Journal of Integrated Information Management

Vol 9, No 2 (2024)

Jul-Dec 2024

A Literature Review on Research Indexes

Konstantina Christopoulou, Evangelia Triperina, Angeliki Antoniou, Manolis Wallace, Dimitrios Kouis

doi: [10.26265/jiim.37724](https://doi.org/10.26265/jiim.37724)

To cite this article:

Christopoulou, K., Triperina, E., Antoniou, A., Wallace, M., & Kouis, D. (2024). A Literature Review on Research Indexes. *Journal of Integrated Information Management*, *9*(2), 14–19. https://doi.org/10.26265/jiim.37724

Journal of Integrated Information Management electronic journal

A Literature Review on Research Indexes

Konstantina Christopoulou[1], Evangelia Triperina[1], Angeliki Antoniou[1], Manolis Wallace[2], Dimitrios Kouis[1]

[1]University of West Attica Department of Archival, Library and Information Studies

[2]University of Peloponnese Department of Informatics and Telecommunications

kchristopoulou@uniwa.gr [ORCID: 0000-0003-4164-2993], evatrip@uniwa.gr [ORCID: 0000-0003-4282-2259], angelant@uniwa.gr [ORCID: 0000-0002-3452-1168], wallace@uop.gr [ORCID: 0000-0002-4629-5946], dkouis@uniwa.gr [ORCID: 0000-0002-5948-9766]

Article Info

Article history:
Received 01 March 2024
Received in revised form 30 May 2024
Accepted 15 July 2024

http://dx.doi.org/10.26265/jiim.v9i2.37724

Abstract:

Purpose – *In this article, we have conducted a literature review (LR) on citation indexes to evaluate their acceptance and usage within the scientific community and the tools and metrics most frequently employed.*

Design/methodology/approach – *The presented LR followed the PRISMA framework and the methodology described by Kitchenham. Based on a set of research questions, several queries were made on the most prominent citation databases to retrieve the respective publications.*

Findings – *According to the outcomes of our LR, researchers utilised all three research databases, showing a preference for Scopus.*

Originality/value – *The paper presents a literature review of the publications related to research databases to gain insights about the current state of searching, retrieval, evaluation, and exploitation of the research publications and the related information by academics.*

Index Terms —citation resource comparison, Google Scholar, citation database, Scopus, Web of Science.

I. INTRODUCTION

Citation indexes or databases provide a reliable source of information for academics and researchers. A large number of citation indexes covers various disciplines **[1]**, while there are also many discipline-focused indexes **[2].** Nonetheless, there are several more that are widely used among the research community. The consistent cataloguing and presentation of the publications and the respective data, as well as the related metrics that accompany them, either in citation indexes as merely a reference or in databases with the full-text publication available, have contributed to the field of Bibliometrics. BBibliometrics corresponds to a set of methods for quantitative analysis of academic outputs and scholarly communications **[3]**, which can be used for books, websites, monographs, conference proceedings, policy statements, and even patents **[4]** and is mainly utilised to find the impact of research publications. At the same time, informetrics has been defined as the discipline that studies information through a quantitative perspective **[5]** by producing, disseminating, and using all forms of information, paying no attention to its formation or origination **[3]**. Likewise, the quantitative study of the field of science **[6]** is referred to as Scientometrics, and it deals with the impact of science on a greater scale. The measurement of research activity and collaboration and its depiction in research metrics facilitate the evaluation of the quality of research. Research activity is captured with metrics such as citation count **[3]**, H-index **[7]**, impact factor, and i10-index **[8]**. In contrast, research collaboration is measured by metrics, such as collaborative index, etc., and presented with co-authorship networks and graphs. The examination of the abovementioned data is of vital importance **[1]**.

Our motivation is to contribute to a more efficient management and assessment of research publications. We are conducting a literature review on the use and acceptance of the citation databases, the involved metrics, and the tools used by the research community to understand the current state of the research publications environment.

In this literature review, we have focused our search on three of the most prominent research indexes: Scopus, Google Scholar, and Web of Science. Scopus was released in 2004 **[9]** and constitutes a citation database, providing some of its services for free. Nevertheless, full access to its content is available only through subscription. The search for scientific publications through Scopus is an effortless task, while it also offers many tools and allows for personalised services during information retrieval.

