ARTICLE

Open and Clarified Process of Compatibility Standards for Promoting Data Exchange

Yoshiaki Fukami[1,2]

Abstract

Standard specifications that realize mutual availability in data distribution are indispensable for cooperation between different fields. On the other hand, the forming standardization processes that allow many different things such as physical objects and services to be connected through the Internet, generate costs and require time to form consensus due to stakeholder diversification. To adapt to social evolution and use of big data generated by a massive amount of distributed data, establishing a method to develop a standard of data specification that involves a large number of diverse industries and stakeholders is necessary. The paper analyzes the evolution of the Standard Developing Organizations (SDOs) management policy for data-related technologies and discusses strategies for encouraging data transactions with rapid standardization processes and early diffusion.

Keywords Standardization · Interoperability · Implementation-oriented policy · GitHub

1 Background

1.1 Data

Standard specifications for realizing mutual availability in data distribution are indispensable for cross-disciplinary collaboration. To realize compatibility among distributed data resources, there must be a common syntax and vocabulary, as well as other specifications. Increase of data transaction is never realized without exchanges among diversified data providers that generate data according to common specifications. Standardization of data specification is therefore

✉ Yoshiaki Fukami
  yofukami@sfc.keio.ac.jp

1   Cyber Civilization Research Center, Keio University, 2-15-45 Mita, Minato, Tokyo 160-8582, Japan

2   Department of Management, Faculty of Economics, Gakushuin University, Toshima City, Japan

indispensable. To formulate standard specifications in a society where all things are connected, however, creates a concern that consensus-building will generate costs and require time due to the diversification of stakeholders, such as firms within diversified industries, data service providers with distributed data, and all data subjects.

To generate big data, such as learning data for artificial intelligence, it is necessary to integrate data provided by a wide variety of entities. Therefore, when data formats and vocabularies are created based on different specifications, difficulties in conversing data and integrated analysis are inevitable. To realize interoperability, standardization is needed. Technical standard specifications are an important element in the cooperation of an unspecified number of entities through data exchanges [1]. However, it is difficult, and reaching consensus among diversified stakeholders with conflicting interests is expensive.

AI and information technologies have advanced and are encouraged to develop novel services rapidly. The suitable specifications and requirement for specifications are also changing and increasing rapidly. To realize innovations with usage of distributed data resources, new standardized specifications for functions must be developed and diffused. However, standardization of technical specifications tends to be implemented by factors other than pure technical content (uncertainty, changes in the competitive environment, etc.) [2]. Moreover, once the specifications are fixed, it is very difficult to revise them because of the direct effect of network externalities [3, 4]. Designing a standardization process is a critical factor to realize compatibility among diversified data resources, making big data with distributed data providers, and increasing data exchange.

Standards bodies have evolved the processes to create specifications to build consensus among various stakeholders and at the same time that early dissemination is achieved. This paper unravels the evolution of the management policy around SDOs and examines governance for the realization of a data distribution society to design a technological platform for data exchange among stakeholders and industries with diversified cultures and conflicting interests.

1.2 The Confusion Over Standards

More applications based on the premise of data linkage among diversified devices and services, such as smart cities, autonomous cars, and other innovations, have been developed and implemented as proof of concept. Where data are exchanged across organizations, such as the spread of Internet of Things (IoT), the increased amount of sensor data that must be analyzed by Artificial Intelligence (AI) and digital transformation will be widely promoted, and the standardization of data will be promoted accordingly. Both the public and private sectors have advocated for this movement.

However, there is some confusion regarding the interpretation and definition of the standards. The term 'standards' can have different meanings and confusing them may complicate the discussion.

1.3 Role of Standards

Wiegmann et al. classified standards into committee-based, market-based, and government-based standards [1]. Cargill also described the committee-based standard as consensus-based [5]. David and Greenstein categorized standards as "sponsored" and "unsponsored," according to whether they were proprietary or public domain. They also used two other categories: standards agreements arrived from within and published by voluntary standards-writing organizations; and mandated standards, which are promulgated by governmental agencies that have some regulatory authority [6]. In recent years, these classifications have stopped being mutually exclusive, and standards are often formed through hybrid processes. In some cases, multiple specifications have been adopted as standards, according to the strategies of national governments, and particular specifications have diffused through market competition. In some cases, specifications that have been agreed upon by a private standardization body are later certified as de jure standards. The International Organization for Standardization (ISO), an organization that develops public standards, includes a fast-track rule that quickly adopts specifications developed by private bodies, such as the World Wide Web Consortium (W3C) and ECMA International, as public standards.

1.4 Evolution of the Standardizing Process

The more applications run on the Web, the more diversified industries engage in the standardization process. The increase in, and diversification of, participants causes delays in the standardization process [7–9]. For such complex systems, implementations and standardization tend to be determined more by political and organizational factors than by technological viewpoints [2].

To shorten the standardizing process and encourage the diffusion of standardized specifications, SDOs have tried to evolve their development processes. The Internet Engineering Task Force (IETF) adopted a "rough-consensus, running-code" policy, a decision-making policy in which consensus is confirmed, not by a rigid agreement formation process but by a gradual method, such as applause, in contrast to the bureaucratic and political approach of the ISO [10]. On the other hand, the W3C introduced an "implementation-oriented policy." Under this policy, no proposed specification is ever certified as a standard without more than two implementation cases. Therefore, proponents are encouraged to promote engineers outside of the W3C to implement and give feedback on the specifications during the very early stages of the process [11].

In recent years, standards have often been formed through hybrid processes. In some cases, multiple specifications have been adopted as standards on a policy basis, and as a result, particular specifications have become widespread through market competition and private consensus standards have been adopted as de jure standards. As I mentioned below, the ISO has a fast-track system that quickly adopts private standard specifications, such as W3C and ECMA International, as public standards.

1.5 The Ambiguity of "Openness"

De facto standards are established through market competition among private companies. Such standardization activities are based on the premise of enclosing technologies. Firms make profits by licensing intellectual property. Therefore, the specifications of de facto standards are managed as proprietary and in a closed manner. On the other hand, the de jure standard is compiled by the governments of each country along with public organizations, such as the ISO and is often a requirement for public sector procurement or customs clearance. However, the status of "public" does not mean "open." Printed ISO specifications are "sold," and the contents cannot be viewed without purchasing.

