Replication: A Longitudinal Dataset to Characterize RFC Deployment

Paper #3, 11 pages

ABSTRACT

Protocol standardization plays an essential role in the successful operation of the Internet and its inherent properties such as reliability, performance and security. Recognizing the relevance of the protocol standardization process, in 2021 McQuistin*et al.* [2] characterized the work of the Internet Engineering Task Force (IETF) in Internet standard development by creating and analyzing a corpus with metadata of more than 8 thousands Request For Comment (RFC) documents—the name used by IETF published documents—, emails exchanged by working groups and connections between authors. Based on experts evaluations of RFCs and their implementation success, McQuistin emails exchanged by working groups and connections between authors. *et al.*build a statistical model to identify factors that impact RFCs implementation.

In this paper, we replicate [2] and build an automated pipeline to create and update a corpus of metadata and features from RFC documents, authors information, RFCs emails exchange, that can be run repeatedly to include new RFCs and emails to enable continued research on the standardization ecosystem. We replicate the multiple manual steps of data cleaning using GPT, and reimplement feature computation with currently available APIs. We also extend the significative attributes of the experts' evaluation by training a Generative AI model to replicate the expert evaluation. Using this corpus, we replicate and extend all analysis done in [2] and explore the discrepancies.

1 INTRODUCTION

Protocol standardization plays an essential role in the successful operation of the Internet and its inherent properties such as reliability, performance and security. However, RFCs and the protocols they describe have varying levels of adoption, and identifying the factors that relate with protocol adoption success is key to improve the standardization process and ultimately enable the Internet properties that the community considers relevant. In addition, monitoring the features of RFC development with more predictive power of success, their evolution overtime and their representativeness in RFCs allows to evaluate the overall health of the standardization process.

Recognizing the relevance of the protocol standardization process, in 2021 McQuistin *et al.* [2] characterized the work of the Internet Engineering Task Force (IETF) in Internet standard development by creating and analyzing a corpus of more than 8 thousands RFC documents and the related emails and connections between authors during the standardization process. The IETF is the main body driving Internet Standard development since the 1980s. The authors study trends in more than 30 years Internet standard development, and find that the complexity of the standardization process and the time for RFCs to get published has increased over time. Additionally, based on experts evaluation of RFC implementation success from [3], they build a statistical model to identify factors that impact RFCs implementation. Their model indicate that RFCs building on existing work, having limited scope and meeting a need have a positive influence in successful deployment.

Although in [2] the authors made the data available at publication time, some metadata needed to compute some of the significant features of the predictive model cannot be reproduced. Indeed, a key component of the metadata of the RFC corpus that feeds a significant feature in the predictive model is the count of inbound citations an RFC gets over time. The authors collect citation counts over time from the Microsoft Academic Knowledge API that was shut down in 2022. In addition, the list of topics extracted from RFCs using Latent Dirichlet Allocation (LDA) is not available in the online documentation of the paper and hence it's not possible to identify the significant topics and analyze their relevance in new RFCs. evaluate and monitor new and current RFC development.

In this paper, we replicate the work of [2] by creating a new RFC corpus of metadata and features computed from RFC text, emails, interactions and authors documents, metadata and features, created using a pipeline that can be run repeatedly to include new RFCs.

We also train a model based on generative Artificial Intelligence (genAI) to extend the most significative features of the expert evaluation and then analyze all newer Standard track RFCs to identify if they have the characteristics or no, providing new features to more RFCs.

Our main contributions are as follow:

- We build an almost fully automated pipeline to productionalize the RFC and standardization process metadata and features corpus, providing up-to-date data to study and monitor the protocol development work of the IETF. Our goal is to run the pipeline once a month to include new RFCs and emails.

- We prove that generative AI can now be trained to learn from the experts' evaluation allowing us to extend the assessment of more complex features of RFCs that convey additional insights into the standardization process.
- We extend the analysis of the target paper by 4 years until December 2024 and find that after the COVID singularity some of the trends identified by the previous authors (such as the increasing diversity of authorship) are being reversed.

To facilitate and support further research, we make our analysis code, tools and computed dataset and summary reports publicly available.

2 BACKGROUND AND MOTIVATION

2.1 The IETF Standardization Process

The IETF is an open standard organization founded in 1986 that develops voluntary standards for the Internet. The IETF publishes Request For Comment documents or RFCs that are created through the collaboration and interaction of the Working Groups and the authors. Since its beginning, the IETF has now issued more than 9,500 RFCs.

The RFC development process is very collaborative, with most of the interaction and work conducted via mailing lists, working group meetings and three large plenary meetings per year. The publication process of an RFC starts when an Internet draft is submitted.After the first post, multiple version might be proposed, each announced to one or multiple mailing lists related to the standard area. Feedback and discussion them primarily takes place via email. For publication in IETF Streams, drafts need to be taken by a working group where further revision and discussion takes place.

Once published, RFC are voluntary standards and therefore not all RFC get to be deploy and much less are widely implemented.

2.2 Target Paper

The target paper [2] was the first work to perform a comprehensive longitudinal analysis of RFC and Internet standardization process. The authors design and develop many features that allow to characterize the process not just based on the final product, the RFC, but also on the many aspects of the interaction between authors and characteristics of authors themselves. The authors use the many features to perform a longitudinal analysis of RFC standardization process, studying the volume, author counts, author diversity, topics, and many more characteristics of the RFC documents and the process over time.

The target paper also looks to explain RFC success by building models based on the expert evaluation of RFCs made by [3]. Using all the features they compute about RFCs, authors and their interaction, their final model is more explainable than prior work.

