Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS

Marios Zervas¹, Alexia Kounoudes², Petros Artemi³, Stamatios Giannoulakis⁴

Abstract

This paper focuses on the transformation of the institutional repository KTISIS into a Current Research Information System (CRIS). KTISIS is the first presentation of a European CRIS implemented with the open-source software DSpace-CRIS. In this paper, we present how KTISIS was tailored according to the DSpace-CRIS features in order to satisfy our own repository needs. Our work aims to keep KTISIS up to date with the current status of technologies, following the guidelines for Next Generation Repositories as published by the COAR Next Generation Repositories Working Group.

Keywords: Institutional Repositories; Scholarly publications; Current Research Information System; Next Generation Repositories;

1. Introduction

Institutional repositories are the means for libraries and research organizations to provide access to and disseminate their research output. For this reason, the Library and Information Services at the Cyprus University of Technology (CUT) has developed the institutional repository KTISIS. Institutional repositories need to constantly adapt their mission, goals, and services to meet those of the University and the community it serves. In the framework of the above, the CUT Library has proceeded with the upgrade of KTISIS using the latest DSpace-CRIS version.

This paper focuses on the transformation of KTISIS into a Current Research Information System (CRIS). KTISIS is the first presentation of a European CRIS implemented with the open-source DSpace-CRIS, a technology that allowed us to maintain independence from vendors and make our system evolve according to the needs of our research community. Based on the researchers needs, the main reason for this project was to enable the researchers to submit their academic profile and work to KTISIS. The most important change in KTISIS was the provision of the Researcher Profile, where the researchers have access to a dedicated set of functionalities that give added value to their work and the repository. In this paper, we aim to present how KTISIS was tailored according to the Dspace-CRIS features in
order to satisfy our own repository needs. DSpace-CRIS allows to disseminate our content and implement Open Science. Therefore, in this paper we will talk about how the relevant open access policies in KTISIS are effectively supported by the new system.

The CUT Library is working to adopt KTISIS to the current status of technologies following the guidelines and the suggested behaviors for Next Generation Repositories as published by the COAR Next Generation Repositories Working Group. The aim of the Next Generation Repositories is to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication on top of which layers of value-added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community. In this paper we will show why KTISIS is an example of such repositories and compare the main features of the COAR concept with our own experience for the specific project. Besides metadata, KTISIS can provide content, link between resources, usage interactions and metrics, navigation, dissemination and preservation. We will discuss how the technology we chose enables us to work towards making KTISIS a Next Generation Repository since our future plans include the implementation of such functionalities, for example comments, notifications, etc.

2. Institutional repositories

Institutional repositories were originally developed to provide a solution for the collection, preservation and dissemination of the research output created at universities and research institutions. The importance of such repositories was quickly recognized since they provided the means for the institutions to showcase their academic work. Using open source software gave the ability to developers worldwide to build custom features and functionalities for the systems.

3. CRIS systems

A Current Research Information System (CRIS) is “a database or other information system to store and manage data about research conducted at an institution” (as defined by Wikipedia). We can say that a CRIS is a picture of the current research activities undergoing in a research organization. The scope of a CRIS system is very broad as it covers people (researcher profiles, CVs, supervisions), research outcomes (publications, patents, research data), funds, projects, professional activities (invited presentations, peer or editorial reviews, etc.), or management information (metrics, bibliometrics, statistics). The content of a CRIS is analyzable and reportable. Content within a CRIS system can be easily connected and linked with other content (Fig. 1).

Fig. 1. A CRIS System.
4. Related Work

The integration of institutional repositories with Current Research Information Systems (CRISs) is a relatively new area of interest which has seen a widespread growth at universities worldwide. A lot has been said about the alleged dichotomy between Institutional Repositories and CRISs which has been caused mainly due to the non-commercial initiatives carried out by Universities and specifically their libraries, in order to freely disseminate the research output using IRs, and the fact that the academics are more aware of the research information management activities of CRISs on which their salaries depend when there is an accurate and complete depiction of their research activities on such systems. Therefore, the academics think higher of CRIS Systems and complaints about the need to update both systems with duplicate information are quite common.