First, the word "open" is used in various contexts. The term "Open Source Software" (OSS) is often used. The Open Source Definition of an Open Source Initiative includes the following:

- Free redistribution;
- The program must include source code and must allow distribution in source code as well as in a compiled form;
- License must not be specific to a product;
- License must be technology neutral.[1]

The Open Source Definition includes not only the freedom to browse and use source code but also the freedom to develop source code. However, even for software and specifications defined as "open," the meaning of "open" may differ. In some OSSs, such as the Android operating system from Google, participation in the development process is restricted to engineers belonging to a specific company. To claim to be "open," it should be assumed that access to deliverables is not restricted. However, even if access to deliverables is open, various rules are established and operative for participation in the development and formulation process. In other words, there are many cases where "closed processes" are in place even for so-called "open standards." There are differences between source codes and software standards. However, this kind of open, free for everyone to participate, might also work for standards, especially compatible ones. Therefore, the hypothesis below can be depicted.

H1: Clarified definition of "open" and diversification in development "process" contributes to rapidity of standardization and meets diversified needs and fast diffusion.

---

[1] The Open Source Initiative (2007). The Open Source Definition version 1.9 http://www.opensource.jp/osd retrieved in December 17th, 2020.

2 Related Works

2.1 The Functions of Standards

Many firms, such as Amazon [12], have opened application programming interfaces (APIs) with their own business models [13], and constitute the API economy [14]. The API economy consists of data with different specifications according to each data provider. Few data providers focus on interoperability among data from different firms. Therefore, there have come to be data aggregators, brokers, and service providers to utilize data from different sources [15, 16]. There have also been attempts to realize data interoperability with standardization such as RDF, XML at the W3C and eXtensible Business Reporting Language (XBRL) [17].

There are various categorizations of standards, and each has different features. Researchers classify standards in two ways: quality/safety standards and interoperability/compatible standards [6, 18, 19]. This paper focuses on compatibility standards that work to realize interoperability. Compatibility standards are highly network external. Therefore, it is difficult to compete with technologically differentiated specifications based on de facto standards, because the direct network effect produces a lock-in effect [4, 20], and switching costs prevent users and complementary goods suppliers from adopting more effective or sophisticated specifications [21]. Excess inertia is locked into nonoptimal technology, such as the QWERTY keyboard [22]. Therefore, the first-mover strategy [23] is effective for compatible standard-setting. Proposers of standards for open systems tend to adopt a strategy of "priming" future expectations [24, 25]. Compatibility is realized only with agreement among stakeholders, including agreement on standards set by market competition. Therefore, standardization must involve a competition–cooperation interplay during multi-firm technology coordination [26].

2.2 User Participation

Perceived usefulness and perceived ease of use contribute to the acceptance of information technology [27, 28]. User participation in development contributes to satisfaction with information systems [29]. User–developer communication is also an important factor for the success of an information system [30]. System developers can recognize the needs of users and establish satisfactory progress through supplier–user [31] and developer–user [32] interactions. Innovation through these kinds of interactions can be referred to as forming innovation communities [33] with users. It is necessary to encourage developers outside of the SDO to engage in development processes.

2.3 Co-opetition

Computers and smart phones have been diffused globally. More and more devices are connected to the Internet. Using Web-based services, users can collaborate

through the Internet. The Internet and related information technologies have enabled distributed collaboration across geographical and organizational barriers. OSS development projects are common for distributed collaborations [34].

Most online collaborations are performed with common objects such as developing one certain software or module of systems. However, not all online collaborations proceed with common objects. Standardization is a typical case of collaboration with conflicting interests. Google, Apple, Microsoft, and IBM competed and cooperated to develop the HTML5 web standard at the W3C; this is a typical case of co-opetition [35, 36], which is not a new phenomenon. However, the advancement of information technologies has led to collaboration among more diversified stakeholders. Development projects, such as OSS or open standards, through collaboration among diversified stakeholders, are private-collective models [37].

There are an increasing number of collaborative developments with complex structures that should be analyzed. If collaboration procedures are open, a project's intellectual property can be publicly and freely available and, at the same time, governable [38]. To address the difficulties of open distributed development, new tools for collaboration have been developed. Communities of OSS developers are used to manage their projects with a newsgroup, Internet relay chat (IRC), and emails. However, such media do not have the necessary functions to facilitate various kinds of proposals, such as bug tracking and sharing change logs. Therefore, tools for efficient collaboration on projects have been invented and adopted. Issue tracker and Bugzilla are services that make lists of issues and bugs to be fixed and share procedures for solving them. Such tools have enabled communities to divide and assign tasks among participants. New collaborative tools provide log data suitable for qualitative analysis, such as statistical analysis, including network analysis. However, analyzing cooperation among stakeholders with conflicting interests by qualitative data analysis is inadequate, simply, because contribution cannot be assessed with the number of posts for mailing lists or lines of code. Data that enable analysis of the context of each action are needed.

Two changes solve these difficulties in analysis. One is the introduction of GitHub. GitHub is a version control system (VCS) used as a web service. The W3C has adopted GitHub, Bugzilla, and other open platforms for standardization processes. GitHub provides rich information on software/standard development projects, such as change logs, participants (who and how), and discussions about all proposals. VCS has enabled leaders of projects to more easily manage proposals of new functions and avoid conflicts in code. Web-based VCS enables the handling of complex collaborative projects among distributed organizations from distributed sites. GitHub also provides log data of who submits proposals, how proposals are evaluated, and procedures of merging codes. The second helpful change is the emergence of discussions over blogs and social network services. Firms more frequently post-press releases on their websites. Moreover, related individuals post their opinions and claims on blogs and SNS. Such articles and posts enable us to recognize the details of discussions more precisely.

