Information-Sharing Pipeline

Ilik, V.; Koster, L.

Routledge
Taylor & Francis Group

Information-Sharing Pipeline

Violeta Ilik and Lukas Koster

Presenters

ABSTRACT

In this article we discuss a proposal for creating an information-sharing pipeline/real-time information channel, where all stakeholders would be able to engage in exchange/verification of information about entities in real time. The entities in question include personal and organizational names as well as subject headings from different controlled vocabularies. Three World Wide Web Consortium–recommended protocols are considered as potential solutions: the Linked Data Notifications protocol, the ActivityPub protocol, and the WebSub protocol. We compare and explore the three protocols for the purpose of identifying the best way to create an information-sharing pipeline that would provide access to the most up-to-date information to all stakeholders.

Introduction

Our longstanding interest in this topic is due to the fact that there is no single place where one can find the most up-to-date information about creators, their institutions, and organizations with which they are affiliated. There are many creators, institutions, and organizations without authority records in the Library of Congress Name Authority File. On the other hand, information about those same creators, institutions, and organizations exists in other data stores such as discipline-specific databases, vendor or publisher databases, databases developed by non-profit organizations, and many more. In most cases these creators, institutions, and organizations have one or more identifiers assigned to them. We also know that information about creators, institutions, and organizations exists in institutional directory databases. In many cases these institutional databases feed into library discovery tools. However, institutional databases do not always synchronize all of this information with outside data stores held by publishers, vendors, non-profit organizations, or open source platforms. The information from all of these databases could be leveraged to support improved discoverability of relevant information about individual authors and institutions. Our article's primary focus is how to exchange and verify data in real time. Below we outline the characteristics of three World Wide Web Consortium (W3C) standards that could enable such an exchange.

Background

We discussed our original idea in a blog post on which we received feedback from experts in the infrastructure field. The current need to solve the identity management/authority control problem led us to initially name the system "Global Distribution System (GDS) system for authors' information exchange." As we discussed in the blog post, the system would be comprised of hubs where all stakeholders would engage in exchange/verification of information about authors/institutions/organizations.[1] We imagined it as a decentralized system that joins together various

software instances so that everyone would be able to see the activities in all of the hubs. Stakeholders in this case include individuals, libraries and other cultural heritage institutions, vendors, publishers, and identity providers, just to name a few. They have knowledge and experience in working with and/or developing various repositories (national or disciplinary), metadata aggregators, personal identifiers systems, publisher platforms, library vendors, and individually or otherwise hosted profiles. We further discussed how the proposed solution to the present challenge is a shared information pipeline where all stakeholders/agents would have access to up-to-date information and also contribute their own. Figure 1 shows a simplified version of the information-sharing pipeline.

As soon as someone updates a profile in a database, anyone interested in that profile could either pull the information or be notified through a subscription service that keeps subscribers informed to changes in a profile.

Interoperability and unique identification of researchers

Everyone wants to uniquely identify a researcher, starting with the researchers themselves, institutions, publishers, libraries, funding organizations, and identity management systems. The most interesting cases for researchers' name disambiguation are the ones with common names. In cases when we are dealing with last names of researchers of Chinese descent, for example, where the top fifty family names comprise 70% of China's population of over one billion, the complications that arise from name disambiguation are difficult.[2]

