

## PAN-EUROPEAN REPOSITORY OF THESES AND DISSERTATIONS AND PAN-EUROPEAN PLAGIARISM DETECTION

Julius Kravjar

The unprecedented changes in the global environment are influencing the role and mode of operation of tertiary education systems all over the world. The importance of knowledge and innovation as drivers of progress is growing. Knowledge is the most important asset of modern society. The ability of a society to generate, adapt and apply knowledge is critical for sustainable economic growth and improved living standards (Salmi, 2015). The skill to analyse and to interpret data (textual data too) is a key factor in building and using knowledge, within both research and business.

“The combination of increased computing power, diminishing prices of hardware and software, [improvement] of wireless and satellite technologies, and reduced telecommunication costs has removed the space and time barriers.” Several new trends in the tertiary education landscape were identified and one of them is the growing impact of open science, big data and open education resources. To successfully fulfill “educational, research, and informational functions in the 21st century, tertiary education [institutions] need to be able to respond effectively to changing education and training needs, adapt to a rapidly shifting tertiary education landscape and adopt more flexible modes of organization and operation” (World Bank, 2002).

Technological advances and reduced costs bring many opportunities and challenges for the future development of higher education. Pan-European Repository of Theses and Dissertations as a rich base for the pan-European plagiarism detection could be one of these challenges in the era of Open Access, Open Data and Open Science.

The aim of the paper is to draw attention to the issues relating to storage of and access to theses and dissertations and their originality check on the EU level in the context of open science, and to provoke further discussions on these issues.

### *Open Science, Open Data, Open Access*

The open access initiative began in 1990s. Open Access is a model for online publishing of scholarly peer reviewed journals, made possible by the Internet (Laakso, 2011). Online publishing becomes a standard. Gradually, open access extended to other areas – to open data and open science – and became a high priority of European institutions.

A broader definition of open access is given by Ramjou, (2016). She understands open access as free on-line access to peer-reviewed scientific publications, access to research data and re-use of research data.

There is a number of open science definitions. Dutch Techcentre of Life Sciences (2016) offers this definition:

Open Science: an umbrella term for a technology and data driven systemic change in how researchers work, collaborate, share ideas, disseminate and reuse results, by

adopting the core values that knowledge should be reusable, modifiable and redistributable. One fundamental requirement for Open Science is that all research data and the associated tools and services should be findable, accessible, interoperable and reusable.

More general definition with less details is offered by C. Ramjou, (2016): “Open science is the transformation and opening up of science, research and innovation through information and communication technologies.”

Mons (2016) stresses the need for new approaches to the scientific data openness to support knowledge generation. “Up to 90% of all data in the world has been generated in the last two years.” Scientific data is in dire need of openness, better handling, careful management, machine actionability and sheer re-use. He sees openness as an enormous opportunity, not a threat but a tool that will help to overcome the paradigms of the last century.

The transition to the Open Science System is supported by EU research and innovation programme Horizon 2020. “The European Commission has taken a big step towards open science in Europe. All projects receiving Horizon 2020 funding are required to make sure that any peer-reviewed journal article they publish is openly accessible, free of charge.” (Open Access in Horizon 2020, 2016)

Open Science, Open Data and Open Access have massive support by all high level European institutions. Council of Europe in the document Council Conclusions on the Transition Towards an Open Science System (2016) accents different dimensions of an Open Science System:

Open access to scientific publications and optimal reuse of research data are of utmost importance for the development of open science. The Open Science Policy Platform established by the European Commission aims at supporting the further development of the European Open Science Policy and promoting the uptake by stakeholders of best practices, including issues such as adapting reward and evaluation systems, alternative models for open access publishing and management of research data (including archiving), altmetrics, guiding principles for optimal reuse of research data, development and use of standards, and other aspects of open science such as fostering research integrity and developing citizen science.

Unnecessary legal, organisational and financial barriers to access results of publicly funded research should be removed as much as possible and appropriate in order to attain optimal knowledge sharing, taking into account when necessary the need for exploitation of results. It is an imperative to remove barriers, and to take the necessary steps for successful implementation in all scientific domains, including specific measures for disciplines where obstacles hinder its progress.

We perceive the initiatives Open Access, Open Data and Open Science as a “big picture” that should be transformed into the real life. In this context, the “smaller picture” could include issues relating to the storage of and access to higher education theses and dissertations and their originality check on the EU level, and their potential implementation in everyday life of the academic community.

