

Preprint of:

Yannis Tzitzikas, Marios Pitikakis, Giorgos Giakoumis, Kalliopi Varouha and Eleni Karkanaki,
How Can a University Take its First Steps in Open Data?,
14th Metadata and Semantics Research Conference (MTSR 2020), Nov 2020

How Can a University Take its First Steps in Open Data?

Yannis Tzitzikas^{1,2}[0000-0001-8847-2130], Marios Pitikakis¹[0000-0003-0440-0266],
Giorgos Giakoumis¹[0000-0002-2853-2346], Kalliopi
Varouha¹[0000-0003-4187-5324], and Eleni Karkanaki¹

¹ University of Crete, Heraklion and Rethymnon, Greece

² Information Systems Laboratory, FORTH-ICS, Heraklion, Greece
{tzitzik,pitikakis}@csd.uoc.gr, geokop7@gmail.com, {kvarouha,
eleni.karkanaki}@uoc.gr

Abstract. Every university in Greece is obliged to comply with the national legal framework on open data. The rising question is how such a big and diverse organization could support open data from an administrative, legal and technical point of view, in a way that enables gradual improvement of the open data-related services. In this paper, we describe our experience, as University of Crete, for tackling these requirements. In particular, (a) we detail the steps of the process that we followed, (b) we show how an Open Data Catalog can be exploited also in the first steps of this process, (c) we describe the platform that we selected, how we organized the catalog and the metadata selection, (d) we describe extensions that were required, and (e) we discuss the current status, performance indicators, and possible next steps.

Keywords: University Open Data · Directive 2013/37/EU.

1 Introduction

Open access has been a core strategy in the European Commission for several years now and aims at improving knowledge circulation and thus innovation. In 2012, the European Commission encouraged all EU Member States, via a Recommendation, to put public-funded research results in the public sphere in order to make science more efficient and strengthen their knowledge-based economy. It is widely recognized that making data and research results more accessible contributes to advancements in science and innovation in the public and private sectors. For example, the EU Open Data Portal³ and European Data Portal⁴ provide access to the European Union open data, categorized by subject and/or application domain.

In Greece, starting from October of 2014 with the law 4305 (amending the law 3448/2006) each public organization is obliged to comply with the Directive

³ <http://data.europa.eu/euodp/en/home>

⁴ <https://www.europeandataportal.eu/>

2013/37/EU⁵ of the European Parliament (amending Directive 2003/98/EC) on the *re-use of the public sector information* and this includes higher education area. Responding to this requirement is a challenging endeavor for a university because (a) a university is a big organization, (b) it comprises schools, departments and units of different characteristics and mindset, (c) the legal framework should be fully respected. The rising question is: how a university could support open data administratively, legally and technically in a *flexible* way that allows for the *gradual* improvement of the open data provided? In this paper we describe what University of Crete has done so far for responding to the requirements of the national legal framework on open data. The key contributions of this (case study) paper are: (a) we detail the steps of the process that we followed (also from an administrative point of view), (b) we show how an Open Data Catalog can be exploited also in the first steps of this process, (c) we describe the platform that we selected, how we organized the catalog and what metadata we included, (d) we describe extensions that were required, (e) we discuss the current status, performance monitoring, and possible next steps. The rest of this paper is organized as follows. §2 describes the general context and related work in national and international level, §3 describes the process that we followed and the envisioned ecosystem, §4 describes the Open Data Catalog, §5 describes supporting related activities, and finally, §6 concludes the paper.

2 Context and Related Work

General Context. [1] contains an interesting discussion and analysis of the challenges the universities face for stewarding the data they collect and hold in ways that balance accountability, transparency, and protection of privacy, academic freedom, and intellectual property, while [8] investigates possible open data partnerships between universities and firms. In general, the potential of open data is unlimited. Just indicatively [4] describes an approach for interlinking educational information across universities through the use of Linked Data principles, while [6] describes an approach for ranking universities using Linked Open Data. Below we list a few indicative universities that follow good practices as regards open data. The Harvard College Open Data Project (HODP) (<https://hodp.org/>) is a student-faculty group that aims to increase transparency and solve problems on campus using public Harvard data, featuring dozens of publicly-available datasets from around Harvard University. The University of Southampton catalog (<https://data.southampton.ac.uk/datasets.html>) provides datasets in (at least) Turtle and RDF/XML formats, as well as an HTML description, and it also offers a SPARQL endpoint. In general, we observe a gap in the literature as regards the processes that a university can follow as regards open data.