We need to be able to pull all the information about researchers from various databases and obtain the best possible aggregated data. Interoperability and reuse of data need to be addressed since we live in an age that has offered us unique opportunities. As Ilik noted, "these revolutionary opportunities now presented by digital technology and the semantic web liberate us from the analog based static world, re-conceptualize it, and transform it into a world of high dimensionality and fluidity. We now have a way of publishing structured data that can be interlinked and useful across databases."[3]
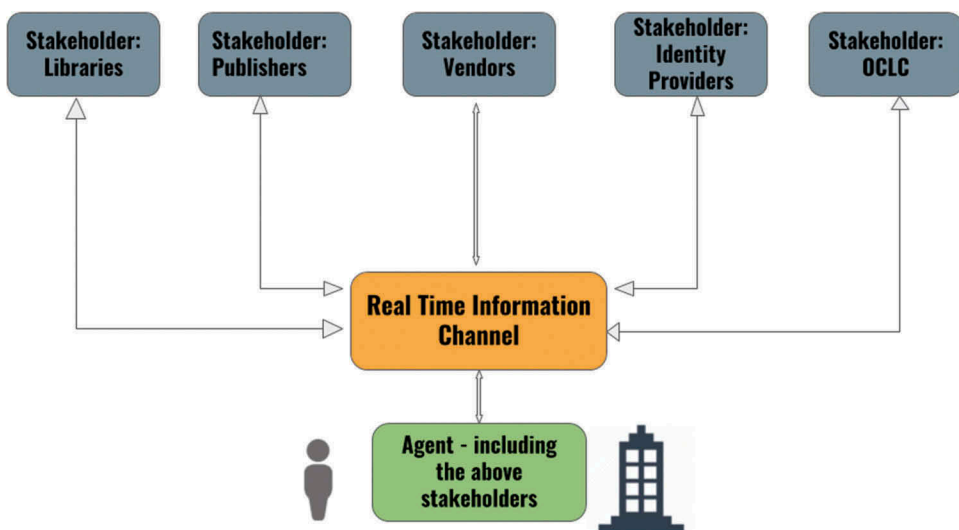

Figure 1. Information-sharing pipeline/real-time information channel.

The problem

Landscape

The OCLC Online Computer Library Center, Inc. (OCLC) Research Registering Researchers in Authority Files Task Group report looked into the landscape that represents various systems that capture data about researchers. According to the report, "rather than waiting until sufficient information is available to create a national authority heading, librarians could generate a stub with an ID that others could augment or match and merge with other metadata."[4]

The stakeholders currently do not utilize the rich information about persons and corporate bodies that is available from various data sources in a dynamic way. The bibliographic utilities and the vendor community, which includes Integrated Library Systems, automated authority control vendors, contract cataloging vendors and publishers, are not yet ready to offer a technical environment that enables us to control the name headings that come from linked data products and services.

Some of this has changed already now that the Authority toolkit, developed by Gary Strawn at Northwestern University to work with OCLC Connexion Client, allows us to semi-manually search outside sources and add information to the Name Authority Record. Some of those sources include Wikidata, Wikipedia, OpenVIVO, the Getty vocabularies, Virtual International Authority File, Medical Subject Heading, LC Linked Data Services, and GeoNames.[5] However, we still cannot execute the control headings from those sources, we just enhance the records with data from these sources.

The solution may not be to try and work within closed and constrained systems, but rather to completely change the way we currently think about and apply the W3C standards and protocols to decouple data from closed systems in a decentralized web.

A decentralized web

The notion of the decentralized web has been the focus of a number of initiatives, publications, and projects in recent years. Most notable of these are Tim Berners Lee's 2009 post "Socially Aware Cloud Storage,"[6] the Solid project,[7] Herbert Van de Sompel's and associates' distributed scholarly communication projects such as ResourceSync,[8] Open Annotation and Memento, Ruben Verborgh's work on Paradigm Shifts for the decentralized web,[9] Sarven Capadisli's dokie.li,[10] and Linked Data Notifications (LDN) protocol, a W3C recommendation. The concept of the decentralized web basically consists of four principles: *Separation or decoupling of data and applications, Control of and access to data by the owner, Applications as views on the data*, and *Exchange of information between data stores by means of notifications.*

Decoupling of data and applications

The current standard situation on the web is that sites store their own and their users' data in a data store that is only accessible for the site's own applications, or via Application Programming Interfaces (API) through the site's own services. These data stores function as data silos that are only accessible via services provided and managed by the data and system owners. This is the situation with the aforementioned categories of systems that hold creators' data.

