Stream: Internet Architecture Board (IAB)
RFC: 9307
Category: Informational
Published: September 2022
ISSN: 2070-1721
Authors: N. ten Oever
University of Amsterdam
C. Cath
University of Cambridge
M. Kühlewind
Ericsson
C. S. Perkins
University of Glasgow

RFC 9307
Report from the IAB Workshop on Analyzing IETF Data (AID) 2021

Abstract

The "Show me the numbers: Workshop on Analyzing IETF Data (AID)" workshop was convened by the Internet Architecture Board (IAB) from November 29 to December 2, 2021 and hosted by the IN-SIGHT.it project at the University of Amsterdam; however, it was converted to an online-only event. The workshop was organized into two discussion parts with a hackathon activity in between. This report summarizes the workshop's discussion and identifies topics that warrant future work and consideration.

Note that this document is a report on the proceedings of the workshop. The views and positions documented in this report are those of the workshop participants and do not necessarily reflect IAB views and positions.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Architecture Board (IAB) and represents information that the IAB has deemed valuable to provide for permanent record. It represents the consensus of the Internet Architecture Board (IAB). Documents approved for publication by the IAB are not candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9307.

Copyright Notice

Table of Contents

Authors' Addresses

1. Introduction

The IETF, as an international Standards Developing Organization (SDO), hosts a diverse set of data about the IETF's history and development, current standardization activities, Internet protocols, and the institutions that comprise the IETF. A large portion of this data is publicly available, yet it is underutilized as a tool to inform the work in the IETF or the broader research community that is focused on topics like Internet governance and trends in information and communication technologies (ICT) standard setting.

The aim of the "IAB Workshop on Analyzing IETF Data (AID) 2021" workshop was to study how IETF data is currently used, to understand what insights can be drawn from that data, and to explore open questions around how that data may be further used in the future.

These questions can inform a research agenda drawing from IETF data that fosters further collaborative work among interested parties, ranging from academia and civil society to industry and IETF leadership.

2. Workshop Scope and Discussion

The workshop was organized with two all-group discussion slots at the beginning and the end of the workshop. In between, the workshop participants organized hackathon activities based on topics identified during the initial discussion and in submitted position papers. The following topic areas were identified and discussed.

2.1. Tools, Data, and Methods

The IETF holds a wide range of data sources. The main ones used are the mailinglist archives [Mail-Arch], RFCs [IETF-RFCs], and the datatracker [Datatracker]. The latter provides information on participants, authors, meeting proceedings, minutes, and more [Data-Overview]. Furthermore, there are statistics for the IETF websites [IETF-Statistics], the working group Github repositories, and the IETF survey data [Survey-Data]. There was discussion about the utility of download statistics for the RFCs themselves from different repos.

There is a wide range of tools to analyze this data produced by IETF participants or researchers interested in the work of the IETF. Two projects that presented their work at the workshop were BigBang [BigBang] and Sodestream's IETFdata [ietfdata] library. The RFC Prolog Database was described in a submitted paper; see [Prolog-Database]. These projects could provide additional insight into existing IETF statistics [ArkkoStats] and datatracker statistics [DatatrackerStats], e.g., gender-related information. Privacy issues and the implications of making such data publicly available were discussed as well.

The datatracker itself is a community tool that welcomes contributions; for example, for additions to the existing interfaces or the statistics page directly, see the Datatracker Database Overview [Data-Overview]. At the time of the workshop, instructions about how to set up a local development environment could be found at IAB AID Workshop Data Resources [DataResources]. Questions or discussion about the datatracker and possible enhancements can be sent to tools-discuss@ietf.org.

2.2. Observations on Affiliation and Industry Control

A large portion of the submitted position papers indicated interest in researching questions about industry control in the standardization process (as opposed to individual contributions in a personal capacity), where industry control covers both a) technical contributions and the ability to successfully standardize these contributions and b) competition on leadership roles. To assess these questions, investigating participant affiliations, including "indirect" affiliations (e.g., by tracking funding and changes in affiliation) was discussed. The need to model company characteristics or stakeholder groups was also discussed.

