

Identifying groups of interest in W3C

Henrique S. Xavier

hxavier@nic.br

NIC.br

São Paulo, SP, Brazil

Beatriz Rocha

biarocha@nic.br

NIC.br

São Paulo, SP, Brazil

ABSTRACT

Let's check who participates in the W3C and what are their interests.

CCS CONCEPTS

• **Information systems** → **World Wide Web**; • **Social and professional topics** → **Management of computing and information systems**; *Computing and business*; • **Applied computing** → *Law, social and behavioral sciences*.

KEYWORDS

world wide web, standards, W3C, SDO, topic modelling, web science

1 INTRODUCTION

The Web has changed significantly the way we communicate, study, work, and live, and today it is an indispensable element to innumerable human activities. Since its creation in 1989, the Web itself has changed significantly, often in unexpected directions. Initially designed as an academic information system for sharing documents in a decentralized way [1], the Web gained the capacity to host interoperable, dynamic, interactive, and real-time applications including Software-as-a-Service (SaaS), e-commerce, social media, and streaming services, while facilitating or allowing for new practices such as monitoring, testing and nudging users [12] and profiting from crowdsourcing, wisdom of the crowds, user-generated content, data mining, and targeted advertising [10]. Also, user activity and content got concentrated in a few for-profit platforms [11].

From the history of the Web, it is clear that foreseeing its development can be extremely valuable both in economic and social terms. Albeit challenging, studying the current state of the Web, its actors and their focus of interest can provide evidence of the Web's possible future directions. Some studies have pursued this path, each using different indicators. [11] analyzed online traffic data from SimilarWeb and identified key industries and pervasive practices encountered by users on the Web, along with the main controllers of the most popular websites, whereas [4] explored the topics covered in the Web Conference series, identifying historical and new trends in Web-related research. These studies confirm the importance of Big Tech US companies in driving research and shaping user experience on the Web, of video streaming services, business models based on crowdsourcing content from users or third parties, and of employing artificial intelligence on the Web, specially as recommender systems. In contrast, once thought as the future of the Web [2], the Semantic Web was found stalled in research and not much used in practice [8].

Our goal is to complement these studies by analyzing the participation of different actors in the World Wide Web Consortium (W3C), the main standards development organization (SDO) for

the Web. The W3C is responsible for many standards and best practices that govern the Web and its members include large and small companies, non-profit organizations, government institutions, and academia. By studying the activity of these actors in W3C, we can gain insights into their interests and priorities, as well as the potential impact of their work on the future of the Web.

The main W3C activities take place in groups¹. These can be split into permanent groups and temporary groups. There are four permanent groups, which have governance and oversight attributions: the Advisory Board (AB), the Technical Architecture Group (TAG), the Advisory Committee (AC) and the Board of Directors (BoD). The temporary groups are created to work on specific topics related to the Web standards and practices and are expected to disband after their work is done. These groups are divided into chartered and non-chartered groups. Only W3C members can participate in chartered groups, while non-chartered groups are open to anyone. The former groups are further divided into Working Groups (WG) and Interest Groups (IG), while the latter are divided into Community Groups (CG) and Business Groups (BG). Despite WGs being the only one that can produce W3C Recommendations (a.k.a standards)², all group types are considered relevant for the development of the Web. The community and business groups, for instance, are seen as incubators for new ideas and technologies and may create momentum for a later standardization³ [7].

Previous publications about the W3C have reported that, although counting with members from academia, government, and civil society, large companies are the main players, followed by smaller ones [3, 6, 7, 9]. The latter pattern is common in other SDOs. According to [7], most standards are submitted to the W3C after they have been developed and implemented, typically when the need for interoperability becomes apparent, and "most salient work within the W3C is led mainly by browser companies such as Google and Microsoft, manufacturers like Intel, and user companies such as Facebook" while "Comcast and Netflix are also important actors particularly within specifications for television." A 2016 analysis of the filiations of the editors of W3C standards revealed the collaboration between universities and large and small companies and the dominance of organizations based in the United States [3]. The latter finding was also highlighted in an analysis of the mailing list of the W3C HTML WG, which also identified negligible participation from Asia, Africa and South America and by women in general [5]. Studies specifically focused on the W3C Encrypted Media Extensions (EME) Recommendation delved into mailing lists activity and participation in related groups, identifying similar patterns [6, 9].