IRs and CRISs initially had different aims. CRIS systems aim to collect data about all aspects of the research activities that are carried out at a university including projects, funding, results and technologies. On the other hand, IRs have grown out of the open access movement aiming to collect and freely provide access to the institutional research outputs.

Today there have been many institutions that have integrated their IR and CRIS systems into one single system, such as the University of Hong-Kong (which has been the point of motivation for this work), while others allow a systematic data transfer between the two systems to keep both up to date.

In Finland, there has been an attempt to integrate administrative research information to institutional repositories, in order to create a CRIS system used by various stakeholders in different organizational units. Even though it was preferable to integrate CRIS and IR into a single system, due to user interface issues, recording conventions and other problems, this was impractical for the specific scenario.

Another approach was presented by the Warsaw University of Technology (WUT). They propose a solution for building a system that combines the functionalities of an IT with the functionalities of a CRIS. The work that they have done has used AI technologies in order to provide attractive features in the system. They have reviewed various approaches, analyzed the problems observed in combining IRs and CRISs and show how these problems can be overcome with a single system that integrates functionalities from both areas. They show that by providing analytical tools for research management is an essential part of the proposed solution.

5. The KTISIS case

5.1. The history behind KTISIS

The Library and Information Services of the Cyprus University of Technology (CUT) has developed the institutional repository Ktisis. It was first developed in 2009 when the main purpose was to collect and preserve the research output of the academic staff and researchers of the university and to collect, digitize and disseminate cultural content. Ktisis was developed using the open source software DSpace. When Ktisis was at the designing stage, the Library defined the set of goals that Ktisis needed to achieve. These goals were:

- To locate and archive together cultural heritage items from private collections.
- To guarantee long-term preservation and access to the data.
- To promote interest and involvement in the digitization process and preservation of cultural heritage.
- To promote open access at the Cyprus University of Technology.

In the following years it was decided that Ktisis should only store the produce of the research activities of the university and not cultural heritage content, and for this reason all the data relating to cultural heritage was moved to another digital library system developed by the Library.
After this development, the staff of the Information Technology Office of the Library continued to update Ktisis by performing the necessary software upgrades according to the available releases of DSpace.

In early 2015 it was decided that the Cyprus University of Technology 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 enhance Ktisis into a CRIS system using DSpace-CRIS, an extension to DSpace, to integrate ORCiD with the repository and to take advantage of all the functionalities provided by a CRIS system.

5.2. Aim of the project

As mentioned earlier, the aim of this project was first to integrate the use of ORCiD in Ktisis. 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. 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 ORCiDs, which can be used by a researcher throughout his career.

Additionally, the Library wanted to extend the scope of the institutional repository Ktisis, by adding value to the system and its content. The idea was to use the benefits provided by DSpace-CRIS, which is free, open source and compliant with open standards. DSpace-CRIS provided us with an effective tool to collect and manage research information such as researchers’ profiles, department pages, project grants, research outputs, metrics, reports and statistics. Since the Library wanted to provide a framework to the researchers to submit information about their research activities (such as Participation in Editorial Boards, Committee Memberships, Participation in Professional Societies or Associations, etc.), and also required to include information about Projects and Patents in Ktisis, DSpace-CRIS was the ideal solution for our project.

5.3. Implementation

5.3.1. Starting the project

For the implementation of this project we collaborated with the company 4Science who provided support and assistance throughout the project lifecycle.

After some initial work which involved the creation of the specifications of the new system and meeting the preconditions set, such as providing SSH and VPN access to the staging and production environments to the 4Science staff, we continued with the redesign of the community hierarchy and data structure. DSpace-CRIS is centered on researchers, therefore we had to change the hierarchy structure to be based around organizational units which are then connected to the researchers.