Discussion about the analysis of IETF data shows that affiliation dynamics are hard to capture due to the specifics of how the data is entered and because of larger social dynamics. On the side of IETF data capture, affiliation is an open text field that causes people to write their affiliation down in different ways (e.g., capitalization, space, word separation, etc). A common data format could contribute to analyses that compare SDO performance and behavior of actors inside and across standards bodies. To help with this, a draft data model was developed during the hackathon portion of the workshop; the data model can be found in Appendix A.

Furthermore, there is the issue of mergers, acquisitions, and subsidiary companies. There is no authoritative exogenous source of variation for affiliation changes, so hand-collected and curated data is used to analyze changes in affiliation over time. While this approach is imperfect, conclusions can be drawn from the data. For example, in the case of mergers or acquisition where a small organization joins a large organization, this results in a statistically significant increase in likelihood of an individual being put in a working group chair position (see the document by Baron and Kanevskaia [LEADERSHIP-POSITIONS]).

2.3. Community and Diversity

The workshop participants were highly interested in using existing data to better understand who the current IETF community is. They were also interested in the community's diversity and how to potentially increase it and thereby increase inclusivity, e.g., understanding if there are certain factors that "drive people away" and why. Inclusivity and transparency about the standardization process are generally important to keep the Internet and its development process viable. As commented during the workshop discussion, when measuring and evaluating different angles of diversity, it is also important to understand the actual goals that are intended when increasing diversity, e.g., in order to increase competence (mainly technical diversity from different companies and stakeholder groups) or relevance (also regional diversity and international footprint).

The discussion on community and diversity spanned from methods that draw from novel text mining, time series clustering, graph mining, and psycholinguistic approaches to understand the consensus mechanism to more speculative approaches about what it would take to build a feminist Internet. The discussion also covered the data needed to measure who is in the community and how diverse it is.

The discussion highlighted that part of the challenge is defining what diversity means and how to measure it, or as one participant highlighted, defining "who the average IETFer is". There was a question about what to do about missing data or non-participating or underrepresented communities, like women, individuals from the African continent, and network operators. In terms of how IETF data is structured, various researchers mentioned that it is hard to track conversations because mail threads split, merge, and change. The ICANN-at-large model came up as an example of how to involve relevant stakeholders in the IETF that are currently not present. Conversely, it is also interesting for outside communities (especially policy makers) to get a sense of who the IETF community is and keep them updated.

The human element of the community and diversity was highlighted. In order to understand the IETF community's diversity, it is important to talk to people (beyond text analysis). In order to ensure inclusivity, individual participants must make an effort to, as one participant recounted, tell them their participation is valuable.

2.4. Publications, Process, and Decision Making

A number of submissions focused on the RFC publication process, on the development of standards and other RFCs in the IETF, and on how the IETF makes decisions. This included work on technical decisions about the content of the standards, on procedural and process decisions, and on questions around how we can understand, model, and perhaps improve the standards process. Some of the work considered what makes an RFC successful, how RFCs are used and referenced, and what we can learn about the importance of a topic by studying the RFCs, Internet-Drafts, and email discussions.

There were three sets of questions to consider in this area. The first question related to the success and failure of standards and considered:

- What makes a successful and good RFC?
- What makes the process of making RFCs successful?
- How are RFCs used and referenced once published?

Discussion considered how to better understand the path from an Internet-Draft to an RFC, to see if there are specific factors that lead to successful development of an Internet-Draft into an RFC. Participants explored the extent to which this depends on the seniority and experience of the authors, on the topic and IETF area, on the extent and scope of mailing list discussion, and other factors, to understand whether success of an Internet-Draft can be predicted and whether interventions can be developed to increase the likelihood of success for work.

