

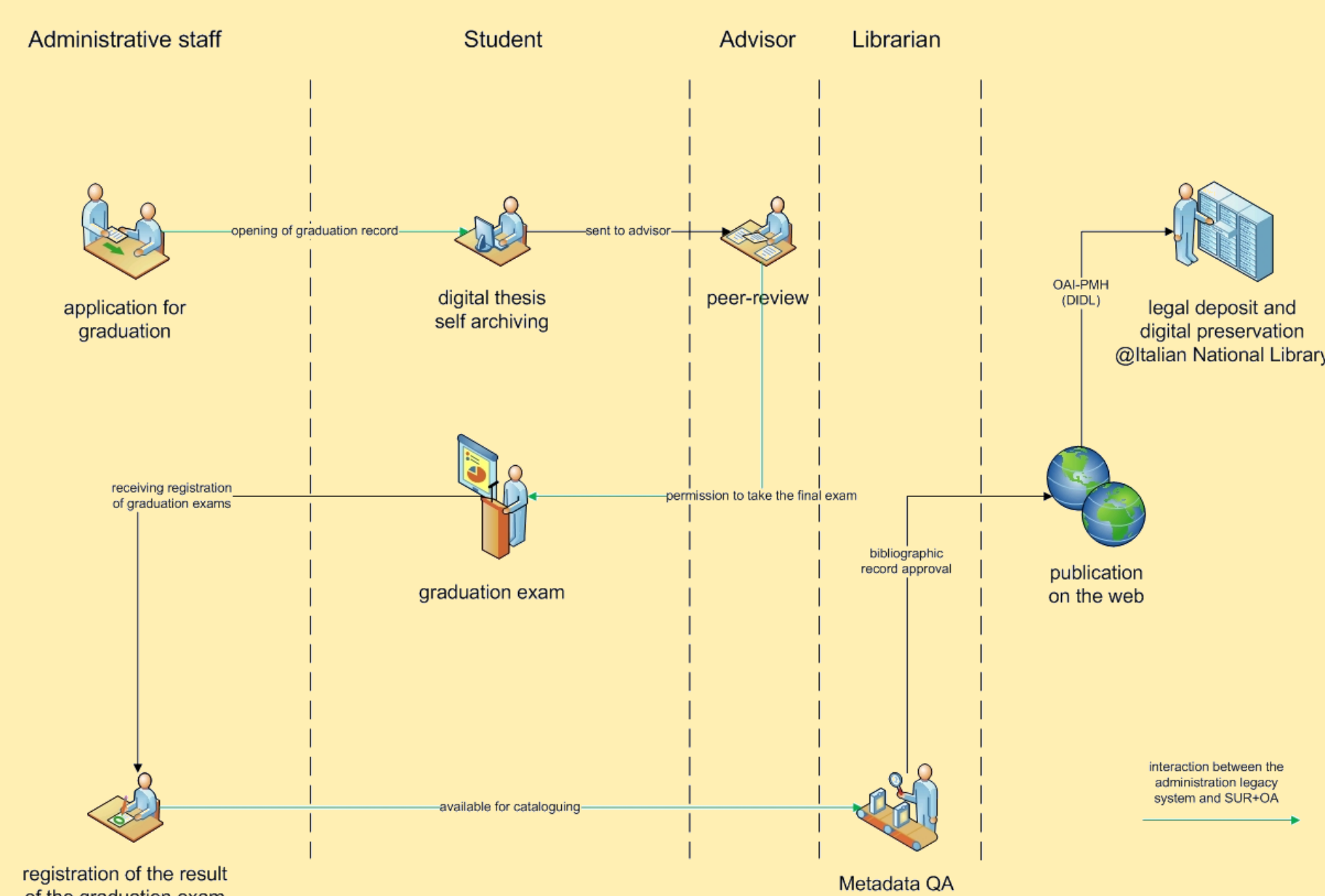
# DSpace & System Integration: the doctoral theses challenge

Andrea *Bollini*, Susanna *Mornati*, Andrea *Pascarelli*, Davide *Vitale*  
*CILEA, Italy*