5.3.2. The transformation

Initially Ktisis was created using DSpace, which is built on a simple data model based on publications. The DSpace-CRIS software approach is more dynamic and flexible and is based on publications, datasets, researcher pages, projects, organizational units and other 2nd level dynamic objects. Only a few entities used in DSpace-CRIS are predefined, and the system gives you the ability to create your own entities according to the needs of your repository.

The data model of DSpace-CRIS entities defines the following 1st level entities: people, organizational units and projects. In the process of redesigning the system, the first entity we had to think of in relevance to the design was the organizational units. We created a new organizational unit which represents the University entity as a whole, and then separate organizational units were created to represent each faculty and department. We also had to create new organizational units to represent all the external affiliations we used in the old system. These values were normalized before they were imported in the new system, so that the data in the system was clear from discrepancies and duplications. Using such a hierarchy provides better statistics for the researchers.
Then we devised a list with the names, the department, faculty and ORCiD of all the University researchers, which was imported into the system and mapped to the correct department and faculty. Consequently, a researcher page was created for each of the researchers. The next step was to follow specific procedures to connect their research output with the profiles. This way the researchers have access to a dedicated set of functionalities, such as editing their personal information, deciding the visibility of their profile, managing their research output, connecting their profile to their ORCiD and managing the synchronization preferences to transfer information between their Ktisis and their ORCiD profile.

5.3.3. New functionalities

5.3.3.1. Researcher Profile Enhancement

The available DSpace-CRIS features were tailored in order to serve our own needs. We wanted to provide to the researchers the ability to enter into the system data regarding their academic activity, a functionality that will be a useful tool for their assessment. This has been achieved with the implementation of the Researcher Pages or Profile. The researcher page offers an aggregated view of all the activities carried on by the researcher, their publications, projects, events, etc.

The researchers are now able to include information in their profile such as Committee Memberships, Contribution in Event Organization, Participation in Editorial Board, Invitation as Guest Speaker, etc. Including such important information in their profiles increases the visibility of the profiles and the researcher’s impact and networking.

We have created these new entities as new collections which are only visible to users connected to the system. Each researcher is responsible for entering his own data into Ktisis by using the “Start a new submission” button and then by selecting the relevant category he wishes to update. At the time being the users are only able to insert new data, not update or delete anything under these categories. This is a functionality that is currently under implementation. However, they can update their Researcher Profile status to make it either visible or hidden. They can also update their picture, their personal information such as name, title, department, personal site, qualifications, etc., their biography, variant names and their ORCiD synchronization preferences.

Additionally, they can manage their publications, by deciding which items to show and which ones to hide, therefore they can select the most representative publications and create lists of the selected contents. When publications are already in the system, the researchers can "claim" a publication (or "disclaim" it) and automatically add it to their profile.

5.3.3.2. Projects

During the transformation of Ktisis into a CRIS system we also made another important addition in the new system with the implementation of Projects. Projects is another managed entity provided by DSpace-CRIS where information about various projects undertaken by the University’s researchers can be found. We have configured Ktisis so that we provide information such as the project title, project abstract, coordinator, start and completion date if available, and the status of the project. We can connect the publications that were published through a project, by looking up the authority value of this project in the system registry and then mapping it to the Dublin Core field (dc.relation) that has been assigned for this purpose. Then these publications are also displayed under the Project’s page.

5.3.3.3. Metrics

Another new functionality in the system is the introduction of metrics. DSpace-CRIS provides a framework to support the collection of metrics about all the entities in the system. We have created scripts to automatically capture bibliometrics from databases such as the Web of Science and Scopus, which are run daily overnight and provide the latest numbers for each item. The number of citations displayed (metric value) is linked to the external page of the database which shows who has cited this item, a very useful feature of the system.
In each case, the rank as percentile of the specific item as compared to other objects in the repository is also shown. For example, if this value is 50 for an item, this means that this item is in the top 50% of the papers in the repository based in the number of the citations in the relevant database.

