Journal of Integrated Information Management

Vol 8, No 1 (2023)

Jan-June 2023

Archives and records management in machine learning technologies context

Ioannis Triantafyllou, Christos Chrysanthopoulos, Yannis Stoyannidis, Anastasios Tsolakidis

To cite this article:

Triantafyllou, I., Chrysanthopoulos, C., Stoyannidis, Y., & Tsolakidis, A. (2024). Archives and records management in machine learning technologies context: a research hypothesis on university records. *Journal of Integrated Information Management*, *8*(1), 7–13. Retrieved from https://ejournals.epublishing.ekt.gr/index.php/jiim/article/view/37908

Journal of Integrated Information Management electronic journal

Archives and records management in machine learning technologies context: a research hypothesis on university records

Ioannis Triantafyllou[1], **Christos Chrysanthopoulos**[2], **Yannis Stoyannidis**[1], **Anastasios Tsolakidis**[3]
[1]Department of Archival, Library & Information Studies, University of West Attica, Athens
[2]Department of History & Archaeology, University of Patras and Department of Archival, Library & Information Studies, University of West Attica
[3]Department of Informatics and Computer Engineering, University of West Attica, Athens

triantafi@uniwa.gr [**ORCID**: 0000-0001-5273-0855], cchrysan@eie.gr [**ORCID**: 0000-0001-9900-1342], ystoyannidis@uniwa.gr [**ORCID**: 0000-0001-9551-8360], atsolakid@uniwa.gr [**ORCID**: 0000-0001-7364-4542]

Article Info

Article history:
Received 01 March 2023
Received in revised form 30 May 2023
Accepted 15 June 2023

http://dx.doi.org/10.26265/jiim.v8i1.4517

Abstract:

Purpose *- This paper explores the goal and potential of integrating machine learning technologies into archives and records management practices. As the volume and complexity of digital records continue to grow, traditional methods of organising, classifying, and managing records face new challenges. Machine learning technologies offer opportunities to revolutionise how records are maintained, accessed, and used.*

Design/methodology/approach *- The relationship between records and archive management and machine learning practices is presented through the literature. This paper proposes a case study implementation of machine learning practices for the subject classification of records at the University of West Attica.*

Findings *- This paper presents a research hypothesis placing the subject classification of records at the center of the discussion. It highlights the necessity of deepening the standardisation of government actions record management processes.*

Originality/value *- By exploring this topic, the paper seeks to contribute to a deeper understanding of the transformative role that machine learning technologies can play in archives and records management and to inform future practices and decision-making in the field. It is also the first theoretical part of an ongoing research project on the subject classification of the University of West Attica records.*

Index Terms — Archives and records management, Machine learning, Computational archival science, University archives, Subject classification

I. Introduction

Records and archives management classification, as a method of identifying and organising the records generated or received throughout business, is advised by archival theory to be based on an examination of the functions and activities of the records creators and reflect them. Methodologically, at the heart of archival theories and practices is the provenance of the records, thus placing it at the center of the archival description. **[1]** According to this theory, the relevance of records is greatly influenced by the context in which they were created, and the organisation and description of these items should be closely tied to their original role. **[2]** The principle of provenance, when applied to appraisal, encourages the use of an organisational method, a "top-down" approach, which has proven to be unsatisfactory because it leaves out the "powerless transactions," which might shed light on the larger social context, from the permanent documentation of society. **[3]** Traditionally the concept of provenance is intimately connected to the concept of original order. In this context, the subject classification of records is subordinate and interests archivists only if it is closely related to a function of the creator. Nowadays, born-digital records have created new needs in recordkeeping and archive management.

Duranti notes that there is essentially no difference between a traditional and a born-digital record because the elements that must be explicit in an electronic record are implicit in any analog/traditional record. The born-digital record has a structure with "date (time and place of creation, transmission, and receipt), an indication of persons (author, addressee, originator, writer, and creator), an indication of action or matter (title or subject), classification code, and any other element required by the creator's procedures or juridical system" **[4].** Although this condition may not

differentiate the traditional from the electronic record, it may create new processes in record management. What role can machine learning and artificial intelligence play in archives and records management?

This paper explores the potential benefits, challenges, and implications of integrating machine learning technologies into university archives and record management. By investigating the research hypothesis, we aim to explore how these technologies can improve the efficiency, accuracy, and accessibility of archival practices, ultimately benefiting academic and other institutions.