Doctoral theses often represent the upper edge of academic research activities. As such, they deserve the highest visibility and dissemination opportunities, for the benefit of their authors and the supporting institutions. Their impact would also contribute to prevent the marginalization of Italian research in the European and world-wide context. Changes in norms and rules are also driving universities towards automation of processes and adoption of digital formats. In this situation universities look for software solutions that can be integrated in their information environment, contribute to the automation of the archiving process and provide theses with high visibility on the net. CILEA fulfils these needs with the OA module of the SURplus suite, built on top of the DSpace software and customized for the specific treatment of dissertations

*two minute drill...*

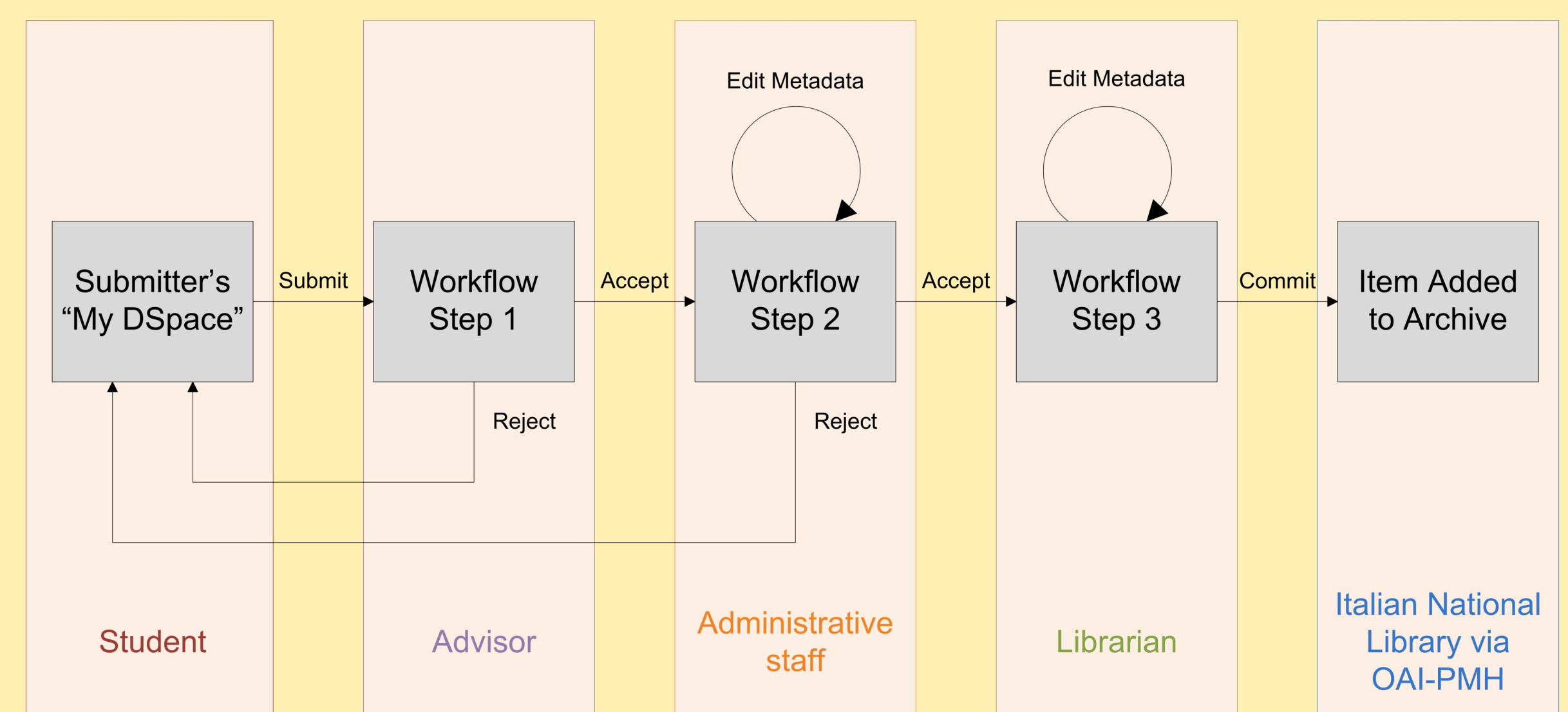
Based on the analysis of several experiences achieved so far, we highlighted a common workflow spanning different Information Systems at universities, from the Students Office Management Software to the Repository of Research Materials to the Researchers' Output Evaluation Software.