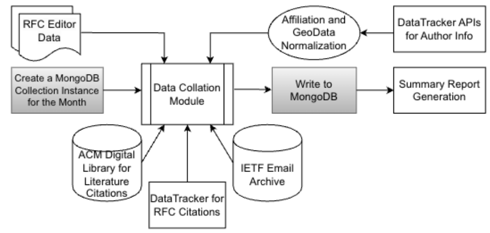

Figure 1: Diagram of datasets used to create the RFC and standardization process dataset

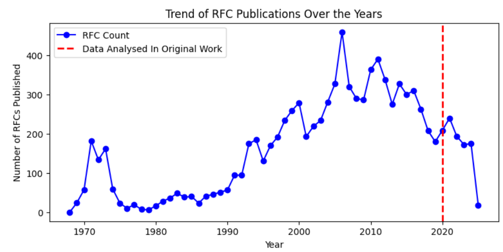

Figure 2: Count of published RFCs per year until end of 2024

2.3 Research Goals

The main goal of this paper is to replicate the target paper by building a pipeline that productionalizes the RFC and RFC standardization process metadata and feature computation, to be able to update the corpus once a month. This corpus of features and metadata provides an unique opportunity to enable further research on Internet standard development as well as to track and monitor the current state of protocol development and its evolution.

With that goal in mind, we focus on automating as much as possible all the manual steps used in the target paper to build the corpus and update the citation API to use ACM citation as the API used by the target paper was retired 4 years ago.

3 DATASETS AND METHODOLOGY

3.1 Data Sources

3.1.1 IETF databases. The IETF has three main databases that cover different parts of the RFC development process: the RFC editor, *(ii)* `datatracker` and email archives. we refer the reader to the target paper for more details on the three databases.

3.1.2 ACM Digital Library. The ACM Digital Library is a comprehensive, digital platform offering access to a vast collection of computing research and publications. It includes full-text articles, bibliographic records, and other materials

```
"""
Which country and continent are these addresses located in?
Simply return a Python compatible dictionary in JSON format corresponding to each input.
Do not modify the input itself.
Make sure the names are in consistent format.
For example, dont use U.S for one address and United States for other.
Write complete names only always.
Santa Clara  0123
FI
{
"Santa Clara  0123" : {"country": "United States", "continent":"North America"},
"FI":{"country": "Finland", "continent": "Europe"}
}
Answer:
"""
```

Figure 3: Geonormalization Prompt

like videos and audio, with a focus on ACM's own publications and a broader selection from other publishers.

The target paper used Microsoft Academic Knowledge (MAK) API to evaluate citations to RFCs. However, the MAK is no longer available. Given the ACM Digital library citation can be scraped from the web, we use ACM citation counts.

3.2 Building the RFC Corpus

Figure 1 shows the pipeline to build the corpus of RFC and RFC publication process metadata and feature. Figure 5 depicts the yearly count of RFCs in our latest dataset, including more than 500 RFCs published after 2021, when the target paper was published.

3.2.1 Geolocalization normalization. The geographical information about the authors is not present in a very consistent manner in the Datatracker. For example, there are many instances where we would have the street address in the "Country" field or the country names might not be in a standard format, for example, at some places we could have the United States or United States of America and United States of America at the others. These inconsistencies make it harder for users to work with and analyze the trends in this data. We used few-shot prompting techniques on the ChatGPT-4 model to perform normalization of the data in these fields. The complete prompt is shown in Fig 3. We found 2187 unique addresses in our RFC corpus of a total of 9,571 RFCs. As part of the validation process, we manually spot checked some couple entries and compared our plots for authorship countries and continents (Fig 16 and Fig 17) against the plots obtained by the original papers. The original authors try to address this issue by a static rule-based method that uses the pycountry Python library and a manually designed hard-coded dictionary. Using the ChatGPT-4 API instead for this task not only offers a more intelligent and flexible alternative to the manual, rule-based approach used in the original paper, but also makes this process scalable and easier to maintain.

3.2.2 Affiliation Normalization. The affiliation names also suffer from inconsistencies similar to those found in the geodata. To address this, we use Chat GPT-4 to normalize them.

```
"""You are an expert in data normalization and entity resolution. Your task is to
normalize raw affiliation strings to standardized organization names.
You will be provided with a mapping dictionary of raw affiliation variants and their corresponding normalized names.
Your goal is to match a given input affiliation string to the correct normalized affiliation,
using the rules implied by the mapping.
Rules to follow:
1. Normalize abbreviations and acronyms to full institution names.
2. Handle punctuation, spacing, and capitalization inconsistencies.
3. Recognize when brand names or departments belong to a parent company or university and map accordingly.
4. Account for multilingual or localized versions of university or organization names.
5. If an affiliation cannot be matched to any known mapping, return "Unknown".

Input:
You will receive a string representing a person's affiliation.

Output:
Return a JSON dictionary with original names as keys and the normalized names as values.
Do not add any additional text in the output please.

Example:
If the input is "UC Berkeley", return "University of California, Berkeley".
If the input is "ATT", return "AT&T".
If the input is "Futurewei", return "Huawei".
If the input is "Independent researcher", return "Unknown".

Use fuzzy matching or logical rules if needed, but ensure high precision.

Task:
Normalize these affiliations:
"""
```

Figure 4: Affiliation Prompt

Previously, the authors implemented a static hardcoded mapping along with a set of predefined rules. While this was a natural go-to approach before the advent of advanced AI models, it faces two major limitations: lack of scalability and the need for frequent manual updates, especially in light of ongoing mergers, acquisitions, and organizational changes in the market. Our approach involved deriving the normalization rules from this hard coded list created by the authors and using these in addition to some examples in our few shot prompt shown in Fig 4 We found 3475 unique affiliation names in our RFC corpus. While our method does address the issues with the original authors approach, given AI models can access the web and have the latest information, we noticed a few other issues while spot checking some entries in the results from GPT-4 normalization. First, sometimes the models may hallucinates and thus forget to apply the rule on certain entries while processing them in bulk. For example, in some cases we saw that it did expand M.I.T to its full form while in others it kept it as is. Next, some mergers and acquisitions were not identified and normalized to parent company. For example, the acquisition of Sun Microsystems by Oracle was not identified and applied during the processing. In the future, we could try other prompting approaches like Chain of Thought Prompting etc or fine-tune a LLM specifically for this purpose.

