Working paper

Crowdsourcing the Digital Parliament – the Case of the Hellenic OCR Team

Fotios Fitsilis, Hellenic Parliament, Athens, Greece

George Mikros, Hamad Bin Khalifa University, Doha, Qatar

Abstract

Parliaments will eventually not evade the digital evolution of every institution to become data-driven organisations. This development constitutes an extraordinary opportunity for the strengthening of legislatures that needs to be embraced rather than a formidable phenomenon that should be delayed by any means. Emerging technologies, the lack of digital strategy, and change management processes put, among other things, parliaments in front of severe obstacles to cope with the issue of its digital transformation. Under certain conditions, crowdsourcing, in other words, the power of the people, can be appropriately channelled and exploited to support representative institutions and their societal stakeholders in managing their change processes. Based on survey findings and structured interviews, this study investigates the case of the Hellenic OCR Team, an innovative crowdsourcing initiative for the processing and analysis of parliamentary data.

Keywords: representative institutions, citizen's engagement, digital platform, textual analytics, parliamentary innovation

1. Introduction and motives

The rise of the digital society has placed parliaments, as representative institutions, in front of new challenges, some of which could be notoriously hard to tackle (Giddings, 2005; Dai and Norton, 2007; Fallon et al., 2011). Addressing those issues could be a new field of labour for active citizens. Citizens' engagement is not meant to alter the representational character of legislatures. Still, it offers unprecedented ways to support and strengthen the evolution of parliamentary functions and procedures, for instance, digitalisation techniques, creation of Big Open Legal Data, the appliance of artificial intelligence tools and services, data-driven decision-making, and other more or less pressuring issues around the digital parliament (Sartor et al., 2011; Fitsilis and Costa, in press).

Regardless of the level of digitisation in parliament, such issues must be tackled and possibly regulated early on. Unfortunately, this is easier said than done. Among others, this can be traced back to the lack of sufficient parliamentary resources to tackle all these issues simultaneously. Surely, inter-parliamentary cooperation and exchange, such as in the framework of the Inter-Parliamentary Union's (IPU) Centre for Innovation in Parliaments, may contribute to a certain extent. Still, the sheer complexity of digital technology and the huge number of potential approaches calls for groundbreaking solutions, at least within parliamentary ecosystems: the power of the people. Harvesting it through sophisticated crowdsourcing methods has proved to be able to frame and analyse many of the above parliament-related challenges.

For the first time, crowdsourcing has been demonstrated in parliamentary matters in the case of the Hellenic Optical Character Recognition (OCR) Team, also referred to as the 'Team', a global

scientific initiative currently spanning 14 countries and four continents. The choice of the prefix 'Hellenic' in a global initiative's name is by no means an oxymoron. It refers to the anti-individualistic ideology of Hellenistic culture in direct analogy to modern crowdsourcing principles (Martin, 1994). Established in 2017 and equipped with unique characteristics, its decentralised character empowered an interdisciplinary team of professionals and organisations that jointly contribute to the greater goal of strengthening the parliamentary institution. A more detailed presentation of the Teams structure, motives and operational patterns is given by Fitsilis and Mikros (2021). Figure 1 shows its basic team structure and operation.

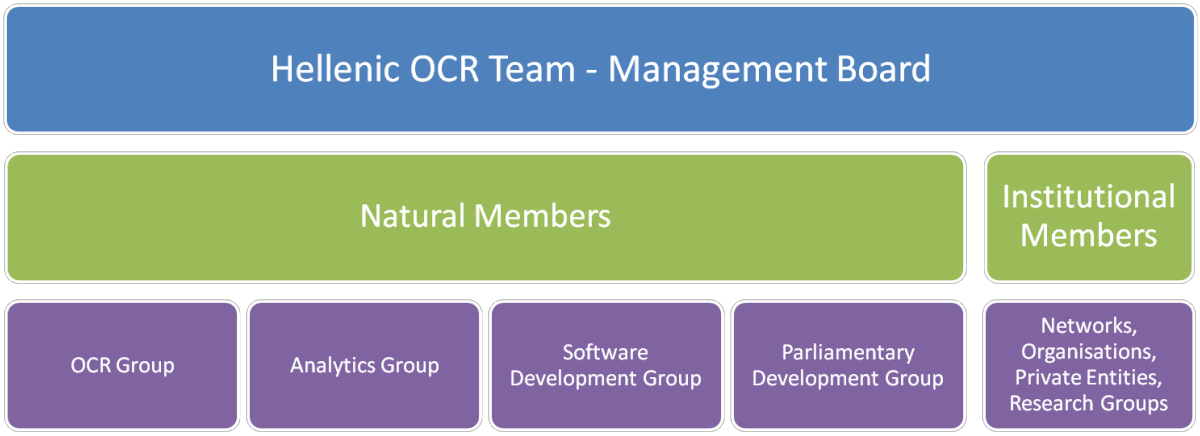

Figure 1. Hellenic OCR Team structure.

It is managed by a three-member management board that coordinates the actions of almost 60 members (experts) and institutional entities (private companies, research groups and organisations, projects, and non-governmental entities).[1] The experts used to be organised in dedicated research groups such as the OCR, the analytics and the software development group. Recently the team transformed into an expert network with a less rigid organisational structure.[2] Despite this substantial change, the research objectives were not conceptually re-arranged and followed the original Team's scope that consisted of:

- Holistic study of parliamentary corpora;
- Design and implementation of digital platforms;
- Citizen's engagement with representative institutions;
- Challenges and application of emerging technologies.

The added value of the Team, though difficult to assess in relevant terms since points of reference with other initiatives cannot be immediately determined and comparative investigations are underway, can be approximated by assessing its impact on three distinct levels: scientific, technological, and societal.