Google Scholar was launched in 2004 **[10]**, and it is available for free; therefore, it has an ever-increasing popularity internationally. It returns a larger amount of results than the other citation indexes, mainly due to its extended volume of content being indexed (see information provided below). Google Scholar provides the user with a much-simplified experience, with fewer advanced search options in comparison to its counterparts. More particularly, the interface of Google Scholar has an advanced search option allowing the user to choose the words or phrases that will be included or excluded, in either the title or in both the title and the content of the article. It also allows filtering the results according to the authors or the Publisher, or the date range in which it was published.

Web of Science is one of the most popular citation databases. It was established in 1993 **[11]** and provides a set of filters and many criteria for content retrieval. The records returned from Web of Science are detailed and of high quality. Also, users can store/export the records in multiple formats.

As mentioned above Google Scholar has the largest number of content items indexed compared to the other two citation databases. Even though it is difficult to assess the magnitude of citation indexes due to their constant growth and increasing popularity, it is estimated that Google Scholar had 389 million records in 2018 and as a direct consequence, it is the most comprehensive academic search engine, while the study was conducted **[12].** More specifically, Google Scholar provided access to 389.000.000 records, Scopus to 72.212.354, whereas Web of Science to 105.519.854 **[12]**. Since then, more than 2.4 billion cited references are available through Scopus (December 2023), providing a time span from 1970 until now **[13]**. However, Clarivate and Google seem to have no up-to-date data regarding the statistics of the available records announced on their websites.

As far as it concerns the methodology followed in this research, the search and the aggregation of the related publications from two out of the three aforementioned citation databases, namely Scopus and Google Scholar, can be facilitated by tools such as the Publish or Perish tool **[14]**, which is freely available. In any case, the search and the retrieval from Scopus and Web of Science, through any tool, let alone Publish or Perish requires a subscription. This particular tool was used because of its reliability, its popularity amongst the academic community, and its openness.

For the screening process of the search results, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guideline **[15]** and the methodology described by Kitchenham **[16]**. The PRISMA guideline is an evidence-based methodology **[17]**, designed to assist the authors in improving the reporting of systematic reviews, enhancing the transparency of their research methods **[18]**. Kitchenham's systematic review activities are organised into three phases, i) the planning (identification of the need for a review and development of the review protocol), ii) the conducting (identification of research, primary studies selection, study quality assessment, data extraction and monitoring, and data synthesis) and iii) the reporting of the review. Several stages of each phase are not necessarily sequential but can be intertwined **[16]**. The publications that are reviewed are the outcome of a set of research questions that be thoroughly described in the Methodology of the paper.

The remaining of the paper is structured as follows: In Section II, the related literature is presented, whereas in Section III, the methodology of our LR on citation databases is thoroughly described. Section IV analyses the results and discussion of the LR, while the conclusions and future work are outlined in Section V.

II. Methodology

The literature review was conducted on three databases Web of Science, Scopus, and Google Scholar. The time frame selected for this review is between 2004 to 2021, following the procedures for conducting a systematic literature review described by Kitchenham (2004) **[16]**, as well as the PRISMA framework **[15]**. This literature review aims to find papers referring to citation databases and their comparison. Due to their interdisciplinary nature, we have focused our research on the most widespread and comprehensive citation databases in the academic community, which correspond to Scopus, Web of Science, and Google Scholar. The review follows the procedures proposed by Moher **[15]**, and more specifically, the process corresponds to the identification of the records, the screening, as well as the assessment of their eligibility, followed by the evaluation of the records according to the inclusion/exclusion criteria, resulting in the research corpus.

A. Planning - Research protocol - Research strategy

In the context of the object of the literature review, the searches carried out in each database were a combination of the keywords "Scopus", "Web of Science", and "Google Scholar". For each database, three queries were carried out (Table 1).

Table 1 – Research keywords per research database

QUERY ID	KEYWORDS
QUERY 1	"Scopus" AND "Google Scholar"
QUERY 2	"Google Scholar" AND "Web of Science"
QUERY 3	"Scopus" AND "Web of Science"

Based on the research questions (Table 2), the inclusion and exclusion criteria were determined to ensure that the review was inclusive and thorough and to eliminate any results that did not satisfy our specific research requirements.
The inclusion criteria were:

- Papers published between 2004 and 2021.
- Papers published in English.
- Papers focused on the comparison of citation databases.