DSpace integration with legacy systems at universities was made possible by introducing a new layer of service to ingest and manage the whole life cycle of items. Through the DSpace public APIs, this service can create items in any submission state (workspace, workflow, published), with or without bitstreams. It can also update (replace and append) metadata, bitstreams and status (forward or rollback) of existing items, based on events. The new layer has two operational modes: batch jobs (PULL interaction) that read instructions and data from temporary database tables, populated at desired frequencies by ETL procedures from legacy systems, or SOAP web services (PUSH interaction) that expose methods for items management. Metadata are expressed in XML as SOAP external attachments to guarantee procedure customization through simple XSL transformations.

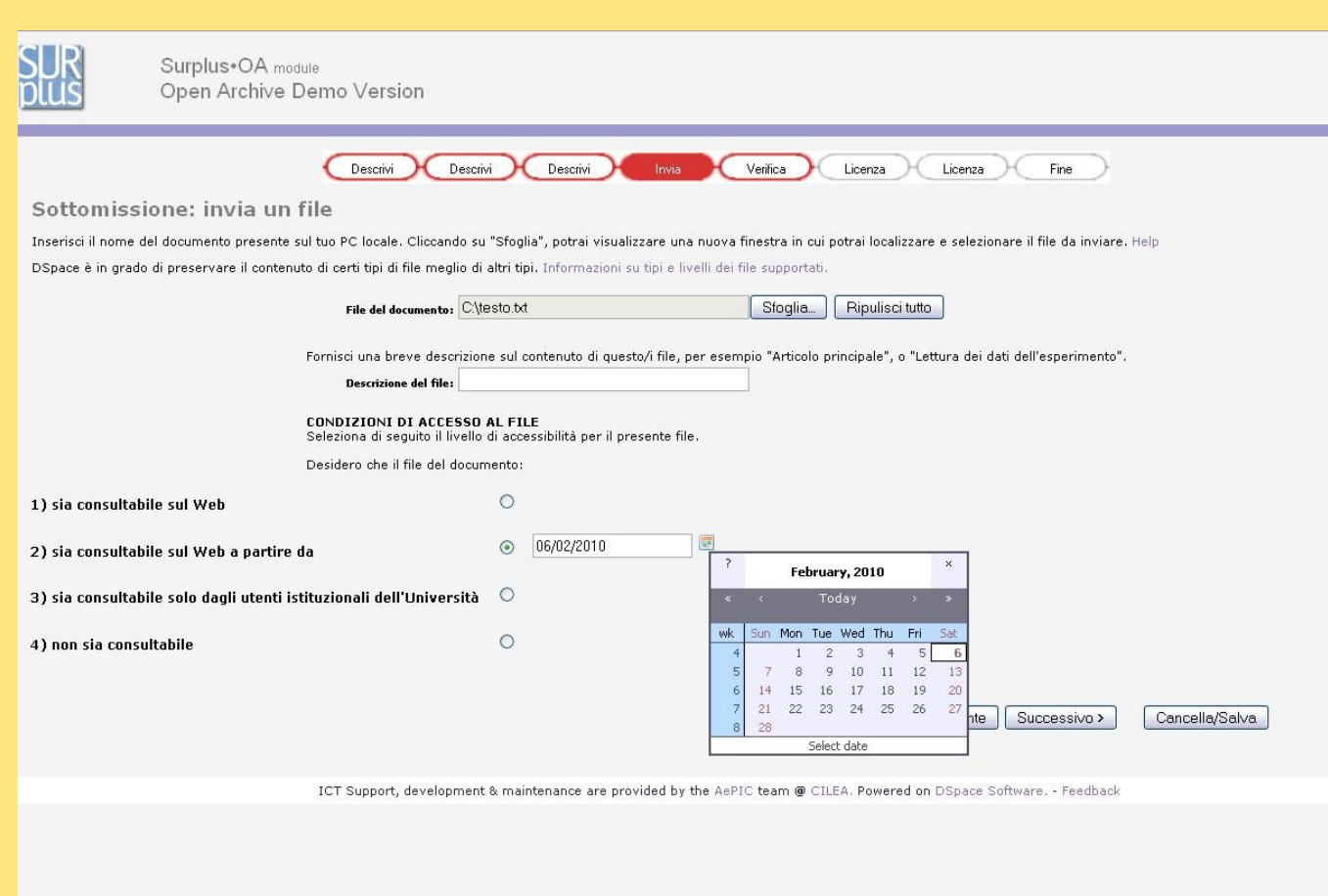
*integration...*

Theses items can be created using the normal DSpace UI or are more often preloaded by an automatic procedure. The student is mapped to the DSpace submitter role. The advisor is responsible of workflow step1, so a final check of the student work is performed before the exam. Notifications were improved to alert the advisor of the approaching deadline. Automatic approval can also be scheduled if no explicit response is provided by the advisor. Administrative staff is responsible of workflow step 2: if no administrative issues arise, theses are archived and items are enriched by other descriptive metadata (e.g. vote) and automatically moved to the next workflow, where librarians can edit metadata for further improvements.



After their publication in the repository, the Italian National Library harvests theses via OAI-PMH. Presently the service is at an experimental phase and the DIDL format has been chosen to represent both metadata and full-text items (bitstreams), since Italian national law requires that doctoral theses must be publicly available at the National Library. Selective harvesting of bitstreams can be easily achieved by enabling the IPAuthentication method and assigning a DEFAULT\_BITSTREAM\_READ action to the National Library Workstations group.

*workflow...*



