

# The Transformation of the Ktisis Repository into a Current Research Information System (CRIS)

Alexia NTINI-KOUNOUEDES<sup>a,1</sup> and Marios ZERVAS<sup>a</sup>  
<sup>a</sup>*Cyprus University of Technology*

**Abstract.** Institutional repositories have played a major role in universities worldwide during the last decade. Such systems are developed with the aim to collect and disseminate the research activities of universities. They provide access to and showcase research outputs, and therefore they have become an essential infrastructure for universities. A repository provides the means to properly preserve research outputs and can also be used for research monitoring and assessment. In this case study, we concentrate on the transformation of Ktisis, the institutional repository of the Cyprus University of Technology, into a Current Research Information System (CRIS). A CRIS system records, processes, and presents metrics and figures related with research activity throughout its life cycle. Particular emphasis is given to the results of research activities (publications, patents, research data) and their connection with the environment within which they were created (researchers, organizations, funded programs and projects, research infrastructures, services). In this case study we will describe the procedures followed in order to transform Ktisis into a CRIS system together with the implementation of the integration of ORCID identifiers within the system. Particular attention will be paid to the challenges we came across throughout the process and how we overcame these difficulties and problems. Ktisis is the institutional repository developed and maintained by the Library and Information Services at the Cyprus University of Technology. Ktisis was created in 2008 using the open source software DSpace after the University's Interim Governing Board made the decision that all the research products of academic members must be deposited in the Library. In subsequent years, the mandatory deposit of undergraduate, MSc and PhD theses was also imposed. In early 2015 it was decided that the Cyprus University of Technology (CUT) must become a member of ORCID and the Library undertook the project to integrate ORCID in its systems. At the same time the Library decided that the best way to move forward was to transform Ktisis into a CRIS system using DSpace-CRIS, an extension to DSpace, in order to integrate ORCID with the repository and to take advantage of all the functionalities provided by a CRIS system.

**Keywords.** repositories, CRIS, ORCID

## 1. Introduction

Institutional repositories have played a major role in universities worldwide during the last decade. Such systems are developed with the aim to collect and disseminate the

---

<sup>1</sup> Alexia Ntini-Kounoudes, Library and Information Services, Cyprus University of Technology, Cyprus; Email: marios.zervas@cut.ac.cy.

research activities of universities. They provide access to and showcase research outputs. As such, they have become an essential infrastructure for universities. A repository provides the means to properly preserve research outputs and can also be used for research monitoring and assessment. In this case study, we concentrate on the transformation of Ktisis, the institutional repository of the Cyprus University of Technology, into a Current Research Information System (CRIS). A CRIS system records, processes, and presents metrics and figures related with research activity throughout its life cycle. Particular emphasis is given to the results of research activities (publications, patents, and research data) and their connection with the environment within which they were created (researchers, organizations, funded programs and projects, research infrastructures, and services)<sup>2</sup>. In this case study we will describe the procedures followed in order to transform Ktisis into a CRIS system together with the implementation of the integration of ORCID identifiers within the system. Particular attention will be paid to the challenges we came across throughout the process and how we overcame these difficulties and problems.

Ktisis is the institutional repository developed and maintained by the Library and Information Services at the Cyprus University of Technology. Ktisis was created in 2008 using the open source software DSpace<sup>3</sup> after the University's Interim Governing Board made the decision that all the research products of academic members must be deposited in the Library. In subsequent years the mandatory deposit of undergraduate, MSc and PhD theses was also imposed. In early 2015 it was decided that the Cyprus University of Technology (CUT) must become a member of ORCID<sup>4</sup> and the Library undertook the project to integrate ORCID in its systems. At the same time the Library decided that the best way to move forward was to transform Ktisis into a CRIS system using DSpace-CRIS, an extension to DSpace, in order to integrate ORCID with the repository and to take advantage of all the functionalities provided by a CRIS system.

