

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 1 December 2022

N. ten Oever
University of Amsterdam
C. Cath

M. Kählerwind
Ericsson
C. S. Perkins
University of Glasgow
30 May 2022

Report from the IAB Workshop on Analyzing IETF Data (AID), 2021
draft-iab-aid-workshop-01

Abstract

The 'Show me the numbers: Workshop on Analyzing IETF Data (AID)' was convened by the Internet Architecture Board (IAB) from November 29 to December 2 and hosted by the IN-SIGHT.it project at the University of Amsterdam, however, converted to an online only event. The workshop was conducted based on two discussion parts and 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.

Discussion Venues

This note is to be removed before publishing as an RFC.

Source for this draft and an issue tracker can be found at <https://github.com/intarchboard/workshop-aid>.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 1 December 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
2. Workshop Scope and Discussion	3
2.1. Tools, data, and methods	3
2.2. Observations on affiliation and industry control	4
2.3. Community and diversity	5
2.4. Publications, process, and decision-making	6
2.5. Environmental Sustainability	7
3. Hackathon Report	7
4. Position Papers	8
4.1. Tools, data, and methods	8
4.2. Observations on affiliation and industry control	8
4.3. Community and diversity	9
4.4. Publications, process, and decision-making	9
4.5. Environmental Sustainability	10
5. Workshop participants	10
6. Program Committee	10
7. Acknowledgments	10
8. Annexes	11
8.1. Annex 1 - Data Taxonomy	11
Authors' Addresses	13

1. Introduction

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

The aim of the IAB Workshop on Analyzing IETF Data (AID) 2021 was to study how IETF data is currently used, understand what insights can be drawn from that data, and to explore open questions around how that data may be further used in 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 submitted position papers. The follow topic areas have been 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 (<https://mailarchive.ietf.org/arch/>), RFCs (<https://www.ietf.org/standards/rfcs/>), and the datatracker (<https://datatracker.ietf.org/>). The latter provides information on participants, authors, meeting proceedings, minutes and more (<https://notes.ietf.org/iab-aid-datatracker-database-overview#>). Furthermore there are statistics for the IETF websites (<https://www.ietf.org/policies/web-analytics/>), working group Github repositories, IETF survey data (<https://www.ietf.org/blog/ietf-community-survey-2021/>) and there was discussion about the utility of download statistics for the RFCs itself from different repos.

There are 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 (<https://bigbang-py.readthedocs.io/en/latest/>) and Sodestream's IETFdata (<https://github.com/glasgow-ipl/ietfdata>)

library; the RFC Prolog Database was described in a submitted paper (see Section 4 below). These projects could be used to add additional insights to the existing IETF statistics (<https://www.arkko.com/tools/docstats.html>) page and the datatracker statistics (<https://datatracker.ietf.org/stats/>), e.g., related to gender questions, however, privacy issues and implication of making such data publicly available were discussed as well.

The datatracker itself is a community tool that welcomes contributions, e.g. for additions to the existing interfaces or the statistics page directly (see <https://notes.ietf.org/iab-aid-datatracker-database-overview> (<https://notes.ietf.org/iab-aid-datatracker-database-overview>)). Instructions how to set up a local development environment can be found, at the time of the workshop, at <https://notes.ietf.org/iab-aid-data-resources> (<https://notes.ietf.org/iab-aid-data-resources>). Questions or any discussion can be issued 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 (vs. individual contributions in personal capacity), where industry control covers both, technical contribution and the ability to successfully standardize these contribution as well as competition on leadership roles. To assess these question it has ben discussed to investigate participant's affiliations including "indirect" affiliation e.g. by funding and changes in affiliation as well as the nessecarity to model company characteristics or stakeholder groups.

Discussions about the analysis of IETF data shows that affiliation dynamics are hard to capture, due to the specifics of how the data is entered but also because of larger social dynamics. On the side of IETF data capture, affiliation is an open text field, which causes people to write their affiliation down in different ways (capitilization, space, word seperation, etc). A common data format could contribute to analyses that compare SDO performance and behavior of actors inside and across standards bodies. To help this a draft data model has been developed during hackathon portion of the workshop which can found as Annex 1 - Data Taxonomy.

Furthermore, there is the issue of mergers and acquisitions and subsidiary companies. There is no authorotative 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

organizations 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 BaronKanevskaja (<https://www.iab.org/wp-content/uploads/2021/11/Baron.pdf>)).

2.3. Community and diversity

High interest from the workshop participants was also on using existing data to better understand who the current IETF community is, especially in terms of diversity, and how to potentially increase diversity and thereby inclusivity, e.g. understanding if there are certain groups or lists 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, how to measure it, or as one participant highlighted to define "who the average IETF is". The question was also raised 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 track conversations as 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. Vice versa, 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 stressed, in order to understand the IETF community's diversity it is important to talk to people (beyond text analysis) and 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 focussed 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 both technical decisions about the content of the standards, but also procedural and process decisions, and questions around how we can understand, model, and perhaps improve the standards process. Some of the work considered what makes a successful RFC, how are RFCs used and referenced, and what we can learn about importance of a topic by studying the RFCs, drafts, and email discussion.