3.2.3 RFC Standard-Requirement Keyword counts. We reproduce the code based on the paper description of this feature, parsing RFC text to count the requirement-setting keywords as described in RFC 2119 [1] and divide by the page count of the RFC. When studying the aggregate values per year (see § 4.1), we realize we obtain slightly higher numbers than the target paper. Unfortunately, the target paper repository does not include the raw counts per RFC nor the list of RFCs that were made available in IETF databases at the time. We attribute the difference with the target paper to RFCs from previous years whose text became available in IETF databases after the target paper publication. We do see new older RFCs added to `datatracker` and `RFC Editor` and consider it the most likely explanation for the overall aggregated statistics differences. We note that the differences are not much and do not affect the overall trend.

3.2.4 Citations. Inbound and outbound citations are useful to quantify the relationship between standards and the impact or relevance of RFCs. The target paper mentions the authors use the metadata from the IETF `datatracker` and Microsoft Academic Knowledge (MAK) API to study the process behind RFCs, including citations from RFCs to RFCs. However, the MAK is no longer available and there is no code to extract the exact counts from RFCs in their public repository. In the next paragraphs, we explain how we compute each citation count.

Outbound Citations. The `Datatracker` provides direct access to the references in an RFC. From those links, the ones that are RFC or drafts of RFCs can be identified as they have links within the datatracker system. We use this capability in our code to count for each RFC, the outbound citations from an RFC to other Internet-drafts and RFC. Our aggregate statistic match perfectly with the graphs of the target paper. However, starting in 2022, we see the outbound citation count to drop significantly (see § 4.1). We manually investigate RFCs from the past few years and find that those RFCs to cite other RFCs and drafts at the same level or more, but that the reference object from in the `datatracker` is not populated. We report this bug in the `datatracker` github repository and will update the corpus once it is fixed.

RFCs' inbound citations. The target paper mentions the authors use the metadata from the IETF `datatracker` to study the process behind RFCs. However, there is no code to extract citations from RFCs in their public repository. We use `datatracker` citation counts from other RFCs and the publication date of RFCs, to compute for each RFC the citations received within one year and two years of publication from other RFCs. When studying the aggregate values per year (see § 4.1), we realize we obtain slightly different median and 75^{th} percentile than the target paper. As for the keyword counts above (see § 3.2.3), we attribute this difference to RFCs from previous years whose text became available in IETF databases after the target paper publication.

Other inbound citations. A significant departure from the target paper is that it used Microsoft Academic Knowledge API the authors used to count inbound citations to the RFCs is no longer available, it was retired at the end of 2021. We implement the inbound citation count using the ACM digital library and build a scraper to get the counts for each RFC. However, the ACM library is for "ACM publications, including journals, conference proceedings, technical magazines, newsletters and books", which represents a much more limited set of works than what the Microsoft Academic Knowledge API used to include. As a consequence, the inbound citation of RFC drops practically to zero for all RFCs after 2004.

3.2.5 RFC topics. Based on the description of the author's methodology, we reproduce the Latent Dirichlet Allocation (LDA) to find the topic distribution in our RFC Corpus. Since the target paper neither outlines the pre-processing steps nor gives the code associated with this modeling methadology, we design our own pre-processing and modeling logic based just on the paper description. Every RFC text data is first preprocessed by cleaning, tokenizing, removing stopwords and lemmatization. A dictionary and a Bag-of-Words model is constructed, filtering out very rare or overly frequent terms to reduce noise. The Gensim library is then used to train LDA model on this corpus specifying 50 topics and using multiple passes for convergence. Hyperparameters like alpha are autotuned and the model outputs a topic distribution for each document, helping identify the themes of every document. The coherence score of our LDA model is 0.5563 and the log perplexity is -10.2017.

3.2.6 Analysis of Email Interactions. To analyze email interactions between RFC authors and working group participants, we reused the code from the target paper to process the IETF email archive (see Section 3.1.1) and store the results in a MongoDB corpus. A key distinction, however, is that while the original study was based on Mail Archive Version 1, our analysis uses the updated Version 2, which the target paper's repository now also reflects. Although most of our plots closely resemble those in the original study, we observed discrepancies in a few. We attribute these differences to the change in Mail Archive versions as well as updates in the underlying data from the IETF Datatracker. Despite efforts to adapt the original code by replacing Mail Archive V1 calls with their V2 counterparts, certain plots could not be reproduced beyond the year 2020 due to cascading programmatic failures—some graphs rely on intermediate outputs from others, which were affected by upstream data and code dependencies.

4 TRENDS AND CHARACTERIZATION

4.1 RFCs and Working Groups

RFCs by Area over time. Our graphs match the ones from the target paper with small change. IETF recently changes the RFC streams. The target paper used some of the previous

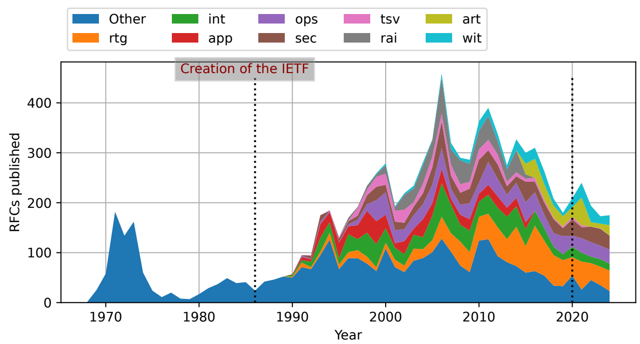

Figure 5: RFCs by area according to RFC streams of 2025. Other includes legacy RFCs, non-working group RFCs, and RFCs from other non-IETF streams.

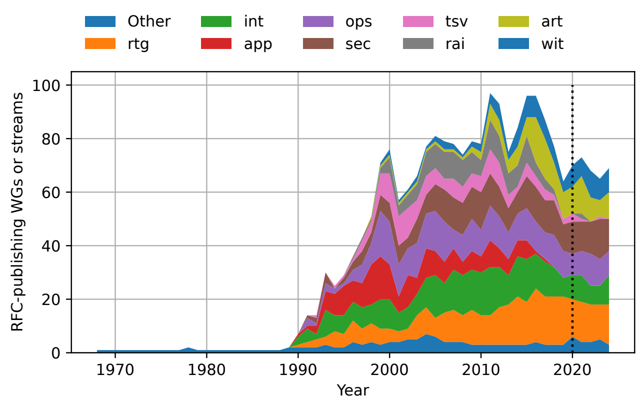

Figure 6: Number of publishing working groups. Other includes legacy RFCs, IRTF research groups, and non-IETF streams.

streams. Unfortunately, we do not have access to the exact count. Also ,datatracker may have had small differences on the past counts a few years ago.

