FAIRness of ETDs - A Conference Report of ETD 2022

Joachim Schöpfel

▶ To cite this version:

Joachim Schöpfel. FAIRness of ETDs - A Conference Report of ETD 2022. 2023. hal-04561126

HAL Id: hal-04561126
https://hal.univ-lille.fr/hal-04561126v1

Preprint submitted on 26 Apr 2024

FAIRness of ETDs – A Conference Report of ETD 2022

Joachim Schöpfel

Univ. Lille, ULR 4073 – GERiiCO

Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication

F-59000 Lille, France

May 16, 2023

Abstract

The Networked Digital Library of Theses and Dissertations (NDLTD), together with the University of Novi Sad and the non-for-profit organization Science 2.0 Alliance organized the 25th International Symposium on Electronic Theses and Dissertations, ETD 2022, at Novi Sad, Serbia, from 07 to 09 September, 2022, at the congress center of the University of Novi Sad. The conference theme was "FAIRness of ETDs and its implications". The following paper provides a short report of the event, of the keynotes, communications, and discussions, along with some perspectives for the research and development on ETDs.

Keywords

Electronic theses and dissertations

FAIR principles

Open science

Research data

Metadata

Interoperability

Accessibility

The Networked Digital Library of Theses and Dissertations (NDLTD[i]), together with the University of Novi Sad[ii] and the non-profit organization Science 2.0 Alliance[iii] organized the 25th International Symposium on Electronic Theses and Dissertations, ETD 2022, at Novi Sad, Serbia, from 07 to 09 September, 2022, at the congress center of the University of Novi Sad. The conference theme was "FAIRness of ETDs and its implications". It was organized as a hybrid event with 79 in-person participants, and 35 remote participants from 19 countries on five continents. The objective of the ETD conferences is to provide opportunities for exchanging ETD experiences with professionals from different countries, as well as for connecting to the good practices around the world. Also, the ETD 2022 conference organizers invited library, graduate school and information technology professionals, graduate students and faculty as well as library and information systems and services representatives to submit proposals on a large range of topics related to ETDs, such as FAIRness, Open Science, research data and source code. The final program with 35 speakers from fifteen countries included six sessions, two workshops and two keynotes.

On FAIR principles and ETDs

In her opening keynote, Violeta Ilik, Dean of Libraries from Adelphi University, Garden City, New York, introduced the FAIR principles in the environment of academic library systems, digital archives and repositories. She described the main purpose of these principles as "keeping up with the best ways to find information", through interconnection, standardization and opening up digital library systems ("operational islands"), in order to "make knowledge open and equitably available to all". She observed that library platforms often fail with the interoperability and the reusability principles, and she asserted that applying FAIR principles would contribute to develop knowledge graphs and trustworthy repositories and software. In the following discussion, Violeta Ilik admitted that the FAIR principles, as a central part of the European Open Science Cloud, may have more relevance

and impact in Europe than in the United States; but she insisted on the universal, global character of these principles and reminded that the US institutions (co)produce global repositories like GenBank or Protein Data Bank which already start to be aligned with these principles. Now, what should librarians do? Her answer was quite clear and simple: "They should define evaluation criteria compliant with FAIR principles for the negotiations with vendors of library systems – this is what we need".

Six other papers focused on the FAIR principles applied to theses and dissertations. Based on a systematic review of PhD dissertations, Alcidina Cunha Costa from the Brazilian Institute of Information in Science and Technology (IBICT) asked the provocative question "Who is keeping an eye on FAIR principles?" The result is disappointing if not discouraging, insofar she identified only ten dissertations submitted between 2016 and 2022, in six countries. She commented that this is not enough; perhaps the principles are "too young"; perhaps one part of the research on accessibility, interoperability etc. does not mention the umbrella term of FAIR. Because of the importance of these principles for the future development of repositories, hopefully there will be more doctoral research on FAIR in the next years. The promotion of best practice guidelines by the US Council of Graduate Schools may contribute to growing awareness and adoption of FAIR principles among PhD students.