The exclusion criteria were the following:

- Master Dissertations
- Presentations

Table 2 – Research questions

ID	Research Question (RQ)
RQ1a:	How many articles provide a comparison between all three citation databases? (Scopus, Web of Science, and Google Scholar)
RQ1b:	How many articles compare two out of three citation indexes?
RQ1c:	What is the percentage of those articles that compare Scopus with other citation indexes?
RQ1d:	What is the percentage of those articles that compare Web of Science with other citation indexes?

RQ1e:	What is the percentage of those articles that compare Google Scholar with other citation indexes?
RQ2a:	How many articles indicate a preference for Scopus?
RQ2b:	How many articles indicate a preference for Web of Science?
RQ2c:	How many of those articles indicate a preference for Google Scholar?
RQ3:	Which bibliometrics were used most commonly?
RQ4:	Which tools helped the authors while carrying out the research for their articles?

B. Data extraction

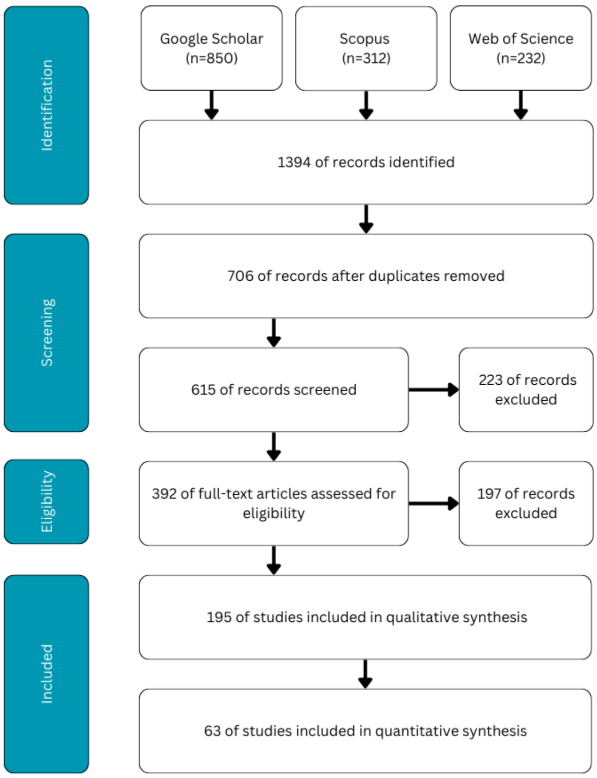

Figure 1 – Selection process of research publications based on the PRISMA framework

Table 3 - Results per keyword search in the field "paper title" per research database

Search ID	Keywords	Nº of results
S1	"Scopus" & "Google Scholar"	62
S2	Web of Science" & "Google Scholar"	57
S3	"Web Of Science" & "Scopus"	193
GS1	"Scopus" & "Google Scholar"	144
GS2	"Google Scholar" & "Web of Science"	122
GS3	"Scopus" & "Web of Science"	584
WoS1	"Scopus" & "Google Scholar"	42
WoS2	"Google Scholar" & "Web of Science"	45
WoS3	"Scopus" & "Web of Science"	145
	Total papers	1394

As mentioned before, the search and the accumulation of the research corpus were performed via Publish or Perish software, for Google Scholar and Scopus, and directly from Web of Science, for research publications spanning from 2004 to 2021. The results per keyword combination in each citation index are presented in Table 3. For the sake of brevity, they will be mentioned as S1, S2, and S3 when referred to Scopus, G1, G2, and G3, when referred to Google Scholar and W1, W2, and W3, respectively, for Web of Science. The queries were performed using the field "paper title".

The results were exported in Excel format for better management and exploration. During the initial search, 1394 papers were retrieved from the three citation indexes mentioned above (850 from Google Scholar, 312 from Scopus and 232 from Web of Science). After applying the deduplication process, we were left with 706 papers. After the screening of the remaining publications for corrupted, non-accessible papers and publications written in other languages, we narrowed them to 392. In the eligibility phase, the papers were reviewed, and based on their relevance to our research queries in terms of both abstract and full text, we retained 195 papers, which were included in the qualitative synthesis (see Figure 1). During the final phase, the remaining publications were examined in relevance to the research questions, and 132 records were excluded, resulting in 63 research publications for analysis.