## Searching for Plagiarism Policies

In addition to open science, open data and open access, we wanted to find a further backing for the Pan-European Repository of Theses and Dissertations and Pan-European Plagiarism Detection. We conducted research concerning plagiarism policies in the EU.

The EU supports projects dealing with plagiarism policies and student plagiarism. An example could be the IPPHEAE project (“Impact of policies for plagiarism in higher education across Europe”, 2013). The project team investigated policies and procedures in place in higher education institutions across Europe for detecting and preventing student plagiarism. In the light of this project, it is a surprise that the websites of the European Commission contain little information on the EU’s plagiarism policies. We found that there is generally too little information about plagiarism.

Firstly, we pay attention to the DG Education, Youth, Sport and Culture. The Policies tab of the site <http://ec.europa.eu/education/> includes Strategic framework and Higher education. There is no mention of plagiarism.

We continued with the resources at the website [http://ec.europa.eu/education/library\\_en](http://ec.europa.eu/education/library_en) where we were interested in two document categories: EU policy and Higher education. In both categories, we worked with EU policy documents. The terms plagiarism and plagiarism policy were absent in these documents.

Further, we continued on the EC level. In the document Digital Futures – A Journey into 2050 Visions and Policy Challenges (2014), plagiarism is mentioned once: “How do we cope with unintended and undesirable effects of pervasive digitization of society such as media addictions, IPR and authenticity, counterfeiting, plagiarism, life history theft? How do we build trust in both artists and audiences?”

There are two documents about intellectual property rights (A Single Market for intellectual property rights, Enhancing the enforcement of intellectual property rights in the internal market) but there is nothing about plagiarism. The document Annual report to the Discharge Authority on internal audits carried out in 2014 states that an IT tool aimed at detecting double funding and plagiarism will be used across all Commission Research Services, striking the right balance between coverage of the riskiest projects and cost of controls. The relevant internal procedures to integrate plagiarism detection into current practices should be developed. In the document Implementation of the Commission Anti-Fraud Strategy (CAFS), plagiarism is mentioned as an internal issue in EC work<sup>1</sup>. Plagiarism is also mentioned in Research ethics and on Horizon 2020 website Ethics (“. . .avoiding fabrication, falsification, plagiarism or other research misconduct”). In the The OLAF report 2015<sup>2</sup>, there is the only sentence about plagiarism:

Examples include the delivery of the same piece of research to several funding authorities within or beyond EU borders, plagiarism – the copying of research which has already been undertaken by others, and the deliberate gross disrespect of the conditions of financial assistance.

<sup>1</sup>There is mentioned that the system URKUND will be used by several DGs for plagiarism check. But there is no information about the transparency of URKUND selection.

<sup>2</sup>OLAF is European Anti-Fraud Office

## Something New on the Horizon?

There are 28 member states of the EU. Higher education institutions in every member state produce different kinds of theses and dissertations every year. Are they all accessible from one website on a national level? No, they are not. Are there all accessible on the EU level? No, they are not. The member states deal with plagiarism like isolated islands. There is advanced and matured technology, but what about parliaments, governments and the public on the national and EU level?

Are the issues relating to the storage of and access to theses and dissertations and their originality check on the EU level irrelevant and unimportant? According to Fact Sheets on the European Union, there are 13 common policies in the EU. The 13th policy is the Culture, Education and Sport. One of the subpolicies is the Higher Education. There is no common policy on storage of and access to theses and dissertations and on plagiarism.

Open access to all EU theses and dissertation from one website could benefit the academic, research, scientific and business community.

The initiatives Open Science, Open Data and Open Access could be a fertile ground for the unborn Pan-European Repository of Theses and Dissertations. Who or what could be the trigger for the Pan-European Repository of Theses and Dissertations and Pan-European Plagiarism Detection initiatives? The best trigger would be if the member states and the European Commission too would identify this need. The value of the Pan-European Repository of Theses and Dissertations project would be higher if it would be followed up by the Pan-European Plagiarism Detection project. These two projects should be closely linked.

## Prerequisites

What could be the prerequisites that would support the Pan-European Repository of Theses and Dissertations and Pan-European Plagiarism Detection projects?