A study from Figshare, presented by Andrew Mckenna-Foster, Mark Hahnel and Maria Cotera (UK), compared the FAIRness of the deposit on Figshare of non-traditional research outputs (supplement materials such as data, code, media etc.) associated with theses and dissertations shared by individual students, with those shared through an institutional repository. Here, FAIRness was defined as richness of metadata which can improve the findability and the reusability of data and documents. According to the Figshare study, records from institutions are FAIRer with significantly longer titles, longer and more detailed descriptions, more references and more keywords. Also, the number of views per month was

significantly higher for records in institutional repositories. Conclusion: ETD related data shared in repositories with institutional oversight appear more findable, interoperable, and reusable than those shared without institutional oversight. The authors interpreted these results as "indicating that graduate student mastery of FAIR sharing is limited but that they benefit from the services attached to institutional repositories". Interesting empirical evidence in favor of institutional control and processing of ETDs and related materials.

Sonja Bezjak and Janez Stebe (Slovenia) presented the activities of the Slovenian Social Science Data Archives (ADP) for the implementation of open data policies at the PhD school of the University of Ljubljana. For nearly ten years and supported by the European CESSDA network, ADP has been promoting the concepts of "open access", "open data", "open science", "data management plan" and "FAIR principles". More recently, ADP has developed and implemented a Data Management Plan (DMP) program for PhD students, including training and promotion, in order to improve their research data practice and contribute to the FAIRization of data related to ETDs; each student must submit a DMP together with her/his PhD project. The described approach is a mix of bottom-up (with European support) and top-down (adoption of CESSDA guidelines); according to the ADP team, data management training of PhD students should be mandatory.

Sangeeta Narang and Manoj Kumar K (India) analysed the metadata quality of 310 ETD records retrieved from institutional repositories from different countries. After a comprehensive introduction into metadata formats, they commented several examples of ETD records. Some findings: the main metadata format is the Dublin Core; 44% records contain an ORCID identifier; only 52% allows version control; and only 15% contain information about intellectual property or licensing (bad for reusability). Also, they observed that metadata are not available in all repositories and finished the presentation with a plaidoyer for the

improvement and standardization of metadata quality of ETDs in institutional repositories, to increase their interoperability and the findability of ETDs.

Another study on the quality of ETD metadata was presented by Cecilia Kasonde Adrian Chisale and Lighton Phiri (Zambia). They assessed nearly 5 million records harvested from the NDLTD Union Catalog against the different elements of the ETD-ms standard format. The results are not really good. One part of the elements, including the legal status, are missing in 30-50% records. The authors of the study noticed a lack of awareness of the ETD-ms standard and recommended that NDLTD should generate missing metadata and produce guidelines for institutional repositories formats and policies, in order to increase metadata standardization and compliance with the FAIR principles, in particular interoperability.

Michael Boock (United States) and Joachim Schöpfel (France) presented results from an international survey coordinated by Behrooz Rasuli (Iran) and conducted along with Brenda Van Wyk (South Africa), on a sample of 100 institutional ETD repositories from universities in 37 countries. The purpose of the study was to assess the characteristics of ETD embargoes. If 57 repositories have published an embargo policy on their web site, most of them lack clarity and do not cover all aspects of embargo procedures (such as, embargo period, criteria, confirmation process, etc.). In order to improve the accessibility of ETDs in institutional repositories, the authors recommended that embargo policies should be standardized and limited to 24 months maximum, and that embargoes should be subject to institutional control and validation. They suggested that NDLTD could contribute to this standardization of embargo policies.

The workshop on best practices of ETD management was led by Meinhard Kettler (ProQuest) and Charles Greenberg (NDLTD) with many helpful recommendations and good practice examples for academic librarians in charge of the processing of PhD dissertations and

related data and other digital objects. In particular, the workshop presented the ETD+ Toolkit[iv], developed by the US Educopia institute in order to help the academic community to train students to ensure the longevity and accessibility of their research outputs. The toolkit consists of an open set of six modules and evaluation instruments that prepare students to create, store, and maintain their research outputs. Like the Slovenian training program (see above), the ETD+ Toolkit will contribute to improve the FAIRness of ETDs, through increase awareness of legal conditions, better metadata quality, improved data organization etc. The toolkit can be adopted and adapted by anyone, for free.

