Fitsilis, F. and V. Bayiokos. 2017.
Implementing structured public access to the legal reports on bills and law proposals of the Scientific Service of the Hellenic Parliament, Greece.
Knowledge Management for Development Journal 13(2): 63-80
http://journal.km4dev.org/

Implementing structured public access to the legal reports on bills and law proposals of the Scientific Service of the Hellenic Parliament, Greece

Fotios Fitsilis and Vasilis Bayiokos

In an era of end-to-end digitalization in parliamentary institutions, open data and established printed and electronic publication formats for contemporary and future content, the old, non-digitized and potentially corroded content still poses a great challenge in terms of preservation and promoting efficient usage. Among its other duties, the Scientific Service of the Hellenic (Greek) Parliament is responsible for the relevant legal reporting on bills and law proposals for the Plenum, the Standing Committees and the recess sections of the Parliament. These reports are publicly available legal documents that are intensively used in several communities within the Parliament and the Judiciary, such as among MPs and associates, legal drafting committees, judges and lawyers. While these reports are placed online in the parliamentary portal accompanying related bills and law proposals, the vast majority of previous reports from the 30-year history of the Scientific Service (1987-2016) remains out of scientific and political reach. The present study highlights the methodology behind the knowledge-capturing and management project to digitalize and scientifically organize this content, which is written in modern Greek. Secondly, it explains collaborations with other related public bodies as an issue of significance for cost-sharing reasons, and the promotion of usability and interoperability. Finally, it presents the specifications of the final online and printed work as well as the distinct steps for the achievement of the project outputs. The paper concludes with an outlook on future work to be conducted.

Keywords: public access; open data; legal reports; Members of Parliament (MPs); legislative operation; law proposals; bills; Hellenic Parliament, Scientific Service; Optical Character Recognition (OCR); digitization; text post-processing; knowledge management; documentation; Greek language; ePublications; ebooks; Open Government Partnership (OGP): Greece

Introduction

Parliaments are the highest representative institutions in any given democratic country. Besides their constitutionally entrusted responsibility to adopt legislation, parliaments have

also other major functions such as parliamentary control. Parliaments generate vast amounts of related content in the form of, for instance, parliamentary debates, Committee, Plenary and Recess Session minutes, which are usually published in the form of parliamentary proceedings. Additional material also includes technical and legal reports that are widely used as a source of information by Members of Parliament (MPs) in the facilitation of political debates.

Throughout decades of parliamentary operation and particularly in the pre-digital era, parliamentary libraries worldwide accumulated an unimaginable thesaurus of parliamentary proceedings in printed form. For years, the only way for scholars, researchers, historians, politicians and other interested professionals to extract information from parliamentary proceedings related to certain political and historical periods was the traditional time-consuming and often frustrating digging into parliamentary archives with the help of parliamentary staffers and dedicated librarians. As a result, and despite the naturally open character of parliamentary proceedings, this search was limited to the few and was barred from the view of a potentially wider public. In recent years, advancements in digitization methods and the setting of widely accepted open data best-practices, led to unparalleled efforts all over the world to open parliamentary archives to the public, leading to greater transparency in the political process. The project this case study refers to is part of the broader national and international tendency to digitize parliamentary material and therefore make it available to the broader public.

The Netherlands are particularly active in their digitization efforts, which in the parliamentary context, comprises the contemporary collection of the published Acts of Parliament (1989-1995) (Raynaert 2008). Marx and Schuth (2010) present a quantitative analysis of a corpus (DutchParl) that contains documents from the parliaments of The Netherlands and Belgium. Another broad digitization joint project was introduced by the Czech and the Slovak parliamentary libraries and included parliamentary documents (1848-Present) from the Czechoslovak, Czech and Slovak parliaments (Malackova and Sosna 2007). A digitization project is also ongoing in the Icelandic parliament for parliamentary documents from the years 1845-1992 (Daðason 2012). In Canada, documents of the Legislative Council of British Columbia (1865-1871) and the Legislative Assembly (1872-1982) have been digitized following a Proof of Concept by Carr-Harris et al. (2011). In Africa, Kenya's digitization efforts include debates from the parliamentary proceedings that date back to the pre-independence Legislative Council (1959) (Amollo 2012). In Greece, the library of the Hellenic Parliament has started approximately a decade ago an ambitious project to fully digitize parliamentary material from 1843 to date.

One cannot contest the fact that digitization needs to be conducted according to established methodologies using open data best practices. A 2007 study commissioned by the Parliamentary Documentation Centre of the European Parliament found that users of electronic library services prefer to sacrifice quality in favour of ease of access (Marcella et

al. 2007). In other words, content needs to be placed where it is mostly expected to be found. Parliaments in recent times are also keen to further involve the public in parliamentary matters. In the UK, a relevant study has concluded that a strategy should involve "engagement using the web and other media… and engagement in communities" (Pullinger and Smith 2010:20). Furthermore, a relevant report on the future of UK Parliament, recommended that the produced documents "should be properly indexed (or tagged if online) to allow them to be searched for their entire contents (rather than just by title or excerpt)" (Allen and Williamson 2010:28). These general directions were followed and put into practice while designing and implementing our project.

Regarding open data, several efforts to establish unified coding and standardized handling of data and content have been attempted. In Greece, a first application has been the OpenGov.gr, the Greek Open Government Initiative, an electronic platform for public consultation of draft legislation and policy initiatives prior to their submission to the Hellenic Parliament. Greece is a member of the Open Government Partnership (OGP 2016), committed to its mission and wider objectives. We can also mention the Global Partnership for Sustainable Development Data (Data4sdgs 2016) as part of the Greek efforts context. These initiatives promote transparency and accountability policies that drive modern Parliaments to open up their work to citizens and facilitate digital access to libraries and other data repositories. As a result, implementation of these policies may truly transform the relation between the State and its citizens. In 2016, the Hellenic Parliament has expressed its commitment to three basic Pillars (HeP 2016): (a) strengthening of Parliamentary Openness and Legislative Transparency, (b) provision of Open Parliamentary Data, and (c) encouragement of Public Participation. These three Pillars are the driving force behind the on-going project to digitize and scientifically organize the legal reports on bills and law proposals of the Scientific Service of the Hellenic (Greek) Parliament.