There were three sets of questions to consider in this area. The first related to success and failure of standards, and considered what makes a successful/good RFC? What makes the process of RFC making successful? And 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 a 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, extent and scope of mailing list discussion, and other factors, to understand whether success of a draft can be predicted, and whether interventions can be developed to increase the likelihood of success for work.

The second question was around 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 around what can be learn 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/or detect ambiguities? What can be learnt by extracting insights from implementations and activities on implementation efforts?

Answers to these questions 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. It discussed what is the IETF's role with respect to climate change both in terms on 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 are IETF meetings, focussing on how much CO2 emissions are IETF meetings 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 carbon costs of an IETF meeting. Initial results suggest that the costs of holding multiple in-person IETF meetings per year are likely unsustainable in terms of carbon emission, although the analysis is ongoing.

Discussion also considered to what extent are climate impacts considered in the development and standardization of Internet protocols? It reviewed the text of RFCs and active working group drafts, looking for relevant keywords to highlight the extent to which climate change, energy efficiency, and related topics are considered in the design of Internet protocols, to show the limited extent to which these topics have been considered. Ongoing work is considering meeting minutes and mail archives, to get a fuller picture, 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, 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 in groups who selected their own respective facilitators. All groups had their own work space and could use their own communication methods and channels, or use IETF's gather.town. Furthermore, daily check-ins were organized during the two hackathon days. At the final day the hackathon groups presented their work in a plenary session.

The objectives of the hackathon, according to the co-chairs, have been met, and the output significantly exceeded expectations. It allowed for 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 to ask participants to submit groups, issues, and questions beforehand (potentially as part of the positions paper 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 (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Benthall.pdf>)

Stephen McQuistin, Colin Perkins The ietfdata Library (<https://www.iab.org/wp-content/IAB-uploads/2021/11/McQuistin.pdf>)

Marc Petit-Huguenin The RFC Prolog Database (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Petit-Huguenin.txt>)

Jari Arkko Observations about IETF process measurements (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Arkko.pdf>)

4.2. Observations on affiliation and industry control

Justus Baron, Olya Kanevskaia Competition for Leadership Positions in Standards Development Organizations (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Baron.pdf>)

Nick Doty Analyzing IETF Data: Changing affiliations (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Doty.pdf>)

Don Le Position Paper (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Le.pdf>)

Elizaveta Yachmeneva Research Proposal (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Yachmeneva.pdf>)

4.3. Community and diversity

Priyanka Sinha, Michael Ackermann, Pabitra Mitra, Arvind Singh, Amit Kumar Agrawal Characterizing the IETF through its consensus mechanisms (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Sinha.pdf>)

Mallory Knodel Would feminists have built a better internet? (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Knodel.pdf>)

Wes Hardaker, Genevieve Bartlett Identifying temporal trends in IETF participation (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Hardaker.pdf>)

Lars Eggert Who is the Average IETF Participant? (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Eggert.pdf>)

Emanuele Tarantino, Justus Baron, Bernhard Ganglmair, Nicola Persico, Timothy Simcoe Representation is Not Sufficient for Selecting Gender Diversity (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Tarantino.pdf>)

4.4. Publications, process, and decision-making

Michael Welzl, Carsten Griwodz, Safiqul Islam Understanding Internet Protocol Design Decisions (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Welzl.pdf>)

Ignacio Castro et al Characterising the IETF through the lens of RFC deployment (<https://dl.acm.org/doi/abs/10.1145/3487552.3487821>)

Carsten Griwodz, Safiqul Islam, Michael Welzl The Impact of Continuity (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Griwodz.pdf>)

Paul Hoffman RFCs Change (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Hoffman.pdf>)

Xue Li, Sara Magliacane, Paul Groth The Challenges of Cross-Document Coreference Resolution in Email (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Groth.pdf>)

Amelia Andersdotter Project in time series analysis: e-mailing lists (<https://www.iab.org/wp-content/IAB-uploads/2021/11/Andersdotter.pdf>)

Mark McFadden Position Paper (<https://www.iab.org/wp-content/IAB-uploads/2021/11/McFadden.pdf>)

4.5. Environmental Sustainability

Christoph Becker Towards Environmental Sustainability with the IETF
(<https://www.iab.org/wp-content/IAB-uploads/2021/11/Becker.pdf>)

Daniel Migault CO2eq: Estimating Meetings' Air Flight CO2 Equivalent Emissions: An Illustrative Example with IETF meetings
(<https://www.iab.org/wp-content/IAB-uploads/2021/11/Migault.pdf>)

5. Workshop participants

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 Kuehlewind, Nick Doty, Niels ten Oever, Priyanka Sinha, Safiqul Islam, Sebastian Benthall, Stephen McQuistin, Wes Hardaker, and Zhenbin Li.

6. 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 Kuehlewind (IAB, Ericsson), Zhenbin Li (IAB, Huawei), and Wes Hardaker (IAB, USC/ISI).

7. Acknowledgments

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

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

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

8. Annexes

8.1. Annex 1 - 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
- 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

Authors' Addresses

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

Corinne Cath
Email: corinnecath@gmail.com

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

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