In addition, the current number of downloads and the number of views is also displayed. This functionality provides the ability to show the statistics for an entity. For each entity it is possible to define which statistics components to show. For example, for a collection we can show the total number of views of this collection, the total number of views of the items in the collection and the total number of downloads of the items in the collection. Internal usage statistics are available at all levels of the underlying hierarchy, including the whole repository. Aggregated statistics of linked items are available, for example in viewing the statistics of an organizational unit, it is possible to include the statistics of the items that have been authored by researchers belonging to that organizational unit. Statistics can me “converted” in metrics including sorting or top items and can be sent to the users periodically either by email subscription or RSS feeds.

We have created facets at the home page of the repository and other pages where the visitor can see in one glance the top most viewed and the top most cited articles in Ktisis. On top of that, when browsing through the system the visitor can see the numbers of each type of entity. For example, number of projects, number of patents, number of departments, number of faculties, etc.

5.4. Ktisis overall numbers and evolution

At the time being and while we are continuously working on adding new data into Ktisis, the figures of the system can be seen in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Ktisis figures.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of research profiles</td>
</tr>
<tr>
<td>No of research output</td>
</tr>
<tr>
<td>No of funded projects</td>
</tr>
<tr>
<td>No of patents</td>
</tr>
</tbody>
</table>

The amount of staff devoted for the transformation of Ktisis into a CRIS system was 5 people: one IT person who worked on all the technical aspects of the system and 4 librarians who worked on testing the system and submitting new data. We also collaborated with a few academics who worked with their researcher profiles to provide us with feedback and inform us with any issues or problems they found.

In Ktisis, there have been a few changes since previous implementations of DSpace-CRIS systems. The implementation of a complete integration between Ktisis and ORCiD is one of the most important evolutions. The system makes it possible to create a new ORCiD for a researcher or use an existing ORCiD record and connect the local profile or create a new one and connect it with the ORCiD registry. Then the integration between ORCiD and Ktisis makes it possible to transfer information between the two systems in both directions. The default behaviour of the system is that when a new Researcher Profile is created it can capture information from the ORCiD registry. From the Researcher Page, via the specific tab “ORCiD preferences”, it is possible to check and connect the local profile to ORCiD.

Another evolution from previous systems is the implementation of Altmetrics. In scientific publishing, Altmetrics are non-traditional metrics which can be used to complement more traditional citation impact metrics. They can explain both the volume and nature of attention that research receives online. Altmetrics can measure how many people have shared or engaged with a scholarly output (e.g., an article, a book, a dataset, software, posters, presentations, videos, and virtually anything else). Examples of this include mentions on Twitter or in the mainstream media; page views and downloads; GitHub repository watchers.6

For the most cited paper (1938 cited times) in Ktisis, the system provides the Altmetrics which show the following information:

• the attention score of the paper which is 11 and means that this paper is in the top 25% of all research output scored by Altmetrics.
• the number of times this paper has been used in blog posts (1)
• the number of Wikipedia pages this paper has been cited in (2)
• the number of readers on Mendeley (905)
• the number of readers on CiteULike (1)

Additionally, a link is provided to the Altmetrics page about this paper.

5.5. Open access and open data policies in Ktisis

Ktisis allows the search and tracing of all scientific publications made by the university while providing access to the full text of the publications. The articles that are being made available in Ktisis are openly accessible and retrievable from any user, except in cases where the open access policy conflicts with third-party (publishers) intellectual property rights. In these cases, only the bibliographic data and the electronic address of the document are available. This applies for the length of time that the restrictive conditions exist. The total number of full-text content available in Ktisis is 1195 (as of October 2018).

For the academic publications the availability of the full text is defined according to the type of the publication. In Ktisis we have defined different availability for articles published in journals, pre-publication articles, books, book chapters, eBooks, conference papers, reports, patents and abstracts. The e-license for such content is the formal agreement between the copyright holder and the Library.