Contributions from Serbia

After an opening workshop on the Serbian landscape of ETD repositories and further developments and projects, three communications provided more insight and a better understanding of the processing of ETDs in Serbia. Mirjana Brković, Tatjana Tošić and Tatjana Zubić gave a detailed overview of the development of ETD management at the University of Novi Sad, with 50,000 students the second largest State University in Serbia. The idea of open access to e-versions of doctoral theses and dissertations appeared in the 90's, and the first digital library of theses and dissertations (DIGLIB) was launched in 2003. From 2005 on, students were required to submit theses and dissertations in PDF which were collected and stored on various external devices (CDs, DVDs, HDs, etc.). Each user could get an e-version by sending a direct request to the librarian. Since 2014, the Digital Library of Theses and Dissertations (PhD UNS) has been introduced as part of a new Current Scientific Research Information System (CRIS UNS) which supports international standards, which is compliant with all requirements prescribed by the university and the government which is connected to the national dissertation system NaRDuS. Today, the NaRDuS system contains records of 6,420 dissertations; about 35% have digital files and are freely available in open access. Concluding the presentation, the authors highlighted the compliance of the Digital

Library with the FAIR principles, especially findability (multilingual metadata) and accessibility (different portals); download and access statistics confirm this statement.

The academic library "Svetozar Marković" of the University of Belgrade has the largest collection of PhD dissertations in Serbia, with more than 30,000 items. Since 2012, all dissertations of the University of Belgrade must be submitted electronically and are archived in the PHAIDRA repository and its subsystem E-theses[v] which contains about 6,600 ETDs. Tatjana Brzulović Stanisavljević and Dragana Stolić provided a detailed description of this hybrid collection and explained how they proceed to valorize the collection and to improve its accessibility. Two initiatives are of particular interest: the digitization-on-demand service, and the retro-digitization of older dissertations from the University of Belgrade and from Serbian scientists in other, foreign universities (Paris, Leipzig, Munich…).

Finally, Biljana Kosanović and Vladimir Otašević from the University of Belgrade presented the Serbian national repository of ETDs, NaRDuS[vi], which contains actually more than 12,000 ETDs, from 18 Higher Education institutions including the Belgrade and Novi Sad universities.

Most of the ETDs are harvested but a small part, from some smaller institutions, has still to be imported manually. Also, the major issue of NaRDuS is standardization and interoperability with national and international institutions, organizations and infrastructures, such as the NDLTD Global Search engine or the European DART and OpenAIRE portals. NaRDuS assigns a persistent identifier (handle) to each dissertation and supports the international ORCID identifier (which is good for findability) and it allows dissemination under open licenses (which is good for reusability). The interconnection with a national CRIS system (DSpace-CRIS) is in the pilot phase.

Other contributions

Joel Azzopardi from the University of Malta presented an in-house system called URECA Online Review System supporting the procedures of research ethics reviews. Regarding doctoral research, URECA considers students working on a dissertation as primary investigators; however, all students' submissions must be endorsed by their supervisors before being processed. The interconnection with data management plans is possible.

Two papers made use of the rich ETD content, yet in quite different ways and for different purposes. First, Brijesh Kumar Verma and Dr. R. Sarangapani (India) reported first results from a long-term citation analysis of dissertations in the field of social work, from 2001 to 2019, providing evidence of the changing use of sources because of the increasing availability of web-based journals and books. The second paper, prepared by a team from Virginia Tech (Aman Ahuja, Chenyu Mao, Chongyu He, Jianchi Wei, Edward A. Fox and William A. Ingram) and presented by William A. Ingram, was based on the assumption that ETDs "contain valuable knowledge that can be useful in a wide range of research areas" and that new tools are needed to discover and access this knowledge. Also, they developed ETD-Topics, "a framework that extracts topics from a large text corpus in an unsupervised way". The results are promising: "representations learnt from topic models can be useful for downstream tasks such as searching and/or browsing documents by topic, document recommendation, topic recommendation, and describing temporal topic trends". Among the perspectives of future research and development are more models, more services and a mapping with ProQuest's ontologies.