The second question focused on decision making.

- How does the IETF make design decisions?
- What are the bottlenecks in effective decision making?
- When is a decision made? And what is the decision?

Difficulties here lie in capturing decisions and the results of consensus calls early in the process, and understanding the factors that lead to effective decision making.

Finally, there were questions regarding what can be learned about protocols by studying IETF publications, processes, and decision making. For example:

- Are there insights to be gained around how security concerns are discussed and considered in the development of standards?
- Is it possible to verify correctness of protocols and detect ambiguities?
- What can be learned by extracting insights from implementations and activities on implementation efforts?

Answers to these questions will come from analysis of IETF emails, RFCs and Internet-Drafts, meeting minutes, recordings, Github data, and external data such as surveys, etc.

2.5. Environmental Sustainability

The final discussion session considered environmental sustainability. Topics included what the IETF's role with respect to climate change, both in terms of what is the environmental impact of the way the IETF develops standards and in terms of what is the environmental impact of the standards the IETF develops.

Discussion started by considering how sustainable IETF meetings are, focusing on the amount of carbon dioxide (CO2) emissions IETF meetings are responsible for and how can we make the IETF more sustainable. Analysis looked at the home locations of participants, meeting locations, and carbon footprint of air travel and remote attendance to estimate the CO2 costs of an IETF meeting. While the analysis is ongoing, initial results suggest that the costs of holding multiple in-person IETF meetings per year are likely unsustainable in terms of CO2 emission.

The extent to which climate impacts are considered during the development and standardization of Internet protocols was discussed. RFCs and Internet-Drafts of active working groups were reviewed for relevant keywords to highlight the extent to which climate change, energy efficiency, and related topics were considered in the design of Internet protocols. This review revealed the limited extent to which these topics have been considered. There is ongoing work to get a fuller picture by reviewing meeting minutes and mail archives as well, but initial results show only limited consideration of these important issues.

3. Hackathon Report

The middle two days of the workshop were organized as a hackathon. The aims of the hackathon were to 1) acquaint people with the different data sources and analysis methods, 2) seek to answer some of the questions that came up during presentations on the first day of the workshop, and 3) foster collaboration among researchers to grow a community of IETF data researchers.

At the end of Day 1, the plenary presentation day, people were invited to divide themselves into groups and select their own respective facilitators. All groups had their own work space and could use their own communication methods and channels. Furthermore, daily check-ins were organized during the two hackathon days. On the final day, the hackathon groups presented their work in a plenary session.

According to the co-chairs, the objectives of the hackathon have been met, and the output significantly exceeded expectations. It allowed more interaction than academic conferences and produced some actual research results by people who had not collaborated before the workshop.

Future workshops that choose to integrate a hackathon could consider asking participants to submit issues and questions beforehand (potentially as part of the position papers or the sign-up process) to facilitate the formation of groups.

4. Position Papers

4.1. Tools, Data, and Methods

Sebastian Benthall, "Using Complex Systems Analysis to Identify Organizational Interventions" [COMPLEX-SYSTEMS]

Stephen McQuistin and Colin Perkins, "The ietfdata Library" [ietfdata-Library]

Marc Petit-Huguenin, "The RFC Prolog Database" [Prolog-Database]

Jari Arkko, "Observations about IETF process measurements" [MEASURING-IETF-PROCESSES]

4.2. Observations on Affiliation and Industry Control

Justus Baron and Olia Kanevskaia, "Competition for Leadership Positions in Standards Development Organizations" [LEADERSHIP-POSITIONS]

Nick Doty, "Analyzing IETF Data: Changing affiliations" [ANALYZING-AFFILIATIONS]

Don Le, "Analysing IETF Data Position Paper" [ANALYSING-IETF]

Elizaveta Yachmeneva, "Research Proposal" [RESEARCH-PROPOSAL]