## 2. Implementation

### 2.1. *The ORCID Project*

The task of identifying researchers and linking them to their research work is difficult and challenging since the researchers can be very active and they frequently move between organizations. Additionally, they often use different variations of their name in their work which means that the information about them needs frequent updating, a complicated procedure in and of itself.

The introduction of CRIS systems comes as a solution for collecting, managing, preserving, analyzing and showcasing the research output of institutions, providing the use of persistent identifiers for uniquely identifying researchers. These identifiers that are being used are the ORCID IDs, which can be used by a researcher throughout his career.

The Library worked on the implementation of ORCID for researchers by creating a portal (<http://library.cut.ac.cy/orcid/>) where the researcher can create an ORCID or connect an existing one with the University's system, thus enabling the authentication,

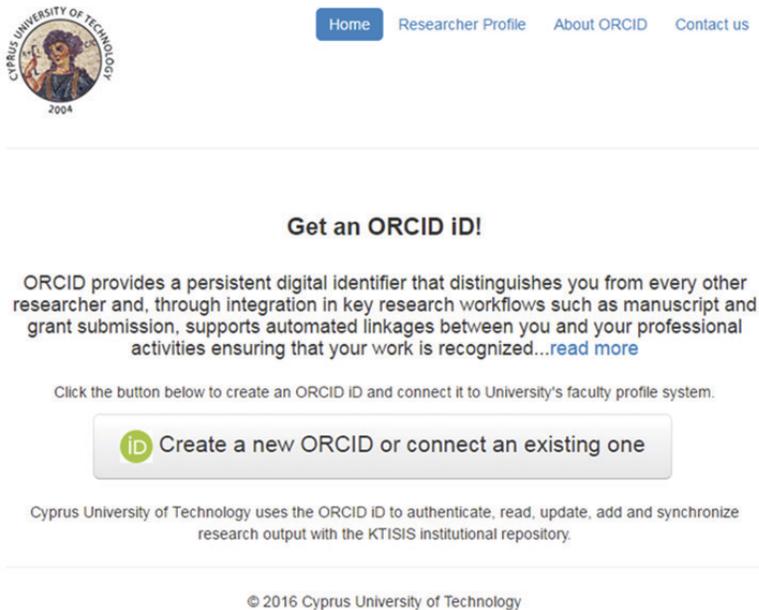
---

<sup>2</sup> <http://metrics.ekt.gr/en/systimata-cris>

<sup>3</sup> <http://www.DSpace.org>

<sup>4</sup> <http://www.orcid.org>

reading, updating, adding and synchronizing of research outputs with the Ktisis repository.



**Figure 1.** The Library's ORCID portal

When a researcher creates an ORCID through the platform, details such as their name, ORCID id, authorization token and ORCID creation date are written in a log file in the system server which can later be used by the Library for various purposes.

At the moment, the latest version of DSpace-CRIS being used in Ktisis does not support the automatic synchronization of the researchers work between DSpace and ORCID. This is a feature of the system that will be available in subsequent releases. Therefore, we use this file to export the ORCID data of the researcher we are interested in. The available options are to export: the entire profile data for a researcher, bibliographic data, research work data, funding data, or affiliations data. These options are the available information that can be exported through the ORCID. We are using an interface where we supply the ORCID with the researcher we are interested in; the system checks in that log file to see if that researcher already has created an ORCID through Ktisis and if a match is found then it proceeds with the export of one of the available options mentioned above.

However, the problem with the exported file is that it is structured in the ORCID-xml format which makes it difficult to process and transform it into a CSV format suitable for importing into DSpace. The Library at this moment is trying to find a way to convert the exported file with the researcher's works data into a CSV file in order to be able to import this data into Ktisis.

The combination of ORCID and DSpace-CRIS is very powerful as it utilizes the ORCID's ability to improve data quality by connecting researchers with their research while also adding significant value to their work.