Qualitative studies, especially case studies, are conducted using various sources. For example, Yin classified data resources for case study into six groups, as follows: (1) documentation, (2) archival records, (3) interviews, (4) direct observations,

(5) participatory observation, and (6) physical artifacts [39]. The entirety of online distributed collaboration activities cannot be observed directly by individuals. It is difficult to conduct interviews with key persons in dispersed locations. However, it is possible to access entire activities, attitudes, and evaluations for proposals of individual participants through the project repository of GitHub, blogs, and other social media. Moreover, such platforms provide data both for qualitative and quantitative analyses. This change enables us to adopt a multi-method research design that provides different perspectives on a particular phenomenon [40]. Rich data that implies a context makes interpretive studies possible.

There are still challenges in utilizing such emerging data. One challenge is finding a method to evaluate the accuracy and orthodoxy of distributed data. Individual participants post their opinions and propose their specifications without the authorization of their parent organizations, making it difficult to evaluate the importance and orthodoxy of each action. Lower credibility for the addressed dates of publications makes it difficult to form an accurate timeline of events. Occasionally, articles on the Web disappear. Although some Web archive services, such as the Internet Archive, exist, there is no common rule for handling archived Web pages. Therefore, it is necessary to establish a method to handle such emerging data in accordance with the various features of the Web.

2.4 Dilemma Between Interoperability and Innovation

The generation of big data by accumulating diverse data is expected to lead to innovation. Since the original meaning of innovation denotes a new combination [41], it can be expected that the combination of previously uncombined data resources will promote innovation.

On the other hand, interoperability or convertibility is necessary for generating machine-readable data resources for AI analysis. Common technological specifications—in other words, standards—play an important role in promoting data exchange for emerging applications.

However, standards also have the ability to prevent innovation [42], because they work by reducing the variety of goods [43]. Moreover, excess inertia causes standards to be locked into once-spread specifications [3]. Interoperability encourages data integration among diversified sources. At the same time, standards prevent data owners from changing their original or industry-specific specifications. Standards cause existing businesses to be more efficiently designed according to industry-specific rules. However, they also interfere with the creation of innovation through new cross-industry data transactions.

3 Scope of this Study, Analysis Target, and Method

In this study, with assumed use cases of exchanging various kinds of data among diversified industries, I analyze the process design and its influence on the standardization of data itself, as well as the protocol technology for data exchange.

Currently, when data are exchanged across organizations, it occurs mostly via the Internet. The World Wide Web occupies an important position as an interface with users. Interoperability among data are realized with common formats, vocabularies, protocols, metadata definitions and so on. Therefore, in this study, I adopt two cases: the IETF, which is responsible for formulating standard specifications for Internet technology, especially data formats such as JSON and protocols such as HTTP; and the W3C, which is responsible for formulating standard specifications, especially vocabularies such as HTML, RDF and DCAT, for Web technology and for inductive and qualitative analysis process analysis [39, 44].

A case study approach [39] is used, because there are few cases of standardization involving diversified participants. Since the phenomenon of interest is emerging and is as yet under-theorized, the inductive case study approach is suitable for our research [39, 44]. This inductive hypothesis-building study attempts to develop generalizable conclusions from a rare event.

As an intern, I conducted fieldwork at the W3C office in Japan from April 2010 to March 2013 and analyzed the flow of the standardization process as defined by the mailing list archives of the working groups (Table 1), meeting minutes, technical documents, and public relations materials. This analysis involved a study of internal documents and emails from the archives issued since the SDO was established. Furthermore, I conducted interviews with individuals from the W3C staff and member organizations, as well as with developers outside of W3C member organizations.

4 Case Study

4.1 IETF

4.1.1 Academic Mode of Standardization

The Internet originated from a network development project by academic researchers sponsored by the US Department of Defense. Currently, the IETF's "Technical Documents Certified as Standard" is referred to as the Request for Comment (RFC). The name RFC was adopted, instead of Standard, to emphasize that it was an informal text, as it was a Pentagon project at the time of the Cold War but escaped confidentiality [45].

In this case, to "escape confidentiality" means that the development of the Internet itself, rather than the Internet standard, was carried out through an open process. Academic researchers commissioned by the Department of Defense as "individuals" conducted a development process through peer reviews, information sharing, and discussions, following the rules of the academic community.

The "rough-consensus, running-code" policy has become widespread as an explicit policy as stakeholders spread to companies due to internationalization and the lifting of the ban on commercial use. Rough consensus is a method of deciding a policy by confirming a loose consensus (specifically, confirmed by applause) rather than by pursuing strict consensus building, and running code is a rule that requires that multiple operating and mutually available implementation examples be

Table 1 Mailing list archives analyzed in this paper

Sum (emails)	237,243
General management of W3C	
w3c-ac-members@w3.org	3932
w3c-ac-forum@w3.org	8754
tag@w3.org	9791
public-new-work@w3.org	483
chairs@w3.org	8717
public-w3process@w3.org	1844
HTML (3.2/4/5), XHTML (1/1.1/2.0)	
html-erb@w3.org	2553
www-html@w3.org	18,128
public-html@w3.org	44,850
w3c-html-wg@w3.org	14,359
w3c-html-cg@w3.org	7352
public-hypertext-cg@w3.org	179
public-xhtml2@w3.org	1852
w3c-xhtml-m12n-review@w3.org	142
Document object model	
www-dom@w3.org	8539
w3c-dom-wg@w3.org	3633
w3c-dom-ig@w3.org	9836
XForms	
www-forms@w3.org	6242
public-forms@w3.org	2574
public-forms-tf@w3.org	107
Device side APIs and related technologies/Web Intents	
member-appformats@w3.org	1123
public-webapps@w3.org	26,768
public-device-apis@w3.org	6997
public-webapi@w3.org	4663
public-web-intents@w3.org	1021
WHATWG	
whatwg@whatwg.org	42,308

presented in advance to promote the proposed specification to the standardization process [10].

A large-scale interoperability test event, called Interop, has come to be held regularly to promote the creation of implementation cases. Interop was first held in the United States in August 1986 and has been held regularly since. Such events have encouraged specification proposals and promoted standardization processes.

Working group discussions are facilitated by mailing lists and face-to-face meetings held three times a year. The venues are located around the United States, Europe, and other regions. This procedure may be evaluated as sufficiently open.

However, the cost of continuously participating in face-to-face meetings three times a year, making a wide network of contacts on the spot, familiarizing oneself with complicated rules, and then proposing specifications to adopt them as standards, is not reasonable for everyone.