¹<https://www.w3.org/groups>

²<https://www.w3.org/standards/types/#x5-standard>

³<https://www.w3.org/community/about/faq>

Using distinct indicators – namely, Recommendation editorships, participation in W3C groups, and activity in mailing lists – the referred works reached an important conclusion about the overall composition of the W3C: the predominance of large companies, especially from the USA and Europe. Presumably, it is the interests of these participants that will guide the evolution of Web standards. However, previous studies did not differentiate the companies (by industry sector, for instance) and did not analyze the topics of interest of these participants. This paper aims to fill this gap by identifying the W3C members’ interests, as expressed in the temporary groups they participate in, and use this information to discover latent topics and interest groups in W3C. Besides revealing the topics being driven by strong Web players, our analysis can also detect fields left to other organizations and that can flourish subject to different interests.

2 DATA SET

The data used in this paper is publicly available on W3C’s website and API and allows for the identification of the organization each W3C group participant is affiliated to. For each organization we collected information on whether it is a member of W3C, and for members we were able to collect their headquarters country. We also used retrieval-augmented generation (RAG) to annotate the sector of each member organization under the following categories:

- **Large enterprises:** for-profit companies with more than 250 employees.
- **Small and medium enterprises (SMEs):** for-profit companies with up to 250 employees.
- **Non-profit organizations:** non-profit organizations, including NGOs and foundations.
- **Government:** government institutions, including intergovernmental organizations.
- **Academia:** universities and research institutions.
- **Technical communities:** technical communities, including standards development organizations (SDOs) and open source projects.

These sectors are mainly based on the multi-stakeholder groups adopted by the Internet Governance Forum (IGF)⁴ and NET mundial⁵, with an extra breakdown in the private sector inspired by [3, 7]. A manual check of 60 organizations indicated an accuracy of 87% for the RAG-annotated sectors.

Table 1 summarizes the data sources used in this paper, along with our own identification scheme and collection date. Given that the member organization headquarter’s countries data were collected several months after the rest of the data, 30 organizations may have left W3C and did not figure in country. Thus, we filled the missing data by manually checking Wikipedia and the organizations’ websites. Fig. 1 presents a graph visualization of our data relations.

At the time of data collection, there were 195 temporary groups in operation: 141 CGs, 43 WGs, 9 IGs, and 2 BGs. The listing data also includes references to task forces, which are teams charged with specific tasks and subordinated to other W3C groups. We opted not to include them in our analysis, as they are not independent.



Figure 1: Graph visualization of the data relations. Datasets with names in red establish connections between entities, while datasets with names in black provide entity properties used to list and filter the data.

Group participants may figure in affiliations with up to two affiliations. In such cases, one is always “W3C invited expert”. We opted to standardize the data by removing this affiliation when it is not the only one. This is because the invited expert affiliation is not an organization, but a status that allows individuals to participate in W3C groups without being affiliated to any organization. After this standardization, we were left with 147 invited experts with no other affiliation, which we classified as “W3C Invited Experts”. Taking W3C and this pseudo affiliation into account, there were 303 organizations with at least one representative in W3C temporary groups at the time of data collection. About 21% of the W3C member organizations were not participating in any group.

3 METHODOLOGY

3.1 Sector annotation with RAG

3.2 Topic modelling

4 ANALYSIS

4.1 Overall demographics

We first assessed the level of participation of member organizations in W3C groups by counting the number of representatives they have assigned to W3C temporary groups. This evaluation might hint us about organizations with the most capacity to influence the development of Web standards. Despite the W3C decision process being consensus-based where each organization has one vote, having representatives in multiple groups may improve the organization’s ability to coordinate efforts across groups. Additionally, a larger number of representatives participating in a given group may sway the concrete consensus-building process in favor of that organization by creating the impression of a larger community support and by improving its arguing capabilities.

Fig. 2 presents the cumulative distribution of the number of representatives per organization for chartered (restricted to W3C members) and non-chartered groups (open to all organizations). The vertical dotted and dashed lines mark the number of organizations required to amass 50% and 80% of the representatives, respectively. We note that the concentration of group participants in a few organizations is larger in chartered than in non-chartered groups (Gini index of 0.68 versus 0.44). In chartered groups, half of all representatives are affiliated to the 20 largest among 397 organizations (5%). By resampling 397 organizations from non-chartered