## 2.2. *The Transformation of Ktisis to CRIS*

Ktisis was created using the open source software DSpace in 2008. In the following years the Library upgraded the system according to new releases, until it was decided that the best way to move forward and satisfy the need for the use of persistent identifiers, and to be able to handle the growing volume and variety of research information, was to transform Ktisis into a CRIS system. The DSpace community had already made available an extension of DSpace, the DSpace-CRIS module, enriching DSpace with CRIS entities and concepts. DSpace-CRIS extends the DSpace data model providing the ability to collect, store, manage, retrieve, expose and exchange data about all research entities (people, organization units, projects, grants, awards, patents, publications, etc.).

The Library collaborated with the company 4Science<sup>5</sup> to transform Ktisis into a CRIS system as previous attempts to perform this task on our own were unsuccessful. The plan was first to clone the live Ktisis system into a testing server so we have a spitting image of the system at that moment. Then all the work needed in order to upgrade Ktisis into a CRIS system was going to be carried out on this staging server and when everything was tested and fine-tuned, the same changes were going to be applied on the live server.

Subsequently, the project was split into 3 phases. The first phase involved the creation of the specifications for the new system and meeting the pre-conditions set, i.e. providing the 4Science staff with SSH and VPN access to the staging environment. The same access was also provided for the production environment.

We asked the Library staff working on Ktisis to stop all activities including the submission of new material. At this point the live system was cloned into the test server so that all the changes would be applied and tested there first.

When these tasks were finished we moved on to Phase 2. The first task in this phase was to configure and create the necessary scripts needed to convert the current repository structure into the structure used in the DSpace-CRIS version. This work was done by the staff of 4Science. When this was completed and applied to the test server, they proceeded with the installation on the staging server of the new configuration leading to the finalized DSpace-CRIS version. A database dump was also taken from the live system that would be imported on the staging server after the installation of the new software was in place.

Before the installation started, we had to perform a number of tasks. We were advised at this point that Tomcat was configured to deal directly on port 80, which is a bit unusual and not the recommended approach since it reduces the performance of the system. The solution to this was to have Apache HTTP run in front of Tomcat. These changes saw the visible improvement of the system performance when the system was up and running. Additionally, it was observed that SOLR<sup>6</sup> was running using the same Tomcat service; therefore, we created a dedicated Tomcat instance for SOLR and noticed again that this provided better performance. SOLR is the indexing server used

---

<sup>5</sup> <http://www.4science.it/en/>

<sup>6</sup> <http://lucene.apache.org/solr/features.html>

as a part of Discovery in DSpace to speed up access to content metadata and statistics and it also provides faceting, search results filtering and “More like this” functionality.

We then had to redesign the community hierarchy and data structure. Since DSpace-CRIS is centered on researchers and their work, we changed the community hierarchy structure so that it was not based around faculties and departments anymore but rather on separate entities for each document type we stored in our system (for example articles, books, conference papers, etc.). Then each item was connected to the researchers that had authored it and was allocated to the correct collection.

The data model of DSpace-CRIS entities defines the following 1st level entities: people, organizational units and projects. The first entity we had to think of in terms of design was the organizational units. An organizational unit was created in order to represent the University as a whole and then other organizational units were created to represent each faculty and department. Additionally, we had to list together all the external affiliations that existed in the system database (as collaborations) and make sure that these values were normalized before they were created in the new system. This was important as we wanted to have a system clear of data discrepancies and duplications. Having such a hierarchy provides better statistics for the faculties and departments and also for individual researchers.

When the new community hierarchy was decided we had to provide the mapping between the community hierarchy in the old system and the new one in order to be able to move the items in the database to the correct location in the new CRIS system. Each collection that existed in the old system was mapped to a collection in the new system, using the collection’s name and handle in the old system.