In addition, as a network layer standard, device-to-device interoperability testing is built into the standardization process in the SDO. It requires developing a prototype for testing and bringing it to the test site. In reality, this hurdle is not insignificant; it is arguably as high as the application and layers above it.

4.1.2 Introduction of GitHub

GitHub is a cloud-based collaboration platform for software development that provides the function of a concurrent version system (CVS) and issue tracker, among other functions. The GitHub system was developed with Git, a version control system originally developed for use with Linux, and used by other open source software development communities.

There are several software programs that provide version management functions with Git as a source code. Among them, GitHub is one of the most popular and has been adopted by SDOs such as IETF, W3C, and Open Geospatial Consortium (OGC), as well as companies and government agencies for specification formulation and demo package creation and distribution. It is also used by individuals and firms to publish source code and platforms for collaborative development.

4.1.3 Changes in IETF

At the IETF, GitHub has come into use informally in discussions on specification development. Based on this situation, an official discussion on how to use GitHub in standardization is needed. The first draft[2] of RFC 8875, entitled "GitHub Configuration for IETF Working Groups" was posted on September 14, 2018. The proposal came to be known as RFC 8875 (Working Group GitHub Administration), on August 27, 2020.[3] The first draft[4] entitled "Using GitHub at the IETF" was posted on February 14, 2019. This proposal has come to be known as RFC 8874 (Working Group GitHub Usage Guidance), on August 27, 2020.[5]

Until then, IETF had been running a version management system for official technical documents called IETF Datatracker which was developed based on Subversion.

---

[2] Draft-cooper-wugh-github-wg-configuration-00—GitHub Configuration for IETF Working Groups https://datatracker.ietf.org/doc/draft-cooper-wugh-github-wg-configuration/00/ retrieved on December 17th, 2020.

[3] Working Group GitHub Administration RFC 8875 https://datatracker.ietf.org/doc/rfc8875/ retrieved on December 17th, 2020.

[4] Draft-thomson-git-using-github-00—Using GitHub at the IETF https://datatracker.ietf.org/doc/draft-thomson-git-using-github/ retrieved on December 17th, 2020.

[5] Working Group GitHub Usage Guidance RFC 8874 https://datatracker.ietf.org/doc/rfc8874/ retrieved on December 17th, 2020.

However, with the publication of the two RFCs, more discussion and editing of documents will be conducted on the GitHub repository than ever before.

Due to the global COVID-19 pandemic in 2020, the IETF has changed the form of its regular meetings from face-to-face to online. With the introduction of GitHub, it is becoming easier for engineers who are unfamiliar with IETF to participate in and contribute to standard development.

4.2 W3C

4.2.1 Establishment as an industrial consortium

Whereas the IETF did not distinguish between members and non-members, took the form of participation on an individual basis, and disclosed everything from the proposal specifications to the content of the discussion, the W3C defined itself as an "industrial consortium." The W3C has adopted a membership system for each group, such as a company or organization. The right to vote and propose specifications are granted only to membership organizations that pay the annual membership fee. The adoption of the paid membership system was aimed at limiting the number of stakeholders participating in the discussion, and promptly formulating and disseminating the specifications required by the industry. However, not all specifications developed by the W3C have come to be effective as standards. For example, in the 1990s, Microsoft and Netscape Communications implemented their own extended specifications in their products to differentiate themselves, resulting in a lack of mutual availability between browsers.

Extensible Hypertext Markup Language (XHTML), which was developed by the W3C as a successor to HTML4.01, has not been widely implemented in browsers, and was overtaken by HTML5, developed by the Web Hypertext Applications Technology Working Group (WHATWG), a grass-roots community of engineers working outside the W3C.

Not only is there fragmentation among members within the W3C, but there has been competition between the W3C and other standards bodies. JavaScript, which was originally proposed to the W3C, is instead being standardized by ECMA International.

The ability to formulate effective standard specifications quickly is an important factor that constitutes a competitive advantage among standardization bodies. As a result, the W3C has evolved to include more diversified stakeholders in the standardization process.

4.2.2 Internal institutionalization of standardization process

Just after the establishment of the W3C (1994–1995), the institutional design of the process was not clearly stated, and there was no working group for institutional design. The Process Editorial Review Board (ERB) was established to design the standardization process in 1996. However, the operating rules, including the method of selecting members, were not clearly stated, and which groups had authority was

not clearly established. The discussion was conducted through the mailing lists of the Advisory Committee, representatives of member organizations, and working group chairs. In other words, only some of the W3C management staff, one representative from each member organization, and the chair of each working group discussed standardization process management through multiple dispersed channels.

In January 2000, a dedicated mailing list (Process-issues@w3.org) was set up to discuss the standardization process. This allows members other than Advisory Committee representatives and working group chairs to participate in the discussion.

4.2.3 Incorporating Non-members into Standardization Process

Implementation-oriented policy, one of the most characteristic features in the W3C standardization process, was adopted in the November 11, 1999 version of the process document. As implementation-oriented policy was introduced, the W3C also began to introduce the implementation principle, which emphasizes the creation of implementation cases and feedback by external third parties. The W3C then began to implement measures to expand the stakeholders who participate in the specification process. The interest group (IG) in which non-members can participate was established in the July 18, 2003 version of the process document.

At the end of the twentieth century, XHTML and HTML5 were competing to be the successor to HTML4.01. XHTML, a new specification with superior in semantic processing by XML technology, advocated by IBM and director Tim Berners-Lee, was declared the successor in 1998. However, XHTML did not diffuse because of its lack of backward compatibility.
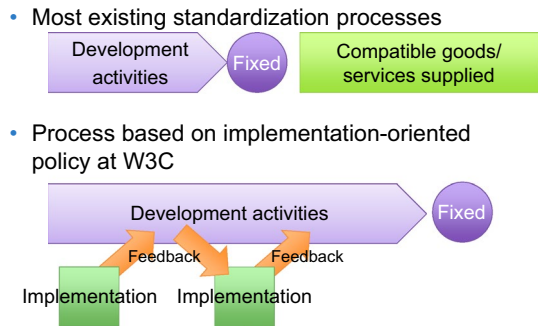
HTML5, which maintains backward compatibility and incorporates new features, was not initially accepted within W3C. Therefore, the specification was developed at the WHATWG, an external grass-roots engineers' community. It was supported by Apple, Google, and other browser vendors besides Microsoft. HTML5-compatible browsers were released, and their proposal had become more popular than XHTML among web content developers and end-users. In other words, the specifications formulated within the W3C did not become widespread and were replaced by specifications developed externally.