Ktisis also provides the ability to choose and apply a Creative Commons License for a submitted publication. This is not mandatory but it is provided as an ability to the copyright holder to determine in an easy and straightforward way which rights he retains and which rights he grants for the benefit of other creators.

5.6. Next Generation Repositories

Repository systems are currently highly used in education and research environments; hence they provide a solid base for a global infrastructure of research. However most of the platforms used in the repositories are using out of date technology which was designed two decades ago. For this reason, repositories have not yet reached their full potential and they are mainly used for publishing the research output of the institutions.

Therefore in 2016, COAR\(^7\) formed a Working Group to “identify the core functionalities for the next generation of repositories. These functionalities include more web-friendly architectures, embedding repositories into the workflow of researchers, open peer review and quality assessment of content, and better impact and usage measures.”

These new functionalities aim to transform repositories into the foundation for a distributed, globally networked infrastructure for scholarly communication. Layers of added value services can be deployed on this infrastructure which make a repository more research oriented, innovative, while providing open peer review and quality assessment of the content and better impact and usage measures.

According to COAR the exact vision for Next Generation Repositories is “to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value-added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community”.

The objectives\(^8\) of Next Generation Repositories are:
• To achieve a level of cross-repository interoperability by exposing uniform behaviors across repositories that leverage web-friendly technologies and architectures, and by integrating with existing global scholarly infrastructures specifically those aimed at identification of e.g. contributions, research data, contributors, institutions, funders, projects.
• To encourage the emergence of value-added services that use these uniform behaviors to support discovery, access, annotation, real-time curation, sharing, quality assessment, content transfer, analytics, provenance tracing, etc.
• To help transform the scholarly communication system by emphasizing the benefits of collective, open and distributed management, open content, uniform behaviors, real-time dissemination, and collective innovation.

The Working Group has focused on the specification of some common behaviors and standards that should be adopted by the repositories. In the transformation of Ktisis into a CRIS system we worked on adopting these
behaviors and standards in the best possible way. The list of behaviors compared to the work that has been in done in Ktisis follows next.

1. **Exposing Identifiers**: As mentioned earlier DSpace-CRIS provides a flexible and robust data model that supports the use of persistent identifiers where possible to create navigable relations between entities. In Ktisis this is achieved by providing identifiers expressed as HTTP(S) URIs. An example is the use of ORCiD identifiers for individual researchers which relate to their publications. These URIs are resolvable and they have associated metadata attached.

2. **Declaring Licenses at a Resource Level**: We are using Creative Commons Copyright Licenses defined at the resource level which is part of the policies defined in Ktisis.

3. **Discovery through Navigation**: The objects in Ktisis are represented as a bundle of resources each with its own HTTP(S) URI. For example, in the landing page of a publication, the PDF version of the paper is also available where possible. Additionally, the appropriate link relation types and format indicators to interlink the web resources that make up a scholarly object are available.

4. **Interacting with Resources (Annotation, Commentary and Review)**: Repositories can increase their value by supporting commentary, annotation and peer review activities. The functionality to allow these activities does not necessarily need to be provided by the repositories themselves but can rather be provided by third party services or tools that specialize in the creation of overlay content. This functionality is part of the future goals for Ktisis.

5. **Resource Transfer**: The vision for next-generation repositories strongly emphasizes a resource-centric paradigm, where resources are not arbitrarily copied from system to system but are, rather, referenced where they are. However, there are use cases where the copying of resources (metadata, content or both) is necessary, generally to avoid the problem of network latency, to support functions which operate simultaneously on large numbers of resources, where those resources are distributed across many repositories. Again, this is a future goal for Ktisis.

6. **Batch Discovery**: Uniform, global, cross-repository discovery of resources is essential to establish repositories as important players in scholarly communication. Batch discovery generally supports search, but also use cases that require content transfer such as text mining and preservation. This is achieved in Ktisis being compatible with aggregators such as OpenAIRE.