II. Machine Learning and Archives

Machine learning [ML] has the potential to significantly impact record management practices, offering new opportunities for efficiency, automation, and enhanced decision-making. Below we present some ways in which machine-learning systems can intersect with record management:

- Automated Classification and Metadata Extraction: ML algorithms can be trained to automatically classify records based on their content or extract relevant metadata from documents. This automation reduces the manual effort required for record classification and metadata entry, saving time and cost. ML models can learn from existing labelled data or use unsupervised learning techniques to identify patterns and make accurate predictions about record classifications. **[5]**
- Intelligent Search and Retrieval: ML algorithms can enhance search and retrieval capabilities within record management systems. By analysing user behavior, query patterns, and contextual information, machine learning models can provide personalised and relevant search results, improving the user experience and increasing the efficiency of information retrieval. Natural language processing techniques can also be employed to understand and interpret user queries, allowing for more accurate and contextual search results. **[6]**
- Record Deduplication and Data Cleansing: ML algorithms can assist in identifying and removing duplicate or redundant records within a collection. By comparing the content, metadata, or other characteristics of records, machine learning models can automatically flag potential duplicates, enabling efficient data cleansing and improving data quality. The above helps reduce storage costs, ensure data integrity, and streamline record management processes. **[7]**
- Predictive Analytics and Decision Support: ML algorithms can analyse historical record data and identify patterns or trends, enabling predictive analytics and decision support capabilities. By leveraging machine learning models, organisations can gain insights into records, such as identifying potential risks, predicting record retention needs, or identifying records requiring special attention or disposition. These insights can support more informed decision-making and aid in developing proactive record management strategies. **[8]**
- Security and Risk Management: ML techniques can be utilised to identify potential security risks or anomalies within record management systems. Machine learning models can analyse user access patterns, identify unusual behaviors, and detect potential security breaches or data leaks. By leveraging these models, organisations can enhance their security measures, identify and mitigate risks, and ensure compliance with data protection regulations. **[9]**

It is important to note that while ML offers significant potential for improving record management practices, careful consideration must be given to data privacy, ethics, and the transparency of algorithms. Organisations should ensure proper training and validation of machine learning models and regular monitoring and evaluation to maintain their accuracy and effectiveness. Additionally, human expertise and judgment remain essential in implementing and overseeing machine learning applications within record management processes.

III. Computational Archival Science and Machine Learning

Computational archival science is an emerging field that explores the intersection of archival science and computational methods, including machine learning. **[10]** It leverages computational techniques and technologies to analyse, process, and manage large-scale archival collections, enabling new insights and capabilities for archival practice. As a subfield of computational methods, machine learning is crucial in advancing computational archival science. Expect automated processing, classification, and analysis of the archival collection; machine learning algorithms, such as optical character recognition (OCR) and image recognition models, can be applied to archival documents. OCR can convert scanned or handwritten text into machine-readable formats, facilitating full-text search and analysis. Image recognition models can identify and classify visual elements, such as photographs, maps, or diagrams, aiding the organisation and contextualisation of archival materials. **[11]**

Machine learning techniques can be utilised to mine and extract valuable information from archival collections. Machine learning models can identify significant entities, events, or subjects by analysing patterns, relationships, and trends within the data. This information extraction process can assist in creating semantic connections and generating metadata, enabling enhanced search, retrieval, and analysis of archival materials. By employing techniques such as clustering, dimensionality reduction, or topic modeling, machine learning models can reveal hidden patterns or

relationships within the data. **[12]** Visualisation tools can then present these findings intuitively and interactively, enabling users to explore archival collections from various perspectives. Algorithms can be trained to identify potential physical or digital records risks, such as deterioration, mold, or data corruption. By leveraging machine learning models, archivists can proactively monitor and address preservation needs, ensuring the longevity and integrity of archival materials. **[13]** Additionally, interdisciplinary collaboration between archival professionals, data scientists, and domain experts is crucial to ensure the ethical and responsible application of machine learning techniques in archival practice. Along these lines have been developed several software tools for archival practices **[Table 1.]**.

Software	Creator	Description
ePADD	Stanford University Libraries	Free and open-source software developed that supports the appraisal, processing, preservation, discovery, and delivery of historical email archives.
BitCurator NLP	BitCurator Consortium	Software for collecting institutions to extract, analyse, and produce reports on features of interest in text extracted from born-digital materials contained in collections.
ArchExtract	Bancroft Library (University of California Berkeley)	A web application that enables archivists and researchers to perform topic modeling, keyword, and named entity extraction on a text collection.