Scientifically, in the past years, the Team's members participated in several research workshops and academic events while also frequently publishing their results in scientific journals and conference proceedings (see, indicatively, Fitsilis et al., 2022; Leventis et al., 2021; Koryzis et al., 2021). On the technology side, apart from utilising existing solutions and assessing the feasibility and applicability of new technologies in the parliamentary workspace, for which the term 'ParlTech' has been coined,

[1] As of July 2022.
[2] This transition is discussed in Section 3 in more detail.

the Team opted to structure its own base of digital solutions while making them available to the wider parliamentary community for experimentation and further development.[3] Furthermore, the members' civic activity can be characterised as a type of public engagement in the form of an expert network, with parallels to other engagement networks such as the International Parliament Engagement Network (IPEN), to which close links are maintained.[4]

Using an internal survey and structured interviews, this contribution will attempt a dissection of the Team's organisation and operational methods. Moreover, its members' deeper goals and motivation will be revealed. This investigation will also highlight any similarities or differences with other crowdsourcing platforms and several good (and perhaps bad) practices for parliament and its administration. Ultimately, cooperation with other networks will be studied and potential synergies discussed.

The next section provides a literature overview of crowdsourcing initiatives from several scientific fields (Section 2). Differences with the Hellenic OCR Team are indicated and possible research gaps are mentioned. The methodology of the study is then presented and its limitations are indicated. Section 3 analyses the empirical evidence and discusses the main findings. It is followed by the conclusions and an attempt to predict the Team's future steps (Section 4).

2. Theory, gaps, and method

Two main areas can be considered for the theoretical framing of the topic around the Team's operation: citizen's engagement and participation patterns and crowdsourcing methods. The former was shaped by the work of Leston-Bandeira (2012; 2016; 2019). However, the role of a particular initiative within this framework is ideally to be considered against other citizens' initiatives from a comparative perspective. This contribution centres on the parameters of harvesting of distributed effort: crowdsourcing. There is no scarcity in literature resources regarding crowdsourcing and its application has been assessed through numerous use cases from several sectors.

Some important yet indicative examples include the report on a system that utilises crowdsourcing to wipe out the errors of multi-version data. Among other features, this system employs a module to determine the 'human workers' with the highest confidence margins (Tong et al., 2014). Sun et al. (2014) describe the use of crowdsourcing to help evaluate the results of product classification. They argue that in large-scale classification problems, crowdsourcing needs to be applied combined with other tools and approaches such as machine learning and the setting of classification rules. Another framework by Lin and Davies (2010) utilises crowdsourcing to improve the social classification structure (folksonomy) through ontology building.

The power of the crowd has also been applied to attempt an increase in the democratic legitimacy of representation systems (Prpić et al., 2015). Finland, for instance, provides a method for the legislative crowdsourcing through the provision of online deliberation tools (Christensen et al., 2015). Landemore (2015) investigated the results of crowdsourcing in constitution-making for the Icelandic

[3] https://github.com/hocrt

[4] IPEN is 'an international network of academics and practitioners across the world to critically examine the effectiveness and impact of public engagement with parliament,' https://ipen-network.org/

use case, where a quasi-representative sample of 100 people ('crowd') was consulted at the beginning of the process. Furthermore, Orozco (2016) sees the rise of crowdsourcing as part of the wider DIY culture and studies legal crowdsourcing ('lawsourcing') that requests people's 'support to achieve a legal objective'. In addition, Oronzo points to the growing market volume for such crowdsourced services.

However, the above studies rarely refer to the parameters related to the qualities and characteristics of human workers. For instance, the effects of paid crowdsourcing and the variations in effort and worker qualifications to assess a search engine's effectiveness were investigated by Kazai (2011). They concluded that increasing pay, reducing effort, and introducing qualification requirements can all help in reducing spam behaviour among workers. They also indicated that due to the interplay of the parameters and their influence on each other, on the task design, and on the output, each such decision needs to be balanced overall, e.g., increased pay may call for additional quality control elements (Kazai, 2011, p. 175). Moreover, Čibej et al. (2015) assessed the crowdsourcing potential in lexicography in dictionary creation at an early stage. They make particular references to crowd motivation, microtask design, and quality control as crucial elements of successful implementation, which are exactly the ones the Hellenic OCR Team pays particular attention to.

Apparently, the Hellenic OCR Team is not the first case crowdsourcing has been used to build and annotate large corpora (see, e.g., Wang et al., 2013). However, there is not another similar example to be found for a crowdsourcing effort that is operating globally and permanently using voluntary effort while applying a training scheme. On the contrary, crowdsourcing is usually monothematic and is conducted on a one-off basis, generally paying less attention to the qualities of the human capital.

The study behind this contribution attempts to tackle two main research questions:

1. What are the profiles and motives of the people engaged in scientific crowdsourcing projects like the Hellenic OCR Team?
2. What are the optimal structure and operational methodology for a crowdsourcing network like the Hellenic OCR Team?

In other words, this contribution investigates both the people, i.e., the 'crowd', crowdsourcing initiatives are relying on as well as the initiatives' organisational aspects. A combination of qualitative and quantitative methods were used to answer these questions. Empirical evidence resulted from a survey directed to Team members (N=50). Institutional members were not included in this research. Consequently, the reported results can be used to draw safe conclusions about the Team's behaviour and extrapolate future trends. The survey consists of 11 questions that were sent to members per email. The full questionnaire can be found in the Appendix. The results were collected from February 2021 until June 2022.

A previous attempt to look into the Team's anthropogeography and background was presented by Fitsilis and Mikros (2021) for a member population of N=39 by analysing the initiative's development over time, the geographic and gender distribution, the academic and the sectoral background of the members. The current investigation deepens this analysis for an expanded members base while also discussing additional research-specific parameters. Responses were treated confidentially and are GDPR compliant.[5]

[5] GDPR: General Data Protection Regulation (EU) 2016/679.

3. Survey results and analysis

3.1. Basic demographics

Empirical evidence for the Hellenic OCR Team's operation and research activities were retrieved from the analysis of the aforementioned survey questions. The first question captures the member's name. In case of follow-up discussions for further specification or for proving additional input, it was necessary to link comments and results with the members.