⁵ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013L0037&from=EN/>

National Level (Greek Universities). In Greece there are 24 universities. 3 of them, namely, University of Crete (<http://opendata.uoc.gr/>), Aristotle University of Thessaloniki (<https://opendata.auth.gr/>) and University of Macedonia (<http://data.dai.uom.gr/data/>) have set up a catalog for open data. University of Macedonia has also selected CKAN platform to develop their catalog which contains 22 organizations and 132 datasets. Other universities have not set up a catalog, but have decided to directly publish their data in the central Greek Government portal for open data. Indeed, in that catalog we can find datasets from 9 universities. The number of datasets per university that are published is rather low, it ranges from 3 to 14. Overall, we can understand that we are still in the first steps of the path of open academic data, as more than the half of Greek Universities ($14/24 = 58\%$) have not published any dataset. However, it is worth mentioning, that in June 2020, a proposal for a “National Plan for Open Science” was released [2], focusing on open science.

3 The Process Followed and the Envisioned Ecosystem

Below we describe the five main steps of the overall process that UoC (University of Crete) has followed.

- \mathcal{S}_1 : A *Task Force for Open Data*, for short TFOD, was appointed (June 2017) which prepared a *feasibility study* about Open Data and the goals/needs of the institution.
- \mathcal{S}_2 : The task force (TFOD) then performed an *inventory of the documents and datasets in the possession of the university*. For this purpose, an Open Data Catalog was installed and customized accordingly (the catalog is described in detail in §4).
- \mathcal{S}_3 : The *ODEDIAD*, acronym in Greek of “Clarity and Open Data Project Management Team” was formally appointed (Dec 2018), in compliance with the circular DHD/F.40/2369 of 24 January 2017 from the Ministry of Administrative Reform and Electronic Governance. To enhance flexibility, a small scheme was decided, comprising a general coordinator, and coordinators for the *administrative* aspect, the *technical* aspect and the *open data catalog* aspect. Moreover, each department/unit of the university has appointed a contact person with whom ODEDIAD communicates. ODEDIAD then (a) *harmonized and characterized* the dataset descriptions collected in step \mathcal{S}_2 , and prepared the *Decision on Open Data* that was formally approved by the University (April 18, 2019) and was submitted to the Ministry of Education and Religious Affairs.
- \mathcal{S}_4 : The Open Data Catalog was improved and *aligned with the decision* (in \mathcal{S}_3), and the Departments/Units were encouraged to start uploading the actual datasets in the catalog.
- \mathcal{S}_5 : The Open Data Catalog became *public* (Oct 18, 2019) and ODEDIAD is responsible for *maintaining and continuously improving* the catalog.

The envisioned ecosystem is shown in Figure 1 in the form of a Use Case Diagram that depicts the main Actors and the main Use Cases. The role of catalog-related actors is discussed in §4.3.

It is worth noting that the process started in a *bottom-up* manner (at step \mathcal{S}_2), for engaging all departments/units of the university, and then we applied a *top-down* harmonization step (at Step \mathcal{S}_3), for achieving completeness and uniformity.

