be Light!

## Indexing Materials from Digital Commons in Apache Solr with OAI-PMH



Solr is a powerful, fast, open source, enterprise level, **full-text search platform**, based on the Lucene Index. Solr is becoming a common piece of infrastructure in libraries and digital collections infrastructure.

**Documents** are added to Solr by "indexing" Solr ready XML files which contain the text will be made searchable.

Finally, **Searching** is done through a separate interfaces that pulls in Solr search results.

[lucene.apache.org/solr](http://lucene.apache.org/solr)

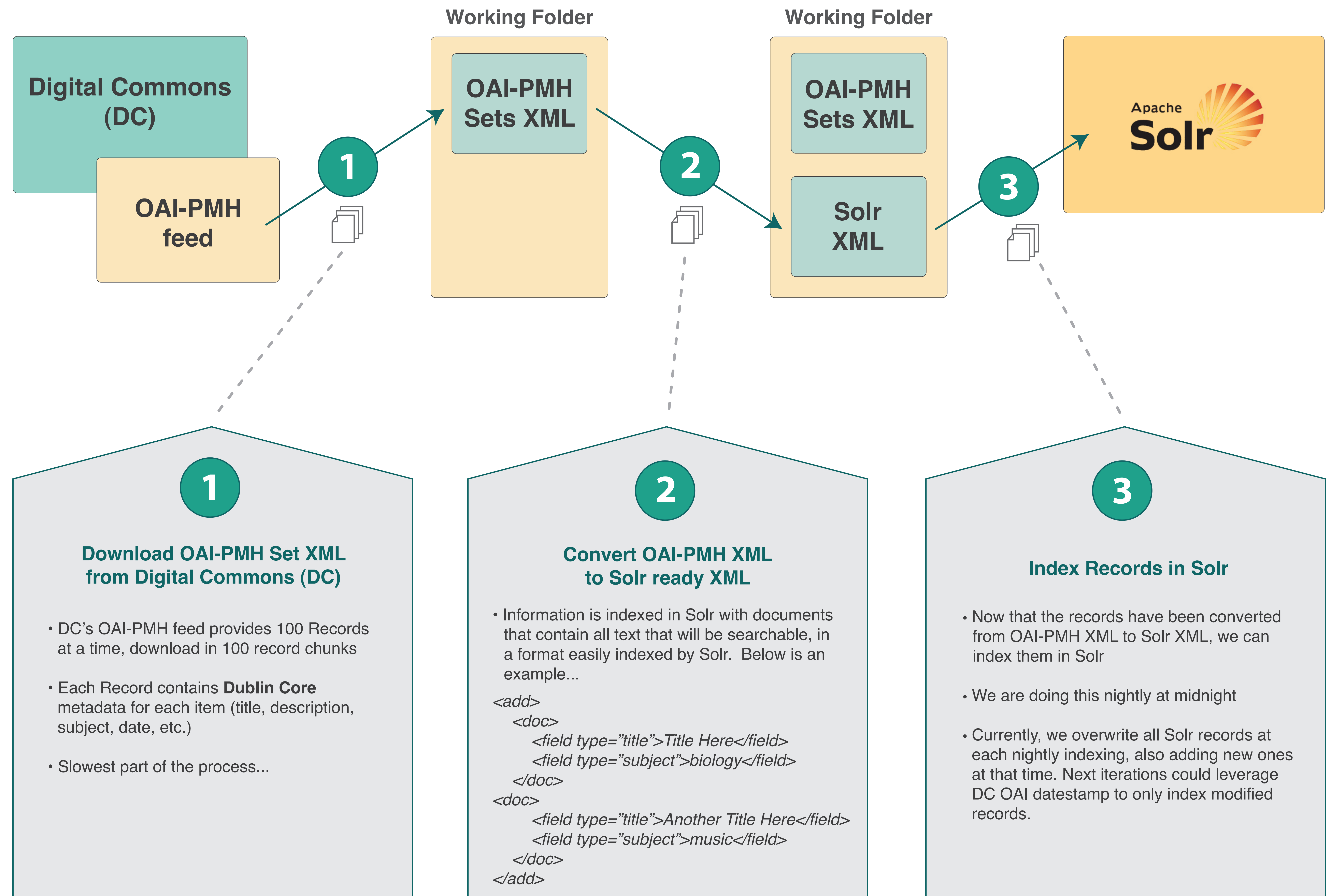


### OAI-PMH

Open Archive Initiative - Protocol for Metadata Harvesting (**OAI-PMH**) is a widely used protocol for sharing structured metadata, in **XML** form, between repositories. Digital Commons (DC) provides metadata records for all items in an institution's repository via an "OAI-PMH feed."

Each **Series** in DC comes through as a **Set** in the OAI-PMH feed. Each **Item** in DC is expressed as a **Record** within a Set. It is these **Records** that are indexed in Solr.

[www.openarchives.org/pmh](http://www.openarchives.org/pmh)



This utility is available on GitHub:  
[github.com/WSULib/dc2Solr](https://github.com/WSULib/dc2Solr)

### Why bother?

- Fully automated, takes approximately **13 minutes for 3,000+ records**
- Potentially pulls Digital Commons records into Library **discovery system** via Solr
- Provides ability to **audit** records from Digital Commons by aggregating them in one place.