Table 1. Representative software for the management of archival material using machine learning

IV. Machine Learning in a University Record Management System: an Experimental Question

University archives and record management are vital in preserving and managing universities' historical, administrative, and cultural records. The connection between the past and the present is made possible by university archives, which also act as a lighthouse for highlighting the contributions made by educational institutions to society. **[14]** Record management in universities involves systematically controlling and coordinating records throughout their lifecycle, from creation to final disposition. Today at the University of West Attica, as in all universities in Greece, there are different software/services for managing bureaucracy and producing records. The Greek government started the Diavgeia project (Transparency Program Initiative) in 2010, intending to restore faith in the democratic system and enable online insights into government spending. Diavgeia has been pivotal in promoting transparency and accountability in the Greek government. By making public administration decisions readily available, the platform has created a culture of openness, allowing citizens to scrutinise government actions and hold public officials accountable. This transparency has helped to reduce corruption, increase public trust, and foster a more accountable government. **[15]** The Diavgeia project has transformed the government information landscape in Greece by introducing transparency, accountability, and open access to public administration decisions. It has empowered citizens, improved governance practices, and inspired similar initiatives worldwide, creating a more open and informed society. **[16]**

The records classification on the platform is essential for effective information retrieval and analysis. However, ensuring accurate and consistent multiple-subject classification can be complex. The uploading of documents to the Diavgeia portal is done manually by university employees in a structured digital environment. The type of university actions and the subject categories are manually selected from the corresponding drop-down lists **[Fig. 1-2]**. This practice creates confusion between the type of document and its subject categories indexing.

Records published on Diavgeia cover a wide range of subjects and topics, reflecting the diverse activities of public administration. The subject matter can vary from legal and regulatory matters to infrastructure projects, public procurement, and personnel issues. Categorising such diverse content can be challenging, as decisions may touch on multiple subjects or fall into ambiguous categories. With this method, among others, a part of the university's administrative function is classified and open. On the other hand, the subject assignment of a document may not always align perfectly with the content or scope of the document. Interpreting and assigning relevant subject categories can be subjective, leading to potential confusion or misclassification.

The University of West Attica has uploaded 73.550 records from 2018 **[17]** until the end of 2022 in the Diavgeia portal **[Fig.3, Table 2, 3]**. The critical question is to what extent the structure of the Diavgeia portal reflects the origin and original order of the university's records. Especially if we consider it a broader service that applies to many different public sector organisations with various characteristics. To evaluate the effects of machine learning technologies on subject classifications, we could collect university records from Diavgeia Portal and implement machine learning model training and prediction.

The Diavgeia portal administrators must continually review and refine their classification system to address the abovementioned challenges. Implementing automated or semi-automated techniques, such as natural language processing or machine learning algorithms, can improve document classification accuracy and consistency. These

technologies can analyse the textual content of documents and suggest appropriate document types and subject categories based on patterns and predefined rules.

Furthermore, feedback mechanisms and user engagement can be vital in identifying and rectifying misclassifications. Users of the Diavgeia portal can report inaccuracies or provide suggestions for improvement, allowing administrators to make necessary adjustments to enhance the classification process.