Working Groups by Area over Time. Here again we have small differences with the graphs from the target paper but the trends are similar.

RFC publication effort. The target paper had already identified that the median number of days from the submission of the first draft to the publication of an RFC had more than doubled from 2001 to 2020, going from 469 to 1,170 days. Figure 7 shows that this trend was made more extreme during pandemic, reaching a peak of 1,702 days (4 years and 8 months) in 2021. In 2022 the time to publication decreased but overall the increasing trend continues and in 2024 the median time to publication was 1,548 (4 years and 3 months). Similarly, the number of drafts per published RFC has continued to increase since 2020, as depicted by Figure 8.

RFC Length. Just like in the target paper, even though the effort it takes for RFCs be published continues to increase, the

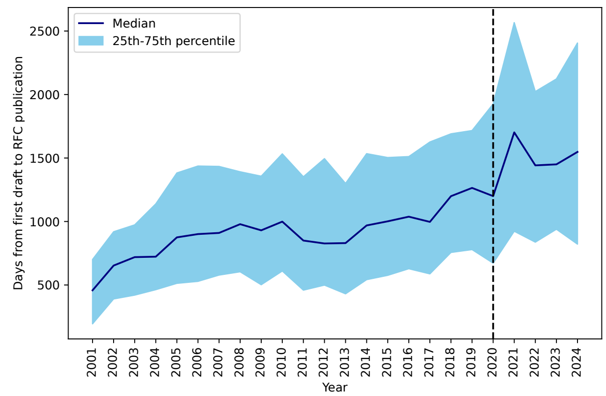

Figure 7: Number of days from first draft to RFC publication

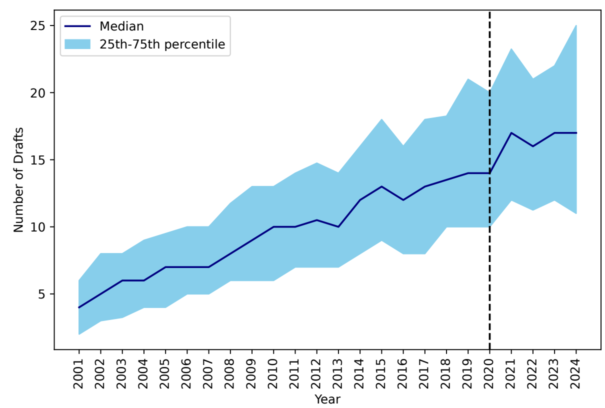

Figure 8: Number of draft per RFC

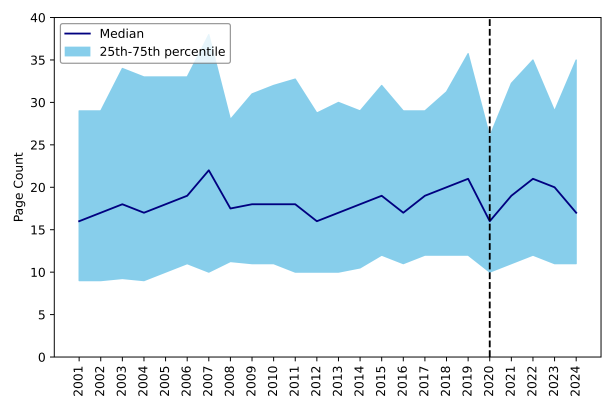

Figure 9: Number of pages per RFC

median page count has continued to stay somewhat stable, as portrayed in Figure 9.

Relationship between RFCs. To capture increasing complexity in newer RFCs as they need to build on top or compare with previous standards, the target paper studies RFCs that directly update or obsolete previous ones and the count of

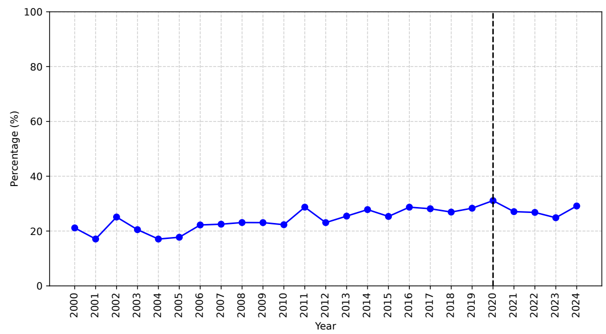

Figure 10: RFCs that update or Obsolete previous RFCs

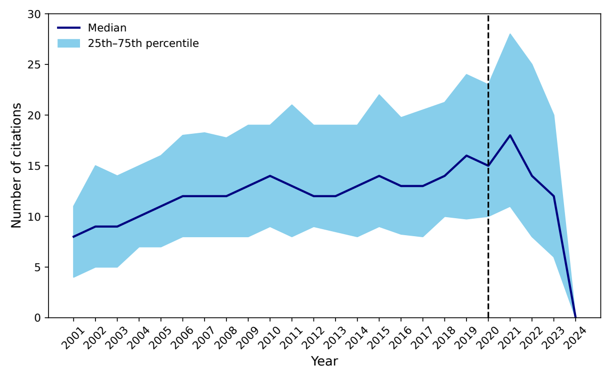

Figure 11: Citations from RFCs to other Internet Drafts (including RFCs) per RFCs