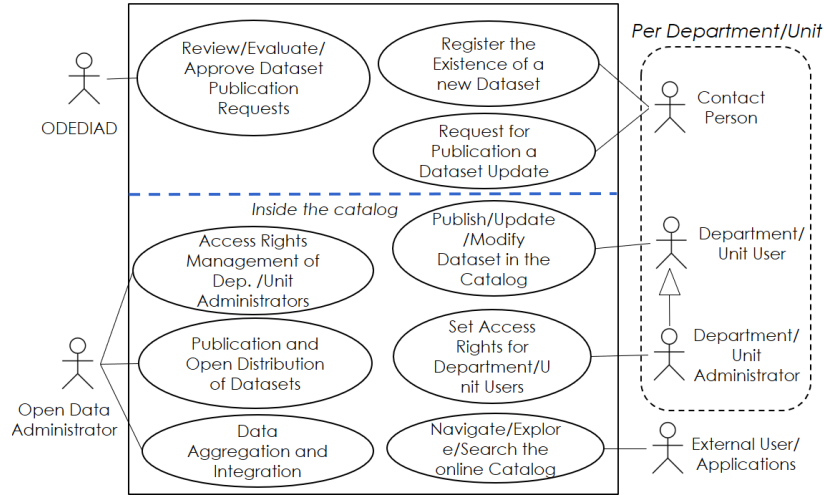


Fig. 1. The Envisioned Open Data Ecosystem of the University of Crete

4 The UoC Open Data Catalog

Here we describe: how the platform was selected (in §4.1), how we customized CKAN (in §4.2), how we organized the catalog in department/units (in §4.3), what metadata we selected to include (in §4.4), what access services are offered (in §4.5), and an analysis of the uploaded content so far (in §4.6).

4.1 Selecting the Platform

For selecting the UoC Open Data Catalog we investigated several options based on the platform characteristics, the technology considerations and the options that were available (Oct 2017). The cost of data catalogs is a key evaluation criterion. Products distributed as “open source” software are generally preferred because they are essentially “free” and can be modified or customized without restriction or licensing fees. However, open source software still requires management costs (hosting, maintenance, updates and security patches, training etc). On the other hand, SaaS products typically are proprietary; a vendor provides the software, the setup and hosting services, at a monthly or annual fee. Under the SaaS delivery model, the vendor is responsible for maintenance, server availability and reliability, scalability and performance according to a contract.

For making our decision we considered the following: a) Self-managed, open source catalogs can provide a high degree of customization and autonomy. Most open source catalogs are designed to run in combination with other open source software and therefore technical proficiency in these areas is required. b) The open data catalog must be hosted on reliable and relatively fast server architecture. Slow response times or periods of unavailability will discourage users. c) Institution policies and/or laws regarding the open data catalog. d) Scalability. As datasets are added, the catalog must be able to handle the additional load and must be easily extensible (have the flexibility to include additional functionality). e) How datasets are managed and stored in the data catalog: “All-in-One” vs “Federated” catalogs. In an “All-in-One” model, datasets are stored in the catalog’s architecture, with the main benefit of hosting and managing from a single platform and thus exercising strong oversight over the entire catalog infrastructure. Alternatively, in “Federated” catalogs datasets can reside on any publicly accessible location and the catalog includes a link (URL) to the dataset, as opposed to including the dataset itself. f) Communities of support, support models or user communities. Below we list and summarize the leading open data platforms.

CKAN (<http://ckan.org>) is a web-based open source management system for the storage and distribution of open data, formally supported by Open Knowledge International (<http://okfn.org>). It can be installed on any Linux server, including cloud-hosted configurations, and it has a modular architecture through which additional or custom features may be added. Many national governments (like the United Kingdom, the United States, Austria, Brazil, Canada, Germany, Netherlands, Norway etc.) use CKAN to power their open data portals and it is also used by international portals such as the EU Open Data Portal (<https://open-data.europa.eu/>).

DKAN (<https://getdkan.org/>) is also open-source and takes a different approach by integrating the open data catalog features of CKAN into a content management system (CMS), namely Drupal. This can be more appealing to organizations that have already invested in Drupal-based websites since DKAN can also be enabled in existing Drupal sites so that anyone can easily start to publish open data in standards compliant ways.

Junar (<http://www.junar.com>) delivers a cloud-based open data platform and manages its content based on the Software-as-a-Service (SaaS) model. This platform is frequently selected because of its ease of deployment and because it provides an “All-in-One” infrastructure. Junar can either provide a complete data catalog or can provide data via an API to a separate user catalog.

JKAN (<https://jkan.io/>) is using Jekyll (<https://jekyllrb.com/>) (a simple static site generator) and allows for a quick deployment of static pages from underlying files. This data portal is based on CKAN and it is aimed at data publishers that need to deploy their data quickly. It is also open source, light-weight and can be easily customized with themes.

The Open Government Platform⁶ (OGPL), like DKAN, is an open-source Drupal-based data catalog, but not designed to be CKAN-compatible at the API level. OGPL was jointly developed by the Government of India and the U.S. Government.

Socrata⁷ is a cloud-based SaaS open data catalog platform that provides API, catalog, and data manipulation tools. One distinctive feature of Socrata is that it allows users to create views and visualizations based on published data and save those for others to use. Additionally, Socrata is proprietary but offers an open source version of their API, intended to facilitate transitions for customers that decide to migrate away from the SaaS model.