- The bottom-up and top-down will;
- Legislative framework;
- Clearly defined need;
- Appointing the responsible institution/institutions for:
  - Mapping and analysing the present state of theses and dissertations collection and issues related to various Acts (Copyright Act, Higher Education Act, Libraries Act, ...);
  - Preparation of recommendations;
  - Definition of system requirements;
  - Procurement;
  - Implementation;
  - Operation;
  - Maintenance;
  - Further development;

- Preparation of General Methodological Guidelines concerning requisites of theses and dissertations, their bibliographic registration, originality check, storage and access; and
- Well-defined infrastructures (human, legislative, organisational, financial, technical and technological).

It is based on our experience with the nationwide central repository and the plagiarism detection system (Kravjar, 2015; Kravjar, Noge 2013; Kravjar, 2013; Kravjar, 2012).

### **The Basic Features of the Pan-European Repository of Theses and Dissertations**

1. Collection of theses and dissertations, metadata from all EU member states to the central repository according to uniform collection methodology.
2. The uniform collection methodology will define the minimal set of documents and metadata.
3. Access to documents stored in the central repository will be as open as possible and as closed as necessary.
4. The user will have the option to view the selected documents
  - (a) in the original language;
  - (b) in the English language; and
  - (c) in the language of his/her choice (in one of the official languages of the EU).

The last feature to view the selected documents in the language of his/her choice (in one of the official languages of the EU) could be very demanding. The feature to view the selected documents in the original language and in English may be sufficient.

### **The Basic Features of the Pan-European Plagiarism Detection System**

1. Every document entering the central repository will be checked for originality by the plagiarism detection system against the central repository and other sources.
2. The system is able to detect cross-language plagiarism.
3. Outputs:
  - (a) Originality protocol that indicates:
    - i. the global similarity percentage of the checked document in comparison to the central repository and other sources;
    - ii. the similarity percentage of the checked document to the documents in which similarity was found;
    - iii. the text of the thesis/dissertation that is similar to other documents;
  - (b) Reports;
  - (c) Statistics.

The value of the system will be higher if higher education institutions delivered feedback to the system if:

- the thesis/dissertation has acceptable similarity index;
- the thesis/dissertation is suspected from plagiarism (unacceptable similarity index);
- in case the thesis/dissertation is suspected from plagiarism (unacceptable similarity index), the relevant higher education institution will inform the system about the measures taken.

## Conclusion

The aim of the paper is to draw attention to the issues relating to storage of and access to theses and dissertations and their originality check on the EU level and to provoke further discussions that could generate new insights into these issues.

Our research showed an unexpected result: there is no plagiarism policy concerning higher education on the EU level. Plagiarism is a “serious disease” that has to be treated on a family level, school level, national level and the European level too.

The existence of the European higher education plagiarism policy could significantly contribute to the fight against plagiarism and could be one of the mainstays. The backing for the Pan-European Repository of Theses and Dissertations and Pan-European Plagiarism Detection is open science, open data, open access and other necessary and important factors are bottom-up and top-down will, clearly defined needs and a regulatory framework.

What could transform the present situation into the operation of the Pan-European Repository of Theses and Dissertations and Pan-European Plagiarism Detection? Concerted and tenacious actions by the higher education sector across all EU Member States towards the European Commission. The legislation on the EU level could speed up the whole process markedly. “Where there’s a will, there’s a way”<sup>3</sup>.

According to Blessinger (2017)

Knowledge creation together with the means to disseminate that knowledge via improved learning platforms has created unprecedented opportunities for lifelong learning. Higher education systems, as centres of knowledge consumption and production, and as engines of economic growth and social development, have taken on new importance in this continually emerging global knowledge society.

The Pan-European Repository of Theses and Dissertations and Pan-European Plagiarism Detection have the real potential:

- to contribute to knowledge dissemination (all theses and dissertation from all EU member states will be accessible for the academic community and the public from one place);
- to raise awareness of plagiarism in an unprecedented way.

Other pan-European benefits (based on Slovak experience with a nationwide central repository of theses and dissertations and a nationwide plagiarism detection system) could include: increased responsibility of students and teachers, greater student autonomy in the creation of papers, improvement in the quality of papers, a higher

<sup>3</sup><http://www.phrasemix.com/phrases/where-theres-a-will-theres-a-way>

level of citations, contributions to improvements in education, better understanding and application of academic ethics, copyright and intellectual property rights (Kravjar, Noge, 2013).