⁴<https://www.intgovforum.org/en/about#about-igf-faqs>

⁵<https://netmundial.br/2014>

Table 1: Data sources

Data ID	Description	Source	Collection date
listing	W3C Groups listing	https://api.w3.org/groups	2024-07-25
groups	Groups properties	https://api.w3.org/groups/{group}	2024-07-25
users	Groups participants	https://api.w3.org/groups/{group}/users	2024-07-25
affiliations	Participants' organization	https://api.w3.org/users/{user_id}/affiliations	2024-07-25
membership	Org. membership status	https://api.w3.org/affiliations/{org_id}	2024-07-26
country	Org. HQ country	https://www.w3.org/membership/list	2025-04-15
sector	Organization sector	<i>Our own annotation using RAG</i>	2025-07-15

groups, we concluded that the probability of reaching the chartered group Gini index by chance is 1.1%, showing that the increased concentration seen in chartered groups requires a sociological explanation. One hypothesis would be that the W3C membership fee selects larger organizations or those more interested in W3C standards.

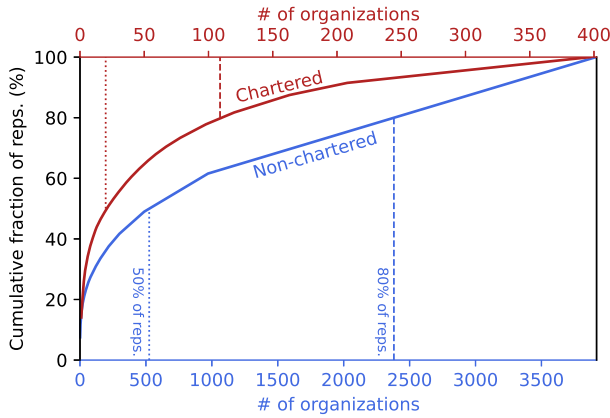


Figure 2: Cumulative fraction of organization representatives participating in chartered (red) and non-chartered (blue) groups by number of organizations considered (ordered by decreasing number of representatives). For chartered groups, the number of organizations is referenced on the top horizontal scale, whereas for the non-chartered groups it is referenced on the bottom.

Fig. 3 shows the number of W3C group participants that are affiliated to the top 20 largest organizations in terms of number of representatives, which all happen to be W3C members. Our hypothesis is that these organizations are the most influential in W3C. Google leads the list by a large amount (693 representatives), followed by Microsoft. Interestingly, apart from W3C invited experts, the top four organizations in plot are major browser vendors.

We aggregated the number of W3C member organization representatives by the country in which the organization's headquarters (HQ) is located. Fig. 4 shows the results for the top 20 countries with most representatives, revealing the strong presence of organizations with HQ in the United States. It amassed 2,245 people,

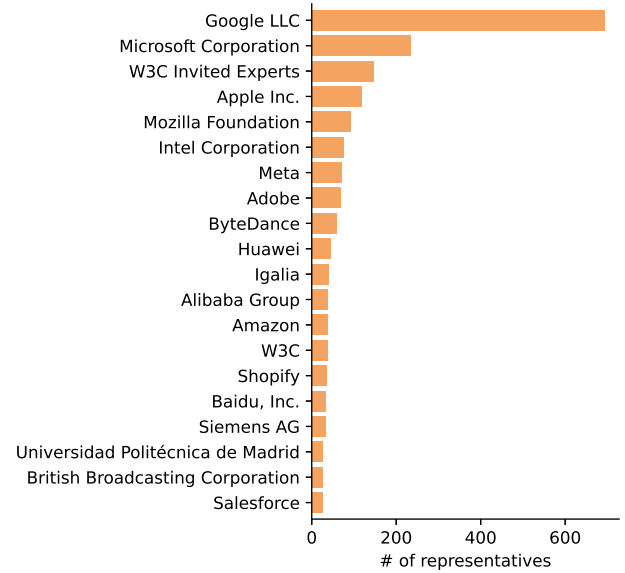


Figure 3: Number of representatives from the 20 largest organizations participating in W3C groups.

comprising 65.7% of all W3C member representatives. This happened both because there are many member organizations in W3C from the United States and because some of these organizations have many representatives. Among the top 5 countries, USA, China, Japan, and UK also emerged as the owners' location of the world's most visited websites [11]. In contrast, Germany was well represented among W3C members but did not figure amidst the 116 most visited domains, whereas Russia was home of highly visited domains but had no members in W3C.⁶

Fig. 5 presents the number of W3C member representatives aggregated by sector. Representatives from large enterprises are by far the most common participant. This is caused by the large number of W3C member organizations that are from this sector and also because such organizations are typically represented by a large team. Conversely, the average Small/Medium enterprise have the smallest number of representatives of all sectors.