At this stage we also had to devise a list of the researchers of the University, their department and faculty and their ORCID IDs. This file was then used in order to import the researcher names into the system, map them to the correct faculty and department and subsequently to create a researcher page for each one of them. Then specific procedures were followed in order to connect their research output already available in Ktisis with their profiles. This way the researchers have access to a dedicated set of functionalities. They can edit their personal information and decide what the visibility of this information will be (public or hidden). The researchers are also able to manage the research output connected to their profiles, such as publications. Furthermore, they can connect their profile with their ORCID and manage the synchronization preferences available in order to transfer information from Ktisis to their ORCID profile, automatically or manually. At the moment only the ability to synchronize the researcher profile is available. The ability to synchronize the researchers’ works is currently under implementation by the developers of DSpace-CRIS.

When the installation was complete and the above mentioned tasks were finished, we had to go through the system in order to test if everything was according to specifications. The Library staff spent a lot of time reviewing the system and we were in constant communication with 4Science in order to fix any problems that were observed. We allowed a full month for the testing procedure as we needed to make sure that everything was in the right order. When both sides were satisfied with the results, we proceeded to Phase 3 which included the application of all the configuration changes to the live system. Again we went through a short period of testing the new system until we were in the position to go live.

### **3. The Day After**

When the new system was live, it was decided that it would be presented to the various departments of the University in order to get feedback from the researchers. We were very keen to find out what they thought about the new Ktisis and the new functionalities it provides. Before we started the transformation of Ktisis to a CRIS system, the Library promoted the ORCID functionality and the benefits that it provides to the researchers. It was at this point that the University senate decided that all the researchers of the University must acquire an ORCID, giving the Library greater motivation to go on with the project.

Most of the feedback we received from the researchers was very positive. They really appreciated the ability to have all the information about them and their research work gathered together under their Researcher Profile. Of course there were cases where some researchers were not very keen with the idea of introducing bibliometrics to the system, showing how they rank up in the databases that are being tracked (Web of Science and Pubmed), but these were isolated cases.

Using the system, we realized that there are great functionalities available for administrators as well. Using the DSpace-CRIS software makes it easy to perform tasks that had to be done manually in the past or through running SQL queries on the database, such as exporting statistics for particular departments or organizational units, or batch updating the records through the import/export functionalities available.

Additionally, the task of submitting new items in the system is also easier now. Using only the publication identifiers, the user is able to search through the available databases. A list of matching publications is subsequently shown in order to proceed with the submission process, thus saving time that was needed to manually enter all the details of a publication.

Another important addition in the new system is the introduction of Projects. Projects is another entity of DSpace-CRIS where information about various projects undertaken by the University's researchers can be found. The system has been configured in such a way so that information such as the project title, project abstract (i.e. details about the nature of the project), and some primary data (such as the project coordinator, project start and completion date as well as the current status of the project) are all displayed. The most important functionality here is that the list of all the publications that were carried out during this project can be seen on the project page. This is achieved by connecting the publication with the project through the Dublin core field that was assigned specifically for this purpose.

### **4. Lessons Learnt**

The project of transforming Ktisis into a CRIS system was not an easy and straightforward task. We have tried in the past to perform the transformation on our own since the software is open source and available for everybody to use. However, since the design and structure of DSpace-CRIS is totally different to the simple and easily customizable DSpace software, we did not succeed in creating the system we longed for. There is a lot of work that needs to be done in the configuration, especially for existing systems and since this is a relatively new release, the documentation available was not enough. After two failed attempts to configure the software on test servers and since we considered this improvement to be of great importance for the

University, we decided to ask for help from the original developers of the software and begin a collaboration with them in order to speed things up. Our experience showed that the transformation of a DSpace system to DSpace-CRIS is a procedure that is achievable but it will require a vast amount of time to reach the desired result.

When we started this project we decided that we would keep the infrastructure as it was. Having followed the guidelines of the University's IT team, the way that the software and files had been installed from the very first DSpace installation was how the installed files and middleware were stored on the NFS storage of our network. After a few weeks of using the system, it was observed that the NFS storage was rather slow when writing to disk, thus worsening the system performance. On top of that the database connections were released more slowly, using more RAM. We were not aware that this infrastructure was making our system less robust with reduced performance. The Library will fix this problem by moving the middleware, such as Postgres and Tomcat, and the DSpace-CRIS installation folders onto the local server partition to reduce these issues. Hopefully, this change will increase the performance of the server and minimize the occasional problem of the server being down due to improperly closed database connections.