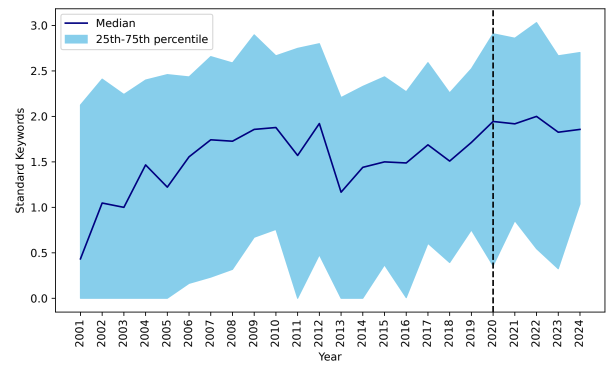

Figure 12: Requirement-setting keywords per page

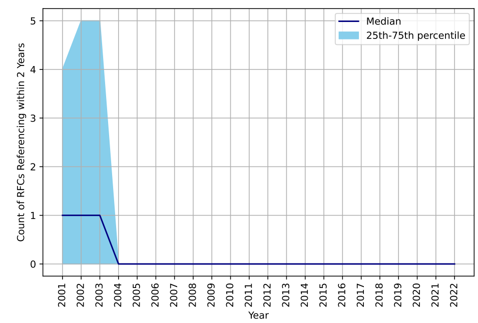

Figure 13: Citations to RFCs received within two years of publication from ACM Digital Library works

citations an RFC has to other RFCs and Internet drafts. Up until 2020, both metrics were steadily increasing. Although the proportion of yearly RFCs that update or obsolete one or more RFCs increased up until 2020 reaching about 30%, since then it has remained fairly stable. As Figure 10 shows, between 2020 and 2024, about 25-30% of RFCs update or obsolete a previously published RFC. The median number of outbound citations from an RFC to other RFCs and Internet draft seems to keep rising after a drop around COVID years. However, starting in 2022, we find that IETF `datatracker` does not populate the datatracker database correctly for references and most RFCs have no identified references in the `datatracker` system. On manual inspection we find this not to be the case ans in the text we see the same or higher level of citation to other RFCs or drafts. As mentioned in § 3.2.4, we report the bug and will update the corpus once the references are fixed.

Requirement-setting language. In the target paper the authors found that the number of requirement-setting keywords such as *SHOULD* and *MUST* had slowly increased between 2001 and 2020, from a median of 0.5 to about 1.5 keywords per page. We find a similar trend though with slightly higher numbers for 2020, reaching almost 2 keywords per page. As mentioned in § 3.2.3, we attribute this different to a few RFCs per year whose text became available in IETF databases after the target paper publication. Since 2020 though, the median count of keywords per page has stabilized at the 2020 level.

Academic impact of RFCs. Given the Microsoft Academic Knowledge (MAK) API is now retired, we use citations counts from the ACM digital library, as explained in § 3.2.4. ACM citations are an order of magnitude lower than the counts from MAK. However, we do see the same trend that RFCs in the early 2000s have many more citations in other works than later RFCs.

The same declining citation trend is present when looking at RFC citations from other RFCs within 2 years of publication.

Summary. Overall the trends identified by the target paper have mostly continued to play out after the COVID singularity. The number of publishing working group has stabilized in the last 5 years while the effort to publish RFCs continues to increase, as revealed by the median days from first draft to publication and median count of draft per RFC continuing to rise.

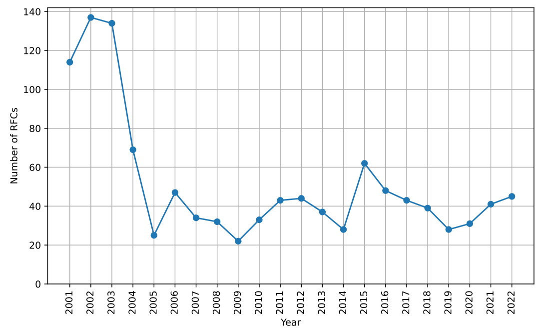

Figure 14: RFCs with Non Zero Citation Count

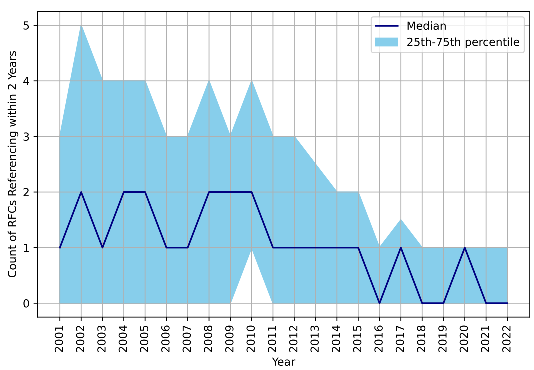

Figure 15: Citations to RFCs received within two years of publication from other RFCs

Field	Valid Count	% of Authors
Author Name	17774	100.00%
Country	13761	77.42%
Continent	13762	77.43%
Affiliation	16600	93.39%

Table 1: Authorship entries with information in `Datatracker`

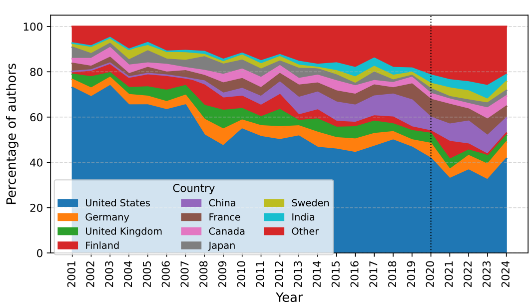

Figure 16: Authorship countries (normalized)

4.2 Authorship

Geodistribution of authors. Figures 16 and 17 show the proportion of authors from countries (top 10 listed) and continents respectively. As described in § 3.2.1, we use GPT instead of a manual hardcoded approach to normalize geolocalization of authors, which leads to some differences in our numbers but our aggregated statistics until 2020 follow the same trend as the target paper.

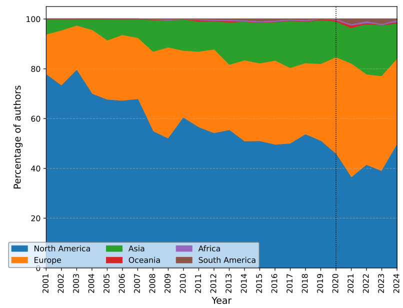

Figure 17: Authorship continents (normalized)

Country-wise, authorship has consistently been dominated by the United States, although after 2013, less than 50% of combined authors of RFC are based in the US. The trend continued to decline after 2020 reaching about 35%, stabilized between 2021 and 2023 and has started to climb back up reaching 42% in 2024. Conversely, authors from China were increasing up until 2019, peaking at 10% but since have started to decline, making only 5% in 2024.