Berners Lee states that there is a possible web architecture where applications can work on top of a layer of read–write data storage, serving as commodity (a good that can be exchanged), independent of the applications. This decoupling of data and applications "allows the user to control access to their data," "allows the data from various applications to be cross-linked," and "allows innovation in the market for applications."[11] Moreover, data are not lost when applications disappear and vice versa. Persistence of data and applications are independent of each other. In this architecture Uniform Resource Identifiers (URI) are used as names for users, groups, and documents, Hyper Text Transfer Protocol (HTTP) is used for data access, and application-independent single sign-on systems are used for authentication. Ruben Verborgh, who works with Berners Lee on the Solid

project, comes to a similar conclusion, but arrives there from the other side. He discusses three paradigm shifts to prepare for if we want to build web applications with a decentralized mindset: (1) End users become data owners, (2) Apps become views, (3) Interfaces become queries.[12]

Control of data and access

In a situation of decoupled data and applications, the service provider no longer controls the data itself nor access to the data. Instead the data owner can decide on the storage location and the reliability of the content and types of access granted to applications and other users. It improves privacy and control.[13] Credentials and access are managed only once, in the data store or data pod, not multiple times in each application. Besides that, trust, provenance, and verification are distributed and are no longer derived from a single actor.

dokie.li, a client-side editor for decentralised article publishing, annotations, and social interactions, is an example implementation of where the concepts of decentralization and interoperability meet toward an architecture where atomic data components are managed by their creators. A Solid-based server compliments tools like dokie.li by using the same set of open web standards.

In the context of creators' authority data, creators would be able to store, verify, and control their own data, coexisting with data pods controlled by other data providers, such as the current system categories described above. However, in practice the number of data pods containing overlapping, identical, or complementary creator data would probably decrease as applications indeed become mere views. Of course this will be a gradual development. For the remaining data stores a real-time information pipeline using notifications is essential.

Applications as views

The logical consequence of decoupling data and applications is that the data can and must be viewed, (re)used, and managed using any number of applications, of any type, including commercial, free, and open source. These applications then act as mere views on the data, because there is no direct and intrinsic relation between data and services in one system. The data service provider can also provide their own applications, as is the case with dokie.li and Solid, but other applications are not excluded. This allows innovation in the market for applications, as Berners Lee and Verborgh say. For this scenario to be reaized, the data must be FAIR (findable, accessible, interoperable, and reusable), and Open if applicable, to be decided by the data owner.

For the creators' authority data ecosystem this would mean that library catalogue systems would directly access the data stores of all the categories of creator authority data systems mentioned above and individual data pods managed by creators themselves.

Exchange of information via notifications

Now, the problem with applications acting as views on distributed data pods obviously lies in keeping track of where the required data reside and when the data are updated. Here the notifications come into play. Instead of a data pull mechanism whereby various applications access all possible data stores in a decentralized web, a data push mechanism is the optimal way of sharing data. All individual data stores will have to notify all interested client applications of any updates. Berners Lee mentions the need for a "publication/subscription" system for real-time notifications.[14]

Verborgh proposes a different perspective with his "interfaces become queries" paradigm shift.[15] In the current situation, data consuming applications send requests to a web API, a custom interface exposed by the data provider. This is not feasible in a decentralized web, where the many individual data pods will not have one standard interface. Instead, decentralized web applications should use declarative queries, which would be processed by client side libraries, translating these queries into concrete HTTP requests against one or multiple data pods. In a later presentation at ELAG

[European Library Automation Group] 2018 elaborating on his post, Verborgh mentions LDN as the way to notify interested parties of data updates.[16] He also predicts the transition from existing data aggregators to a network of caching nodes in the decentralized web.

For the creators' authority data situation this would mean a decentralized network of caching nodes or hubs exchanging and synchronizing information via a publish/subscribe notifications system. One of these is LDN, others are WebSub and ActivityPub, which we will discuss below.

At the same ELAG 2018 conference, a presentation entitled "Pushing SKOS" was given by Felix Ostrowski about distributing controlled vocabularies containing authority data in general, using a publication/subscription model with LDN and WebSub.[17] This idea shares similarities with the information-sharing pipeline for creators data proposed here.

Who holds the data we need?

Data we are interested in are diffused among all of these stakeholders: individuals, libraries and other cultural heritage institutions, vendors, publishers, and identity providers.

What is becoming more and more obvious is that the technologies to implement some changes in how we do business are already available, but the adoption and implementation is taking time. The resistance is based on the, in our opinion, fear of the new unknown. Experts see the benefit, but unfortunatelly the stakeholders are comfortable with the status quo and are not receptive to this big change. We still want to control our systems, re-entering the same data over and over again, when in fact what we should do is remember Verborgh's paradigm shift, and enter data only once.