Our review revealed that the format in which metadata are published depends highly on the open data portal that publishes it. Open data portal software frameworks are either built on their own standards or use an already existing standard. The two predominant platforms – CKAN and Socrata – have each been developed on their own respective frameworks that are coming from major standards such as Dublin Core⁸ and RDFS⁹. The platforms tend to use either one standard to generate metadata or a combination of a few. The most commonly used metadata standards in Open Data catalogs are based on some version of the Data Catalog Vocabulary (DCAT¹⁰) standard. DCAT has gained popularity due to its flexibility and elegant design, and aims at improving the data catalogues’ interoperability so applications can easily consume metadata (even from multiple catalogs).

CKAN stores the datasets as a folder that hosts datasets or resources. Metadata are presented at a dataset level using Dublin Core, DCAT and INSPIRE geospatial format¹¹. Junar uses the RDF metadata standard as presented in Dublin Core and DCAT, it does not support structural metadata. JKAN is based on CKAN, therefore it supports the same metadata standards as previously mentioned for CKAN. Socrata is based on the RDF metadata (Dublin Core and DCAT) with enrichment from custom metadata fields.

4.2 Adopting and Customizing CKAN

For the UoC Open Data Catalog we have selected CKAN to publish, store and manage open datasets. A dataset is a “unit” of data and contains two things: *metadata* (i.e. information about the data) and any number of *resources*, which hold the data itself. CKAN can store a resource internally, or store it simply as a link (i.e. the resource itself could be elsewhere on the web). It allows the creation of custom metadata fields, supports Dublin Core/DCAT for publishing metadata and provides data in open, non-proprietary formats such as CSV, XML, and JSON.

⁶ <http://ogpl.github.io/index-en.html>

⁷ <https://www.tylertech.com/products/socrata>

⁸ <http://dublincore.org/documents/dces/>

⁹ <https://www.w3.org/TR/rdf-schema/>

¹⁰ <https://www.w3.org/TR/vocab-dcat/>

¹¹ <https://inspire.ec.europa.eu/data-specifications/2892>

CKAN provides a *full-text search* for dataset metadata as well as filtering and sorting of results. It is possible to restrict the search to datasets with particular tags, data formats etc or targeted search within an organization/department. When a dataset is found and selected, CKAN displays the dataset page, which includes the name, description, and other information about the dataset, and links to and brief descriptions of each of the resources that belong to the dataset. On-site data/resource preview is also available for known data formats (like csv, xls, xlsx, rdf, xml, rdf+xml, owl+xml, atom, rss, json, geojson, png, jpeg, gif).

CKAN offers flexibility to integrate or embed the data catalog with other websites, add additional pages, layouts, color schemes, logos and generally easy customization. It has a large community of users and support, and offers extensibility with many additional or custom features developed regularly via extensions. CKAN also provides multi-language support and internationalization.

Plug-ins. CKAN provides a DCAT extension/plugin¹² to export and harvest RDF serializations of datasets based on DCAT. The Data Catalog Vocabulary (DCAT) is "an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web". The extension defines the mapping of metadata for CKAN datasets and resources to the corresponding DCAT classes, mainly `dcat:Dataset` and `dcat:Distribution`, which is compatible with DCAT-AP v1.1. DCAT-AP was designed to meet the metadata publishing needs in the context of the European Commission's Interoperability Solutions for European Public Administrations (ISA) programme: "Improving semantic interoperability in European eGovernment systems".

Other CKAN extensions that are used for the UoC Open Data catalog are:

- A geospatial viewer¹³ for CKAN resources, which contains view plugins to display geospatial files and services in CKAN. More specifically, we are using the Leaflet GeoJSON viewer plugin.
- The pages extension¹⁴ that provides an easy way to add simple pages to CKAN. Using this extension we added a "featured datasets" page and Twitter feed page (for embedding Twitter content, news, dissemination activities etc)
- An LDAP plugin¹⁵ which provides LDAP authentication for CKAN and integration with our existing UoC LDAP user service. This extension allows us to import existing username/full name/email/description, and add LDAP users to a given organization automatically.
- An extension¹⁶ that integrates Google Analytics into CKAN.