The W3C finally ended the development of XHTML and decided to adopt HTML5 as the successor to HTML4.01. At that time, the issue was how to handle the activities of the WHATWG, which is an external community, and which edited the drafts of the specification.

Eventually, the W3C decided to treat the entire WHATWG activity as a community group (CG) of the SDO. A CG is a new type of group introduced in W3C and is defined as an open forum for discussions in the pre-stage of the standardization process in the working group.

The competition between XHTML and HTML5 has resulted in non-members being more involved in the specification process than ever before. In addition, as the activities of WHATWG were merged into the standardization process of the W3C as they were, more collaboration tools developed outside of the W3C began to be used more in the standardization process of the W3C, where most tools had previously been developed by themselves.

Fig. 1 Implementation-oriented policy of the standardization process at the W3C

- Most existing standardization processes

Development activities → Fixed → Compatible goods/ services supplied

- Process based on implementation-oriented policy at W3C

Development activities → Fixed

Feedback Feedback

Implementation Implementation

4.2.4 Standardizing Process Update to Open More

The major revision items of the process document that define the specification development process in W3C are as follows.

- Introduction of implementation-oriented policy in the November 11, 1999 edition;
- Introduction of an Interest Group, where non-members can participate, in the April 17, 2003 edition;
- Revision of the W3C Process Interest Group, a "public" group to discuss the revision of the process document, was established at the Technical Plenary and Advisory Committee annual meeting in November 2011;
- The forum for process document revision discussions was moved to GitHub on April 21, 2017.

With the introduction of an implementation-oriented policy, implementation cases can be developed long before the completion of the standardization process, and rapid dissemination can be realized. At the same time, feedback from various external stakeholders responsible for implementation can be reflected in the standard specifications (Fig. 1).

Not only that, but the introduction of the new policy also opened up discussions on rule revisions, which are the basis of standardization bodies, such as the operating policy of the standard specification development process.

The decision to revise the process document in an interest group that non-members can participate in means increases diversity in the group of participants who review rules. In other words, specifications tend to meet the needs of diverse industries and stakeholders.

4.2.5 Open Collaborative Project for Testing Program Development

Newly introduced specifications never completely alleviate all bugs and problems. The more specifications there are, the more possibilities there will be of conflicts among specifications. Moreover, installations of new functions to standards

sometimes hinder interoperability among browsers, because not all browser vendors can efficiently implement the proposed specifications. Therefore, test programs are indispensable for software standardization.

Test programs that check whether there are bugs or conflicts with other standards are indispensable in developing software standards such as software programs. They are required not only to detect problems with specifications themselves but also to check whether web browsers are compatible. The W3C has made a test program development project open to the public.

The development of a testing program is an important procedure in standardization. The W3C team staff has developed testing programs for XHTML and earlier versions of specifications. On the other hand, HTML5 is a much larger specification because its role has been extended to that of a runtime environment for applications. This means it is impossible for the W3C team to develop test programs for the entire specification.

The solution was proposed by engineers of member companies. A collaborative development project of the HTML5 testing programs was proposed by Paul Irish from Google and Divya Manian from Adobe in 2011.[6] The project, known as "Move the Web Forward," was designed to allow developers outside of W3C to participate. The goal of the project was stated on the website, as follows:

Our goal is to make it easy for anyone to get started contributing to the platform, whether that's learning more about how it works, teaching others, or writing specs.[7]

In July 2012, the project held a hackathon focusing on developing a test program with Adobe's support.[8] Not only employees of Adobe, but also the W3C team, engineers from major browser vendors of Google, Mozilla, and Microsoft took part in the event, as lecturers or other staff members.[9,10] The hackathon included events called "Test the Web Forward" in China, France, Australia, Japan, and so on (Fig. 2).

The W3C began to sponsor the event from October 2012,[11] and the event has been one of the official activities of the W3C since October 2013.[12,13]

The test programs developed at these hackathons are stored and shared on GitHub. The W3C tends to develop tools, including teleconference systems, on their

[6] Irish, P. (2011, December 3). Moving the Web Forward. Paul Irish—I Make the Www Fun. Retrieved June 25, 2014, from http://www.paulirish.com/2011/moving-the-web-forward/.

[7] Adobe Systems (2012). Move the Web Forward. Retrieved June 25, 2014, from http://movetheweb forward.org/.

[8] Test the Web Forward http://rhauck.github.io/testtwf-2013/#/10.

[9] World Wide Web Consortium (2012). Speakers. *Test the Web Forward San Francisco.* Retrieved June 25, 2014, from http://testthewebforward.org/events/2012/sanfrancisco.html#speakers.

[10] World Wide Web Consortium (2012). Experts. *Test the Web Forward San Francisco.* Retrieved June 25, 2014, from http://testthewebforward.org/events/2012/sanfrancisco.html#experts.

[11] Berjon, R. (2012, October 1). Test the Web Forward. W3C Blog. Retrieved June 25, 2014, from http://www.w3.org/blog/2012/10/test-the-web-forward/.

[12] Langel, T. (2013, October 30). Welcoming Test the Web Forward to W3C. W3C Blog. Retrieved June 25, 2014, from http://www.w3.org/blog/2013/10/welcoming-testtwf-to-w3c/.