4.3. Community and Diversity

Priyanka Sinha, Michael Ackermann, Pabitra Mitra, Arvind Singh, and Amit Kumar Agrawal, "Characterizing the IETF through its consensus mechanisms" [CONSENSUS-MECHANISMS]

Mallory Knodel, "Would feminists have built a better internet?" [FEMINIST-INTERNET]

Wes Hardaker and Genevieve Bartlett, "Identifying temporal trends in IETF participation" [TEMPORAL-TRENDS]

Lars Eggert, "Who is the Average IETF Participant?" [AVERAGE-PARTICIPANT]

Emanuele Tarantino, Justus Baron, Bernhard Ganglmair, Nicola Persico, and Timothy Simcoe, "Representation is Not Sufficient for Selecting Gender Diversity" [GENDER-DIVERSITY]

4.4. Publications, Process, and Decision Making

Michael Welzl, Carsten Griwodz, and Safiqul Islam, "Understanding Internet Protocol Design Decisions" [DESIGN-DECISIONS]

Ignacio Castro et al., "Characterising the IETF through the lens of RFC deployment" [RFC-DEPLOYMENT]

Carsten Griwodz, Safiqul Islam, and Michael Welzl, "The Impact of Continuity" [CONTINUITY]

Paul Hoffman, "RFCs Change" [RFCs-CHANGE]

Xue Li, Sara Magliacane, and Paul Groth, "The Challenges of Cross-Document Coreference Resolution in Email" [CROSS-DOC-COREFERENCE]

Amelia Andersdotter, "Project in time series analysis: e-mailing lists" [E-MAILING-LISTS]

Mark McFadden, "A Position Paper by Mark McFadden" [POSITION-PAPER]

4.5. Environmental Sustainability

Christoph Becker, "Towards Environmental Sustainability with the IETF" [ENVIRONMENTAL]

Daniel Migault, "CO2eq: Estimating Meetings' Air Flight CO2 Equivalent Emissions: An Illustrative Example with IETF meetings" [CO2eq]

5. Informative References

[ANALYSING-IETF] Article 19, "Analysing IETF Position Paper", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Le.pdf>.

[ANALYZING-AFFILIATIONS] Doty, N., "Analyzing IETF Data: Changing affiliations", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Doty.pdf>.

[ArkkoStats] "Document Statistics", <https://www.arkko.com/tools/docstats.html>.

[AVERAGE-PARTICIPANT] Eggert, L., "Who is the Average IETF Participant?", November 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Eggert.pdf>.

[BigBang] BigBang, "Welcome to BigBang's documentation!", <https://bigbang-py.readthedocs.io/en/latest/>.

[CO2eq] Migault, D., "CO2eq: Estimating Meetings' Air Flight CO2 Equivalent Emissions: An Illustrative Example with IETF meeting", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Migault.pdf>.

[COMPLEX-SYSTEMS] Benthall, S., "Using Complex Systems Analysis to Identify Organizational Interventions", 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Benthall.pdf>.

[CONSENSUS-MECHANISMS] Sinha, P., Ackermann, M., Mitra, P., Singh, A., and A. Kumar Agrawal, "Characterizing the IETF through its consensus mechanisms", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Sinha.pdf>.

[CONTINUITY] Griwodz, C., Islam, S., and M. Welzl, "The Impact of Continuity", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Griwodz.pdf>.

[CROSS-DOC-COREFERENCE] Li, X., Magliacane, S., and P. Groth, "The Challenges of Cross-Document Coreference Resolution in Email", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Groth.pdf>.

[Data-Overview] "Datatracker Database Overview", for the IAB AID Workshop, <https://notes.ietf.org/iab-aid-datatracker-database-overview#>.

[DataResources] "IAB AID Workshop Data Resources", <https://notes.ietf.org/iab-aid-data-resources#>.

[Datatracker] IETF, "Datatracker", <https://datatracker.ietf.org/>.