In addition, some custom made extensions were created for the purposes of our UoC Open Data Catalog. For example, normally each dataset is owned by an "Organization" in a CKAN instance. A plugin was created to rename the default "Organization" of CKAN into "Department". Another plugin was developed to easily export the list of all datasets from a "Department". Some

¹² <https://github.com/ckan/ckanext-dcat>

¹³ <https://github.com/ckan/ckanext-geoview>

¹⁴ <https://github.com/ckan/ckanext-pages>

¹⁵ <https://github.com/NaturalHistoryMuseum/ckanext-ldap>

¹⁶ <https://github.com/ckan/ckanext-googleanalytics>

required additional metadata fields were inserted with another plugin and finally the appearance/theme of the catalog was customized.

4.3 Organization of the Catalog

Currently (July 2020), there are 47 “departments” (16 academic, 22 administrative and 9 other) inserted into the catalog that correspond to the actual structure of the University of Crete. We have created another “department” for the UoC OpenData Task Force, which contains instructions and reference material shared between the OpenData team, the dataset administrators and the catalog users.

Each catalog “department” has its own administrator and multiple members (which can be either editors or simple users). Depending on each user’s role in the department, users can perform different actions. A department administrator can edit the department’s information, add/edit a dataset or add individual users to the department, with different roles depending on the level of authorization needed. An editor can create a dataset owned by that department. By default, a new dataset is initially private, and visible only to other users in the same department. When it is ready for publication, it can be published by the department administrator (this may require a higher authorization level within the department). A simple user in a department can only view private (and public) datasets. During the S_2 inventory phase, all users (contact persons, local administrators, department/unit users) were allowed to insert datasets. The uploaded datasets/resources were reviewed by the ODEDIAD and the University DPO in order to be compliant with the data protection directive. After phase S_5 , when the catalog became public, new datasets/resources can still be uploaded to the catalog but need to be reviewed, evaluated and approved by ODEDIAD and the DPO before they become available for public consumption and use. In addition, there is an annual validation and update procedure of all the UoC open datasets.

4.4 Selected Metadata

The metadata fields (per category) for each set of documents, information and data are shown in Table 1.

4.5 Access Services

The user can *search* the catalog, can *browse* the catalog by department, and can *explore* in a faceted-search manner the available datasets (through tags, file formats, access rights) as shown in Fig. 2 (right). Apart from the above *programmatic access* is supported. The native CKAN API allows users, providers, consumers and developers to write code that interacts with the CKAN site and its data. CKAN’s Action API is a RPC-style API that exposes all of CKAN’s core features to API clients. All of a CKAN website’s core functionality (everything you can do with the web interface and more) can be used by external code that calls the CKAN API.

Table 1. Metadata fields: General, Right-related, for Digital Files, for Physical Files

General fields	
1	Title
2	Description
3	File Owner / Location (e.g. Office, Department, Address)
4	Quantity of data (indicative) e.g. number, size in GB, number of documents etc.
5	Refresh rate (approx.)
6	Is it geospatial data? (YES/NO)
Rights of use and licences	
7	Availability (YES/NO)
8	If already available, way of accessibility (e.g. website, on request etc.)
9	If available through on request (either electronic or printed application) method of accessibility (e.g. by mail, in person, electronically etc.)
10	Is there a privacy/personal data restriction? (YES/NO)
11	Are the other restrictions? (national security issues, tax secrecy etc.) (YES/NO)
12	Available through fees? (YES/NO)
13	Available through licensing? (YES / NO / If YES, type of licence)
14	File Type (Physical/ Digital / Physical and Digital)
For digital file/dataset	
15	Automatically machine processable & editable format (e.g. txt, csv, html, xml, rdf, odt, ods, doc, xls, docx, xlsx)
16	Non automatically processable format (e.g. jpeg, tiff, gif, pdf, scanned documents etc)
17	Access through URL (e.g. website if available)
18	API (if available)
For physical file/dataset	
19	Format (e.g. A4 document, photographic archive etc)
20	Location (e.g. file cabinet 5 in room 123)
Additional information	
21	Content topic or tag (e.g. health, environment, economy, specific science etc.)