In Slovakia, we have 7 years of experience with the operation of a nationwide central repository of theses and dissertations and a nationwide plagiarism detection system. In April 2010, “the SK ANTIPLAG system (a central repository of theses and dissertations, a plagiarism detection system, a comparative corpus, local repositories of theses and dissertations) started routine operation after a preparatory phase. The use of SK ANTIPLAG is mandatory for all Slovak higher education institutions operating under the Slovak legal order.” (Kravjar, 2015). It is an example of an unparalleled and unprecedented implementation of such a system on a national level that has no predecessor, it is a unique implementation worldwide (Kravjar, 2015).

The implementation of the SK ANTIPLAG system significantly raised awareness of plagiarism among students: “The responses from Slovak students demonstrated the highest level of understanding about plagiarism within the whole Europe.” (Plagiarism Policies in Slovakia, 2013). More information about SK ANTIPLAG can be found in the conference paper of Kravjar (2015).

## References

- DTL – Dutch Techcentre for life Sciences. (2016). Vision on Open Science. Retrieved from <http://www.dtls.nl/5825-2/>
- European Commission. (2009). Enhancing the enforcement of intellectual property rights in the internal market. Retrieved from [http://ec.europa.eu/internal\\_market/iprenforcement/docs/ip-09-1313/communication\\_en.pdf](http://ec.europa.eu/internal_market/iprenforcement/docs/ip-09-1313/communication_en.pdf)
- European Commission. (2011). A Single Market for intellectual property rights. Retrieved from [http://ec.europa.eu/internal\\_market/copyright/docs/ipr\\_strategy/COM\\_2011\\_287\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/ipr_strategy/COM_2011_287_en.pdf)
- European Commission. (2014). Digital Futures – A Journey into 2050 Visions and Policy Challenges. Retrieved from [http://espa.eu/orbis/sites/default/files/generated/document/en/Digital%20Futures\\_final\\_report.pdf](http://espa.eu/orbis/sites/default/files/generated/document/en/Digital%20Futures_final_report.pdf)
- European Commission. (2015). Annual report to the Discharge Authority on internal audits carried out in 2014. Retrieved from <https://ec.europa.eu/transparency/regdoc/rep/1/2015/EN/1-2015-441-EN-F1-1.PDF>
- European Commission. (2016). Implementation of the Commission Anti-Fraud Strategy (CAFS). Retrieved from [http://ec.europa.eu/anti-fraud/sites/antifraud/files/cafs\\_implementation\\_2015\\_en.pdf](http://ec.europa.eu/anti-fraud/sites/antifraud/files/cafs_implementation_2015_en.pdf)
- European Commission. (2016). Open Access in Horizon 2020. Retrieved from <http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess>
- European Commission. (2016). The OLAF report 2015. Retrieved from [http://ec.europa.eu/anti-fraud/sites/antifraud/files/olaf\\_report\\_2015\\_en.pdf](http://ec.europa.eu/anti-fraud/sites/antifraud/files/olaf_report_2015_en.pdf)
- European Commission. (2017). DG Education, Youth, Sport and Culture. Retrieved from <http://ec.europa.eu/education/>
- European Commission. (2017). DG Education, Youth, Sport and Culture. Retrieved from [http://ec.europa.eu/education/library\\_en](http://ec.europa.eu/education/library_en)
- European Commission. (2017). Ethics. Retrieved from <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/ethics>