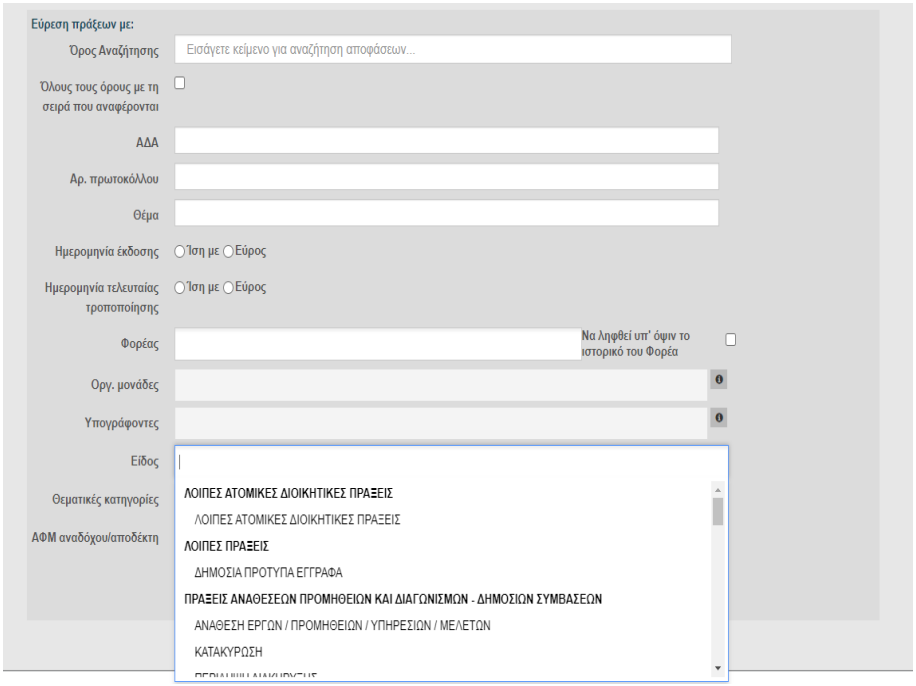

Fig. 1. Type of Government Actions on Diavgeia portal

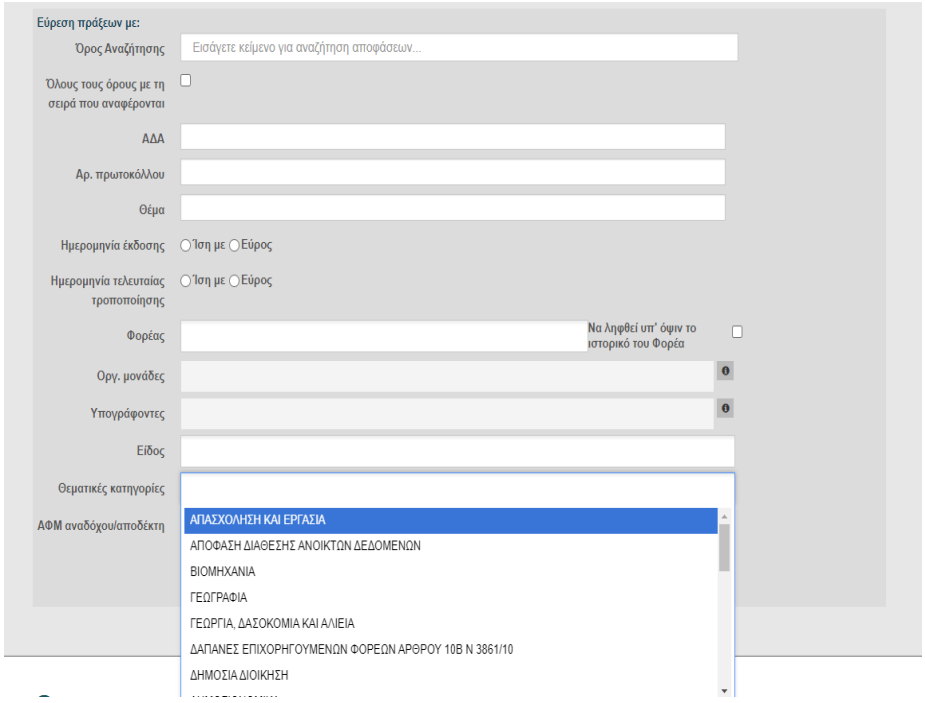

Fig. 2. Subject Categories of Government Actions on Diavgeia portal

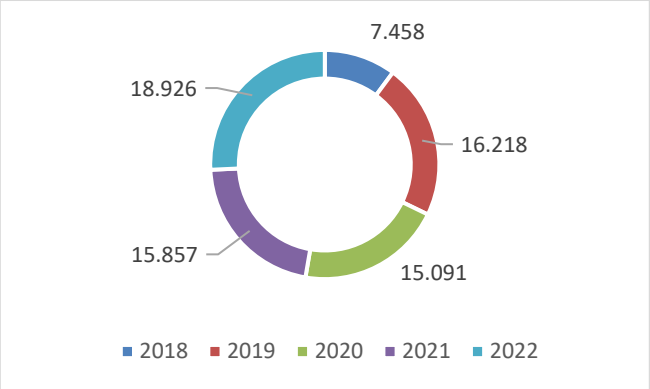

Fig. 3. Total number of entries on Diavgeia portal by the University of West Attica per year