Two other papers analysed the implementation of ETD infrastructures, in different environments and on different levels. Lazarus Matizirofa (University of Pretoria) and Daisy Selematsela (Witwatersrand University) described the development of ETD repositories in South Africa by all HE institutions and the challenges of a new national open science policy

which requires the interconnection with a national ETD portal[vii] funded by the National

Research Foundation and the Committee for Higher Education Libraries of South Africa

(CHELSA). Particularly interesting was the fine description of the different stakeholders and

the assessment of different weaknesses and threats to the initiative, including lack of staff and

skills, funding, political and institutional support, and so on.

In her closing keynote, Michelle O'Hara presented a case study from the University of

Glasgow which decided to update its hardcopy thesis deposit policy and to shift to e-only

theses. She gave a brief overview of the challenges facing the Glasgow project team, in

particular the need to integrate new systems and existing platforms with issues of security and

interoperability, and the modelling of a new digital workflow for PhD dissertations in the

university archival system which should be able to support non-digital items like models and

artworks. The Glasgow solution may not fit with challenges and requirements from other

institutions; but their lessons learned, their respect of the FAIR principles and their

engagement with standards will be helpful for other ETD project teams.

Poster session

Twelve posters have been presented at the poster session, with time for explanations,

questions and discussion. They were from Serbia, North Macedonia, Nigeria, India,

Bangladesh and Sri Lanka, and they covered the whole range of the conference topics:

institutional repositories, national infrastructures, academic collections and digital libraries,

particular initiatives and projects, tools and procedures for ETD management, international

visibility, scientometrics, plagiarism and originality check, metadata and persistent identifiers,

policies and best practices. We can't present here all posters as they deserve; instead, I would

like to mention one poster from Nigeria on a rather unusual topic with high relevance for

people like me teaching Master and doctoral students: Emmanuel Olorunleke Eseyin from the

Nigerian Institute of Social and Economic Research (NISER) presented results of a study on

students' post plagiarism experience and resilience for quality research activities, with 378 postgraduates from four Nigerian Universities. The poster contained empirical evidence on how plagiarism affects students' research activities, on causes of plagiarized research, on challenge faced in overcoming plagiarism and on resilience strategy. An interesting, realistic, and inspiring contribution to the conference.

Technology fair

Opening the "technology fair" sponsor session, Meinhard Kettler (Germany) from ProQuest (Clarivate) presented results from a project on bibliographic coupling of 1.1 million theses and dissertations with full text with 149 million references. Similarity between documents is calculated based on the comparison of cited sources. The project's goal is to improve the retrieval performance of the ProQuest Dissertations and Theses Global database with a new recommender system. Kettler also announced the progressive integration of theses and dissertations in Clarivate's Web of Science and confirmed the possibility of future inclusion of other national ETD repositories, like the National Repository of Dissertations in Serbia (NaRDuS).

A new digital work environment for students called Digital Commons was presented by Pavel Milasevic from Elsevier. This platform allows "administrators from different graduate programs or departments to jointly manage electronic theses and dissertations instantly through a central system (and) gives students visibility and exposure for their research and scholarship (…) Faculty mentors and other stakeholders can also track impact to demonstrate the success of their departments and student research initiatives".

Roman Piontek from EBSCO described three new initiatives on linked data (Zepheira, a data service to normalize, connect and enrich catalogue or repository data and transform them into BIBFRAME and facilitate their discoverability on a multitude of platforms), on statistics (Panorama, a platform that can regularly harvest and visualize statistical or analytical

data from third-party systems and thus compare and consolidate otherwise disparate data such as data from a library system, an authentication system, a campus system etc.) and in the field of open source (FOLIO, a modular platform for library services).