- European Commission. (2017). Fact Sheets on the European Union. Sectoral policies. Retrieved from <http://www.europarl.europa.eu/atyourservice/en/displayFtu.html?ftuld=theme5.html>
- European Commission. (2017). Research ethics. Retrieved from <http://ec.europa.eu/research/swafs/index.cfm?pg=policy&lib=ethics>
- IPPHEAE Project Consortium. (2013). Impact of policies for plagiarism in higher education across Europe. Retrieved from <http://plagiarism.cz/ippheae/>
- IPPHEAE Project Consortium. (2013). Plagiarism Policies in Slovakia. In *Conference Proceedings Impact of Policies for Plagiarism in Higher Education Across Europe*. Retrieved from <http://plagiarism.cz/ippheae/files/D2-3-26%20SK%20IPPHEAE%20MENDELU%20Survey%20SlovakiaNarrative%20FINAL.pdf>
- KRAVJAR, J. (2012). Barrier to Thriving Plagiarism. In *Conference Proceedings of The 5th International Plagiarism Conference*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.7921&rep=rep1&type=pdf>
- KRAVJAR, J., NOGE, J. (2013). Strategies and Responses to Plagiarism in Slovakia. In *Conference Proceedings Plagiarism across Europe and Beyond*, pp. 201–215. Retrieved from <http://ippheae.pefka.mendelu.cz/files/proceedings.pdf>
- KRAVJAR, J. (2013). The Path to an Award-Winning Antiplagiarism System for Higher Education Institutions. 2013 <https://www.linkedin.com/pulse/path-award-winning-antiplagiarism-system-higher-julius-kravjar> (originally [http://www.scilogs.com/scientific\\_and\\_medical\\_libraries/the-path-to-an-award-winning-antiplagiarism-system-for-higher-education-institutions/](http://www.scilogs.com/scientific_and_medical_libraries/the-path-to-an-award-winning-antiplagiarism-system-for-higher-education-institutions/))
- KRAVJAR, J. (2015). SK ANTIPLAG is bearing fruit. In *Conference Proceedings Plagiarism Across Europe and Beyond* (pp. 147–163). Retrieved from [http://plagiarism.pefka.mendelu.cz/files/proceedings\\_15.pdf](http://plagiarism.pefka.mendelu.cz/files/proceedings_15.pdf)
- KRAVJAR, J. (2016). Pan-European Plagiarism Detection System ICT Proposers' Day 2016. Retrieved from <https://ec.europa.eu/digital-single-market/events/cf/ict-proposers-day-2016/item-display.cfm?id=19221>
- KRAVJAR, J. (2016). Pan-European Plagiarism Detection System. ICT Proposers' Day 2016. Retrieved from [https://ec.europa.eu/digital-single-market/events/cf/ict-proposers-day-2016/document.cfm?doc\\_id=36047](https://ec.europa.eu/digital-single-market/events/cf/ict-proposers-day-2016/document.cfm?doc_id=36047)
- LAAKSO, M., WELLING, P., BUKVOVA, H., NYMAN, L., BJORK, B., HEDLUND, T. (2011). The Development of Open Access Journal Publishing from 1993 to 2009. PLOS ONE, June 10, 2011. Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020961>
- MONS, B. (2016). Realising the European Open Science Cloud: First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud (Preface, p. 5). Retrieved from [https://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf)
- RAMJOU, C. (2016). Open Access / Open Science. ICT Proposers' Day 2016. Retrieved from <https://ec.europa.eu/digital-single-market/events/cf/ict-proposers-day-2016/item-display.cfm?id=19132>
- SALMI, J. (2015). Study on Open Science – Impact, Implications and Policy Options. Retrieved from [https://ec.europa.eu/research/innovation-union/pdf/expert-groups/rise/study\\_on\\_open\\_science-impact\\_implications\\_and\\_policy\\_options-salmi\\_072015.pdf](https://ec.europa.eu/research/innovation-union/pdf/expert-groups/rise/study_on_open_science-impact_implications_and_policy_options-salmi_072015.pdf)
- SALMI, J. (2016). Tertiary Education and the Sustainable Development Goals. In Search of a Viable Funding Model. Retrieved from <http://report.educationcommission.org/download/877/>
- The Council of the EU. (2016). Council Conclusions on the Transition towards an Open Science System. Retrieved from <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>

World Bank. (2002). Constructing Knowledge Societies: New Challenges for Tertiary Education. Retrieved from [http://siteresources.worldbank.org/INTAFRREGTOPTEIA/Resources/Constructing\\_Knowledge\\_Societies.pdf](http://siteresources.worldbank.org/INTAFRREGTOPTEIA/Resources/Constructing_Knowledge_Societies.pdf)

### *Copyright statement*

Copyright © 2017. Author(s) listed on the first page of article: The authors grant to the organizers of the conference “Plagiarism across Europe and beyond 2017” and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the conference proceedings. Any other usage is prohibited without the express permission of the authors.

### **Author**

Julius Kravjar, Slovak Centre of Scientific and Technical information, Slovak Republic