In order to build an inclusive playground and work together to exchange/verify information about entities in real time, we need to break down the walls. Once we manage to do that, all stakeholders can engage in exchange/verification of information about entities.

Sarven Capadisli noted that the social paradigm shift is 25–50 years behind the technical shift and this has presented significant obstacles for the efforts to decentralize the web.[18]

Existing standards/protocols

We will look at similarities and differences between three W3C protocols, LDN, ActivityPub, and WebSub for the purpose of identifying which one fits with the use case of creating an information-sharing pipeline that will enable all stakeholders to have access to the most up-to-date information.

Linked Data Notifications

The design principles of LDN are that "data on the Web should not be locked in to particular systems or be only readable by the applications which created it. Users should be free to switch between applications and share data between them."[19] Applications generate notifications about activities, interactions, and new information, which may be presented to the user or processed further. LDN supports autonomy so any resource can have an Inbox anywhere; gives an identifiable unit that means that notifications have URIs; is reusable in the sense that a notification can contain any data and can use any vocabulary; and maintains a separation of concerns among sender, consumer, and receiver.[20] According to the W3C, "Linked Data Notifications is a protocol that describes how servers (receivers) can have messages pushed to them by applications (senders), as well as how other applications (consumers) may retrieve those messages. Any resource can advertise a receiving endpoint (Inbox) for the messages. Messages are expressed in Resource Description Format (RDF), and can contain any data, as noted on the W3C site and Capadisli et al. paper."[21,22]

This allows for more modular systems, which decouple data storage from the applications that display or otherwise make use of the data. As previously mentioned while describing the decentralized web and the Solid project, decoupling of data from the applications is an important and

necessary step toward a decentralized web. The protocol is intended to allow senders, receivers, and consumers of notifications, which are independently implemented and run on different technology stacks, to seamlessly work together, contributing to decentralization of our interactions on the web. This specification enables the notion of a notification as an individual entity with its own URI. As such, notifications can be retrieved and reused.

It is important to remember that the LDN protocol is a simple protocol for delivery only. It is used to publish notifications on the web to another user that the receiver has *not explicitly asked* for.[23]

There is no subscription option. No further specification of the payload is needed; it is a Pull approach for retrieval of notifications from an inbox. Senders and consumers discover a resource's Inbox Uniform Resource Locator (URL) through a relation in the HTTP Link header or body of the resource (see Figure 2). LDN completely decouples the three roles that an application (actor) can perform: sender, receiver, consumer. Those applications can be different. And as mentioned before, an application can have more than one role.
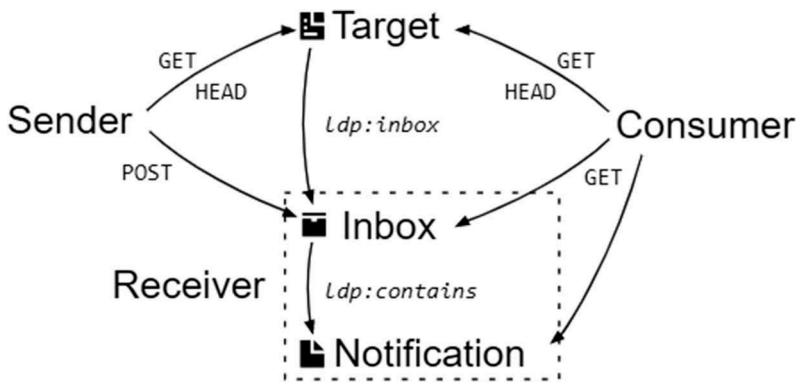

Figure 2. Overview of linked data notifications.