⁶Note that Yandex participate in non-chartered groups, but not as a W3C member.

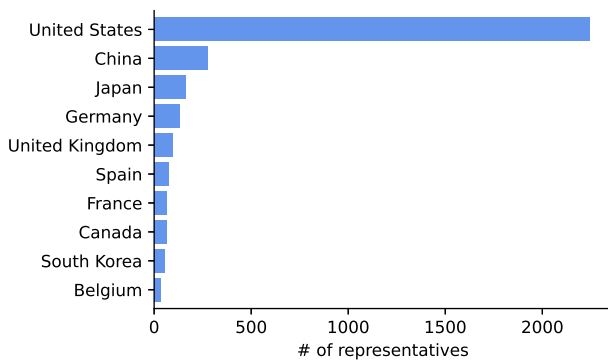


Figure 4: Number of W3C member representatives aggregated by member HQ country, for the top 10 countries with most representatives.

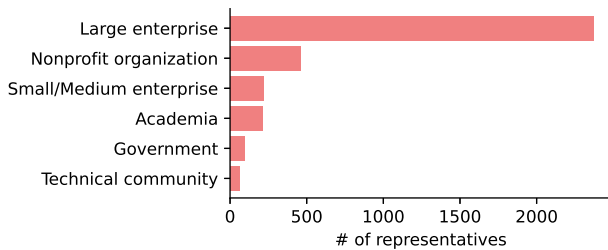


Figure 5: Number of W3C member representatives aggregated by member sector.

4.2 Interest modelling

5 SUMMARY AND CONCLUSIONS

REFERENCES

- [1] Tim Berners-Lee. 1990. World Wide Web: Information management - a proposal. (1990). <https://cds.cern.ch/record/1405411>
- [2] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American* 284, 5 (2001), 34–43. <https://www.scientificamerican.com/article/the-semantic-web/>
- [3] Jonas Gamalielsson and Björn Lundell. 2016. On Involvement in Open Standards: How Do Organisations Contribute to W3C Standards Through Editorship?. In *Open Source Systems: Integrating Communities*. Springer International Publishing, Cham, 57–70.
- [4] Damien Graux and Fabrizio Orlandi. 2022. Through the Lens of the Web Conference Series: A Look Into the History of the Web. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) (WWW '22). Association for Computing Machinery, New York, NY, USA, 3458–3464. <https://doi.org/10.1145/3485447.3512281>
- [5] Harsh Gupta. 2016. (Lack Of) Representation of Non Western World in process of creation of Web standards. arXiv:1609.01996 [cs.CY] <https://arxiv.org/abs/1609.01996>
- [6] Harry Halpin. 2017. The Crisis of Standardizing DRM: The Case of W3C Encrypted Media Extensions. In *Security, Privacy, and Applied Cryptography Engineering*, Sk Subidh Ali, Jean-Luc Danger, and Thomas Eisenbarth (Eds.). Springer International Publishing, Cham, 10–29.
- [7] Alison Harcourt, George Christou, and Seamus Simpson. 2020. Internal governance of the IETF, W3C and IEEE: structure, decision-making and internationalisation. In *Global Standard Setting in Internet Governance*. Oxford University Press, Oxford, UK, Chapter 3.
- [8] Pascal Hitzler, Krzysztof Janowicz, and Aidan Hogan. 2020. The Semantic Web: Two decades on. *Semant. Web* 11, 1 (Jan. 2020), 169–185. <https://doi.org/10.3233/SW-190387>
- [9] Scott Mason. 2019. 'Lowering your standards'?: assessing the procedural legitimacy of the World Wide Web Consortium (W3C). Ph. D. Dissertation. Keele University, Keele, UK. <https://keele-repository.worktribe.com/output/413817/lowering-your-standards-assessing-the-procedural-legitimacy-of-the-world-wide-web-consortium-w3c>
- [10] Tim O'Reilly. 2005. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. (2005). <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html> Accessed on 2025-04-24.
- [11] Henrique Xavier. 2024. The Web Unpacked: A Quantitative Analysis of Global Web Usage. In *Proceedings of the 20th International Conference on Web Information Systems and Technologies - WEBIST*. INSTICC, SciTePress, 183–190. <https://doi.org/10.5220/0012905900003825>
- [12] Shoshana Zuboff. 2019. *The age of surveillance capitalism: the fight for a human future at the new frontier of power* (1 ed.). Public Affairs.

A ANNOTATED DATA

Received 19 April 2024