4.6 Analysis of the Uploaded Contents

After several rounds of discussions with all the academic department contact persons (during \mathcal{S}_3), we arrived to a commonly agreed upon collection of datasets (homogenization/harmonization) that every department should provide openly (and preferably using in a machine processable format). This collection includes the 22 datasets for the 16 Departments of UoC that are shown in Table 2. Similarly, for the 5 Schools of UoC, a 5 list of common datasets were identified. The rest of the University units had not any common dataset.

The Open Data Decision of the University of Crete, that contains the harmonized descriptions of all datasets of the organization is publicly available¹⁷. As mentioned in §3 (Step \mathcal{S}_4), following the issuance of this decision, the Open Data Catalog was improved and aligned with that decision. Subsequently, the

¹⁷ <https://rb.gy/0rqbdu>

Table 2. The Datasets of an Academic Department

1	Open Courses	12	Internship Programs
2	Course Outlines	13	Photo Gallery
3	Undergraduate and Postgraduate Study Guides	14	Events, Summer Schools, Meetings, Conferences, Competitions
4	Course Schedules	15	Alumni information (after explicit consent of the graduates)
5	Admission and Study Statistics	16	Announcements and News
6	List of Publications of the Teaching, Scientific and Research staff	17	List of Personnel
7	MSc and PhD theses	18	Elections of collective bodies
8	Internal/external Evaluations and Accreditation	19	Financial information and data
9	Distinctions and Achievements	20	Decisions on committees establishment
10	Building facilities, maps, infrastructure	21	Decisions related to studies
11	Promotional/Dissemination material	22	Decisions on Tender Notices

Departments/Units were encouraged to start uploading the actual datasets in the catalog or add links to the related web resources in case they are available in other systems. Currently (July 2020), the UoC Open Data Catalog¹⁸ hosts 509 descriptions of datasets, and 104 (20,76%) of them contain at least one resource or external link to a web page or institutional system. In machine processable formats (CSV, JSON, GeoJSON, XLS), we have 9 (1,8%) datasets. Note that some datasets are offered in more than one formats.

Some especially useful datasets, are those that provide statistics about the UoC student population from 2006 up to now (gender, age, nationality, duration of studies etc.). Another, is the "Research Directory" that contains information about all labs of the University, categorized by research field, department, position, or name. Finally, the Computer Science Department has posted many datasets in CSV and JSON format. Currently, the number of *tags* inserted into the CKAN catalog are 118.

5 Supporting and Ongoing Activities

Training and Outreach Activities. A manual with examples of how to use the catalog was created and published in the unit of ODEDIAD in the catalog (also for the needs of the step \mathcal{S}_2) as described in §3). The catalog became public on Oct 18, 2019 (Figure 2 show the first page of the UoC Open Data catalog website) and all members of the university were informed through email and public announcements on the UoC website. To support also notification services to those interested in the updates of the catalog, a twitter account (https://twitter.com/UoC_OpenData) was created on April 2020. The engagement of the community is rather low, however we have to note that no outreach action took place during the covid-19 pandemic period.

¹⁸ <http://opendata.uoc.gr/>

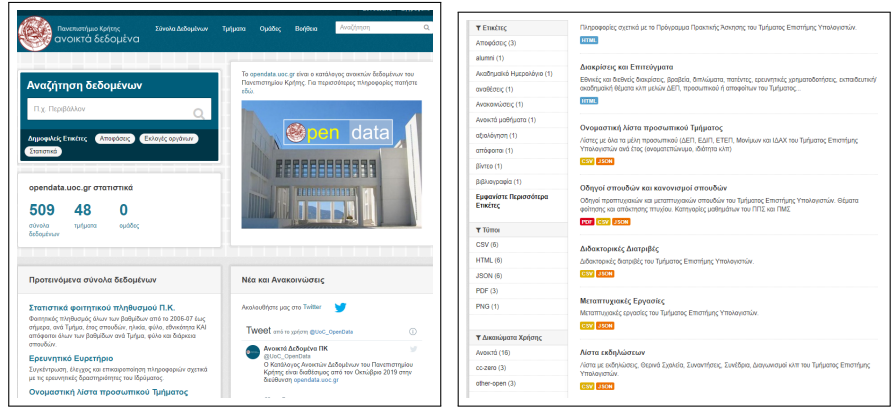


Fig. 2. The UoC Open Data Catalog website

Monitoring and KPIs. We decided a number *Key Performance Indicators* (KPIs) for measuring and monitoring the performance. Their definition, and 2020 value, are described next.

