

## Open Repositories 2013 Proposal

**Proposed by:**

Sarah Forbes, Scholarly Communications Librarian – [sforbes@utsc.utoronto.ca](mailto:sforbes@utsc.utoronto.ca)  
Kelli Babcock, Special Projects Librarian – [kbabcock@utsc.utoronto.ca](mailto:kbabcock@utsc.utoronto.ca)

**Institution:**

University of Toronto Scarborough Library  
Room AC270  
1265 Military Trail  
Toronto, ON, Canada  
M1C 1A4

**Submission Title:**

*Reproducing Metadata Between TSpace and Bioline International: A Study of Managing Interoperability*

**Type:**

User Group Presentation (DSpace)

**Description:**

This DSpace user group presentation will describe the practical measures taken to incorporate the records of Bioline International into the TSpace repository. It will outline problems encountered, lessons learned and the broader benefits of improving networks of scholarly output through interoperability.

Drawing from the Confederation of Open Access Repositories' *The Current State of Open Access Repository Interoperability* report (2012), the presentation will explore the value of open access repositories, such as TSpace, becoming "an interconnected repository network – a network that can provide unified access to an aggregated set of scholarly and related outputs that machines and researchers can work with in new ways". Improving the interoperability between Bioline and TSpace will allow for expanded access to Bioline content. TSpace's ability to ingest Bioline records has far-reaching implications for improving the global visibility of the under-represented content submitted to Bioline from the developing world.

Bioline International, directed by University of Toronto Scarborough (UTSC) faculty member Leslie Chan, is an open access platform for biomedical journals published in developing countries. Bioline is a long-standing collaboration between UTSC and the Reference Center on Environmental Information (Centro de Referência em Informação Ambiental, or "CRIA"), based in Brazil.

In 2012, the University of Toronto Scarborough Library's Digital Scholarship Unit committed its services to supporting the Bioline International project. UTSC Special Projects Librarian Kelli Babcock has been coordinating the Bioline International project since May 2012. Many improvements have been made to the Bioline system over the past year - such as streamlining content ingest work-flow and XML creation. Bioline looking at ways of improving its metadata interoperability with other repositories in 2013 - for example, the use of Extensible Stylesheet

Language Transformations (XSLT). Interoperability is essential to Bioline's ability to re-distribute its metadata to other open access repositories and improve the visibility of its member journals. The TSpace repository is a partnership between the University of Toronto communities and the University of Toronto Libraries. Using the DSpace software, its content consists of collections produced by the communities, which are managed, preserved and distributed by U of T Libraries through TSpace.

TSpace contains the Bioline International Legacy Collection, which is made up of Bioline-hosted content published between 1990 and 2006. The Bioline International Legacy Collection was created by retrieving records from the Bioline server to the TSpace server. This retrieval included the metadata of each article (in qualified Dublin Core format), the full text of the article in PDF/HTML format, and any images associated with the HTML versions of the article. In the past, the process of retrieving records from the Bioline server and depositing these into TSpace was laborious due to the inconsistency of Bioline's metadata and inconsistency of full text file formats. The retrieval of records to populate the Bioline International Legacy Collection in TSpace involved a great deal of manual work as well as customized short scripts that tackled small problems in uniquely identifying records and verifying metadata. The harvesting of Bioline to TSpace stopped in 2006.

One impediment to interoperability that can slow down the process of sharing records is communication. In 2013, Bioline is working with UTSC Scholarly Communications Librarian Sarah Forbes to renew the retrieval of Bioline records into a newly created Bioline International community in TSpace. Some of the work being done includes frequent consultation between "content provider" (Bioline) and "repository administrators" (TSpace), reviewing old retrieval and script processes, and examining record ID lists to avoid duplication. As part of this process, we recognized that it was important to keep each other informed of our system's updates, such as TSpace's planned update to the current DSpace 3.1 release.

More broadly, we are looking to identify and create new procedures for managing the interoperability of Bioline to TSpace that could have implications for other repositories that may be ingesting content from either Bioline or TSpace. The presentation will highlight some of the lessons learned regarding this process, including the human resources requirements and sustainability planning involved in interoperability projects of special OA collections and institutional repositories. Using Bioline and TSpace as an example, we will share our experiences with others in the DSpace community to share ideas on how to resume and sustain relationships between repository manager and content provider.