[13] Hauck, R. (2013, October 30). Test the Web Forward: Now Part of the W3C. Web Platform Team Blog. Retrieved June 25, 2014, from http://blogs.adobe.com/webplatform/2013/10/30/test-the-web-forward-now-part-of-the-w3c/.
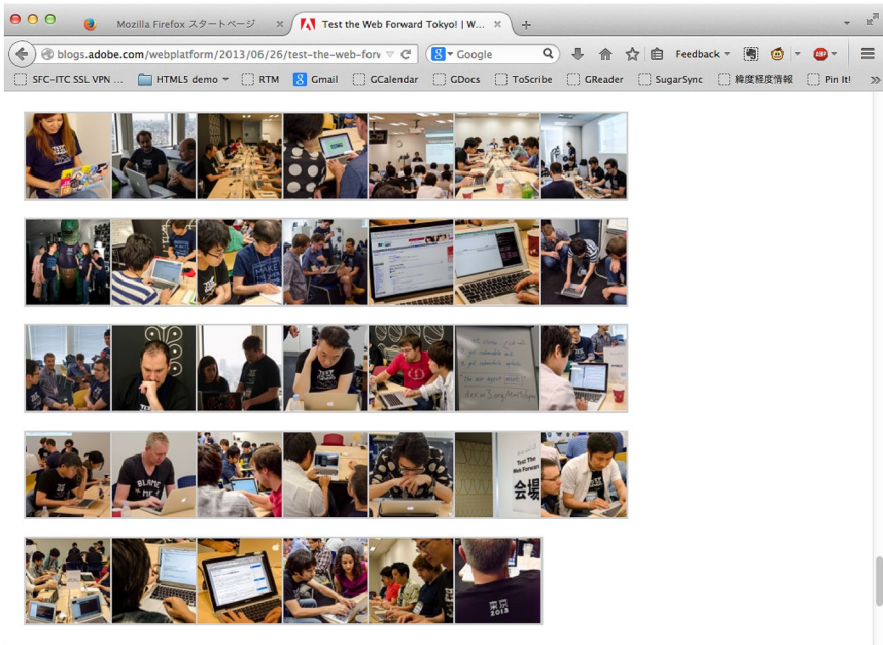
Fig. 2 Pictures from Test the Web Forward, Tokyo, 2013. Hauck, R. (2013, June 26). Test the Web Forward Tokyo! Web Platform Team Blog. Retrieved June 25, 2014, from http://blogs.adobe.com/webplatform/2013/06/26/test-the-web-forward-tokyo/

own. On the contrary, Move the Web Forward succeeds in escaping from the "not-invent-here" syndrome.

Codes developed through Test the Web Forward are stored and managed in the repository of w3c/web-platform-tests on GitHub[14] (Fig. 3). This repository also includes educational materials for test program development.[15]

The W3C had already utilized Bugzilla, a web-based bug tracker, in the process of specification development.[16] Extending areas, which are open to the public via the mailing lists of Bugzilla and GitHub—both of which are tools developed through OSS projects—increase the diversity of stakeholders in the standardization process. Test the Web Forward changes the role of developers outside of the W3C from users simply offering feedback to collaborators in test program development. Participation in developing activities works to promote emerging specifications.

Generally, there are high thresholds for joining OSS developers' communities [46]. Engineers need to achieve a very high level of domain knowledge and experience. Originally developed tools were used throughout the standardization

---

[14] Web-platform-tests GitHub. Retrieved June 25, 2014, from https://github.com/w3c/web-platform-tests

[15] Rebecca Hauck, R. (N.D.). Retrieved June 25, 2014, from http://adobe.github.io/web-platform/presentations/test-the-web-forward-tutorial/index.html#/

[16] W3C's Main Page of Bugzilla. Retrieved June 25, 2014, from https://www.w3.org/Bugs/Public/.
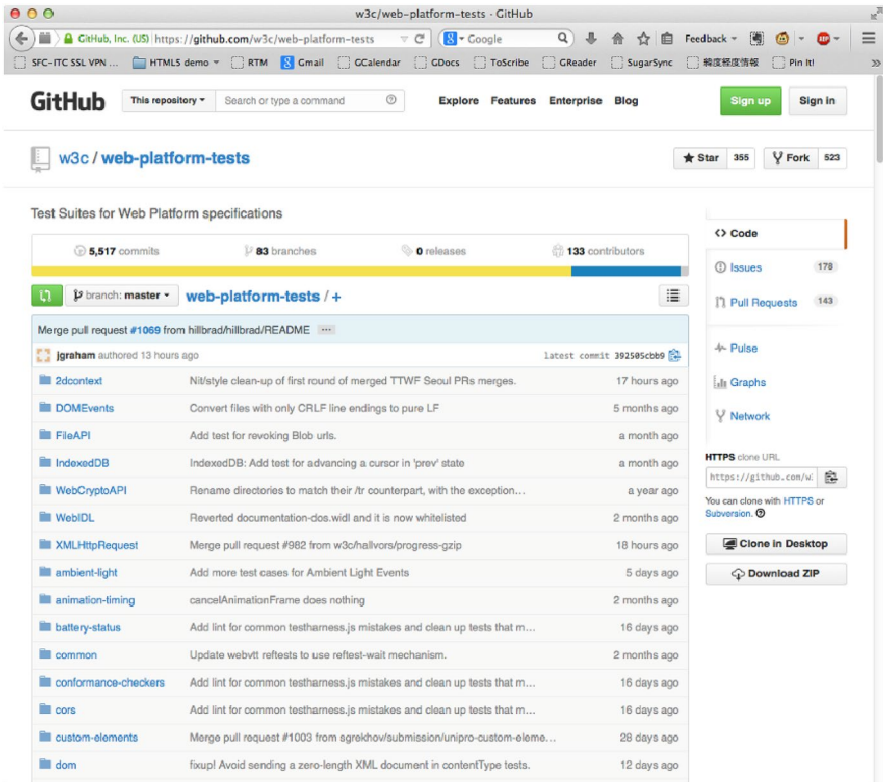
Fig. 3 Repository of w3c/web-platform-tests on GitHub.Web-platform-tests GitHub. Retrieved June 25, 2014, from https://github.com/w3c/web-platform-tests

process at the W3C. Although the standardization process is open, there are still obstacles for beginners to contribute, including the manipulation of tools and recognition of abbreviations. The W3C and other open SDOs face the same difficulties as OSS projects.

There are many cases where users take part in the development process of software through the use of beta versions to find bugs. On the other hand, the use of GitHub in the development process of HTML5 allows a collaboration among producers and users for the coding of test programs that had previously been done only by producers.