This project the paper is named after is conducted by the Directorate of Scientific Supervision of the Scientific Service and represents a significant contribution to the development of sustainable methods in the area of content-specific preservation. In order to analyze the project methodology and the digitization procedures, first we give a general overview of our Scientific Service, followed by the drafting of legal reports process. This project has been a pilot for the application of new methods and technologies. Along with the major users of our reports we present the external stakeholders and collaborations that made possible the know-how acquired.

The core of this paper consists in the detailed description of the digitization procedures which are presented and visualized with the help of tables and flowcharts. Particular attention is paid to the clear setting of specifications for the project outputs, which are expressed in the form of structure, layout and indexing of printed and electronic publications. We conclude by presenting the lessons learned during the implementation of this on-going project with a nominal duration of two years (2016-2017) and an outlook with possible and envisaged

research options as a project follow-up.

The Hellenic Parliament Scientific Service

General information on research services

Research services are often referred to as parliamentary institutes (a typology is presented by Miller et al. 2004). In the Hellenic Parliament, the President of the Scientific Council, an integral part of the Scientific Service, scientifically overviews the Parliament's Library. Parliamentary research services are more important than ever, mainly due to the steep increase of available information in the past decades, the complexity of the globalized world and the rapidly changing information society, which directly reflect into more complex legislation and the need for more technical, and not just political, parliamentary debates. Looking back, in Blischke (1981) one may find the early structure and operation of the research services in the German Federal Parliament (Bundestag). An overview of parliamentary research services in Central and Eastern Europe is offered by Robinson and Gastelum (1998). At about the same time, Michalowski and Nawrocka (1999) reported on increased workload and greater responsibility for the parliamentary research services due to technological evolution on the one hand and budgetary restriction on the other. A more recent study on the outlook for parliamentary libraries and research services has found that within this new and complex environment MPs expected research services "to support them in new ways, and to assist the parliament adopt new technologies" (Missingham 2011:52).

Founding and structure

A Scientific Service within the Hellenic Parliament has been foreseen in art. 65 par. 5 of the Greek Constitution, according to which "a scientific service to the Parliament may be established through the Standing Orders to assist Parliament in its legislative work". The relevant provision has been activated in 1987 through art. 160 of the parliamentary Standing Orders. Currently, the Scientific Service consists of the Scientific Council, which answers directly to the Speaker of the Parliament, and three directorates: the First Directorate of Scientific Studies, the Second Directorate of Scientific Studies, and the Directorate of Scientific Supervision. With the assistance of the latter, the President of the Scientific Council also exercises the scientific supervision of the directorates of scientific studies. In addition, the President of the Scientific Council exercises the scientific supervision of the Parliament's Library, the Benaki library, the collections of political personalities and the Directorate of IT & New Technologies. The art. 160-162A of the Standing Orders (SO), as well as its special internal Rules of Procedure (RoP), regulate the operation of the Scientific Service (HeP 2010).

The legal reporting of the Scientific Service

The Hellenic Parliament's Scientific Service has the duty to draft reports on bills and law

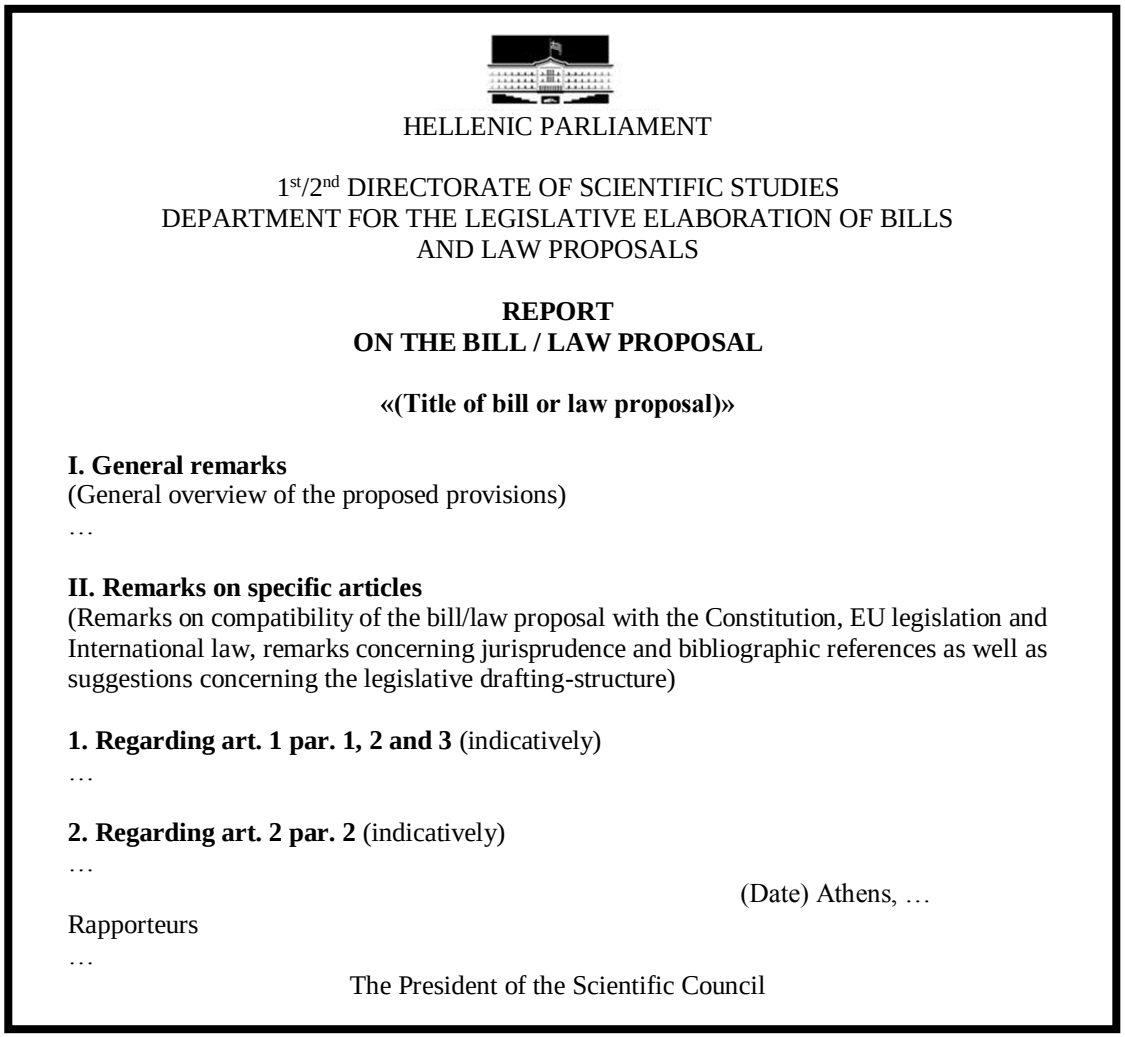
proposals. The legal basis for the legal-technical elaboration of bills and law proposals are provided in art. 92 par. 1&2 and art. 162 of the parliamentary SO, as well as the relevant provisions of the special internal RoP. Figure 1 presents a general overview of specific elements and internal drafting procedures. The bills or law proposals are committed to the Scientific Service by the Speaker of the Parliament. Notably, a relevant proposal from the competent Standing Committee (SC) is a sufficient but not necessary condition for the bill/law proposal to be committed for elaboration to the Scientific Service. The Scientific Service elaborates the text of the bill or law proposal from a legal-technical aspect within the time limit set by the Speaker of Parliament and prepares a relevant report. Reports are drafted on the text of the bill or law proposal as elaborated and voted by the competent SC.