The publications within the defined time range were divided chronologically into three categories, as shown in Table 3. The first category includes articles from 2004 to 2009, with at least 50 citations. The second category comprises articles published between 2010 and 2015, with at least 10 citations. The third category corresponds to the articles published between 2016 to 2021. In the last category, all the retrieved publications were considered irrespective of their citations.

Table 4 – Categorisation of the selected publications

Period	Citations	Publications
2004-2009	≥50	8
2010-2015	≥10	6
2016-2021	-	49

As shown in Figure 2, most of the publications have been published in journals, whereas several papers were published in conference proceedings. There were fewer publications in other forms, such as chapters in books, reports, preprints, and PhD dissertations. Furthermore, it is evident that from 2016 to 2021, there was an increase in publications concerning citation databases.

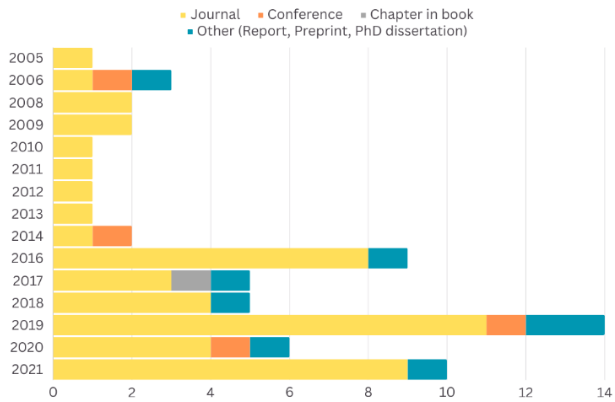

Figure 2 – Distribution of the papers included in the quantitative synthesis

III. Results and Discussion

One of the most important outcomes of our literature review is an overview of the papers that concern citation database research. In the following section, we answer each research question according to the information that was evident in the research corpus.

RQ1a How many articles provide a comparison between all three citation databases?

Most of the research publications focused their comparison on two citation index databases (see below). Nevertheless, 34.92% of the papers (22 out of 63) included all three citation indexes in their comparisons (see Figure 3).

RQ1b How many articles compare two out of three reference indexes?

In addition, 41 out of the 63 papers that were studied performed a comparison between two citation indexes, which is 65.08% of the total papers (see Figure 3).

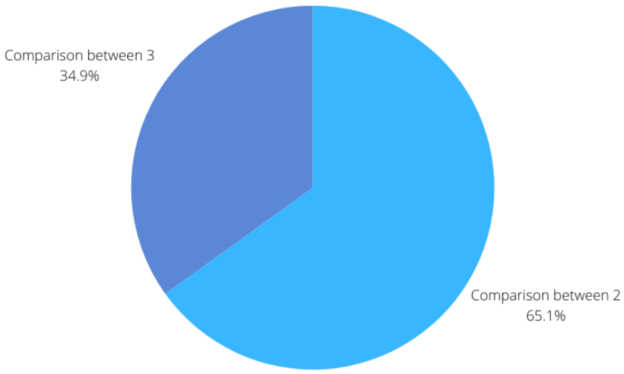

Figure 3 - Comparison between research databases in the research corpus

RQ1c How many articles were compared to Scopus with other citation indexes?

Scopus was the most compared citation index, with 60 articles including it in their study. This is estimated at 95.24% of the total papers studied, which depicts the reception of this particular research database from the research community.

RQ1d How many articles compared Web of Science with other citation indexes?

Web of Science is the second most frequently compared citation index, with a percentage of 93.65%, which means that 59 papers compared WoS to one or more citation indexes.

RQ1e How many articles compared Google Scholar with other citation indexes?

The citation index that was the least compared to the others was Google Scholar, with 29 papers including it in their comparisons, giving a 46.03%, which was expected, given the fact that it is newer compared to the other two.

RQ2a Which citation index is the most preferred by the scientific community?

Scopus seems to be the citation index that received the most positive feedback, with 20 papers (31.75%) indicating their preference, and stating its positive features and the quality of the search results in their publications.

RQ2b How many articles indicate a preference for Web of Science?

and

RQ2c How many of those articles indicate a preference for Google Scholar?