The open process encourages developers outside of organizations to contribute to increasing implementation cases and provide feedback. An increase in implementation cases and improvement in the specification leads to further diffusion.

More developers outside of organizations have come to take part in standard development activities. Testing the Web Forward is one of the reasons for this. In an interview, Philippe Le Hégaret, Interaction domain leader of the W3C, stated as follows:

But that was major change that we do, so if you compare specific on HTML4, we never tested HTML. We didn't have a testbed. But we didn't have any implementations, report for HTML4. Because that was really in the late, ah, 1990s. You know. Nowadays, we're trying to ship HTML5, and we need other test, which improves test for the feature that comes the differentiation from the HTML4, which we didn't other test at the top.[17]

Test the Web Forward is an expansion of collaborative activities among the W3C, member organizations, and external developers. Hackathon and other events increase interface contact among producers and users of the specifications. The collaborative development of test programs leads to an increased understanding of specifications among developers and encourages further diffusion.

4.2.6 Introducing GitHub in the standardization process.

Other than the Revising W3C Process Document Interest Group, certain working groups have been using GitHub for specification development work since before 2017. At the W3C and the IETF, collaboration tools such as IRC, Bugzilla, and the issue tracker have long been used in specification development. With the introduction of GitHub, a popular tool adopted by many collaborative software development projects, it is recognized as clarifying differences between specification versions and revising specifications according to feedback through pull requests. In addition, as collaboration methods come to be similar to many OSS projects, it can be expected to have effects such as reducing barriers to entry for engineers outside of the W3C.

Implementation-oriented processes encourage outside developers to take part in the standard development process. The introduction of tools frequently used in OSS development is one of the factors that reduce barriers to contributing.

Evans & Wolf [47] pointed out the principles of success in collaboration as deploying pervasive collaborative technology, keeping work simple, building communities of trust, thinking modularity, and encouraging teaming. The W3C has satisfied these requirements by implementing OSS-based CVS and bug-trackers and increasing developers' participation in collaboration. Such tools enable modularizing tasks, increasing the transparency of contribution, and the cooperation of distributed engineers.

Groups of the W3C conduct their activities with various tools for communication, such as Google Spreadsheet and CryptPad. Among these groups, the Revising W3C Process Community Group, which is in charge of updating the standardization process, uses GitHub and existing official tools. Their method could spread to all groups. Therefore, I analyzed the method of GitHub utilization and its influence in this CG.

Before the introduction of GitHub, discussions were held mainly through the mailing list, with the draft revised on HTML. In 2016, just before the introduction of GitHub, the maximum number of contributors per month was 16, which occurred in May.

---

[17] Interview with Philippe Le Hégaret on August 22, 2012.

Fig. 4 Issues labeled with color tags on repository of W3C Process Document Revision.Issues w3c/
w3process GitHub Retrieved December 25, 2020, from https://github.com/w3c/w3process/issues

The GitHub repository for Process Document revision was set up on April 21, 2017. As of December 18, 2020, 19 contributors had participated in the GitHub repository. Therefore, the number of contributors did not change much before and after the introduction of GitHub.[18]

With the introduction of GitHub, the status of updates, such as document revision proposals and issue submissions, has become clearer. Issues are labeled with color-coded tags that help identify relationships among them in the discussion as a whole (Fig. 4). In addition, the roles of W3C management staff and member companies have been clarified more than ever.

While there were three issues proposed using the issue tracker, which was originally used between November 2011 and March 2017, five issues were raised on GitHub between April 2017 and June 2018. In short, the number of issues discussed has increased due to the transition of the issue management tool from the issue tracker to GitHub. The introduction of GitHub has led to an increase in participants and revitalization of discussions.

5 Discussion

This study analyzes the expansion of developers' participation in standardization. With two cases analysis shows the hypothesis that Clarified definition of "open" and diversification in development "process" contributes to rapidity of standardization and meets diversified needs and fast diffusion is proved.

Many scholars have analyzed open source software development projects as examples of open collaborative innovation. Standardization activities tend to be considered as only building consensus among stakeholders with diverse interests. However, standardization activities among distributed and diversified industries and stakeholders have become more important for users to enjoy the results of IoT- and AI-based services with their massive amount of distributed data. Standardization management policy is necessary to ensure that diversified stakeholders can reach consensus rapidly.

The IETF and W3C are independent of governments and businesses. Therefore, their standards are not mandatory, and not all specifications developed there are diffuse. Moreover, they sometimes fail to coordinate members with different proposals and implementation cases.

Nonetheless, even though compatibility standards tend to be greatly influenced by network externalities, the IETF and the W3C have changed their standardization processes to be more open, encouraging wider and more diverse stakeholders to join in the process to diffuse specifications more and ensure effectiveness as a standard.