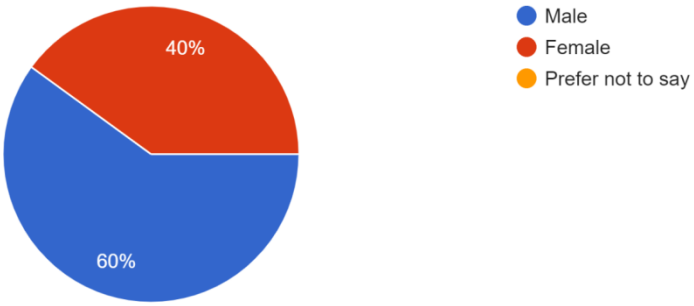

Figure 2. Hellenic OCR Team gender distribution.

Since the Team's establishment in 2017, gender distribution has been observed to be fairly even within the Team. Currently, it is slightly in favour of male members, who comprise 60% of the population, i.e., 30 in raw frequency (Figure 2). Non-binary responses were not recorded. This shift in recent years can be mainly attributed to the increase in memberships from technical disciplines such as engineering (including software engineering and data science) that stems from the Team's strategic option to strengthen its technical background and, in particular, the software development group.[6]

#	Country	Members	% (N=50)	#	Country	Members	% (N=50)
1	Greece	31	62%	8	Serbia	1	2%
2	Italy	4	8%	9	Argentina	1	2%
3	United Kingdom	2	4%	10	Belgium	1	2%
4	Germany	2	4%	11	Brazil	1	2%
5	Qatar	2	4%	12	Canada	1	2%
6	Luxembourg	1	2%	13	Cyprus	1	2%
7	Netherlands	1	2%	14	Finland	1	2%
					Totals	**50**	**100%**

Table 1. Country of residence.

Table 1 shows aggregated information that represents the country of residence at the moment of the completion of the survey and is not automatically updated. In the context of a high-mobility

[6] As of March 2022, women make up 16.5% of all engineers, https://www.wes.org.uk/content/wesstatistics/

environment, such as in the case of Europe, it is realistic to expect changes in the members' residence that cannot be captured here. Nonetheless, to the best of the authors' knowledge, such changes were limited and did not alter this metric's basic outcomes. There is no surprise that more than 6 out of 10 reside in Greece, or Hellas, giving full credits to the Team's name, the Hellenic OCR Team. At the same time, it is also extraordinary that almost 4 out of 10 reside outside the birthplace of the initiative. This highlights the Team's decentralised organisational scheme that equally handles and engages all members in its scientific actions and projects. It also underlines the highly international scope of the Hellenic OCR Team research. Most of the members reside in the EU (84% or 42 members). When including the UK and Serbia, the numbers climb to 90% and 45 members, respectively. The remaining 10% is distributed in three additional continents: Asia (Qatar), North America (Canada), and South America (Argentina and Brazil) increasing their overall number to four continents.

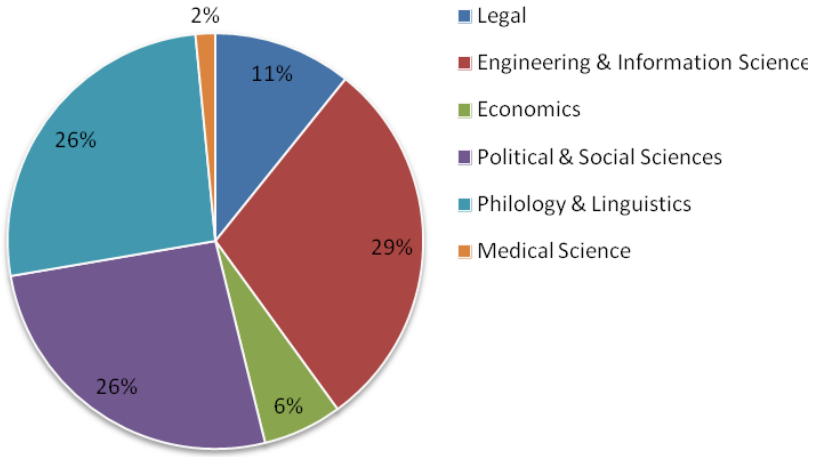

Figure 3. Scientific background (in percentage; N=65).

To determine their scientific background, members were able to provide more than one response. The results are depicted in Figure 3. Overall, N=65 selections were provided by the 50 members. Some of these were manually merged into wider disciplines. For instance, Italian philology was added to the philology/linguistics domain, and social/health policy, psychology, and methodology were merged with political science to form a new sub-group for political and social sciences. Additionally, machine learning, computational linguistics, and data science were linked to engineering and information science. What stands out is that the most represented disciplines that account for almost 82% of the selections are: a) engineering and information science (29.2% or 19 selections), b) political and social sciences (26.1% or 17 selections), and c) philology and linguistics (26.1% or 17 selections). This almost even distribution of major scientific fields proves that the Hellenic OCR Team is not a narrow, one-sided group of experts but, on the contrary, a diverse initiative that is able to cover several research aspects of parliamentary studies. Furthermore, the Team is composed of legal, economics, and medical experts.

.

As already mentioned, in recent years, there has been a clear shift to be observed toward the more technical faculties. When the initiative started, the legal and philology faculties were overrepresented within the Team, which was necessary due to the necessity to process, validate and study the generated parliamentary corpora. With the Team's expansion and the increase in the volume of available corpora, additional disciplines became necessary to analyse the political, financial, and technical parameters of parliamentary discourse and workspace. Lately, within the past couple of

years, a sharp increase in the members of engineering and informatics faculties has been visible. This can be attributed, for once, to the Team's focus (and the related public exposure) on the development of digital platforms and services. The increase also matches the network's research priorities, in particular in regards to emerging technologies and digital platforms for parliament, and constitutes a trend that is expected to continue in the next few years.