Type of University Action (single)	Number
<type>ΕΓΚΡΙΣΗ ΔΑΠΑΝΗΣ</type>	30.268
<type>ΑΝΑΛΗΨΗ ΥΠΟΧΡΕΩΣΗΣ</type>	12.808
<type>ΛΟΙΠΕΣ ΑΤΟΜΙΚΕΣ ΔΙΟΙΚΗΤΙΚΕΣ ΠΡΑΞΕΙΣ</type>	5.412
<type>ΑΝΑΘΕΣΗ ΕΡΓΩΝ / ΠΡΟΜΗΘΕΙΩΝ / ΥΠΗΡΕΣΙΩΝ / ΜΕΛΕΤΩΝ</type>	5.134
<type>ΕΓΚΡΙΣΗ ΠΡΟΥΠΟΛΟΓΙΣΜΟΥ</type>	5.108
<type>ΚΑΝΟΝΙΣΤΙΚΗ ΠΡΑΞΗ</type>	4.968
<type>ΟΡΙΣΤΙΚΟΠΟΙΗΣΗ ΠΛΗΡΩΜΗΣ</type>	4.901
<type>ΠΡΑΞΗ ΠΟΥ ΑΦΟΡΑ ΣΕ ΣΥΛΛΟΓΙΚΟ ΟΡΓΑΝΟ - ΕΠΙΤΡΟΠΗ - ΟΜΑΔΑ ΕΡΓΑΣΙΑΣ - ΟΜΑΔΑ ΕΡΓΟΥ - ΜΕΛΗ ΣΥΛΛΟΓΙΚΟΥ ΟΡΓΑΝΟΥ</type>	3.950
<type>ΣΥΜΒΑΣΗ</type>	494
<type>ΠΕΡΙΛΗΨΗ ΔΙΑΚΗΡΥΞΗΣ</type>	269
<type>ΥΠΗΡΕΣΙΑΚΗ ΜΕΤΑΒΟΛΗ</type>	105
<type>ΠΡΟΚΗΡΥΞΗ ΠΛΗΡΩΣΗΣ ΘΕΣΕΩΝ</type>	56
<type>ΙΣΟΛΟΓΙΣΜΟΣ – ΑΠΟΛΟΓΙΣΜΟΣ</type>	39
<type>ΔΙΟΡΙΣΜΟΣ</type>	20
<type>ΠΡΑΞΗ ΠΟΥ ΑΦΟΡΑ ΣΕ ΘΕΣΗ ΓΕΝΙΚΟΥ - ΕΙΔΙΚΟΥ ΓΡΑΜΜΑΤΕΑ - ΜΟΝΟΜΕΛΕΣ ΟΡΓΑΝΟ</type>	11
<type>ΚΑΤΑΚΥΡΩΣΗ</type>	3
<type>ΠΡΑΞΕΙΣ ΧΩΡΟΤΑΞΙΚΟΥ - ΠΟΛΕΟΔΟΜΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ</type>	2
<type>ΔΩΡΕΑ - ΕΠΙΧΟΡΗΓΗΣΗ</type>	1
<type>ΕΓΚΥΚΛΙΟΣ</type>	1