At a continental level, authorship from North-America declined until 2021, going from 75% in 2001 to almost 35%. However, after staying below 40% until 2023, it increased to almost 50% of authors in 2024. Conversely, European authorship increased consistently until 2021, reaching almost 45% but has since declined and in 2024 it European authors were 33% of all authors. Authors from Asia have represented between 14 and 18% of all authors for the past decade. Since 2020, there has been the authors from South America have significantly increase, going from about 0.5% to 2% between 2021 and 2023. In 2024, authors for South America, Africa and Oceania combined represented less than 2%. As identified by the target paper, while the IETF has become more representative over time, further efforts are needed to continue that path.

Affiliations. Figure 18 shows the top 10 affiliations from 2001 to 2024 by proportion of authors each year. In § 3.2.2 we explain how we use GPT to automate the normalization of affiliation taking into account variation is spelling, subsidiaries and merged companies. Our final result have small differences with the target paper. As information in the `datatracker` has changed over time even for older RFCs, we cannot directly compare metadata per RFC and find the specific reasons for disagreements. However, we compare the affiliation mapping file from the target paper with GTP results and we find that GPT groups together more names or abbreviations used by authors of the same affiliation, which

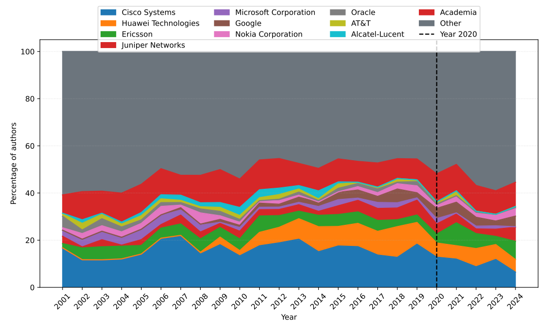

Figure 18: Authorship affiliations (normalized)

would also explain some differences in results. Nonetheless, we still find the same top 10 affiliations and the top 4 are the same organizations and in the same position.

Up until 2020, Cisco remained a consistent employer of RFC authors, as identified in the target paper. However, since then the number of authors from Cisco has declines and in 2024 it represented only about 6%. Similarly, authors with Huawei and Microsoft affiliation have continued to decline in the past few year. On the other hand, authors from Ericsson, Juniper, Google and Nokia have moderately increased in the past 4 years.

Overall, the combined share of the top 10 affiliations in 2024 is almost 35%, very close to the 2020 level. Similarly, the share of academic affiliations has remained fairly stable in the past years, between 10-12%.

Academia. Figure 19 shows the share of authors from the top 10 academic affiliations. As described in § 3.2.2, we use GTP to normalize names of authors affiliation. As a results, our counts by RFC and globally do not exactly match those of the target paper. The target paper revealed that academic institutions heavily involved in RFCs and Internet standard have changed since 2001, with less authors coming from MIT, Columbia, ISI/USC and UCL in the last decade and the rise of Tsinghua University. Since 2020 however, there are almost no author from Tsinghua University and many more authors from John Hopkins University (JHU). As a consequence, when comparing to the top 20 academic affiliations considering RFCs only until 2020, in 2024, JHU is now part of the 10 and UCL is no longer in the top 10. Figure 28 in appendix shows the share of authors per instution for the top 10 institution up until 2020.

New authors. Figure 20 depicts the percentage of authors each year that have not previously authored an RFC. The stable trend on about 30% new authors every year starting around 2017 and identified by the target paper continues to hold.

Summary. Studying authorship of RFCs up until December 2024 confirms the trends identified by the target paper that the authorship of RFCs has changed and diversified over time since 2001. However, in the past 4 years there appears to be some reversal of the diversifying trend, with an uptake of authors from North America.

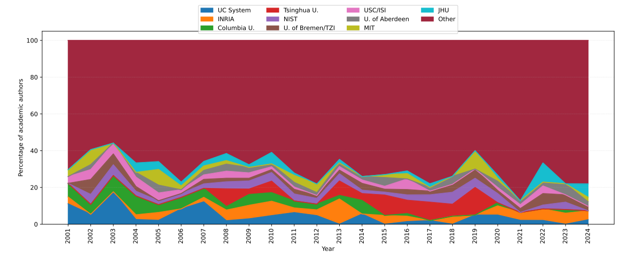

Figure 19: Authorship academic affiliations (normalized)

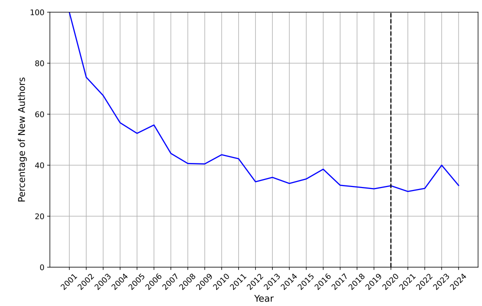

Figure 20: Percentage of new authors per year

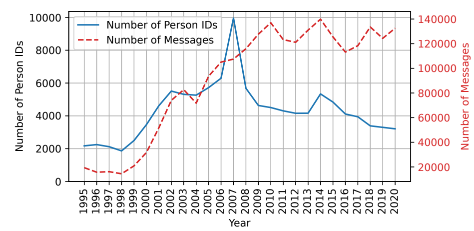

Figure 21: Number of Person IDs exchanging emails each year

4.3 Email interactions

Volume of email. Figure 21 shows the number of emails messaged sent since 1995.

Figure 22 depicts the relative email volume by type of email account. The volume of emails was in an upward trend up until 2020 and then had a significant decline in 2021, likely due to the pandemic. However, it has not come back to pre-pandemic levels.

Discussion of drafts. Figure 23 reveals the number of draft mentions in emails per year.

Contribution duration.