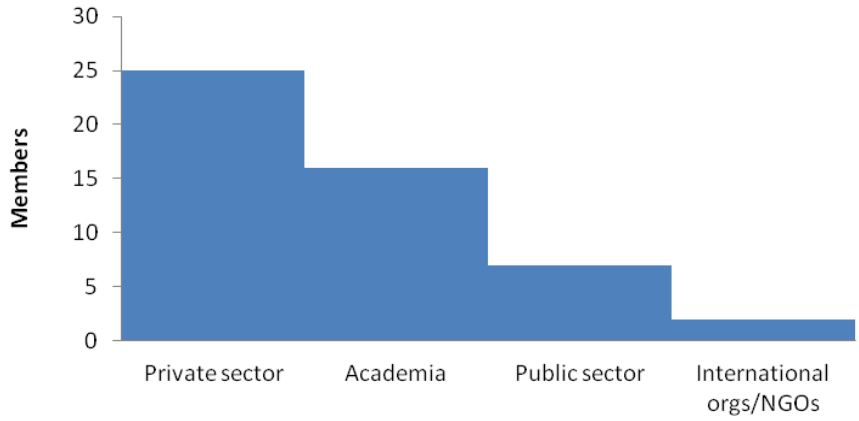

Figure 4. Working sector.

Figure 4 visualises the working sector's experts originate from. When studying the working sector members originate from, one finds that 50% (25 members) work in the private sector. This is astonishing when thinking that the initiative is dealing with a public institution *per se* and that this is a volunteering activity. The understanding of the motivational patterns is highly interesting as it could be used to strengthen the Team's normative and operational framework, thus attracting more members with similar moral and guiding principles. An additional 32% (16 members) originate from the academic sector, including students of all grades (graduates, post-graduates, Ph.D. students). Moreover, 14% (7 members) come from the public sector, and 4% (2 members) work in international organisations and non-governmental institutions. As outlined in Fitsilis and Mikros (2021, p. 5), the Hellenic OCR Team needed to rely on the technical expertise of the private sector to develop open source digital solutions for parliament as well as on the scientific/research skills of academia to study and analyse the transformed parliamentary data. On the supply side, i.e., Hellenic OCR Team members, provide several reasons for joining the team (see also question 7 below 'Why did you join the Hellenic OCR Team?')

3.2. Involvement metrics

One must keep in mind that this is a volunteering crowdsourcing initiative. There are no wages nor fees paid to the experts, though there are several incentives that the Hellenic OCR Team provides to its members. Figure 5 shows the weekly time experts are ready to invest or are already investing in the Team's activities. According to the Figure, the large majority, i.e., 41 out of 50 members (82%), invest an average of up to 4 hours weekly which is fairly low time investment compared to other international crowdsourcing initiatives, e.g., an average worker in Amazon's Mechanical Turk spends double time reaching 8 hours per week (Felstiner 2011, p. 167). The number of time invested naturally decreases with higher involvement. Ultimately, only 3 members invest more than 8 hours a week into the initiative. These can be pinpointed as the 3 members of the management board.

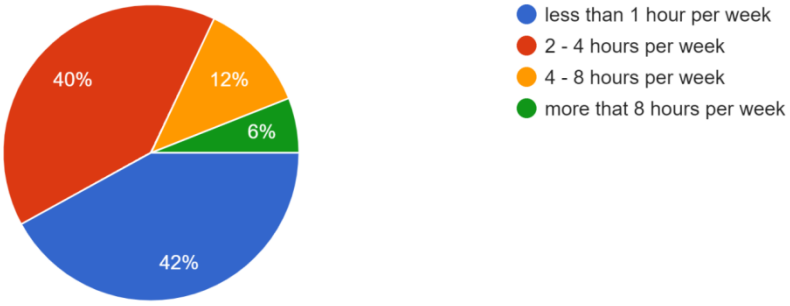

Figure 5. Time investment.

Certainly, the cumulative effort that experts put into the Team is difficult to be estimated, nor is it evenly distributed over time. On the occasion of scientific projects, for instance, involvement may peak, followed by large inactivity periods. *Vice versa*, there are activities such as OCR and validation of large corpora that demand a more constant effort. The minimum effort that is asked from members is to follow the Team's announcements and the ongoing scientific debates that are happening within the initiative. When members' count was low, this happened via email communication that quickly became impractical as the numbers of members and projects rose. Therefore, communication migrated to an online digital platform (Slack) offering, among others, workflow management, instant messaging, and data exchange.

These results should be read in conjunction with the results on the reasons for joining the Team (Question 7). The option that gathered the most responses was 'because of the interesting research topics' (39 out of 50 or 78%). This pure scientific curiosity is exactly why the Hellenic OCR Team was built in the first place and, certainly, belongs to the drivers of innovation and progress. This option is closely followed by the one to 'join a scientific community' (72%, 36 from 50), which can be distantly associated with the 'need to belong' and interact with peers within social groups (see, for instance, Baumeister and Leary, 1995). The co-founders identified this untapped potential early on and, based upon it, the Team's main expansion strategy. These two high-ranking selections are followed by more 'selfish' options with a high frequency of occurrence. These are the gain of technical know-how (50%, 25 of 50) and the strengthening of scientific background (52%, 26 of 50), followed by the added value to the CV (40%, 20 of 50). These choices can be justified by the fact that several members are students or young professionals in the early stages of their careers wishing to acquire new skills and knowledge to improve their employment or promotion chances.

The frequent option of networking (50%, 25 of 50) deserves a special mention in a globalised working environment. The power and prospects of networking for personal and professional development are widely acknowledged (see, for instance, Jacobs et al., 2019). Additionally, such high favourability of networking was expected since the Team advocated early on its strategy to gradually transform (from January 2022) into an expert network with a global presence (on this, see the discussion for Question 9B on the Team's transformation into an expert network).

The Hellenic OCR Team actively responded to those 'requests'. First, it did so by strengthening the capacity of its members through training. In the case of OCR and text validation, a standardised two-day training course has been developed and absorbed by 16 of 50 members (32%), even by a few that eventually decided not to join the initiative. Moreover, networking provenly paid off for a couple of student members directly finding employment in the Team's institutional members and in several

cases of graduate students that received recommendations and, eventually, joined postgraduate university courses across Europe.

Other minor options are also related to more noble goals, including: 'Help dissipate scientific knowledge to democratic institutions' and 'to promote better parliaments and democracies worldwide'. Crowdsourcing initiatives such as the one discussed here are driven by a sincere will of the members engaged in working for a higher purpose and contributing to common good objectives (Proulx et al., 2011).