Researchers of 17 publications expressed their preference for the Web of Science, scoring the same percentage with Google Scholar (26.98%).

Meanwhile, it is worth mentioning that 29 papers do not mention any preference, with many of them advising the researchers to utilise as many citation indexes as possible to get the best results.

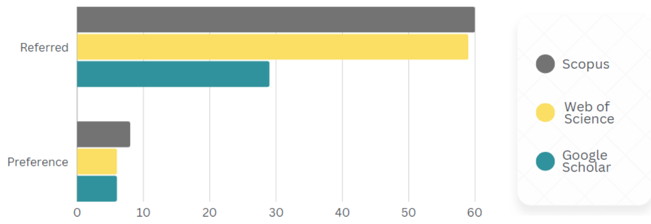

Figure 4 – An overview of the total publications that included a research database in their comparison and the respective indication of preference between the alternatives

RQ3 Which bibliometrics were used most commonly?

The most popular metric that was taken into consideration in the papers that were analysed, was the h-index, with 10 records mentioning it in their content. H-index, which was proposed by J.E. Hirsch in 2005 [7], is considered to be reliable for the qualitative evaluation internationally. Other bibliometric indicators that have been referred by the papers are the Mentor-Index, the SCIE and SSCI Indexes, H-classics, the Relative Citation Ratio (RCR) and the JIF and SJR indicators.

RQ4 Which tools helped the authors while carrying out the research for their articles?

A variety of tools were mentioned in the research corpus. Among them, excel and SPSS were the most used software as far as it concerns the statistical analysis. Publish or Perish and Classic Papers by Google were also popular tools regarding the collection, the organisation, and the study of publications. Other tools discussed in the publications were CiteSearch by the ACM Digital Library, HistCite by Clarivate, VOSviewer (a software tool for constructing and visualising bibliometric networks), the Sapiro-wilk test of data set normality, and several APIs.

IV. Conclusions and Future Work

The LR was based on the PRISMA framework, using a set of research questions for all the citation indexes that were taken under consideration (Scopus, Web of Science, and Google Scholar). After having analysed the research findings, we concluded that there is no clear indication as to which is the best in its field. This is understandable, as each citation index has its unique features, filtering methods, and search criteria. For this reason, the majority of the researchers employ a variety of citation indexes during their research, to benefit from their unique features. However, we should keep in mind that these citation indexes were created in different years, affecting their credibility, their acceptance, as well as their usage over the past years. Another consideration to bear in mind is that all citation indexes are constantly enriching their content and improving the search tools that they provide.

Furthermore, we discovered that different languages affected our search outcomes more than expected. Articles written in different languages appeared in the search results, even when in some cases we have specifically searched for articles written only in English. In the before mentioned cases, the abstracts were written in English. Consequently, that indicates a weakness into categorising the content properly when it comes to language preference. Future work lies in the study of the influence of open science in conducting, capturing, and disseminating research.

REFERENCES

[1] Bar-Ilan, J., Levene, M., & Lin, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics, 1*(1), 26-34. https://doi.org/10.1016/j.joi.2006.08.001.

[2] Torres-Salinas, D., Robinson-Garcia, N., Miguel Campanario, J., & Delgado Lopez-Cozar, E. (2014). Coverage, field specialisation and the impact of scientific publishers indexed in the Book Citation Index. *Online Information Review, 38*(1), 24-42, https://doi.org/10.1108/OIR-10-2012-0169.

[3] Das, A. K. (2015). *Research evaluation metrics* (Vol. 4). UNESCO Publishing, ISBN 978-92-3-100082-9.

[4] Cooper, I. D. (2015). Bibliometrics basics. Journal of the Medical Library Association: JMLA, 103(4), https://doi.org/217. 10.3163/1536-5050.103.4.013.

[5] Qiu, J. P. (2007). Informetrics. Hubei: Wuhan Publisher. https://doi.org/10.1007/978-981-10-4032-0.

[6] Hess, D. J. (1997). Science studies: An advanced introduction. NYU press. ISBN 9780814790953.

[7] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences, 102*(46), https://doi.org/10.1073/pnas.0507655102.