[DatatrackerStats] IETF, "Statistics", <https://datatracker.ietf.org/stats/>.

[DESIGN-DECISIONS] Welzl, M., Griwodz, C., and S. Islam, "Understanding Internet Protocol Design Decisions", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Welzl.pdf>.

[E-MAILING-LISTS] Andersdotter, A., "Project in time series analysis: e-mailing lists", May 2018, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Andersdotter.pdf>.

[ENVIRONMENTAL] Becker, C., "Towards Environmental Sustainability with the IETF", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Becker.pdf>.

[FEMINIST-INTERNET] Knodel, M., "Would feminists have built a better internet?", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Knodel.pdf>.

[GENDER-DIVERSITY] Baron, J., Ganglmair, B., Persico, N., Simcoe, T., and E. Tarantino, "Representation is Not Sufficient for Selecting Gender Diversity", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Tarantino.pdf>.

[IETF-RFCs] IETF, "RFCs", <https://www.ietf.org/standards/rfcs/>.

[IETF-Statistics] IETF, "Web analytics", <https://www.ietf.org/policies/web-analytics/>.

[ietfdata] "IETF Data", Internet Protocols Laboratory, commit c53bf15, August 2022, <https://github.com/glasgow-ipl/ietfdata>.

[ietfdata-Library] McQuistin, S. and C. Perkins, "The ietfdata Library", <https://www.iab.org/wp-content/IAB-uploads/2021/11/McQuistin.pdf>.

[LEADERSHIP-POSITIONS] Baron, J. and O. Kanevskaia, "Competition for Leadership Positions in Standards Development Organizations", October 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Baron.pdf>.

[Mail-Arch] IETF, "Mail Archive", <https://mailarchive.ietf.org/arch/>.

[MEASURING-IETF-PROCESSES] Arkko, J., "Observations about IETF process measurements", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Arkko.pdf>.

[POSITION-PAPER] McFadden, M., "A Position Paper", <https://www.iab.org/wp-content/IAB-uploads/2021/11/McFadden.pdf>.

[Prolog-Database] Huguenin, P., "The RFC Prolog Database", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Petit-Huguenin.txt>.

[RESEARCH-PROPOSAL] Yachmeneva, E., "Research Proposal", <https://www.iab.org/wp-content/IAB-uploads/2021/11/Yachmeneva.pdf>.

[RFC-DEPLOYMENT] Castro, I., Healey, P., Iqbal, W., Karan, M., Khare, P., McQuistin, S., Perkins, C., Purver, M., Qadir, J., and G. Tyson, "Characterising the IETF through the lens of RFC deployment", November 2021, <https://dl.acm.org/doi/abs/10.1145/3487552.3487821>.

[RFCs-CHANGE] Hoffman, P., "RFCs Change", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Hoffman.pdf>.

[Survey-Data] IETF, "IETF Community Survey 2021", 11 August 2021, <https://www.ietf.org/blog/ietf-community-survey-2021/>.

[TEMPORAL-TRENDS] Hardaker, W. and G. Bartlett, "Identifying temporal trends in IETF participation", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/Hardaker.pdf>.