ActivityPub

The W3C describes the ActivityPub protocol as "a decentralized social networking protocol based upon the ActivityStreams 2.0 Terms. It provides a client to server API for creating, updating and deleting content, as well as a federated server to server API for delivering notifications and content."[24] W3C further defines the features of ActivityPub and describes the two layers it provides:

A server to server federation protocol (so decentralized websites can share information)

A client to server protocol (so users, including real-world users, bots, and other automated processes, can communicate with ActivityPub using their accounts on servers, from a phone or desktop or web application or whatever)

In ActivityPub, a user is represented by "actors" via the user's accounts on servers. The core Actor Types include: Application, Group, Organization, Person, and Service. Every Actor has:

An inbox: How they get messages from the world

An outbox: How they send messages to others

In the client server setting, notifications are always published to the sender's outbox. Actors wishing to receive a sender's notifications must send a request to that sender's outbox. In the federated server setting, notifications are published directly to the receiving actor's inbox. For the purpose of our case ActivityPub plays well with the idea of each system keeping their data intact but sharing the relevant information with other systems that relate to each actor type (people, corporate bodies). However, subscriptions are only available for federated servers. In ActivityPub the sender must be aware of the receiving federated server in a Followers list. On the other hand, non-federated server actors must be aware of all senders.

ActivityPub specializes in LDN as the mechanism for delivery of notifications by requiring that payloads are ActivityStreams2. *Inbox* endpoint discovery is the same. LDN receivers can understand requests from ActivityPub federated servers, but ActivityPub servers cannot necessarily understand requests from generic LDN senders. It is important to note that ActivityPub reuses LDN's Inbox mechanism. This means that tools are not only interoperable over sets of standards, but some of these standards themselves are designed to work with each other out of the box. Another important fact that positions LDN as a great solution for the problem we discuss is that any Linked Data Platform (LDP) implementation is an LDN Receiver out of the box. LDP "defines a set of rules for HTTP operations on web resources, some based on RDF, to provide an architecture for read-write Linked Data on the web."[25] Fedora Commons,[26] a flexible, modular, open source repository platform with native linked data support, is an example of a service that passes the LDP conformance test.[27]
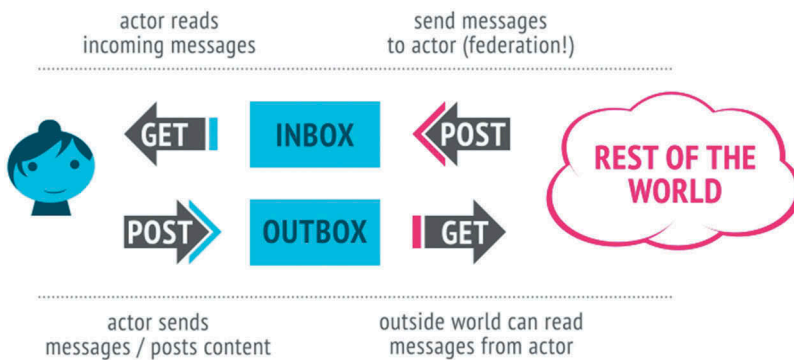
Figure 3 shows the simple flow of the data.

Figure 3. ActivityPub: Illustration of data flow.
https://www.w3.org/TR/activitypub/illustration/tutorial-2.png

WebSub

WebSub provides a common mechanism for communication between publishers of any kind of web content and their subscribers, based on HTTP web hooks. Subscription requests are relayed through hubs, which validate and verify the request. Hubs then distribute new and updated content to subscribers when it becomes available. The W3C document defines the important terms used in WebSub and they are described below.[28]

A subscriber is an entity (person or application) that wants to be notified of changes on a topic. The topic is the unit of content that is of interest to subscribers, identified by a resource URL. The topic is owned by a publisher. The publisher notifies one or more hubs, servers that implement both sending and receiving protocols. The hub notifies all subscribers that have a subscription to a specific topic. A subscription is a unique key consisting of the topic URL and the subscriber's callback URL.

In WebSub a subscriber does not have to be aware of all publishers that own the topics of interest. Also, a publisher does not have to be aware of all interested subscribers. Here the hub is a caching node in the decentralized web.

WebSub would provide an environment where each party could post its evolving version of a description on a channel to which all parties subscribe. Each party would be able to gather the information they need from that channel. In this environment there is no central/correct/unique version of the data. There are many versions that are informed by work being done in different institutions/applications that manage and use identity information. This is a real-time information channel fed by and consumed by institutions and applications that manage and use identity information.