- K1: Number of Datasets in the Open Data Decision of the University of Crete (April 2020 value: 501).
- K2: Number (and Percentage) of Datasets in the Open Data Catalog with at least one resource or external link to a web page or institutional system (April 2020 value: 104 (20,76%)).
- K3: Number (and percentage) of Datasets in the Catalog available in machine-processable form (April 2020: 9 (1,8%)).
- K4: Number of Visits to the Catalog (we started counting the visitors on April 2020). (April 2020 value: not available).

Possible Next Steps. Our short term goal is to increase K2 and K3. To this end, we have categorized the possible next steps into three main categories:

- *Training and Good Practices.* Good practises for each individual type of dataset (news, personnel lists, etc), for increasing the percentage of datasets that are in a machine-processable format. This is also related to suggestions for increasing the interoperability of the information systems of the university, so that some information to be automatically exported in the catalog without any human intervention.
- *Advancing the Services of the Catalog.* Currently we are working on providing a bulk and/or automated dataset upload (and export) procedure, combined with a custom data interface to other internal UoC systems. Another important advancement concerns the *search* service: since the one offered by CKAN exploits only the metadata of the datasets, we investigate the provision of an additional search service over the *contents* of the datasets. Another challenging direction is the automatic production of a *Knowledge Graph (KG)* based on the contents of all datasets, an endeavor that requires applying methods for large scale integration (see [7] for a survey). The availability of such a KG, would also enable

the application of the *multi-perspective keyword search* services described in [5], for providing a user friendly search system over RDF data. In the future, the KG could be enriched with scholarly data (e.g. as in [3]) for covering also the research output of the institution.

- *Connections with External Systems.* This includes communication and integration with <http://data.gov.gr/>, which is the central directory of public data of all Greek government agencies and contains more than 10K datasets and 340 organizations, as well as connection with repositories of scientific data, specifically HELIX (the Hellenic Data Service)(<https://hellenicdataservice.gr>) with aims to be the national e-Infrastructure for data-intensive research.

6 Concluding Remarks

The provision of an organizational and technical framework for Open Data in a big organization, like a university, is a challenging task. In this paper we described the *five-steps process* that the University of Crete has followed and the outcomes of this process so far, as well as the forthcoming steps. In brief, we adopted a mixed, *bottom-up and top-down* approach, in which the catalog played a central role in the entire process. Even if the majority of the data are not in a common, easily processable form, the catalog contains the *placeholders* for the *entire spectrum of the datasets* in the possession of the organization, and we have realized even in its current form serves as a *global index* of the various resources that are published in the various websites of the university. To set up priorities and plan our next actions, we have identified a few KPIs. We hope this will be useful in other organizations that have similar obligations and characteristics.

References

1. Christine L Borgman. Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Tech. LJ*, 33:365, 2018.
2. Spiros Athanasiou et al. National plan for open science. Technical report, January 2020. <https://doi.org/10.5281/zenodo.3908953>.
3. Michael Färber. The microsoft academic knowledge graph: A linked data source with 8 billion triples of scholarly data. In *International Semantic Web Conference*, pages 113–129. Springer, 2019.
4. Miriam Fernandez, Mathieu d’Aquin, and Enrico Motta. Linking data across universities: an integrated video lectures dataset. In *International Semantic Web Conference*, pages 49–64. Springer, 2011.
5. Giorgos Kadilierakis, Christos Nikas, Pavlos Fafalios, Panagiotis Papadakis, and Yannis Tzitzikas. Elas4RDF: Multi-perspective Triple-centered Keyword Search over RDF using Elasticsearch. In *(Demo Paper) Proceedings of the 17th Extended Semantic Web Conference (ESWC’2020)*, 2020.
6. Rouzbeh Meymandpour and Joseph G Davis. Ranking universities using linked open data. In *LDOW*, 2013.
7. Michalis Mountantonakis and Yannis Tzitzikas. Large-scale semantic integration of linked data: A survey. *ACM Computing Surveys (CSUR)*, 52(5):1–40, 2019.
8. Markus Perkmann and Henri Schildt. Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy*, 44(5):1133–1143, 2015.