From the standpoint of developing an environment for massive and diversified data exchange, it is more important than ever to gather the support of more diverse stakeholders to develop common specifications for data such as common syntax, vocabulary, and other specifications. This is achieved through opening standardization processes and adopting popular collaboration tools such as GitHub. The meaning of "open" should not be limited to the free use of deliverables; the openness of process involvement is at least as important.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Wiegmann, P. M., de Vries, H. J., & Blind, K. (2017). Multi-mode standardisation: A critical review and a research agenda. *Research Policy, 46*(8), 1370–1386.
2. Tushman, M. L., & Rosenkopf, L. (1992). Organizational determinants of technological change: Toward a Sociology of technological evolution. *Research in Organizational Behavior, 14*, 311–347.
3. Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *The RAND Journal of Economics, 16*(1), 70–83.
4. Katz, M. L., & Shapiro, C. (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy, 94*(4), 822–841.
5. Cargill, C. F. (1989). *Information technology standardization: Theory, process, and organizations*. Digital Press.
6. David, P. A., & Greenstein, S. (1990). The economics of compatibility standards: An introduction to recent research. *Economics of Innovation and New Technology, 1*(1–2), 3–41.
7. Funk, J. L., & Methe, D. T. (2001). Market- and committee-based mechanisms in the creation and diffusion of global industry standards: The case of mobile communication. *Research Policy, 30*(4), 589–610.
8. Simcoe, T. (2006). Delay and de jure standardization: Exploring the slowdown in Internet standards development. In S. M. Greenstein & V. Stango (Eds.), *Standards and public policy* (pp. 260–295). Cambridge Univ Press.
9. Farrell, J., & Simcoe, T. (2012). Choosing the rules for formal standardization. *RAND Journal of Economics, 43*(2), 235–252.
10. Russell, A. L. (2006). 'Rough consensus and running code' and the internet-OSI Standards War. *IEEE Annals of the History of Computing, 28*(3), 48–61.
11. Fukami, Y. (2019). How communities of users appear in standardization processes like ghost. In *Proceeding of Standing Conference on Organisational Symbolism* 2019
12. Evans, P. C., & Basole, R. C. (2016). Economic and business dimensions: Revealing the API eco-system and enterprise strategy via visual analytics. *Communications of the ACM, 59*(2), 26–28.
13. Zachariadis M, Ozcan P (2017). The API Economy and Digital Transformation in Financial Ser-vices: The Case of Open Banking. SWIFT Institute Research Paper Series
14. Tan, W., Fan, Y., Ghoneim, A., Hossain, M. A., & Dustdar, S. (2016). From the service-oriented architecture to the web API economy. *IEEE Internet Computing, 20*(4), 64–68.
15. Immonen, A., Palviainen, M., & Ovaska, E. (2014). Requirements of an open data based business ecosystem. *IEEE Access, 2*, 88–103.
16. Janssen, M., & Zuiderwijk, A. (2014). Infomediary business models for connecting open data pro-viders and users. *Social Science Computer Review, 32*(5), 694–711.
17. Yoon, H., Zo, H., & Ciganek, A. P. (2011). Does XBRL adoption reduce information asymmetry? *Journal of Business Research, 64*(2), 157–163.
18. Grindley, P. (1995). *Standard strategy and policy: Cases and stories*. Oxford: Oxford University Press.
19. David, P. A., & Shurmer, M. (1996). Formal standards-setting for global telecommunications and information services: Towards an institutional regime transformation? *Telecommunications Policy, 20*(10), 789–815.
20. Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal, 99*(394), 116–131.
21. Klemperer, P. (2014). Markets with consumer switching costs. *The Quarterly Journal of Economics, 102*(2), 375–394.
22. David, P. A. (1985). Clio and the economy of QWERTY. *The American Economic Review, 75*(2), 332–337.

23 Lieberman, M. B., & Montgomery, D. B. (1988). First-mover advantages. *Strategic Management Journal, 9*(S1), 41–58.

24. Farrell, J., & Saloner, G. (1986). Installed base and compatibility: Innovation, product preannouncements, and predation. *The American Economic Review, 76*(5), 940–955.

25. Garud, R., Jain, S., & Phelps, C. (1998). Technological linkages & transience in network fields: New competitive realities. In P. Shrivastava, A. S. Huff, J. E. Dutton, J. P. Walsh, & A. S. Huff (Eds.), *Advances in strategic management: Organizational learning and strategic management Vol: 14* (pp. 205–237). Emerald Group Publishing Limited.

26. Ranganathan, R., Ghosh, A., & Rosenkopf, L. (2018). Competition-cooperation interplay during multifirm technology coordination: The effect of firm heterogeneity on conflict and consensus in a technology standards organization. *Strategic Management Journal, 39*(12), 3193–3221.

27. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

28. Gefen, D., & Keil, M. (1998). The impact of developer responsiveness on perceptions of usefulness and ease of use. *ACM SIGMIS Database, 29*(2), 35–49.

29. Ives, B., Olson, M. H., Science, S. M., & May, N. (1984). User involvement and MIS success: A review of research. *Management Science, 30*(5), 586–603.

30. McKeen, J. D., Guimaraes, T., & Wetherbe, J. C. (1994). The relationship between user participation and user satisfaction: An investigation of four contingency factors. *MIS Quarterly, 18*(4), 427.

31. Habermeier, K. F. (1990). Product use and product improvement. *Research Policy, 19*(3), 271–283.

32. Leonard-Barton, D., & Sinha, D. K. (1993). Developer-user interaction and user satisfaction in internal technology transfer. *Academy of Management Journal, 36*(5), 1125–1139.

33. Lynn, L. H., Reddy, N. M., & Aram, J. D. (1996). Linking technology and institutions: The innovation community framework. *Research Policy, 25*(1), 91–106.

34 Lakhani, K. R., & Panetta, J. A. (2007). The principles of distributed innovation. *Innov Technol Gov Glob, 2*(3), 97–112.

35. Brandenburger, A. M., & Nalebuff, B. J. (1996). *Co-opetition*. New York: Doubleday Business.

36. Casadesus-Masanell, R., & Yoffie, D. B. (2007). Wintel: Cooperation and conflict. *Management Science, 53*(4), 584–598.

37. von Hippel, E., & von Krogh, G. (2003). Innovation model: issues for organization science open source software and the "private-collective" innovation model: Issues for organization science. *Organization Science, 14*(2), 209–223.

38. O'Mahony, S. (2003). Guarding the commons: How community managed software projects protect their work. *Research Policy, 32*(7), 1179–1198.

39. Yin, R. K. (2014). *Case study research: Design and methods*. Sage Publications.

40. Becker, J., & Niehaves, B. (2007). Epistemological perspectives on IS research: A framework for analysing and systematizing epistemological assumptions. *Information Systems Journal, 17*(2), 197–214.

41. Schumpeter, J., Becker, M. C., & Knuden, T. (2002). New translations: Theorie der wirtschaftlichen Entwicklung. *American Journal of Economics and Sociology, 61*(2), 405–437.

42. Baldwin, C., & von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science, 22*(6), 1399–1417.

43. Farrell, J., & Saloner, G. (1986). Standardization and variety. *Economics Letters, 20*(1), 71–74.

44. Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review, 24*(4), 691–710.

45. O'Neil, J. (1991). Oral history interview with Stephen Crocker, Charles Babbage Institute, University of Minnesota. https://conservancy.umn.edu/handle/11299/107234. Accessed 18 Oct 2018.

46. Von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: A case study. *Research Policy, 32*(7), 1217–1241.

47. Evans, P., & Wolf, B. (2005). Collaboration rules. *Harvard Business Review, 83*(7), 96–104.