Another problem that we observed throughout the use of the new system is that we have to be very careful with the submission process of new items. We have already been through the procedure of normalizing the external organizational units that have been created in the system in order to eliminate data duplications and discrepancies. However, we noticed that there are cases when new items are submitted where the submitter (authorized library staff) does not pick up the value from the drop down list that shows the available organizations but instead types in a new one. This has caused problems with the data since the numbers of items belonging to an organizational unit is not correct. It is an easy task to fix the data but difficult to locate the problem. The data can easily be fixed by exporting all the items for the specific "invalid" organizational unit in a CSV format, make the necessary changes of fixing the name of the unit and then importing the file back to the system using the import functionality. Then the invalid organizational unit can be deleted from the administrator user interface.

One of the most important lessons that were learnt throughout this project is that the DSpace-CRIS configuration is messy and very delicate. What we have done in order to avoid any problems on the live system is kept the staging environment in place so that any changes we wish to do are first applied there. If no problems occur, we proceed with applying the change on the live environment. What we plan to do in the near future is to again clone the live system into the staging server as to sync the two systems – since a number of changes were applied directly on the live server, causing a few problems.

## **5. Future Work**

One of the tasks that will be undertaken soon is the creation of translation files for the Greek language for the key messages that are missing from the message catalogue. This is important to us since Ktisis comes both in Greek and English languages and we want to have a proper Greek version of the system. The files produced will be made available to the community through 4Science, the developers of DSpace-CRIS.

Another task that we will work on in collaboration with the DSpace-CRIS developers is the implementation of the capability to import a researcher's work from ORCID to DSpace-CRIS, and vice-versa. This is very important for the Library since using this functionality will make it easier to import all the research output of the University's academics and have a complete listing of their publications.

Additionally, we will continue the effort to keep the researchers' publications up to date by exporting the data from databases such as Web of Science and Scopus and importing them to the system. This, in combination with the new functionality of importing publications directly from a researchers ORCID profile will keep Ktisis up to date.

The Library has put a lot of effort on this project and will continue to work in order to keep the system up to date following new releases of the software and implementing any new functionalities that will be made available in future releases.

## References

- [1] David T. Palmer, Andrea Bollini, Susanna Mornati, Michele Mennielli, Achieving Visibility with a CERIF Compliant Open Source System. *Procedia Computer Science*, Volume 33, 2014, Pages 118-123. 12th International Conference on Current Research Information Systems, CRIS 2014 – Managing data intensive science – The role of Research Information Systems in realising the digital agenda.
- [2] Nick Sheppard (2010). Learning How to Play Nicely: Repositories and CRIS. Friday, 30 July 2010, Ariadne Issue 63. Retrieved from <http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt>
- [3] euroCRIS - <http://www.eurocris.org/>
- [4] Miika Nurminen (2014). Preparing for CRIS: Challenges and Opportunities for Systems Integration at Finnish Universities. Paper presented at Open Repositories 2014, At Helsinki, Finland
- [5] Josh Brown (2016). ORCID + Pure: a “happy marriage” that makes researchers' lives easier. Collaboration allows researchers and universities to automatically connect their curated data in Pure with a researcher's ORCID iD. Retrieved from <https://www.elsevier.com/connect/orcid-pure-a-happy-marriage-that-makes-researchers-lives-easier>
- [6] CERIF - the Common European Research Information Format. European Scene by Keith G Jeffery (<http://ercim-news.ercim.eu/en68/european-scene-qsupport-of-the-research-processq/cerif-the-common-european-research-information-format>)
- [7] <http://duraspace.org/now-available-first-release-dspace-cris-module-dspace-3>
- [8] <https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home>
- [9] <http://www.doria.fi/handle/10024/97722>