3.3. Motivation and objectives

To find more about the Team's dynamics, it is significant to study the specific motives of the members. For this, their activity and research plans have been studied (Question 8A). The responses collected could be perceived as partially overlapping. Leaving aside 'OCR of parliamentary texts' (26% or 13 responses) and 'data and text analytics' (40% or 20 responses) that constitute the Team's baseline activity, the top-rated responses are to 'work on a paper' (54% or 27 responses), 'participation in sub-group activities' (50% or 25 responses), and 'participation in a research project' (52% or 26 responses). Their similar frequency cannot be a coincidence since these options correlate: research projects within the team are conducted within dedicated sub-groups and almost always lead to scientific publications and/or conference presentations (see, indicatively, Koryzis et al., 2021; Leventis et al., 2021; Fitsilis et al., 2022).

In one of the explanatory responses (see Question 9B), a member underlined the above with particular clarity:

> 'Having worked extensively in OCR text analysis and text analytics, I think it would be a great opportunity to be able to work on a very specific topic (with a group of people) and to publish the results of our respective research and be presented by ourselves at a conference…'

Additionally, single responses tackle specific operational issues such as 'digital identification' or 'web development', and research activities: 'data visualisation' and 'webscraping script'.

The members were given the opportunity to further specify their responses to their activity plans (Question 8B). Of 50 members, 29 provided their comments that were related to several aspects of the Team's activity. These are mostly linked to their ongoing (by the time of completing the survey) activities on parliamentary digitalisation processes via OCR, corpus analysis, and state-of-the-art topics such as rule-as-code, recommender systems, and cyber security. However, there are also responses referring to desired and future activities that offer significant insights into the group's dynamics and need to be taken into account during planning.

According to members, work on parliamentary diplomacy needs to be further reinforced. A sub-group on parliamentary diplomacy currently counting four members was already established in 2019. Original work on the Hellenic Parliament's use of digital media and a study of its response to the 2019 Turkey-Libya Memorandum of Understanding on maritime boundaries in the Mediterranean Sea resulted in a working paper (Fitsilis and Stavridis, 2021). Others wish to engage in novel research directions and/or topics such as psychology research and the use of terms of socio-political interest by political parties in Greece. In particular, one member mentioned:

> '…I would find [it] useful and beneficial to contribute with further analysis upon the socio-economic approach based on the socio-linguistic analysis results from the text, as well as to broaden my research scope in regards to the parliamentary dialogue and the structural base of such political methods in a democratic state'.

Both directions are screened for their originality and feasibility. They can be materialised through the building of ad-hoc research groups, provided there is a threshold of interest among members. On the one hand, psychology research has been considered early on as a possibility to investigate niche domains of representation and parliamentarism, such as the issue of democratic processes in confined environments (see, for instance, the surrounding discussion by Stone, 1993, and Carrère et al., 1991).

On the other hand, linguistics research, be it traditional or computational, constitutes one of the main Team's expertise. The use of language in parliamentary corpora has the potential to provide substantial added value in the understanding of political discourse and to highlight the links of parliamentary caucuses with specific societal spaces. The availability of unified and verified corpora allows for interlinking several –formerly distant– areas of research, e.g., history, political science, social psychology, and others, thus opening new horizons in the understanding of parliamentary information and discourse. Various research questions can be formulated about the corpus, including how critical foreign affairs issues are formulated and what concepts are most relevant. Moreover, how specific critical socio-political issues like the refugee crisis are being framed in the broader public debate and its social consequences. The application of advanced text mining methods like sentiment analysis and network analysis can also help answer complex questions linking specific politicians, political parties' affiliations, and their position on current critical legislature issues. This kind of layered approach can uncover deeply rooted links between individual actions, ideologies, and social interaction, offering plausible interpretative models of political action and public engagement.

Yet, one needs to keep in mind that several members might not possess the necessary expertise nor the capacity to perform the required tasks, as put forth in one of the responses:

> '…I would very much like to participate in something more. However, because I do not have the scientific training required by the research work of the team, I could, perhaps, help with some training or guidance'.

Training does belong to the Team's essential features but in a structured form, it is mainly offered to the members of the OCR group, for which a particular training methodology has been developed. So far, capacity building on other occasions has been conducted using a training-on-the-job approach. In the light of the Team's transformation into an expert network, the training of new members and the re-training of existing ones remains a challenge. Preparing a Training Needs Analysis (TNA) and a limited series of online training modules[7] are considered possible approaches to fill the gap. These can either be prepared in-house or with the help of the academic institutions some members are working for.

From the remaining opinions, the call for additional research on the 'Parliament of the Future' stands out. Parliamentary evolution and the future workspace of parliaments have been already objects of study (Williamson and Fallon, 2011; Weber et al., 2019). Such studies mainly deal with the potential use of digital tools or elaborated/new institutional processes to tackle societal and administrative

[7] An offline approach is considered obsolete due to the global nature of the network. For the same reason, the initially monthly physical meetings have been limited to twice a year.

complexity. Still, the Parliament of the Future is a rather fuzzy than well-defined concept, to which the Hellenic OCR Team attempts to bring some clarity. There is no scarcity of resources due to the planned widening of the Team's digital scope. Accordingly, members wish to participate in a '[t]echnical teamwork on a dev[elopment] or research project, especially an open source one' and '...some dev[elopment] / programming activity depending on what needs come up.' Interestingly, more experts offer indications about the technologies they want to get involved with. For example, they are 'interested in developing advanced visualization methods and applications … [and] would like to be engaged in … cyber security activities,' they intend to conduct research or write a 'paper on certain fields of interest like AI, Machine Learning and Data Science,' and they wish to promote their '[a]cademic development on the digital transformation in the legislative …'.[8] Those experts were linked to the relevant working groups or ad-hoc projects to take advantage of their drive and motivation and create win-win situations for the Team.

3.4. The future of the initiative

The Hellenic OCR Team is an already 5-year-old initiative in its prime, but, nonetheless, it can improve in many ways, and the members provided several possible options (Question 9A). Almost three-fourths of the members think the initiative should become publicly more visible (74% or 37 from 50). The Team already operates a website and has a social media presence in terms of a company profile on LinkedIn. Moreover, it engages a digital identity expert with professional experience designing logos, banners, videos, and other promotional features.[9] A widening of public exposure that could be possible through further engagement in other social media networks has been considered but rejected due to its time-consuming nature. However, the optimal marketing and communication mix for the Team to approach significant target audiences is still to be determined. Currently, there are preliminary discussions with member companies to prepare a special study.

