Enhancing DSpace to Synchronize with Sources having Distinct Updating Patterns

Manuela Klanovicz Ferreira, Afonso Comba de Araujo Neto, Raquel Almeida, Zaida Horowitz, Janise Silva Borges da Costa, Caterina Groposo Pavão
Universidade Federal do Rio Grande do Sul - UFRGS - Centro de Processamento de Dados, Rua Ramiro Barcelos, 2574, 90035-003 Porto Alegre, RS, Brasil, Phone: +55 (51) 3308-5029/3308-5071, E-mail: comissao@cpd.ufrgs.br

Abstract
Lume is the Digital Repository of Federal University of Rio Grande do Sul (UFRGS - Universidade Federal do Rio Grande do Sul) and aims at fostering the dissemination of digital content produced in the context of this University. The data sources for the repository can be classified into three groups: institutional cataloging systems, institutional event's management system, and special archives which items are submitted manually in an individual basis. The poster presents a detailed diagram that shows Lume's customized usage of DSpace, highlighting the developed extensions that allow the synchronization of Lume with the several data sources that have distinct updating patterns, and how the provision of embedded videos was enabled.

Lume - Digital Repository of UFRGS
Lume is the Digital Repository of UFRGS and was implemented in 2008 using DSpace - Institutional Repository System. Its purpose is to preserve, disseminate, and increase the use and visibility of documents produced by the University that are preserved due their scope and/or historical characteristics.
UFRGS is more than 100 years old and covers several fields of knowledge, with 90 undergraduate courses, 71 master programs, and 68 doctoral programs distributed on four centers in Rio Grande do Sul, Brazil.

Using DSpace to Centralize and Disseminate the Information
The University has several information systems with distinct purposes, responsible for the management of a fair amount of digital information. However, that information was, before Lume, visible only to systems managers. In recent years, the interest of disseminating digital data to the public has increased significantly, as a means to promote the Universities intellectual production. To take advantage of the data archiving and management already in place in the source systems, the importing of these data to the institutional repository was automated by several DSpace extensions implemented by UFRGS.
Lume has two main data sources for which importing was automated:

- *Institutional cataloging systems*: this set includes the *SABi* (the library automation system, based in a private closed system) which records are updated every day by librarians, with a fair number of corrections and exclusions that have to be accounted for; and the *Acervo Fotográfico* (Photographic Archive), the system used to catalog photos taken every day in several University departments. For these two systems, the importing of data to the repository is made on a daily basis, allowing the inclusion of new items on Lume and the update of information of items imported before.
- *Institutional event's management system*: the main goal of this system is to provide a flexible environment that can be used to manage essential information related to several types of events promoted by the University. The use cases covered by this system include submission processes, advisor's reviews, approval of external institutions, registration of participants, organization of the presentation sessions, and choice of awarded works, among others. For this system, items are imported on demand, at the end of each event, and only after all the data is completely committed into the system. The online availability of the work presented in these events has the additional advantage of avoiding the necessity to produce event's proceedings using physical support.

Additionally to the aforementioned imports of data, individual submission of new items directly into DSpace is available for some members of University, allowing them to publish information not yet being managed by any other system. These members use the institutional LDAP to authenticate and have permission to manage a few special communities.
Finally, to allow streaming of video embedded directly in the items web pages, DSpace is being integrated with the Kaltura Video Platform, which proved to be a robust tool capable of enabling direct video streaming seamlessly to almost all existing clients' platforms for most existing videos formats.

Conclusions
Integrating distinct systems is not an easy task. Even when managing common data, each system is designed for a specific administrative purpose, so they have distinct architectures and data models. In this sense, the adoption of patterns and standard communication interface in the development of information systems is essential to simplify its interoperability.