Interactions based on contribution duration. Figure 24 depicts the distribution of contribution duration of junior-most, senior-most and all other authors or RFCs.

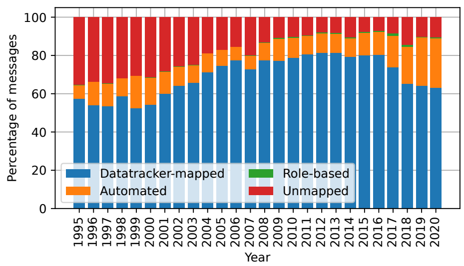

Figure 22: Number of emails exchanged per year category: datatracker-mapped contributors or role-based emails, automated email addresses and unmapped emails.

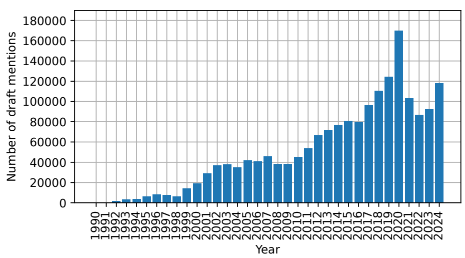

Figure 23: Number of draft mentions in each year found in the mail lists

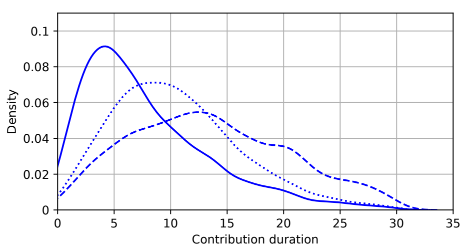

Figure 24: Contribution duration distribution of authors of RFCs by author seniority category in RFC authorship.

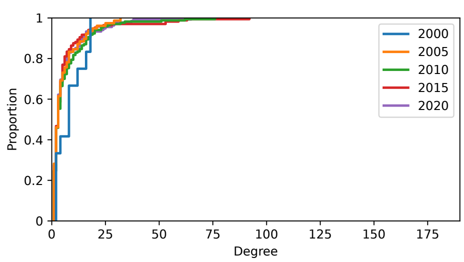

Figure 25: CDF of annual degree of interaction of RFC authors.

5 RFC SUCCESS MODELING

In the target paper, the authors leverage the dataset and metadata from the RFC corpus that they built to understand the factors that might impact the success of an RFC. They use the evaluation of RFC characteristics and deployment success of Nikkhah *et al.* [3] as the ground truth of successful deployment.

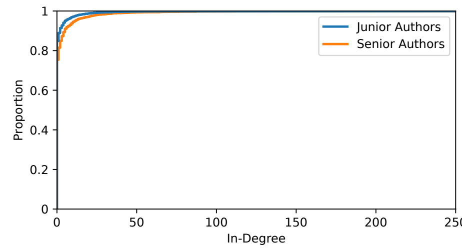

Figure 26: CDF of degree of email interactions between senior contributors to junior and senior contributors.

In [3], experts evaluate 251 Standard Track RFCs published between 1983 and 2011 to determine whether or not it had been successfully deployed and then assess different aspects of the RFC documents including the area, scope, type, changes to other RFCs and whether or not it relates to scalability, security, performance issues and whether or not it adds value to the protocol or involves some network effect, which they express in 20 features.[1] To identify RFC success factor, the target paper builds on the work of [3] by augmenting RFC's features using the metadata derived from RFCs, authors and email interaction datasets that we replicate in the previous sections.

In this section, we first replicate the model of the target paper. Then we explore the use of Large-Language-Models (LLMs) and generative Artificial Intelligence (genAI) to train models to perform some of the experts evaluation in [3] and study the evolution of RFC with those characteristics over time.

5.1 Reproducing the RFC success model

To reproduce the model in the target paper, we first reproduce the features from the RFC analysis using the data in our MongoDB database and then train classifications models to predict if an RFC was evaluated as successfully deployed in [3].

Reproducing features. The features based on RFCs, emails and authors interactions can only be computed for RFCs where the `datatracker` metadata is available. Of the 251 RFCs evaluated in [3], in line with the target paper we are able to compute all the features for 155 RFCs. We refer the reader to the target paper and the code for a full list of the 177 features that we reproduce.

Reproducing the model. To balance the number fo feature for the limited ground truth available, the target paper takes multiple feature reduction steps that we replicate: *(i)*

[1]We refer the reader to [3] for more details about those features and how the evaluation was done.

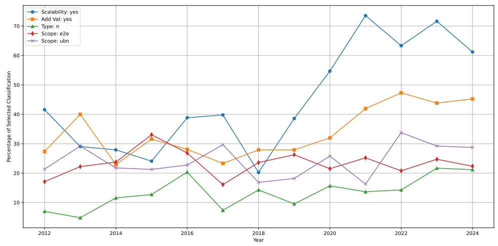

Figure 27: Trends in Significant Expert Features Over Time

we reduce the topic features (50) and interaction features (54) using the χ^2 test to select only the top 5 features in each group;*(ii)* we use the Variance Inflation Criterion to remove co-linearity by dropping all features with a value above 5; *(iii)* We apply forward Feature Selection (FS) to select features with high predictive value, by starting with the most predictive feature and on each iteration adding the feature that improves the AUC[2] score the most until there is no improvement.

Given the target paper files include the values of the features for the 155 RFCs with all feature, we reproduce the models using that data and then we replicate using the features we compute using our data. We also reproduce the model in [3] and we obtain the same results at the target paper, with the best model yielding an $F_1 = 0.761$ and $AUC = 0.650$.

5.2 Extending the RFC Success Model

5.2.1 AI Tagging Methodology. We design few-shot prompts for each input feature based on the original dataset provided by [3], and use the ChatGPT-4o model to perform the classification tasks. The prompts are constructed using the definitions and annotation guidelines specified by the original authors. As part of our evaluation strategy, we compare the model's classification outputs with the ground-truth labels from the original dataset, which consists of 251 RFCs.