The high-level protocol flow is shown in Figure 4, taken from the W3C site.

ResourceSync Framework Specification—Change Notification is based on WebSub. ResourceSync Change Notifications can be used to create/update/delete links when information about a new or updated description is sent via the URI of the description. The nature of the change (create, update, or delete) and the associated URI are sent through Change Notification Channels[29] as Change Notifications.[30] These notifications "are sent to inform Destinations about resource change events, specifically, when a Source's resource that is subject to synchronization is created, updated, or deleted" as described in the ResourceSync Framework Specification (ANSI/NISO Z39.99–2017).[31]

ResourceSync Change Notification specification describes an additional, push-based capability that a Source can support. It is aimed at reducing synchronization latency and entails a Source sending notifications to subscribing Destinations. The push-based notification capability is aimed at decreasing the synchronization latency between a Source and a Destination that is inherent in the pull-based capabilities defined in the ResourceSync core specification.

In order to implement the publish/subscribe paradigm, WebSub introduces a hub that acts as a conduit between Source and Destination. A hub can be operated by the Source itself or by a third party. It is uniquely identified by the hub URI. WebSub's topic corresponds with the notion of channel used in this specification. A topic is uniquely identified by its topic URI. Hence, per set of resources, the Source has a dedicated topic (and hence topic URI) for change notifications.
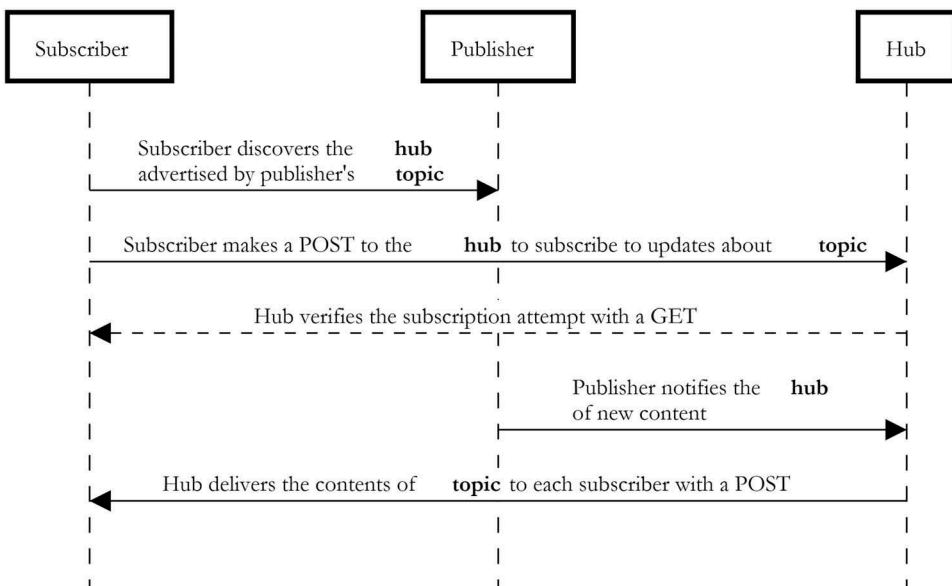

Figure 4. WebSub flow diagram.
https://www.w3.org/TR/websub/#x2-high-level-protocol-flow