The reports on bills and law proposals are divided into two sections. An introduction with a short presentation of the scope and the content (provisions) of the bill or law proposal, followed by references to the relevant legal framework, e.g. relevant national, EU, international framework, relevant case-law (national, EU, international), references to the relevant legal theory. In the second section, concrete legal remarks on the proposed provisions of the bill or law proposal are made. The work includes bibliographical references as well as comparative research, especially with regard to EU law being transposed into the Greek legal order. The relevant case-law on national, EU and international level is also studied in order to locate whether similar issues have already been dealt with by court decisions, especially Supreme Courts' decisions. Special attention is paid to the decisions of the Hellenic Council of State (Symvoulio tis Epikrateias), the Court of Justice of the European Union, and the European Court of Human Rights.

The rapporteurs usually have to deal with technical questions such as legal gaps (lacunae of law), antinomies, contradictions or other wording deficiencies in the proposed provisions. The legal reports are drafted using non-imperative wording. As highlighted above, considerable legal research is necessary while drafting, and the legal entities of the public sector, including ministries, universities, research institutes, public libraries et al., are legally obliged to provide to the Scientific Service any requested information or data necessary in exercising its constitutional role. Parliamentary debates, and the results of the public consultation of the draft bill/law proposal are also taken into account while drafting the legal report.

The length of the reports vary between 2 and 40 pages, with an average length being about 8 pages, depending on plurality and complexity of provisions contained in the bill/law proposal. As mentioned earlier, limited time is a major obstacle towards producing more detailed reports. Legal reports referring to bills or law proposals entering the parliament via urgent procedure may limit the time available for reporting to 24 hours. The legal reports of the Scientific Service are uploaded to the Hellenic Parliament's website and are essential documents that accompany almost any bill or law proposal entering the Parliament with the exception of those containing ratification of treaties or large public contracts.

However, depending on the nature of the bill/law proposal, there are also several other reports that accompany, as legal prerequisites, legislation entering the Parliament. These are the Explanatory Report, the report of the General Accounting Office of the State, the Regulatory Impact Assessment (RIA) report, the Public Consultation Report, the special report of the competent Minister and the Minister of Finance, the opinion provided by the Court of Audit (CoA) and the reasoned opinion of the Economic and Social Committee (OKE). The particular importance of the legal reports of the Scientific Service lies in the fact that they are among the ones that are not produced by governmental organs, thus counterbalancing governmental superiority in public debate and the information of MPs.



HELLENIC PARLIAMENT

1st/2nd DIRECTORATE OF SCIENTIFIC STUDIES
DEPARTMENT FOR THE LEGISLATIVE ELABORATION OF BILLS AND LAW PROPOSALS

REPORT
ON THE BILL / LAW PROPOSAL

«(Title of bill or law proposal)»

I. General remarks
(General overview of the proposed provisions)
…

II. Remarks on specific articles
(Remarks on compatibility of the bill/law proposal with the Constitution, EU legislation and International law, remarks concerning jurisprudence and bibliographic references as well as suggestions concerning the legislative drafting-structure)

1. Regarding art. 1 par. 1, 2 and 3 (indicatively)
…

2. Regarding art. 2 par. 2 (indicatively)
…

(Date) Athens, …

Rapporteurs
…

The President of the Scientific Council

Figure 1. Template of a Scientific Service legal report

Collaborations and communities

The Scientific Service is cooperating with a large network of institutional and academic stakeholders in exercising its constitutional tasks. For the present project, there is close cooperation with the Greek National Documentation Center (NDC), especially in the domain of ePublishing (see: ekt.gr).

MPs and their direct associates directly benefit from the reports as they represent a rich source of legal, and thus solid, argumentation for or against legal provisions contained in the bills or law proposals. The Judiciary also belongs to the frequent users of this type of legal information and it is not uncommon for the Hellenic Council of State to refer to the reports of the Scientific Service in its decisions. However, several other communities, such as lawyers, members of legal drafting committees and political journalists have little use for these reports (apart from the ones referring to current legislation activities), mainly due to the present unstructured and un-indexed form of past reports, which does not allow for a broad and sensitive legal search. Publishing the legal reports in a structured and scientifically organized way will significantly increase dissemination of the huge thesaurus of legal knowledge the Scientific Service has accumulated to a greater audience. In this way, this publication project is a great service to the normalization and plurality of political discourse in both the Parliament and in society.