Finally, Katya Levchenko and Silvana Pierri from Turnitin gave a presentation of the four core products of their portfolio of anti-plagiarism software (iThenticate, Similarity, Feedback Studio and Originality) which contribute to academic integrity in many institutions worldwide.

NDLTD Leadership Award

The 2022 NDLTD Leadership Award was given to Libio Huarato who is the director of the institutional repository of the Universidad Peruana de Ciencias Aplicadas S.A.C. Libio Huarato is member of the NDLTD Board of Directors and was host-organizer of the conference ETD 2012 in Lima, Peru. The award speech traces his career: "Since the 1980s, he has promoted access to theses and academic and research information through technological platforms. He was a pioneer in the use of the Cybertesis methodology in Peru, specifically at the Universidad Nacional Mayor de San Marcos. During the following years he became the advisor to several universities that began to use this methodology. Subsequently, he continued to promote technologies that allow the academic production of higher education entities to be brought closer and made visible, thus, he was also the promoter of the use of free software DSpace for the implementation of thesis repositories. These activities led him to participate in working groups convened by the National Council for Science, Technology, and Innovation (Concytec), which resulted in national application standards and meetings that allowed the promotion of good practices".

Perspectives

In order to facilitate the coming to Novi Sad, five travel grants have been awarded.

The ETD 2022 conference has been supported by three local authorities (University of Novi Sad, Government of Vojvodina, and Serbian Ministry of Science) and by five commercial sponsors (Clarivate-ProQuest, Elsevier, Turnitin, EBSCO, and Levi9). The social program comprised a reception at the university congress center, a conference dinner at the Terasa restaurant in the old Petrovaradin fortress high above the Danube and a walking tour around the historical town center of Novi Sad.

More information is available on the conference web site[viii], while all presentation materials including extended abstracts, posters, slides, pre-recorded videos are available on Zenodo[ix]. Extended versions of selected conference papers will be published by the *J-ETD* journal.

The presentations and posters of the ETD 2022 conference were rich, stimulating and thought-provoking, as were the discussions with the speakers and the other participants. Too rich, in fact, some short "take-away messages". Let's give it a try all the same. This is what I took back to France:

- FAIR principles may have been born in Europe but have become a global issue.

- FAIR principles have been developed for research data management but are also relevant for ETD management.

- Even those who don't speak about FAIR principles contribute to increasing compliance with these principles, through improved metadata quality, growing openness and availability, persistent identifiers and so on.

- Even if local, regional and national conditions are quite different, and even if one size still does not fit all, global and international discussion and cooperation are required because more than ever many challenges and issues are more or less common to all ETD projects.

In a way, Novi Sad marks a turning point or, at least, a new stage of the NDLTD initiative. The first purpose of NDLTD was the transition from print to digital. The second challenge was open access. The new challenge is interconnection of ETD infrastructures and machine-readability, and the way to go there is compliance with FAIR principles, opening perspectives on new infrastructures based on text and data mining, like those presented from the Virginia Tech team. Therefore, the Novi Sad conference was very timely and will go down in the history of NDLTD. A big thank you for the local organizers!

The next NDLTD symposium on electronic theses and dissertation will be organized by INFLIBNET, the Indian Information and Library Network Centre, at Ghandinagar, Gujarat, India, in November 2023.

---

[i] NDLTD https://ndltd.org/
[ii] University of Novi Sad http://www.uns.ac.rs/
[iii] Science 2.0 Alliance http://sci2zero.org/
[iv] ETD+ Toolkit http://educopia.org/etdplustoolkit
[v] PHAIDRA etheses https://eteze.bg.ac.rs/
[vi] NaRDuS https://nardus.mpn.gov.rs/?locale-attribute=en
[vii] National ETD Portal http://www.netd.ac.za/
[viii] ETD 2022 https://etd2022.uns.ac.rs/
[ix] Zenodo community https://zenodo.org/communities/etd2022