Table 2. Number of entries per Type of University Action

Subject of University Action (multiple)	Number
<subject>ΠΑΡΑΓΩΓΗ, ΤΕΧΝΟΛΟΓΙΑ ΚΑΙ ΕΡΕΥΝΑ</subject>	42.147
<subject>ΕΠΙΣΤΗΜΕΣ</subject>	38.513
<subject>ΟΙΚΟΝΟΜΙΚΕΣ ΚΑΙ ΕΜΠΟΡΙΚΕΣ ΣΥΝΑΛΛΑΓΕΣ</subject>	14.674
<subject>ΕΠΙΚΟΙΝΩΝΙΑ ΚΑΙ ΜΟΡΦΩΣΗ</subject>	14.553
<subject>ΔΗΜΟΣΙΟΝΟΜΙΚΑ</subject>	6.980
<subject>ΑΠΑΣΧΟΛΗΣΗ ΚΑΙ ΕΡΓΑΣΙΑ</subject>	650
<subject>ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ</subject>	299
<subject>ΟΙΚΟΝΟΜΙΚΗ ΖΩΗ</subject>	8
<subject>ΔΑΠΑΝΕΣ ΕΠΙΧΟΡΗΓΟΥΜΕΝΩΝ ΦΟΡΕΩΝ ΑΡΘΡΟΥ 10Β Ν 3861/10</subject>	4
<subject>ΕΝΕΡΓΕΙΑ</subject>	4
<subject>ΕΠΙΧΕΙΡΗΣΕΙΣ ΚΑΙ ΑΝΤΑΓΩΝΙΣΜΟΣ</subject>	2
<subject>ΑΠΟΦΑΣΗ ΔΙΑΘΕΣΗΣ ΑΝΟΙΚΤΩΝ ΔΕΔΟΜΕΝΩΝ</subject>	1
<subject>ΕΥΡΩΠΑΪΚΗ ΈΝΩΣΗ</subject>	1
<subject>ΥΓΕΙΑ</subject>	1
<subject>ΒΙΟΜΗΧΑΝΙΑ</subject>	0
<subject>ΓΕΩΓΡΑΦΙΑ</subject>	0
<subject>ΓΕΩΡΓΙΑ, ΔΑΣΟΚΟΜΙΑ ΚΑΙ ΑΛΙΕΙΑ</subject>	0
<subject>ΔΙΑΤΡΟΦΗ ΚΑΙ ΓΕΩΡΓΙΚΑ ΠΡΟΪΟΝΤΑ</subject>	0
<subject>ΔΙΕΘΝΕΙΣ ΟΡΓΑΝΙΣΜΟΙ</subject>	0
<subject>ΔΙΕΘΝΕΙΣ ΣΧΕΣΕΙΣ</subject>	0
<subject>ΔΙΚΑΙΟ</subject>	0
<subject>ΚΟΙΝΩΝΙΚΑ ΘΕΜΑΤΑ</subject>	0
<subject>ΜΕΤΑΦΟΡΕΣ</subject>	0
<subject>ΠΕΡΙΒΑΛΛΟΝ</subject>	0
<subject>ΠΟΛΙΤΙΚΗ ΖΩΗ</subject>	0

Table 3. Number of entries per Subject of University Action

The absence of standardised subject categories across all public administration entities in Greece poses a challenge to the consistent categorisation of Diavgeia. Different entities may use different classification schemes or terminologies, making establishing a uniform and comprehensive subject taxonomy difficult. This lack of standardisation can hinder effective information retrieval and cross-referencing of related records. Public administration decisions often involve complex subject matter that requires in-depth understanding and expertise to categorise accurately. Decisions may involve technical, legal, or specialised terminology that requires domain knowledge for proper classification. Ensuring that subject categories capture the nuanced aspects of the decisions can be challenging, particularly when limited contextual information is provided. In addition, the subject categories on Diavgeia must adapt to the evolving nature of public administration and address emerging topics or issues. New policy areas, technological advancements, or societal changes may introduce subjects not previously accounted for in the categorisation scheme. Regular updates and adjustments to the subject categories are necessary to ensure the relevance and coverage of the platform. To address these challenges, Diavgeia needs to establish a well-defined and comprehensive subject taxonomy that reflects the breadth of public administration activities. This taxonomy should be developed in consultation with relevant stakeholders, including public administration entities, subject matter experts, and platform users. Regular reviews and updates of the subject categories are essential to accommodate changes and evolving needs. Additionally, providing guidelines and training to public administration entities regarding subject categorisation can help improve

consistency and accuracy. Encouraging feedback and collaboration from users of Diavgeia can also assist in refining the subject categories and addressing any ambiguities or gaps.

The proposed research aims to investigate the impact of machine learning technologies on the classification of records at the University of West Attica. A similar methodology has been applied to other aspects of academic activity, such as the repository of the scientific activity of university members. **[18, 19]** It has to be mentioned that the etchings of applying machine learning practices to archival practices differ in their use in digital libraries and repositories. Striking a balance between AI technologies and human expertise will be crucial in leveraging the benefits of AI while upholding the core principles and values of archival practice. **[20]**

By exploring the potential effects, this research seeks to improve record management practices and enhance the accessibility and retrieval of archival materials within the university. The research question is clearly stated, focusing on the specific context of the University of West Attica and the application of machine learning technologies. Potential research outcomes could include improved accuracy and consistency in subject categorisation, enhanced efficiency in record management processes, and increased discoverability of records within the University of West Attica. The research may also identify challenges, limitations, or ethical considerations associated with applying machine learning technologies in subject classification. The research has practical implications for the University of West Attica and similar institutions. If the study demonstrates positive effects, machine learning technologies could significantly streamline and automate the subject categorisation process, reducing manual effort and improving the overall productivity of record management staff. The findings can inform the development of guidelines and best practices for implementing machine learning services for subject categorisation within the university's record management framework. In addition, it is possible to test methodologies applied in a different context and produced valid results. **[21]**