Digitization and publishing of legal reports

The project of digitization and publishing of legal reports comprises the output of 16 full years of Scientific Service operation (2000-2015). Initial assessment showed that the original reports were available in the form of high quality printed material and electronic documents in .pdf format. To ensure the quality of the end product, specifications were particularly strict and included a 100% accuracy in wording and references between original and published versions. As a consequence, the project managers had to design an especially rigid and reproducible digitization process with extensive OCR post-processing. In the Hellenic Parliament such an end-to-end digitization project was designed and implemented for the first time and particular attention was paid in the exact documentation of project flows, so that they can serve as a pilot for further related activities within the Scientific Service and the Hellenic Parliament at large.

As expected, indexing, along with proofreading of the OCR digitized texts, was time-consuming and a manual approach will be followed to complete the first volumes of the publishing project. Other institutions, such as the Swedish parliament library, tested specialized software in order to determine "whether manual indexing of the parliament documents could be supplemented or even substituted by automatic indexing " (Dura 2003: 539). These software methods compare keywords assigned manually and automatically. Such methods offer remarkably high levels of term relevance (in the above article, 87.6% were relevant to those assigned manually), but could clearly accelerate the indexing process by creating a basic structure to be further elaborated by indexers. Table 1 displays the number of laws and the related legal reports per year, from 2000 to 2015.

During this 16-year range, 1574 laws (mean: μ=98, standard deviation: σ=24) have been passed through the Parliament and 739 legal reports (μ=46, σ=11) have been drafted. That means that 47% of laws are actually accompanied by a legal report. Of particular relevance to

the project was the number of pages of the legal reports to be processed. The total number was 6680 pages (μ=418, σ=226). The page count was calculated post-OCR after insertion of the entire text into a word processor. Word count of all reports amounts to 1.825 million. The last column contains the average page number of a legal report throughout the years. A slow increase of the mean report length can be detected until it reaches between 2005 and 2009 a stable plateau of 7 to 8 pages. What is then visible is an abrupt rise in the mean report length in the year 2010, which remains stable at a level of 12-15 pages in the years to date. This phenomenon can be directly linked to the entering of Greece to the so called "Memoranda of Understanding (MoU) era", an era which is related to numerous reform efforts and the associated complex legislative activities that require amending of a large number of existing legislative provisions.

Table 1. Number of laws and legal reports per year

Year	Law IDs	Number of laws	Number of legal reports	Number of pages	Average Page number
2000	2783 - 2878	96	31	160	5
2001	2879 - 2981	103	40	224	6
2002	2982 - 3092	111	43	218	5
2003	3093- 3217	125	54	237	4
2004	3218 - 3303	86	38	195	5
2005	3304 - 3430	127	53	385	7
2006	3431 - 3523	93	47	379	8
2007	3524 - 3629	106	48	349	7
2008	3630 - 3732	103	49	328	7
2009	3733 - 3813	81	28	224	8
2010	3814 - 3906	93	68	787	12
2011	3907 - 4035	129	53	811	15
2012	4036 - 4102	67	42	601	14
2013	4103 - 4224	122	56	689	12
2014	4225 - 4319	95	61	713	12
2015	4320 - 4356	37	28	380	14
Total		**1574**	**739**	**6680**	

Creating an editable corpus

The entire project and the underlying end-to-end process is depicted as a flowchart divided in two parts contained in Figure 2 and Figure 3, respectively. Three years have been initially allocated to this project, which begun with the collection and the preliminary analysis of the

relevant legal reports of 16 years of parliamentary operation (2000-2015). The Scientific Service keeps record of all its reports in digital .pdf format and the collection and organization of the relevant electronic documents could proceed within days. Special attention was paid in the qualitative control of content and morphology (layout) of the .pdf documents, in order to determine, if necessary, the amount of pre-processing activities. As a result, the entire document batch (739 reports with 6680 text pages) was found to be of very high quality with little to no necessity for corrective measures, i.e. re-scanning of documents.

However, analysis displayed that the available electronic documents are divided into two large categories: editable and image .pdf files. Image .pdf files from the years 2000-2008 (2475 pages) were found to represent 37% of the total amount of pages. The text in an image pdf cannot be simply extracted and a suitable Optical Character Recognition (OCR) process needs to be applied in. On the other hand, editable .pdf files from the years 2009-2015 (4205 pages) represented (63%) of the available material. The text from these documents can be extracted in a fairly simple way using the copy and paste functions of a given pdf reader and a word processor, respectively.

Hence, having determined the nature and characteristics of the pdf files, a twofold approach was taken in order to extract the text. For the post-2009 reports the text was easily transferred into a word processing application, one document after another. For the ones prior to 2009 and given the high image quality of the given material a commercial HP® ScanJet 8270 Document Flatbed Scanner with standard embedded OCR software proved to be well-suited, thus providing nearly flawless text recognition. This was revealed during the time consuming step of manual proofreading and fault correction of the OCR-generated text against the original content.

According to the given specifications, manual proofreading was necessary, as the text from the original legal documents needed to match the content in the new printed and electronic publication format with 100% accuracy, leaving only some space for layout modifications. The legal reports on bills and law proposals of the Scientific Service of the Hellenic Parliament make extended use of legal and technical abbreviations, which put hurdles in the work of the proofreaders and excluded wide use of standard thesaurus tools. The most usual errors in the text were found to appear within abbreviations. In some occasions “0” (zero) appears instead of “Ο”, i.e. in “0λ.” instead of "Ολ.", which stands for plenary session, and in "C0M" instead of "COM", which stands for Commission. Also, some letters, particularly in the beginning of the words, are separated by space from the rest of the word, i.e. “Γ ιώργος” instead of “Γιώργος”.