7. **Collecting and Exposing Activities**: Repositories should be able to actively and in real-time collect and expose activity (e.g. information about changes, additions, comments, annotations, peer-reviews, access, downloads, etc.) pertaining to scholarly objects they host. Authors of the scholarly object involved in an activity, other repositories, and a variety of consuming applications that keep the pulse on scholarship as it happens should be able to receive metadata about activity not only retrospectively through harvesting, but also in real-time. To that end, notification mechanisms need to be put in place. This functionality is a future goal for Ktisis.

8. **Identification of Users**: Repositories should support the creation of overlay content such as annotation, commentary, peer review, as well as other interactions with the scholarly objects they host. Inviting users to identify themselves by means of identifiers that have global reach (HTTP(S) URIs) when interacting with objects in this manner can lead to constructive conversations and the creation or reinforcement of social connections. This is again a future goal for Ktisis where user identification will be implemented to support commenting and interaction with social media.

9. **Authentication of Users**: Requiring users to identify themselves by means of identifiers that have global reach (HTTP(S) URIs) when interacting (e.g. annotation, commentary, review) with scholarly objects hosted by a repository can lead to constructive conversations and the creation or reinforcement of social connections. Like the previous recommended behavior, Ktisis will implement the authentication of users to enable commenting and feedback.

10. **Exposing Standardized Usage Metrics**: Repositories should be able to share user interaction data to enable the development, deployment and evaluation of innovative value-added global services over repositories. Collecting standard metrics is important in order to optimize, operate, and enhance the repository and demonstrate the value
of the repository to authors and other stakeholders. This has been achieved in Ktisis with the implementation of metrics and statistics.

11. **Preserving Resources**: Open access means not just that you can have access to things today, but also into the future. Not every repository needs to run its own preservation processing stack, but rather we need common standards, protocols and interoperability that will enable us to build these services for repositories in a collective way. Additionally, it is necessary to preserve the complex interconnection of resources, which involves preservation activities at various levels including the resource, metadata and information graph. This is achievable in Ktisis by making sure that the scholarly output and all resources will be available in the long term by preservation and keeping copies of the system daily.

To sum up, Ktisis, besides metadata, provides content, links between resources, usage interactions, metrics, navigation, dissemination and preservation. The technology we have chosen to use enables us to work in the points mentioned above for next generation repositories, such as comments, notifications, etc. as future goals for our system.

### 6. Benefits and Conclusion

The benefits of transforming Ktisis into a CRIS system using DSpace-CRIS are many. First, the system provides a flexible data model to describe all the entities that populate the research environment and their meaningful links. DSpace-CRIS is open-source and it provides a technology that allows us to maintain independence from vendors and make our system evolve according to the needs of our community. Our publications are safely and easily managed as before, but now we have the advantage to link them to relevant data such as authors, projects, metrics, networks, statistics, etc. Now we have an institutional repository that can collect, manage, preserve and disseminate all information about our university’s research and its performance.

One of the most important benefits of using DSpace-CRIS in our system is the provision of persistent identifiers. The flexible data model of the software takes advantage of persistent identifiers whenever possible to create relations between all entities of the system, for example a researcher’s ORCID with a project ID.

We also make great use of another powerful feature of the software, the ability to import records from external databases such as Web of Science, Scopus, etc. This interoperability saves a lot of time for the population of the system.

Overall, the transformation of Ktisis into a CRIS system has provided the academic community of our university a lot of benefits and functionalities which allow the researchers to keep track of their work while enabling them to build a full academic profile which will assist them in their assessment process. We will continue to work with adding new features into the system as mentioned earlier, providing all the necessary technologies and behaviours for the repository to be considered as a Next Generation Repository. Our Library staff is dedicated to providing the required amount of time and keeping up with the new relevant technologies.

### References

5. https://www.4science.it/en/
Author name / Procedia Computer Science 00 (2018) 000–000