WebSub protocol and ResourceSync Change Notification work well together. As shown in Figure 5 the Source submits notifications to Hub, the Destination subscribes to Hub to receive notifications, the Hub delivers notifications to Destination, and Destination unsubscribes from Hub. It is important to remember that WebSub has a higher bar to send/receive and it specializes in subscriptions.
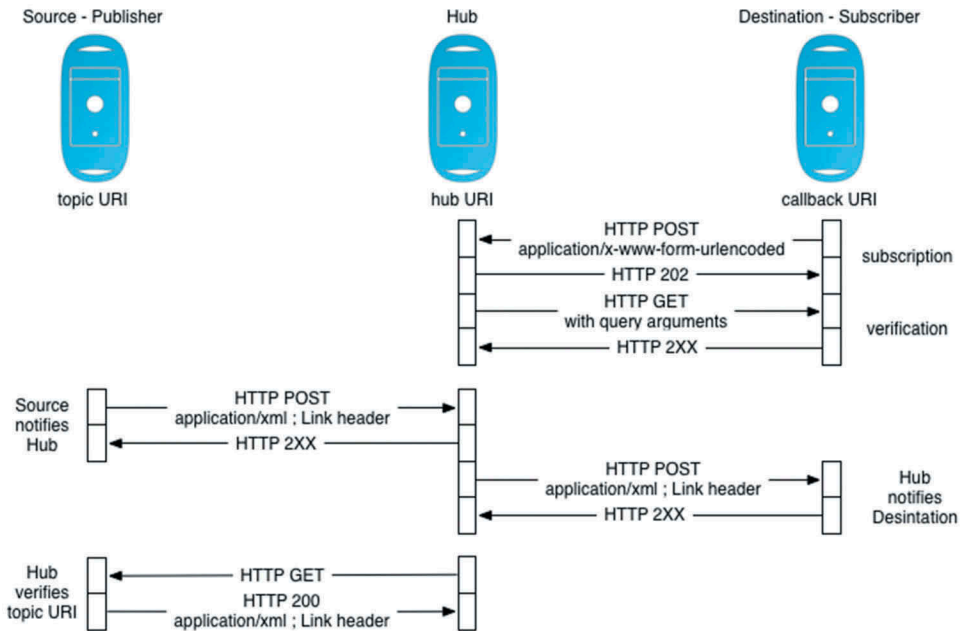

Figure 5. ResourceSync change notifications—WebSub as transport protocol HTTP interactions between Source, Hub, and Destination.

http://www.openarchives.org/rs/notification/1.0.1/notification

Conclusion and recommendation

All three protocols have the capacity to provide a technical solution to the problem of creators' authority data being unavailable to all stakeholders. As previously mentioned, the holdup is the simple yet hard social paradigm shift that lags behind the technological shift. Organizations that manage identity information (OCLC, ORCID, the Library of Congress, DBpedia, WikiData, to name a few) need to come together to deploy an information-sharing pipeline. Consumers of such data need to be involved, including higher education and library system vendors.

One obvious benefit would be that all stakeholders would be able to look up information about authors in one place, since the information that comes from various data sources would be synchronized. Next, even if the information about a person is stored on a personal server, a copy of the data can always be found in the Information-Sharing Hubs or in other data pods and would not be lost after the person for any reason ceases to maintain the personal server.

All organizations that manage identity information (Library of Congress, OCLC, ORCID, DBpedia, WikiData, libraries, museums, archives, library system vendors) should have a clear interest in deploying an information-sharing pipeline. The most important motivator for all of these organizations to agree on an information-sharing pipeline is that all of them would need to work with only one information exchange method, which saves everyone time and money, while enhancing information accuracy.

Acknowledgments

Notes

1. Violeta Ilik, "Real Time Information Channel," *Violeta's Blog*, October 19, 2017, https://ilikvioleta.blogspot.com/2017/10/real-time-information-channel.html (accessed June 11, 2018).

2. *Cultural Diversity: A Resource Booklet on Religious and Cultural Observance, Belief, Language and Naming Systems*, (London: HM Land Registry), archived from the original (PDF) on January 13, 2006, https://web.archive.org/web/20060113025139/http://www.diversity-whatworks.gov.uk:80/publications/pdf/hmlandregistryculturaldiversity.pdf (accessed January 4, 2019).

3. Violeta Ilik, "Cataloger Makeover: Creating Non-MARC Name Authorities," *Cataloging & Classification Quarterly* 53, no. 3–4 (2015): 382–98, doi:10.1080/01639374.2014.961626.

4. Karen Smith-Yoshimura et al., *Registering Researchers in Authority Files* (Dublin, Ohio: OCLC Research, 2014), 23, https://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf (accessed January 4, 2019).

5. Gary L. Strawn, "Authority Toolkit: Create and Modify Authority Records," http://files.library.northwestern.edu/public/oclc/documentation/ (accessed July 29, 2018).

6. Tim Berners-Lee, "Socially Aware Cloud Storage," August 17, 2009, https://www.w3.org/DesignIssues/CloudStorage.html (accessed July 29, 2018).