V. Conclusion

The dynamic relationship between archival science and machine learning practices and, more broadly, AI is ongoing and can reshape theoretical and methodological schemes. By embracing these technologies within the framework of the Records Continuum model, organisations can effectively manage born-digital archives, ensuring their long-term preservation and facilitating their valuable contribution to research, historical documentation, and organisational memory. **[22]**

In conclusion, this theoretical paper investigated the application of machine learning technologies in archives and records management, with a particular focus on government actions. The research hypothesis was that integrating machine learning technologies can significantly improve the efficiency, accuracy, and accessibility of the University of West Attica records. Through automated technologies, ongoing refinement of classification systems, and user engagement, efforts can be made to minimise confusion and improve the overall effectiveness of the portal in delivering reliable and well-categorised information to the public. By embracing these technologies, academic institutions can enhance the efficiency, accuracy, and accessibility of their records, ultimately benefiting researchers, administrators, and the wider society. As the field of archival science continues to evolve, further research and collaboration between archivists, data scientists, and stakeholders will be vital to fully realise the potential of machine learning in transforming university records management practices.

VI. References

[1] Sweeney, S. (2008). The ambiguous origins of the archival principle of "provenance". *Libraries & the Cultural Record*, *43*(2), 193-213. https://www.jstor.org/stable/25549475

[2] Hensen, S. L. (1993). The first shall be first: APPM and its impact on American archival description. *Archivaria*, *35*, 64-70. https://www.archivaria.ca/index.php/archivaria/article/view/11886

[3] Duranti, L. (1998). *Diplomatics: New Uses for an Old Science. Lanham*, MD: Scarecrow Press, in association with the Society of American Archivists and Association of Canadian Archivists, 177.

[4] Duranti, L. (2001). The impact of digital technology on archival science. *Archival Science*, *1*(1), 39-55. https://doi.org/10.1007/BF02435638

[5] Pandey, N., Sanyal, D. K., Hudait, A., & Sen, A. (2017). Automated classification of software issue reports using machine learning techniques: an empirical study. *Innovations in Systems and Software Engineering*, *13*, 279-297. https://doi.org/10.1007/s11334-017-0294-1

[6] Yakel, E., & Torres, D. (2003). AI: Archival intelligence and user expertise. *The American Archivist*, *66*(1), 51-78.

[7] Manghi, P., Mikulicic, M., & Atzori, C. (2012). De-duplication of aggregation authority files. *International Journal of Metadata, Semantics, and Ontologies*, *7*(2), 114-130. https://doi.org/10.17723/aarc.66.1.g022h85pn51n5800

[8] Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). Organisational business intelligence and decision making using big data analytics. *Information Processing & Management*, *58*(6), 102725. https://doi.org/10.1016/j.ipm.2021.102725

[9] Fennelly, L. J. (2014). *Museum, archive, and library security*. Butterworth-Heinemann.

[10] Stančić, H. (2018). Computational archival science. *Moderna Arhivistika*, *1*(2), 323-330. Computational-Archival-Science.pdf (researchgate.net)

[11] Traub, M. C., Van Ossenbruggen, J., & Hardman, L. (2015). Impact analysis of OCR quality on research tasks in digital archives. In *Research and Advanced Technology for Digital Libraries: 19th International Conference on Theory and Practice of Digital Libraries, TPDL 2015, Poznań, Poland, September 14-18, 2015, Proceedings 19* (pp. 252-263). Springer International Publishing. https://doi.org/10.1007/978-3-319-24592-8_19

[12] Jo, E. S., & Gebru, T. (2020, January). Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 306-316). https://doi.org/10.1145/3351095.3372829

[13] Thibodeau, K. (2018, December). Computational Archival Practice: Towards A Theory for Archival Engineering. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 2753-2760). IEEE. https://doi.org/10.1109/BigData.2018.8622174

[14] Chrysanthopoulos, C., Drivas, I., Kouis, D., & Giannakopoulos, G. (2023). University archives: the research road travelled and the one ahead. *Global Knowledge, Memory and Communication*, *72*(1/2), 44-68. https://doi.org/10.1108/GKMC-08-2021-0128

[15] Karamagioli, E., Staiou, E. R., & Gouscos, D. (2014). Government spending transparency on the internet: an assessment of Greek bottom-up initiatives over the Diavgeia Project. *International Journal of Public Administration in the Digital Age (IJPADA)*, *1*(1), 39-55. https://doi.org/10.4018/ijpada.2014010103

[16] Mastora, A., Koloniari, M., & Monopoli, M. (2021). The Government Information Landscape in Greece. *IFLA Professional Reports*, 64-82. The Government Information Landscape in Greece - ProQuest

[17] The University of West Attica (UNIWA) was founded in March 2018 by the National Law 4521.