In some reports the Greek letter “β” appears as “6”, while the presence of non-existing symbols due to impurities/irregularities in the original file has to be sorted out. Proofreading is time consuming and a huge issue for large digitization projects. Therefore it needs to be minimized. Reynaert acknowledges the need of automated OCR post-correction systems,

while presenting an overview of current standards, tools and formats (2014:161). A proofreading step needs to be also accomplished for the post 2009 documents, even though of much shorter intensity and duration, as pre-formatted or rich, e.g. text that has styling information beyond the minimum of semantic elements, may be also transmitted during the transfer operations.

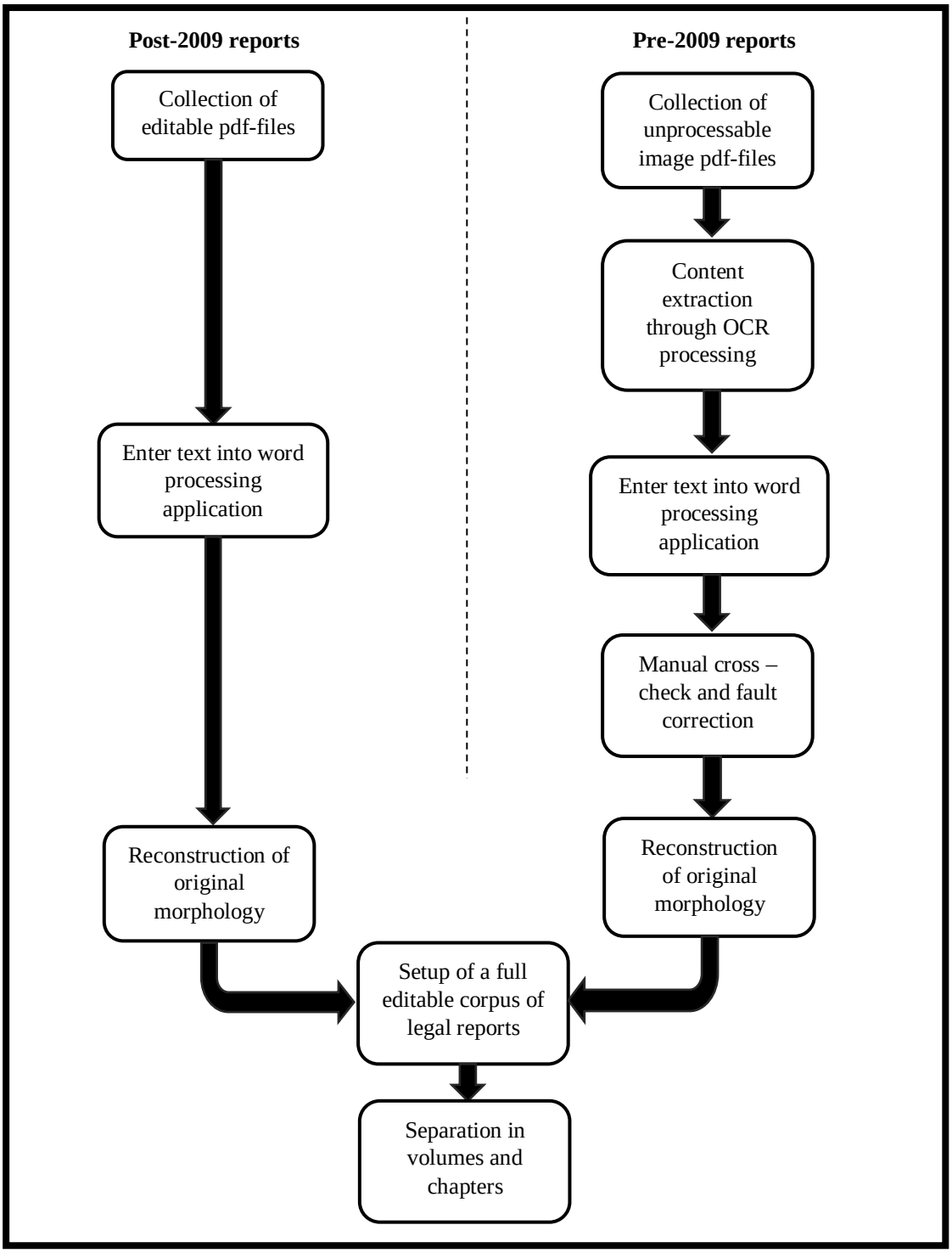


Figure 2. The first part (part A) of the HeP Scientific Service publishing project: creating an editable corpus of reports

Having available a full editable corpus of legal reports, initially separated into 16 distinct text files (one per year), they were combined into volumes that correspond to the ones for printed publication. For reasons of cost-effectiveness, printed volumes comprise legal reports from one to three years of parliamentary activity. In the case of electronic publications, the separation step can be omitted as the need for distinct volumes is not applicable. The content in electronic publications can be indexed and searched in its entity. At this point the text format, e.g. headings, fonts, font size, bold/italic elements etc., and document layout were modified to match the given specifications. The first part of the project is estimated to take one year to complete. According to our estimations, in this first project work time is mainly utilized by OCR (30%) and proofreading (50%) tasks. The other 20% belongs to other activities, mainly layout and management.

Indexing and publishing

Legal texts contain a multitude of interrelated legal terms and references to constitutional provisions, laws, decrees, EU treaties and directives, as well as court decisions of all instances. As a consequence, one of the major differences between legal and other scientific publications is that the first ones need to be equipped with several detailed indexing tables. According to the current practice for legal publications and internal discussions it has been decided for the present publication project to include four types of indexing tables, e.g. for legal terms, domestic law, EU law and codes. For indexing domestic law, separate tables for provisions within the constitution, laws and different decrees are provided. There are also different EU law tables for treaties and directives. Finally, code indexes may include provisions of the Civil Code, the Code of Civil Procedure and the Penal Code.

Table 2 presents an indicative version of the format of the different indexing tables. For presentation purposes the terms that are used are in English. However, the original language of the indexes and legal reports is Greek. The type of referencing is following the practice used in the joint publication of the Constitution and the Standing Orders of the Hellenic Parliament, which was also developed by the Scientific Service (HeP 2010). Indexing of legal terms contains hierarchical relations up to the 2nd-level degree, as seen in the example of the general term “amendment”, i.e. the term “legal amendment” is broader that the term “amendment of (a legal) provision”.