Appendix A. Data Taxonomy

```
A Draft Data Taxonomy for SDO Data:

Organization:
  Organization Subsidiary
  Time
  Email domain
  Website domain
  Size
          Revenue, annual
          Number of employees
  Org - Affiliation Category (Labels) ; 1 : N
    Association
    Advertising Company
    Chipmaker
    Content Distribution Network
    Content Providers
    Consulting
    Cloud Provider
    Cybersecurity
    Financial Institution
    Hardware vendor
    Internet Registry
    Infrastructure Company
    Networking Equipment Vendor
    Network Service Provider
    Regional Standards Body
    Regulatory Body
    Research and Development Institution
    Software Provider
    Testing and Certification
    Telecommunications Provider
    Satellite Operator

Org - Stakeholder Group : 1 - 1
    Academia
    Civil Society
    Private Sector -- including industry consortia and associations;
    state-owned and government-funded businesses
    Government
    Technical Community (IETF, ICANN, ETSI, 3GPP, oneM2M, etc)
    Intergovernmental organization

SDO:
  Membership Types (SDO)
  Members (Organizations for some, individuals for others...)
  Membership organization
    Regional SDO
      ARIB
      ATIS
      CCSA
      ETSI
      TSDSI
      TTA
      TTC
    Consortia

Country of Origin:
```

```
  Country Code

Number of Participants

Patents
  Organization
  Authors - 1 : N - Persons/Participants
  Time
  Region
  Patent Pool
  Standard Essential Patent
    If so, for which standard

Participant (An individual person)
  Name
  1: N - Emails
    Time start / time end

  1 : N : Affiliation
    Organization
    Position
          Time start / end

  1 : N : Affiliation - SDO
    Position
    SDO
    Time

  Email Domain (personal domain)

  (Contribution data is in other tables)

Document
  Status of Document
          Internet Draft
          Work Item
    Standard
  Author -
    Name
          Affiliation - Organization
    Person/Participant
        (Affiliation from Authors only?)

Data Source - Provenance for any data imported from an external data set

Meeting
  Time
  Place
  Agenda
  Registrations
    Name
    Email
    Affiliation
```

Appendix B. Program Committee

The workshop Program Committee members were Niels ten Oever (Chair, University of Amsterdam), Colin Perkins (Chair, IRTF, University of Glasgow), Corinne Cath (Chair, Oxford Internet Institute), Mirja Kühlewind (IAB, Ericsson), Zhenbin Li (IAB, Huawei), and Wes Hardaker (IAB, USC/ISI).

Appendix C. Workshop Participants

The Workshop Participants were Bernhard Ganglmair, Carsten Griwodz, Christoph Becker, Colin Perkins, Corinne Cath, Daniel Migault, Don Le, Effy Xue Li, Elizaveta Yachmeneva, Francois Ortolan, Greg Wood, Ignacio Castro, Jari Arkko, Justus Baron, Karen O'Donoghue, Lars Eggert, Mallory Knodel, Marc Petit-Huguenin, Mark McFadden, Michael Welzl, Mirja Kühlewind, Nick Doty, Niels ten Oever, Priyanka Sinha, Safiqul Islam, Sebastian Benthall, Stephen McQuistin, Wes Hardaker, and Zhenbin Li.

IAB Members at the Time of Approval

Internet Architecture Board members at the time this document was approved for publication were:

Jari Arkko

Deborah Brungard

Lars Eggert

Wes Hardaker

Cullen Jennings

Mallory Knodel

Mirja Kühlewind

Zhenbin Li

Tommy Pauly

David Schinazi

Russ White

Quin Wu

Jiankang Yao

Acknowledgments

The Program Committee wishes to extend its thanks to Cindy Morgan for logistics support and to Kate Pundyk for note-taking.

We would like to thank the Ford Foundation for their support that made participation of Corinne Cath, Kate Pundyk, and Mallory Knodel possible (grant number, 136179, 2020).

Efforts put in this workshop by Niels ten Oever were made possible through funding from the Dutch Research Council (NWO) through grant MVI.19.032 as part of the program 'Maatschappelijk Verantwoord Innoveren (MVI)'.

Efforts in the organization of this workshop by Colin Perkins were supported in part by the UK Engineering and Physical Sciences Research Council under grant EP/S036075/1.

Authors' Addresses

Niels ten Oever
University of Amsterdam
Email: mail@nielstenoever.net

Corinne Cath
University of Cambridge
Email: corinnecath@gmail.com

Mirja Kühlewind
Ericsson
Email: mirja.kuehlewind@ericsson.com

Colin Perkins
University of Glasgow
Email: csp@csperkins.org