[18] Vorgia, F., Triantafyllou, I., & Koulouris, A. (2017). Hypatia Digital Library: A text classification approach based on abstracts. In *Strategic Innovative Marketing: 4th IC-SIM, Mykonos, Greece 2015* (pp. 727-733). Springer International Publishing. https://doi.org/10.1007/978-3-319-33865-1_89

[19] Triantafyllou, I., Vorgia, F., & Koulouris, A. (2019). Hypatia Digital Library: A novel text classification approach for small text fragments. *Journal of Integrated Information Management*, *4*, 16-23. https://doi.org/10.26265/jiim.v4i2.4420

[20] Cushing, A.L. and Osti, G. (2023), ""So how do we balance all of these needs?": how the concept of AI technology impacts digital archival expertise", *Journal of Documentation*, *79*(7), pp. 12-29. https://doi.org/10.1108/JD-08-2022-0170

[21] Triantafyllou, I., Drivas, I. C., & Giannakopoulos, G. (2020). How to Utilise my App Reviews? A Novel Topics Extraction Machine Learning Schema for Strategic Business Purposes. *Entropy*, *22*(11), 1310. https://doi.org/10.3390/e22111310

[22] Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2021). Archives and AI: An overview of current debates and future perspectives. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, *15*(1), 1-15. https://doi.org/10.1145/3479010

VII. AUTHORS

Ioannis Triantafyllou holds a PhD from the National Technical University of Athens, Department of Electrical & Computer Engineering, and is currently an Associate Professor at the Department of Archives, Library and Information Studies at the University of West Attica. He has previously worked as a research associate in many European and Greek research projects at the Institute of Language & Speech Processing (ILSP / Athena RC). Recently, he participated in the CrossCult research program (Horizon2020) as a research team member. He specialises in Digital Libraries, Data & Text Mining, Text Classification & Clustering, Ontologies & Metadata, Linked Data, Information Extraction, Text & Information Retrieval, Automated Summary & Text Synthesis, Translation Memories, etc.

Christos Chrysanthopoulos is a PhD candidate at the University of Patras (GR), Department of History and Archaeology (former Department of Cultural Heritage Management and New Technologies). His doctoral research focuses on "Digital Public History". He is an Academic Fellow on Archival Science at the Department of Archival, Library and Information Studies at the University of West Attica. He works as a Special Scientific-Technical Staff [Humanities, Projects and grants management specialist] at the Institute of Historical Research of the National Hellenic Research Foundation. He obtained his BA in History (University of Thessaly Department of History, Archaeology and Social Anthropology), and he continued with a fast-track graduate entry for a second BA in Archival, Library and Information Studies (University of West Attica). He received a Master's Degree (M.Sc.) in Modern and Contemporary History (the Panteion University of Social and Political Sciences Department of Political Science and History) and a Master of Education (M.Ed.) in Adult Education (Hellenic Open University-School of Humanities).

Yannis Stoyannidis studied History at the Department of History, Archaeology and Social Anthropology of the University of Thessaly, from which he received his Master of Arts (M.A.) and his PhD. His research interests include archives management, social history, urban history and the history of institutions, and he has published several articles in scientific journals and conference proceedings. He has participated in various research programs concerning social history, industrial heritage and archive management. Currently, he is an Assistant Professor at the Department of Archival, Library and Information Studies, University of West Attica.

Dr. Anastasios Tsolakidis received his PhD degree in computer science from the University of Limoges, France, in 2015. His research interests lie in the fields of Visual Analytics, Decision Support Systems, Business Intelligence and E-health. During his PhD studies, he has been collaborating with the Quality Assurance Unit of the Technological Educational Institute of Athens, as Data Scientist and since the July 2017 he has been working as Business Intelligent Analyst at "e-Government Center for Social Security (IDIKA SA)" at the sector of E-Health.