For all legal texts (constitution, laws, treaties, codes etc.) indexing goes down to article level. In the case of constitutional provisions, the relevant paragraph is also noted. All indexes contain the relevant page(s) they refer to. A transparent and effective strategy for indexation can be of great help both for the indexers as well as for the scholars and practitioners who are seeking legal information and advice within multi-volume works. After having determined the indexing strategy and communicated it to the indexers in form of detailed guidelines (do’s and don’ts with several examples) the actual indexation process may begin. The indexation

Table 2. Primary set of indexing tables

i. Index of legal terms

Legal term	Page
administrative act	151
amendment	
legal	45, 66, 113
provision	8 f.
award (of a public contract)	102 f.
corporate entity	15, 110 f.
Independent State Body	25 f., 30 f., 61 f., 147 f.
(…)	

ii. Domestic Law (Constitution, Laws & Decrees)

a. Constitution

Article	Paragraph	Page
art. 1		
	par. 3	13
art. 2		
	par. 2	8, 56, 103
	par. 4	102, 112
art. 5		
	par. 2	189
(…)		

b. Laws

Law/Year	Article	Page
146/1914	art. 1	61, 205
	art. 14	38
602/1915		89, 115, 235
5060/1931	art. 31-33	43
	art. 55	100, 104
(…)		

c. Legal Decrees

Decree/Year	Page
2493/1953	240
1943/1954	11, 26, 34, 78
3323/1955	62, 205
4111/1960	79
(…)	

d. Presidential Decrees

Decree/Year	Page
774/1980	61
671/1982	330
496/1987	16, 23, 34, 77
410/1988	62, 207
(…)	

iii. EU Law

a. EEC Treaty

Article	Page
5	52
7	220
9	12, 22, 35, 48
33	82, 244
(…)	

b. Directives

Directive	Page
2000/43/EC	240
2000/78/EC	41, 53, 84, 118
2004/113/EC	35
2006/54/EC	27
(…)	

iv. Codes

a. Code of Civil Procedure

Article	Page
1	32
11	340
79 f.	21, 46, 55, 99
90	67, 115
(…)	

b. Civil Code

Article	Page
127	69
247	236
249	19, 36, 64, 79
250	44, 111
(…)	

Three indexers and two full time people have been allocated in this process. For a given volume, an indexer is solely responsible for the creation and filling of the indexing tables. Weekly meetings between the project manager and the indexers are scheduled to supervise and control indexing operations. Regular cross-checks are also needed in order to fine-tune indexing efforts and ensure indexing homogeneity, particularly between volumes.

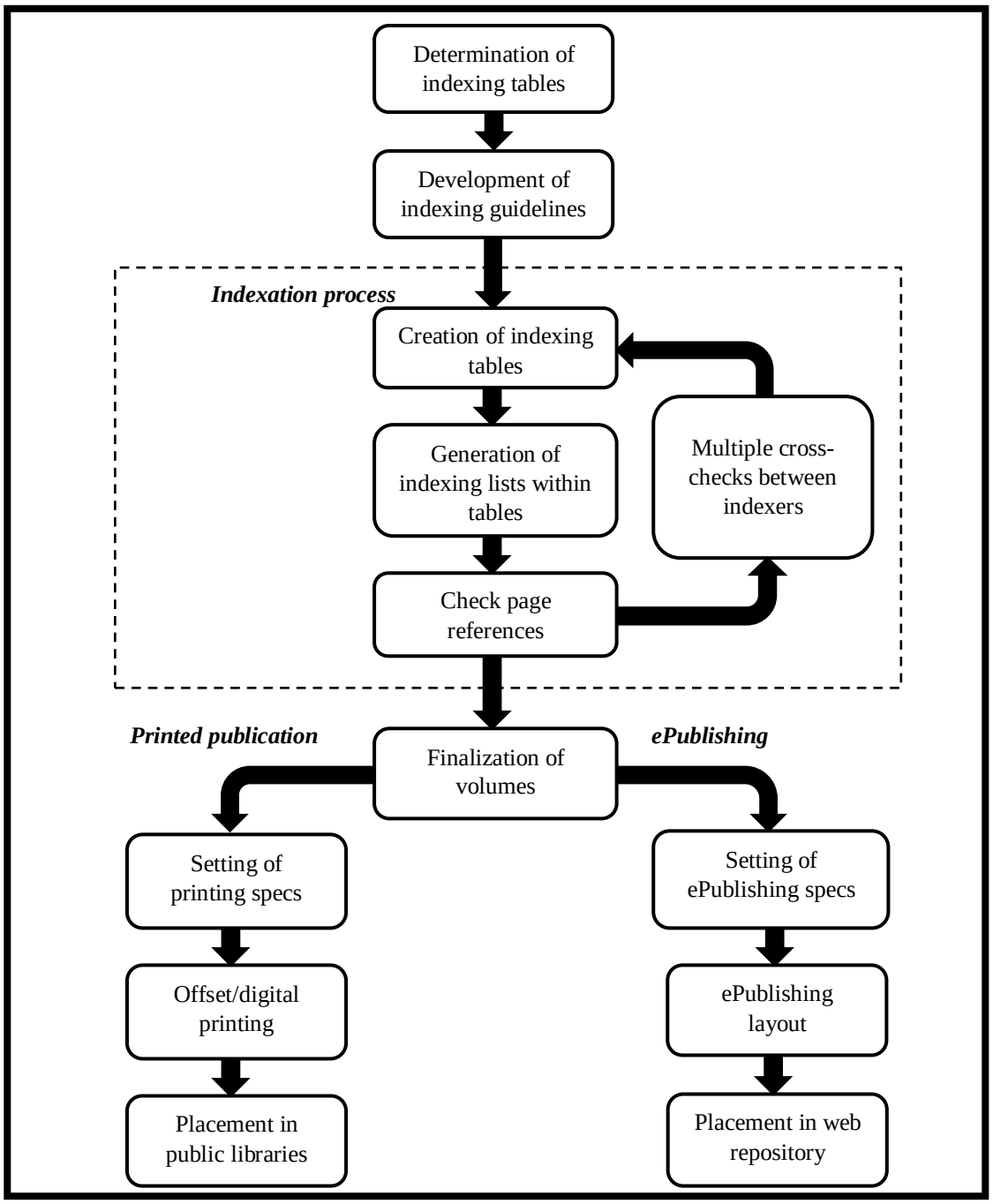


Figure 3. The second part (part B) of the HeP Scientific Service publishing project-indexing and publishing