The expansion of the member base is not unrelated to public visibility. While roughly one-third (32% or 16 from 50) wishes to see more members, their management is also linked with additional administrative overhead. Since the Team's foundation in 2017, admissions have followed a linear trend. Exceeding the 50 members' mark is a point where the day-to-day Team management becomes difficult to handle. Therefore, though administrative efficiency has grown with the introduction of digital platforms and tools, the decision was taken to change the organisational model into an open and independent expert network.[10] This was a landmark decision that affected both the Team's routine and its operational practice.

A significant part favours a conversion of the Team into a legal entity (38% or 19 from 50). Though not of imminent priority, the changing of the Team's nature has been considered and discussed several times among members. A legal identity would certainly provide additional possibilities for attracting grants and implementing national or international (EU) funded projects. At the same time, it would deprive the Team of degrees of freedom to operate and expand in any scientific area without any bureaucracy or binding frameworks other than the agreements reached among members and the

[8] Besides technological innovation, the Team also deals with procedural advances as a main -and equally significant- driver of parliamentary evolution.

[9] See, for instance, our teaser video, https://youtu.be/gKVV1scHzAY, as well as a video message that was developed in the framework of the Team's social responsibility, https://youtu.be/GrhVPQlvahI

[10] Despite its current form as an expert network, the term 'Team' will continue to be in use as it is an inseparable part of the initiative's name that remains unchanged.

management board. While a legal form is not off the table, the decision was made to transform the Team into an 'expert network' from 2022 onwards. Apart from the above re-conceptualisation efforts, which marked the gradual transition away from standard sud-structures (the groups shown in Figure 1) to a more laissez-faire organisational type, this was also a rebranding effort to indicate the global outreach and interoperable character of the initiative.

As a research-driven initiative, the Hellenic OCR Team relies on its Team's interests to grow and advance scientific research around parliaments. While several members are involved in research efforts, the survey provided a clear call to 'follow more personalised research interests of members' (36% or 18 from 50). In effect, follow-up discussions with members were conducted, revealing that the Team's dedication to legislatures might not offer the scientific range or depth some members had hoped for. Among others, it was this survey outcome to increase the personalisation features of the initiative that sped up the transformation into the expert network, within which ad-hoc groups can be formed to instantly tackle new research activities or action items that appear in the (scientific) horizon.

Minor personal opinions that cumulatively attracted for 8% of member opinions include 'more online networking opportunities,' 'follow active topics in the parliamentary research area (...) on a project basis,' and ' get involved in more funded research projects and increase (...) dissemination activities'. These can be linked to various degrees with the above major selections about the Team's future. One of those comments dealing with 'eat your own dog food' projects attracted particular interest and sparked a discussion that showed that internal use of the Team's own products could be, in fact, one of the reasons for its growth. This is because the Hellenic OCR Team does not simply reproduce foreign parliamentary research or advocates for third-party tools and services but invents and develops methods and applications in-house that are applied in its own projects. Hence, if those are good enough for the Team's high standards that are constantly scientifically scrutinised by the parliamentary community, they can surely meet the expectations of parliamentary scholars and practitioners that do not have the expertise or the resources to develop their own.

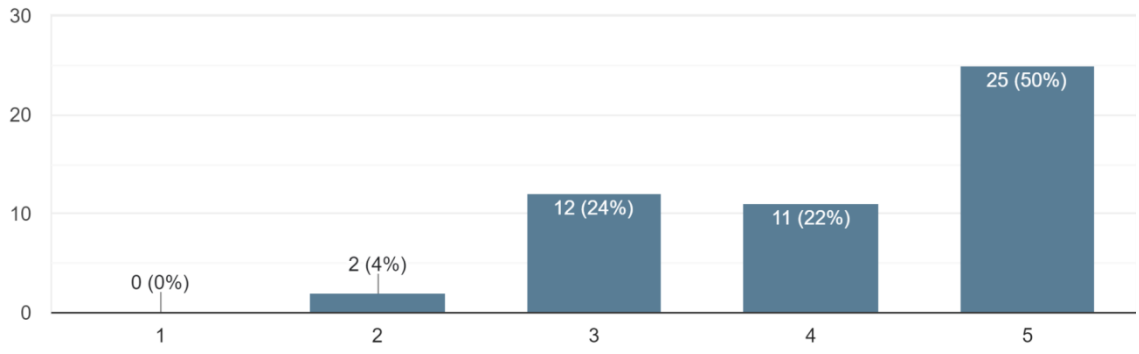

Figure 6. Tendency to transformation.

A 5-point Likert scale was used to capture the members' agreement with the transformation into an expert network (Question 9B; see Figure 6). Overall, the transformation was backed by members: 72% (27 of 50) highly or fully agree with it. The rest 24% (12 of 50), represent a moderate position, and only 2 members are rather cautious (2: little agreement).

4. Conclusions and outlook

This article presented and discussed the evolution, growth, and possible future steps of the volunteering initiative called the Hellenic OCR team. Empirical evidence was drawn from a survey of the entire population of the initiative (N=50) at the moment of the study's submission. Analysis of the pertinent literature proves the unique character of the initiative compared to other types of crowdsourcing activities. This can be particularly attributed to a triplet of characteristics that are nowhere else to be found: its volunteering nature, the members' training, and its permanent character. Furthermore, the Team's geographic, disciplinary, and sectoral diversity is reflected in its research efforts and ongoing projects, which is evident in its scientific publications.