7. The Solid Project, "What is Solid?" 2017, https://solid.mit.edu/ (accessed July 29, 2018).

8. Open Archives Initiative, "Resourcesync Framework Specification – Table Of Contents," February 22, 2017, http://www.openarchives.org/rs/toc (accessed July 29, 2018).

9. Ruben Verborgh, "Paradigm Shifts For the Decentralized Web," December 20, 2017, https://ruben.verborgh.org/blog/2017/12/20/paradigm-shifts-for-the-decentralized-web/ (accessed July 29, 2018).

10. Sarven Capadisli, dokie.li, https://dokie.li/ (accessed July 29, 2018).

11. Berners-Lee, "Socially Aware Cloud Storage," August 17, 2009.

12. Verborgh, "Paradigm Shifts," December 20, 2017.

13. Ibid.

14. Berners-Lee, "Socially Aware Cloud Storage," August 17, 2009.

15. Verborgh, "Paradigm Shifts," December 20, 2017.

16. Ruben Verborgh, "The Delicate Dance of Decentralization and Aggregation," *International Conference of the European Library Automation Group (ELAG)*, June 5, 2018, http://slides.verborgh.org/ELAG-2018/#inbox (accessed July 29, 2018).

17. Felix Ostrowski and Adrian Pohl, "Pushing SKOS," *International Conference of the European Library Automation Group (ELAG)*, June 6, 2018, http://repozitar.techlib.cz/record/1241/files/idr-1241_1.pdf (accessed July 29, 2018).

18. Sarven Capadisli, "Enabling Accessible Knowledge," *CeDEM 2015, Open Access* (Danube University Krems, 2015), http://csarven.ca/presentations/enabling-accessible-knowledge/?full#trouble-in-paradigm-shifts (accessed July 29, 2018).

19. Sarven Capadisli, "Linked Data Notifications," *Scholastic Commentaries and Texts Archive (SCTA)* (Basel, June 6, 2018), http://csarven.ca/presentations/linked-data-notifications-scta/ (accessed July 29, 2018).

20. Ibid.

21. World Wide Web Consortium, "Linked Data Notifications," May 2, 2017, https://www.w3.org/TR/ldn/ (accessed July 29, 2018).

22. Sarven Capadisli et al., "Linked Data Notifications: A Resource-Centric Communication Protocol" (14th International Conference, Extended Semantic Web Conference (ESWC), Portorož, Slovenia, 2017), http://csarven.ca/linked-data-notifications (accessed July 28, 2018).

23. Ibid.

24. World Wide Web Consortium, "ActivityPub," January 23, 2018, https://www.w3.org/TR/activitypub/ (accessed July 29, 2018).

25. World Wide Web Consortium, "Linked Data Platform 1.0," February 26, 2015, https://www.w3.org/TR/ldp/ (accessed August 29, 2018).

26. DuraSpace, "Fedora," https://duraspace.org/fedora/ (accessed August 29, 2018).

27. W3C Working Group Note, "Linked Data Platform Implementation Conformance Report," December 2, 2014, https://dvcs.w3.org/hg/ldpwg/raw-file/default/tests/reports/ldp.html (accessed August 29, 2018).

28. World Wide Web Consortium, "WebSub: Definitions," January 23, 2018, https://www.w3.org/TR/websub/#definitions (accessed July 29, 2018).
29. Open Archives Initiative, "ResourceSync Framework Specification—Change Notification: Notification Channels," July 20, 2017, http://www.openarchives.org/rs/notification/1.0.1/notification#NotificationChannels (accessed July 29, 2018).
30. Open Archives Initiative, "ResourceSync Framework Specification—Change Notification," July 20, 2017, http://www.openarchives.org/rs/notification/1.0.1/notification#ChangeNoti (accessed July 29, 2018).
31. Open Archives Initiative, "ResourceSync Framework Specification (ANSI/NISO Z39.99-2017)," February 2, 2017, http://www.openarchives.org/rs/1.1/resourcesync (accessed July 29, 2018).

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Violeta Ilik, Head of Digital Collections and Preservation Systems, Columbia University Libraries.

Lukas Koster, Library Systems Coordinator, University of Amsterdam.