Because of differences in-volume length and indexing speed, the publishing work is designed

with a pipeline approach in order to avoid procedural bottlenecks and improve overall performance. When a volume is finalized, it is prepared both for printed and electronic publication (ePublication).

In the case of a printed publication, text and layout for a given volume need to be aligned in the word processor environment according to the printing specifications. The electronic document is then exported as a pdf file and imported to the QuarkXPress™ design, layout and printing software currently in use by the Publications and Printing Directorate of the Hellenic Parliament. The volumes shall be printed in a WYSIWYG (What You See Is What You Get) manner and distributed to the major communities, public libraries and stakeholders, e.g. parliamentary groups and MPs. Printed and ePublications may differ in terms of layout, functionality and availability.

Open data and user access

In the present project, printed and ePublications must follow an identical layout. However, functionality shall be enhanced due to the wide use of markup tags and hyperlinks for legal terms, references, citations and indexes according to the recommendations by Allen and Williamson (2010:28) and the general requirements for open parliamentary data, to which the Hellenic Parliament has committed itself (HeP 2016). Placement of the ePublications in the Hellenic Parliament website as well as in web repositories, such as the epset.gr digital repository of the National Information System for Research and Technology, provide for universal availability of our publications.

Public access in free open repositories containing various formats such as HTML, web, and print PDF should provide for increased visibility of this publication series. To further extend the possibilities for utilizing published data, we are converting the original documents into eXtensible Markup Language (XML 2013). The use of XML offers a set of advantages that make it particularly interesting in scientific publications, as pointed out by Murray-Rust and Rzepa (2002:95 f.), particularly the fact that it makes textual data, meta data (titles, keywords, legal terms, references, etc.) and procedures publicly available. By becoming fully accessible to the scientific community, researchers and/or interested individuals, maximal transparency and the possibility of external validation and exploitation is ensured.

However, it remains open whether an XML-first principle shall be deployed, meaning, whether the legal reports shall be first converted into XML and then this XML should be used to prepare the output form. Equally open remains the decision regarding format validation using an appropriate Document Type Definition (DTD). While existing legal reports need to be transferred ex-post one-by-one in XML syntax, the use of DTD is crucial for speeding up the publishing process of future reports, as the responsibility to conform to the publication requirements is moved towards the drafter's side. The choice of the open standards for legal documents, which derive from the original XML format, is also under consideration. Among the solutions considered, we count LegalDocumentXML specifications, which provide "a

common legal document standard for the specification of parliamentary, legislative and judicial documents" (OASIS 2017). It is at hand that publication of open data in the form of XML needs to be in line with the overall digital strategy of the Hellenic Parliament, which follows the general OGP and the resulting national guidelines. A self-consistent open data policy is currently under development.

Conclusions and outlook

In 2017, the year following its 30th anniversary (1987-2016), the Scientific Service of the Hellenic Parliament decided to publish the legal reports on bills and law proposals it has produced to date. By doing so these would be made available to the wider public, thus enabling further study and research activities. Both printing and electronic publications were considered in order to be in line with the contemporary call for the "trinity" of big, open and linked data. At first, the last 16 years (2000-2015) of legal reports have been included in the project, since the material could be most easily collected and processed. However, from the very beginning of this project, it was clear that it was meant to be a pilot project for future digitization endeavors within the Hellenic Parliament and possibly beyond. This is why the underlying processes have been carefully designed and much time has been invested in their fine-tuning during real-time operation.

In order to fully comply with the strict specifications, a robust digitization and verification process has been developed, which included OCR and post-processing of the available digital images. Indexing has been found to be particularly time consuming, and issues of consistency within a project of this magnitude that includes several volumes need to be addressed. At the time of reporting, a manual approach to indexing is taken. However automatic indexing using Natural Language Processing (NLP) and Natural Language Information Retrieval (NLIR) methods could be tested in the preparation of future volumes, as shown by Dura (2003).

The scientific outlook following this project is particularly rich. Assuming that the project will be completed on-time and according to the strict specifications, the possibility is at hand to extend it to include the rest of the legal reports from the early days of the Scientific Service, from 1987 to 1999. However, these are available as image pdf files of variable quality and an automated, reliable and customized OCR engine should be developed in cooperation with partner research institutions, in order to achieve reliable and reproducible OCR results.

Making digital content available immediately enables linking with countless, former distant, areas of research, e.g. history, political science, linguistics and others, based on post-processing of available electronic text data. For instance, following digitization of parliamentary proceedings in The Netherlands, a whole new chapter for quantitative political research has been opened (Piersma et al. 2014). In the present context, digitization of parliamentary proceedings could drive the study of parliamentary debates and questions from

the point of view of computational linguistics. Hence, analysis of the text style and of the political behavior over a broad time scale (1843 to date) could be conducted revealing patterns and evolution of political discourse.

Current discussions with NDC also include the development of an enhanced web repository with ePublishing functions. This digital service will contain an appropriate repository infrastructure tailored to the needs of the Scientific Service of the Hellenic Parliament and its major clients, thus offering open, organized and standardized access to legal reports on bills and law proposals. At the same time, it will support the need for on-demand printed and electronic publications of selected subsets of reports, and will automate and standardize to a great extent the daily and periodic workflows of the Hellenic Parliament Scientific Service.