Scientific research is pushed through academic (participation in publications and conferences) and personal incentives (strengthening of CV, building-up of know-how, advancement of computer programming skills, and more). From January 2022, the initiative moved away from a rigid organisational scheme in favour of an expert network with ad-hoc project management and implementation features. During the pandemic, a clear shift to more technical matters in parliamentary research such as AI and digital platforms for the parliamentary workspace, could be observed - a trend that is expected to continue. The new form of the initiative, together with the accumulated advanced technical expertise of the members, enable the further development of the Team's landmark projects, such as the flexible platform for the parliamentary workspace that integrates existing and novel digital solutions through an agent-oriented approach based on Enterprise Integration Patterns (Hohpe and Woolf, 2004; Leventis et al., 2021). This is the precursor for the building of an ecosystem of apps and services with the potential to advance interoperability (Fitsilis and Kalogirou, 2021) and strengthen the position of representative organisations such as parliaments in any given inter-institutional environment.

In addition, the aforementioned digital platform will boost parliamentary data analytics by implementing dedicated modules for specific needs and configurations. For instance, we are expecting new large-scale language models like GPT-3 (Brown et. al., 2020) and open-source multilingual initiatives like BLOOM, to introduce innovative applications in the parliamentary document management cycle, including efficient multilingual text summarization, data-intensive decision-making support, real-time public engagement in open consultation platforms, etc.

Concluding, the Hellenic OCR Team is a unique volunteer-based crowdsourcing initiative in Greece that leads the way in Parliamentary innovation. It will continue to evolve and expand its members base with an approximate rate of 10 members per year. The profile of its members is highly interdisciplinary and multinational. All engaged members are highly motivated either due to high-impact research conducted in the Team or due to unique social networking opportunities offered by the multiversity of the developed network. New opportunities arise through the transformation of the Team into an Experts' network. This transformation seems to be the optimal strategy for the development of this initiative since it maintains the flexibility of an international interdisciplinary scientific crowdsourcing team and further enhances its professional dimensions since it creates a pool of experts that can dynamically form task forces and engage in complex projects that require unique specializations. Furthermore, its transformation into an expert network has lowered the management board's administrative burden, which can now handle strategic development and scientific tasks more efficiently.

Yet, in a globalised world with interconnected parliaments and International Parliamentary Initiatives (IPIs), one initiative, however innovative and far-reaching, is unlikely able to make the difference. The authors believe that true advancements can only originate through extensive cooperation and collaboration schemes. Therefore, team members are linked with the Centre for Innovation in Parliament[11] and actively participate in IPEN and the Inter Pares project.[12]

References

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., . . . Amodei, D. (2020). *Language Models are Few-Shot Learners*. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Carrère, S., Evans, G.W., & Stokols, D. (1991). Winter-Over Stress: Physiological and Psychological Adaptation to an Antarctic Isolated and Confined Environment. In: Harrison A.A., Clearwater Y.A., McKay C.P. (eds) *From Antarctica to Outer Space*. Springer, New York, NY, https://doi.org/10.1007/978-1-4612-3012-0_21

Christensen, H. S., Karjalainen, M., & Nurminen, L. (2015). Does crowdsourcing legislation increase political legitimacy? The case of Avoin Ministeriö in Finland. *Policy & Internet*, 7(1), 25-45. https://doi.org/10.1002/poi3.80

Čibej, J., Fišer, D., & Kosem, I. (2015). The role of crowdsourcing in lexicography. Electronic lexicography in the 21st Century: linking lexical data in the digital age. *Proceedings of the eLex*, 2015, 70-83. Retrieved on July 5, 2022 from https://elex.link/elex2015/proceedings/eLex_2015_05_Cibej+Fiser+Kosem.pdf

Dai, X., & Norton, P. (2007). The internet and parliamentary democracy in Europe. *The Journal of Legislative Studies*, 13(3), 342-353. https://doi.org/10.1080/13572330701500763

Fallon, F., Allen, B., & Williamson, A. (2011). *Parliament 2020: Visioning the Future Parliament*. Hansard Society, London.

Felstiner, A. (2011). Working the Crowd: Employment and Labor Law in the Crowdsourcing Industry. *Berkeley Journal of Employment and Labor Law*, 32(1), 143–203. http://www.jstor.org/stable/24052509Fitsilis, F. & Costa, O. (in press) Parliamentary Administration Facing the Digital Challenge. In Christiansen, T., Griglio, E. & Lupo, N. (Eds.), *The Routledge Handbook of Parliamentary Administrations*. Routledge.

[11] The Centre is an IPU facility that offers 'a platform for parliaments to develop and share good practices in digital transformation strategies, and practical methods for building capacity,' https://www.ipu.org/our-impact/strong-parliaments/setting-standards/centre-innovation-in-parliament

[12] Inter Pares is an EU-funded project 'to strengthen representative democracy through support to Parliaments worldwide,' https://www.inter-pares.eu/

Fitsilis, F., & Kalogirou, V. (2021). Interoperability Legal Patterns: a concept for Greece. In: *14th International Conference on Theory and Practice of Electronic Governance* (pp. 485-487). https://doi.org/10.1145/3494193.3494261

Fitsilis, F., Koryzis, D., & Schefbeck, G. (2022). Legal Informatics Tools for Evidence-Based Policy Creation in Parliaments. *International Journal of Parliamentary Studies*, 2(1), 5-29. https://doi.org/10.1163/26668912-bja10031

Fitsilis, F., & Mikros, G. (2021). Development and Validation of a Corpus of Written Parliamentary Questions in the Hellenic Parliament. *Journal of Open Humanities Data*, 7, 18. DOI: http://doi.org/10.5334/johd.45

Fitsilis, F., & Stavridis, S. (2021). The Hellenic Parliament's use of digital media in its response to the 2019 Turkey-Libya Memorandum of Understanding on maritime boundaries in the Mediterranean Sea: a preliminary assessment. *GreeSE – Hellenic Observatory Papers on Greece and Southeast Europe* 163, Hellenic Observatory, LSE.

Giddings, P. (2005). *The Future of Parliament: issues for a new century*. NYC: Palgrave Macmillan.

Hohpe, G., & Woolf, B. (2004). *Enterprise integration patterns: Designing, building, and deploying messaging solutions*. Addison-Wesley Professional.