[8] Noruzi, A. (2016). Impact Factor, h-index, i10-index and i20-index of Webology. *Webology, 13*(1), 1-4, Available at: http://www.webology.org/2016/v13n1/editorial21.pdf.

[9] Deis, L., & Goodman, D. (2005). Web of Science (2004 version) and Scopus. The Charleston Advisor, 6(3). Retrieved from http://www.charlestonco.com/comp.cfm?id=43.

[10] Mayr, P., & Walter, A. K. (2007). An exploratory study of Google Scholar. *Online information review, 31*(6), 814-830. https://doi.org/10.1108/14684520710841784.

[11] Qiu, J., & Lv, H. (2014). An overview of knowledge management research viewed through the web of science (1993-2012). Aslib Journal of Information Management, 66(4), 424-442. https://doi.org/10.1108/AJIM-12-2013-0133.

[12] Gusenbauer, M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics 118, 177–214 (2019). https://doi.org/10.1007/s11192-018-2958-5

[13] Elsevier. (n.d.). www.elsevier.com. https://www.elsevier.com/products/scopus/content

[14] Harzing, A.W. (2007) Publish or Perish, available from https://harzing.com/resources/publish-or-perish

[15] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group*, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement", Annals of internal medicine, vol. 151(4), pp.264-269, 2009. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

[16] B. Kitchenham, "Procedures for performing systematic reviews", Keele, UK, Keele University, vol. 33, pp. 1-26, 2004. ISSN:1353-7776.

[17] PRISMA statement. (n.d.). PRISMA Statement. https://www.prisma-statement.org/

[18] Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. bmj, 372, https://doi.org/10.1136/bmj.n160.

Konstantina Christopoulou is a Ph.D. researcher and a member of the Information Management Research Lab at the Department of Archival, Library, and Information Studies of the University of West Attica. She graduated from the Department of Informatics and Telecommunications of the University of Peloponnese, earning her BSc diploma in Informatics and Telecommunications, majoring in Informatics. Furthermore, she continued her studies in the same Department, earning her MSc diploma in Computer Science, with merit. She has also gained her Pedagogy and Teaching Competence (PDE) certification from the Hellenic Open University and is currently working as a High School Computer Science Teacher.

Evangelia Triperina holds a PhD in Computer Science from the University of Limoges (France), with a thesis entitled "Visual interactive knowledge management for multicriteria decision making and ranking in linked open data environments". She is a Department of Computer Engineering graduate of TEI of Athens. She holds an MSc in Information Technology, Image Synthesis and Computer Graphics from the University of Limoges (France). She has worked in European research projects at GRNET, Agro-Know Technologies and the University of West Attica. She is currently a PostDoc Researcher at the the Department of Archival, Library, and Information Studies of University of West Attica.

Angeliki Antoniou is an Assistant Professor in the Department of Archival, Library & Information Studies at the University of West Attica, a collaborating researcher at the ATHENS Research Center and at the Department of Information Studies at the University College London. She holds a degree in Preschool Education (University of Athen), a degree (BSc) in Clinical with Social Psychology (University of Kent at Canterbury, UK), a postgraduate degree (MSc) in Human Interaction with Ergonomics (University College London, UK), a music degree, a Piano Diploma, and a PhD in Educational Technologies (University of Peloponnese, Department of Computer Science and Technology).

Dr. Manolis Wallace [http://gav.uop.gr/wallace/], Associate Professor at the Department of Informatics and Telecommunications and Director of the Knowledge and Uncertainty Research Laboratory (ΓΑΒ LAB) [http://gav.uop.gr/], has a long experience in the management of educational and research organisations from his previous tenures at the University of Indianapolis and at the Foundation of the Hellenic World. He now transfers this experience to Tripolis and the University of Peloponnese where he leads an interdisciplinary group of researchers specialising on cultural and educational informatics.

Dimitrios Kouis received his Diploma in Computer Engineering and Informatics from the University of Patras and his PhD from National Technical University of Athens (NTUA) in 1994 and 2004 respectively. His scientific interests include Library Networks, Digital Publishing, Scholarly Communication topics, Software development, Content Management, IT middleware platforms, meta-data modelling etc. He has been involved in several European and national projects and has published more than 30 articles in journals and conferences. Currently, he is an assistant professor at the Department of Archival, Library and Information Studies, University of West Attica.