References

Allen, B. and Williamson, A. (2010) Parliament 2020: visioning the future parliament: report on the UK Parliament Focus Groups, Hansard Society: London.

Amollo, B. A. (2012) Digitization for libraries in Kenya, presented at the 2nd International Conference on African Digital Libraries and Archives (ICADLA-2), University of Witwatersrand, Johannesburg, South Africa, 14th – 18th November, 2011. Available from: http://wiredspace.wits.ac.za/bitstream/handle/10539/11485/Amollo%20-%20Digitization%20for%20libraries%20in%20Kenya.pdf (Accessed 28 November 2016).

Blischke, W. (1981) 'Parliamentary Staffs in the German Bundestag', Legislative Studies Quarterly, 6(4), 533-558.

Carr-Harris, M., et al. (2011) British Columbia Government Publications Digitization Project: Proof of Concept. Available from: https://dspace.library.uvic.ca/bitstream/handle/1828/7463/BCGovDocs_proposal_Final_Corrected_03_28_11.pdf (Accessed 28 November 2016).

Daðason, J. F. (2012) Post-correction of Icelandic OCR text, thesis at the University of Iceland. Available from: http://skemman.is/stream/get/1946/12085/30520/1/Post-Correction_of_Icelandic_OCR_Text.pdf (Accessed 28 November 2016).

Data4sdgs (2016) The Data Ecosystem and the Global Partnership, Global Partnership for Sustainable Development Data web site. Available from: http://www.data4sdgs.org/who-we-are/ (Accessed 6 December 2016).

Dura, E. (2003) Natural language in information retrieval, presented at the International Conference on Intelligent Text Processing and Computational Linguistics, Springer Berlin Heidelberg, 537-540.

HeP (2010) Standing Orders and Hellenic Constitution, Hellenic Parliament. Current version available from: http://www.hellenicparliament.gr/en/Vouli-ton-Ellinon/Kanonismos-tis-Voulis/ (Accessed 14 December 2016).

HeP (2016) Public consultation for the 2016-2018 Action Plan for the Open Government Partnership, Hellenic Parliament web site. Available from: http://diafaneia.hellenicparliament.gr/ElectronicDeliberation/DeliberationActionPlan2016-2018OpenGovernmentPartnership (Accessed 6 December 2016).

Malackova, E., and Sosna, K. (2007) 'The Joint Czech and Slovak Digital Parliamentary Library', IFLA journal, 33(3), 251-257.

Marcella, R. et al. (2007) 'The information needs and information-seeking behaviour of the users of the European Parliamentary Documentation Centre: A customer knowledge

study', Journal of Documentation, 63(6), pp.920-934.
Marx, M., and Schuth, A. (2010) A Corpus of Parliamentary Documents in Dutch. Available from: http://dare.uva.nl/document/2/88437 (Accessed 28 November 2016).
Michalowski, J. and Nawrocka, E. (1999) 'New developments in parliamentary research services and technology', INSPEL, 33, 20-28.
Missingham, R., 2011 'Parliamentary library and research services in the 21st century: A Delphi study', IFLA journal, 37(1), 52-61.
Miller, R., et al. (2004) Parliamentary libraries, institutes and offices: The sources of parliamentary information, World Bank Institute: Washington, D.C. Available from: http://siteresources.worldbank.org/PSGLP/Resources/ParliamentaryLibrariesInstitutesandOffices.pdf (Accessed 28 November 2016).
Murray-Rust, P. and Rzepa, H. S. (2002) 'Scientific publications in XML-towards a global knowledge base', Data Science Journal, 1, 84-98. Available from: http://datascience.codata.org/articles/10.2481/dsj.1.84/galley/176/download/ (Accessed 24 March 2017).
OASIS (2017) LegalDocumentML Technical Committee, OASIS web site. Available from: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legaldocml (Accessed 29 March 2017).
OGP (2016) What is the Open Government Partnership?, Open Government Partnership web site. URL: http://www.opengovpartnership.org/about (Accessed 6 December 2016).
Piersma, H. et al. (2014) 'War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History', Digital Humanities Quarterly, 8(1).
Pullinger, J. and Smith, E.H. (2010) The public engagement strategy in the UK Parliament since 2006, presented at the IFLA conference, Gotenberg, Sweden. Available from: http://www.ifla.org/past-wlic/2010/141-pullinger-en.pdf (Accessed 28 November 2016).
Reynaert, M. (2008) 'Non-interactive OCR post-correction for giga-scale digitization projects', International Conference on Intelligent Text Processing and Computational Linguistics, Springer: Berlin, Heidelberg, 617-630.
Reynaert, M. (2014) 'On OCR ground truths and OCR post-correction gold standards, tools and formats', Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, ACM, 159-166. Available from: http://repository.ubn.ru.nl/bitstream/handle/2066/134274/134274.pdf (Accessed 28 November 2016).
Robinson, W. H. and Gastelum, R. eds. (1998) Parliamentary libraries and research services in Central and Eastern Europe: Building more effective legislatures (Vol. 87). Walter de Gruyter.
XML (2013) Extensible Markup Language 1.0 (Fifth Edition). W3C Recommendation 26.11.2008. Available from: http://www.w3.org/TR/xml/ (Accessed 24 March 2017).

About the authors

Fotios Fitsilis, Dr.-Ing. Dipl.-Wirt.Ing., is Head of the Department for Scientific Documentation & Supervision of the Directorate for Scientific Supervision in the Scientific Service of the Hellenic Parliament and Special Scientific Associate to its Scientific Council. In addition, he is Visiting Professor at the Universidad Complutense de Madrid, Spain, on issues related to the function of parliamentary control. Email: fitsilisf@parliament.gr

Vasilis Bayiokos, Dipl. in Eng., exec. MBA, is Head of the Directorate for Scientific Supervision in the Scientific Service of the Hellenic Parliament and Special Scientific Associate to its Scientific Council. In addition, he is certified auditor of the Independent Audit Authority of Greece, responsible for verifying the effective functioning of the management and control system for all Operational EU Programmes, 2007-2013.
Email: b.bayiokos@parliament.gr