Jacobs, S., De Vos, A., Stuer, D., & Van der Heijden, B. (2019). "Knowing Me, Knowing You" the Importance of Networking for Freelancers' Careers: Examining the Mediating Role of Need for Relatedness Fulfillment and Employability-Enhancing Competencies. *Frontiers in psychology*, 10, 2055. https://doi.org/10.3389/fpsyg.2019.02055

Kazai, G. (2011). In Search of Quality in Crowdsourcing for Search Engine Evaluation. In: *Advances in Information Retrieval*. ECIR 2011. Lecture Notes in Computer Science, vol 6611. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20161-5_17

Koryzis, D., Dalas, A., Spiliotopoulos, D., & Fitsilis, F. (2021). Parltech: Transformation framework for the digital parliament. *Big Data and Cognitive Computing*, *5*(1), 15. https://doi.org/10.3390/bdcc5010015

Landemore, H. (2015). Inclusive constitution-making: The Icelandic experiment. *Journal of Political Philosophy*, 23(2), 166-191. https://doi.org/10.1111/jopp.12032

Leston-Bandeira, C. (2012). Studying the relationship between Parliament and citizens. *The Journal of Legislative Studies*, *18*(3-4), 265-274. https://doi.org/10.1080/13572334.2012.706044

Leston-Bandeira, C. (2019). Parliamentary petitions and public engagement: an empirical analysis of the role of e-petitions. *Policy & Politics*, *47*(3), 415-436. https://doi.org/10.1332/030557319X15579230420117

Leston-Bandeira, C. (Ed.). (2016). *Parliaments and citizens*. Routledge.

Leventis, S., Fitsilis, F., & Anastasiou, V. (2021). Diversification of Legislation Editing Open Software (LEOS) Using Software Agents—Transforming Parliamentary Control of the Hellenic

Parliament into Big Open Legal Data. *Big Data and Cognitive Computing*, 5(3), 45. https://doi.org/10.3390/bdcc5030045

Lin, H., & Davis, J. (2010). Computational and crowdsourcing methods for extracting ontological structure from folksonomy. In *Extended Semantic Web Conference* (pp. 472-477). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13489-0_46

Martin, L. H. (1994). The Anti-Individualistic Ideology of Hellenistic Culture1. *Numen*, *41*(2), 117-140. https://doi.org/10.2307/3270256

Orozco, D. (2016). The Use of Legal Crowdsourcing ('Lawsourcing') as a Means to Achieve Legal, Regulatory and Policy Objectives. *American Business Law Journal*, 53(1), 145-192. https://doi.org/10.1111/ablj.12074

Prpić, J., Taeihagh, A., & Melton, J. (2015). The fundamentals of policy crowdsourcing. *Policy & Internet*, 7(3), 340-361. https://doi.org/10.1002/poi3.102

Proulx, S., Heaton, L., Jane Kwok Choon, M., & Millette, M. (2011). Paradoxical empowerment of produsers in the context of informational capitalism. *New review of hypermedia and multimedia*, *17*(1), 9-29. https://doi.org/10.1080/13614568.2011.552646

Sartor, G., Palmirani, M., Francesconi, E., & Biasiotti, M.A. (2011) *Legislative XML for the semantic web: principles, models, standards for document management* (Vol. 4). Springer Netherlands.

Stone, C.N. (1993) Urban Regimes and the Capacity to Govern: A Political Economy Approach, *Journal of Urban Affairs*, 15:1, 1-28. https://doi.org/10.1111/j.1467-9906.1993.tb00300.x

Sun, C., Rampalli, N., Yang, F., & Doan, A. (2014). Chimera: Large-scale classification using machine learning, rules, and crowdsourcing. Proceedings of the VLDB Endowment, 7(13), 1529-1540. https://doi.org/10.14778/2733004.2733024

Tong, Y., Cao, C. C., Zhang, C. J., Li, Y., & Chen, L. (2014). Crowdcleaner: Data cleaning for multi-version data on the web via crowdsourcing. In: *2014 IEEE 30th International Conference on Data Engineering* (pp. 1182-1185). IEEE. https://doi.org/10.1109/ICDE.2014.6816736

Wang, A., Hoang, C. D. V., & Kan, M. Y. (2013). Perspectives on crowdsourcing annotations for natural language processing. *Language resources and evaluation*, *47*(1), 9-31. https://doi.org/10.1007/s10579-012-9176-1

Weber, K. M., Gudowsky, N., & Aichholzer, G. (2019). Foresight and technology assessment for the Austrian parliament—Finding new ways of debating the future of industry 4.0. *Futures*, *109*, 240-251. https://doi.org/10.1016/j.futures.2018.06.018

Williamson, A., & Fallon, F. (2011). Transforming the future parliament through the effective use of digital media. *Parliamentary Affairs*, *64*(4), 781-792. https://doi.org/10.1093/pa/gsr028

Appendix – Questionnaire

ID	Question	Note
1	Member's name	[text input]
2	Gender	Options: male-female-prefer not to say
3	Country of Residence	[list of countries]
4	Your scientific background	Multiple choices possible; options: legal/engineering-information science-economics-political science-philology/linguistics-other[text input]
5	Working sector	Options: public sector-private sector-academia-international organisation/NGO - other[text input]
6	How much time do you (intend to) invest in the Hellenic OCR Team?	Options: less than 1 hour per week - 2-4hours per week - 4-8 hours per week - more than 8 hours per week
7	Why did you join the Hellenic OCR Team?	Multiple choices possible; options: networking-to acquire technical knowhow-to join a scientific community-because of the interesting research topics-to add value to my CV-to strengthen my scientific background-other [text input]
8A	What are your research/activity plans within the team till the end of the year?	Multiple choices possible; options: work on a paper-conference presentation-participation in sub-group activities-participation in a research project-OCR of parliamentary texts-data and text analytics-other [text input]
8B	Specify your above options	[text input]
9A	How can the Hellenic OCR Team become even better in the future?	Multiple choices possible; options: change research focus-convert into a more solid legal entity-follow more personalised researests of members-become publicly more visible-further expand its member base
9B	Do you agree with the team's transformation into an expert network?	Options: 1-5; rating: 1 - not at all; 2 - little; 3 - moderate; 4 - highly; 5 - fully