5.2.2 RFCs with significant expert features over time. Firgure 27 shows the count of RFCs that have some of the features that the group of experts evaluated in the RFC and that our models and the model of the target paper find significant, such as if the protocol in the RFC impacts scalability, security, performance or if it add value. We note that more RFCs over time are showing those characteristics.

6 CONCLUSIONS

In this paper we replicate the work of [2] with the goal of building a pipeline that would allow us to frequently update the corpus with metadata and features about RFCs and RFC development process to enable further research and monitoring on Internet standard development.

[2]AUC: the area under the ROC curve.

Feature Name	Coef.	P>|Z|
Avg-author - senior (in messages)	-0.2373	0.203
Senior-author - senior (in* messages)	-0.1223	0.307
Topic 0	-3.3222	0.117
Topic 18	-0.9610	0.523
Topic 22	**28.3888**	**0.078**
Topic 30	41.7506	0.176
Topic 32	**21.4984**	**0.069**
Updates others (Yes)	0.2877	0.514
Obsoletes others (Yes)	**1.5315**	**0.000**
Keywords per page	**0.1404**	**0.023**
Page count	**0.0067**	**0.012**
Inbound ACM citations	0.0115	0.167
Inbound RFC citations, 2 years	**0.0508**	**0.005**
-00 draft mentions (normalised)	5.8503	0.482
Change to others (CO)	0.000	1.000
Adds value (AV)	**0.7828**	**0.009**
Security (SCRT)	0.3830	0.253
Scalability (SCAL)	**0.8755**	**0.100**
Performance (PERF)	0.5108	0.323
Has continent diversity (Yes)	-0.2007	0.528
Has affiliation diversity (No)	**0.8362**	**0.012**
Has an academic author (Yes)	0.000	1.000
Has a consultant author (Yes)	-1.3863	0.215
Has author in Europe (No)	**0.6190**	**0.008**
Has author in N. America (No)	0.000	1.000
Has author in Asia (Yes)	-0.7885	0.144
Has author from Cisco (Yes)	0.2877	0.594
Has author from Ericsson (Yes)	-0.6931	0.423
Has experienced author (No)	0.4700	0.410
Area (INT)	-0.1671	0.683
Area (OPS)	0.5108	0.323
Area (RTG)	0.4274	0.198
Area (SEC)	0.2231	0.638
Area (TSV)	0.5108	0.323
No incumbent	**0.6061**	**0.039**
Has incumbent	-0.2007	0.655
Scope, End-to-end (E2E)	**0.5878**	**0.035**
Scope, Local (L)	1.3863	0.215
Scope, Unbounded (UB)	**-1.0986**	**0.033**

Table 2: Logistic regression w/o feature selection. Statistically significant rows ($p \leq 0.1$) are highlighted.

A ETHICS

There are no ethical considerations in this paper. It uses only datasets that are already publicly available and does not identify any individual person in any of the analysis.

Feature Name	Coef.	P>|Z|
Scope, Unbounded (UB)	**-1.9101**	**0.003**
Has author in Asia (Yes)	**-1.4687**	**0.019**
Obsoletes others (Yes)	**2.2327**	**0.000**
Has a consultant author (Yes)	-1.8933	0.145
Has author from Ericsson (Yes)	-0.6570	0.449
Topic 0	**-4.9466**	**0.071**
Inbound ACM citations	0.0115	0.167

Table 3: Statistical analysis using logistic regression w/ feature selection. Statistically significant rows ($p \leq 0.1$) are highlighted.

Model	F_1	AUC	$F_{1_{macro}}$
Most frequent class	0.757	0.500	0.379
Baseline	0.753	0.611	0.597
Baseline + FS	0.761	0.655	0.601
Most frequent class	0.724	0.500	0.362
Baseline	0.687	0.557	0.554
Baseline + FS	0.731	0.638	0.591
Logistic regression all feats	0.685	0.669	0.662
Logistic regression all feats + FS	0.808	0.682	0.722
Decision tree all featsc + FS	0.802	0.721	0.716

Table 4: Classifier scores on the entire dataset (251 RFCs, above) and those with our features available (155 RFCs), with or without feature selection (FS).

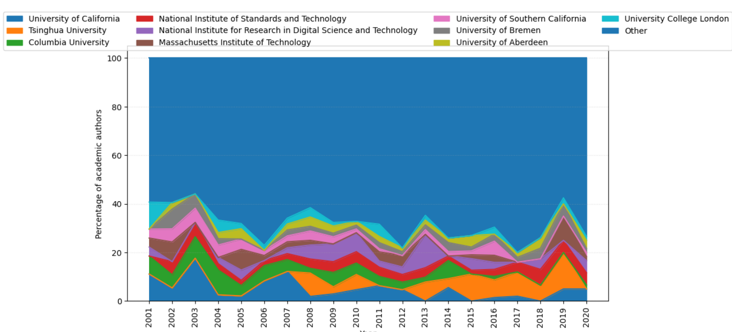

Figure 28: Authorship academic affiliations (normalized) using RFCs published up until December 2020.

B ADDITIONAL FIGURES

B.1 Academic Authorship until 2020

REFERENCES

[1] Scott O. Bradner. 1997. *Key words for use in RFCs to Indicate Requirement Levels*. Request for Comments RFC 2119. Internet Engineering Task Force. https://doi.org/10.17487/RFC2119 Num Pages: 3.

[2] Stephen McQuistin, Mladen Karan, Prashant Khare, Colin Perkins, Gareth Tyson, Matthew Purver, Patrick Healey, Waleed Iqbal, Junaid Qadir, and Ignacio Castro. 2021. Characterising the IETF through the lens of RFC deployment. In *Proceedings of the 21st ACM Internet Measurement Conference* (Virtual Event) *(IMC '21)*. Association for Computing Machinery, New York, NY, USA, 137–149. https://doi.org/10.1145/3487552.3487821

[3] Mehdi Nikkhah, Aman Mangal, Constantine Dovrolis, and Roch Guerin. 2017. A Statistical Exploration of Protocol Adoption. *IEEE/ACM Trans. Netw.* 25, 5 (Oct. 2017), 2858–2871. https://doi.org/10.1109/TNET.2017.2711642