Specific access policies need to be applied to PhD theses full-text files. Most universities allow students to choose among a pre-determined set of policies from Open Access to limited access (e.g. restricted to staff), via intermediate solutions such as an "embargo" policy\* (restricted access for a defined period of time). The screenshot illustrates 4 options: Open Access, limited embargo, limited access, no access. Limited access can be regulated via authentication systems such as Shibboleth that allow groups to represent shared categories beyond the borders of a single institution (e.g. researchers in Italy).

*what we need...*

\* The embargo functionality is based on the patch proposed by Leiden University, The Netherlands, available here:  
[http://wiki.dspace.org/index.php/Embargo\\_on\\_Bitstream\\_\(JSP\)](http://wiki.dspace.org/index.php/Embargo_on_Bitstream_(JSP))

## *...what is shared?*

**DS-309.** Bug Fix for the management of Shibboleth default groups. To be included in the next DSpace release 1.6.0. A patch to configure Shibboleth unique identifiers different from the email address is foreseen.

**DS-206.** Capacity to show field content for reading only during the submission/workflow process. Useful when data come from administrative legacy systems and must not be modified. To be included in 1.6.0.

**DS-107.** Bug fix for the DIDL crosswalk. Released in DSpace 1.5.2.

**DS-271.** Configurability of oai\_dc format. A patch to configure the DIDL format is foreseen.

**DS-236.** The SUR+OA module draws some data from the authoritative personnel database (e.g. thesis advisor). At the moment this is achieved through a local customization of DSpace, while the DS-236 patch aims at building a framework where these needs can be addressed in a pluggable way. The patch has been developed by Larry Stone,

our contributions are on the JSPUI/PostgreSQL version and in some improvements to the base code where we have added the management of variants and the support to mixed browsing indexes (where "certified" and non-certified terms live together).

**DS-217.** Capacity to use substitution variables in the submitter license. Data on uploaded files and their access policies can be included. To be included in a minor release after 1.6.0.

**DS-267.** Proposal to modify the storing mechanism of policies during the submission phase. The patch addresses problems that arise when an item is removed and then published again (access policies were lost). A related patch can be released to manage embargo policies on single bitstreams, that can be highly configurable (number, order and default contents can be defined for each collection). A patch to configure if the bitstream upload is mandatory/optional on a collection basis is also foreseen.



