SHARE Notification Service Architectural Overview

The SHARE notification service will gather information about research release events through both a direct push protocol and a harvest strategy. The service will then notify consumers of these events through free subscriptions to predefined channels of notices and by allowing searches of its digest of research release events.

Research release events describe the release of publications, datasets, and other results of scholarly research. Sources that report these events include repositories, publishers, and academic profiles. Sources may also be consumers of event notifications, though consumers would also include research information systems, funding agencies, and our own SHARE registry.

The value of the SHARE notification service is that each source only needs to report new research release events once in order to reach a whole range of stakeholders quickly and efficiently. Similarly, those interested in staying abreast of research release events only need to subscribe to the notification service rather than keeping track of an ever-changing universe of sources.

The highest priority task of the service would be to transform reports of events into a consistent set of notifications for consumers. The service will devote little effort to cleaning up these event descriptions; it will focus instead on passing them along quickly and reliably. For example, there may be multiple events related to the same resource, yet the notification service would not de-duplicate these events or relate them to one another. Some event descriptions may include grant information and others may not; the service will not fill in these gaps. Event descriptions may refer to the same author in different ways, with slightly different forms of name or with standard identifiers; the service will not try to regularize these variations. The point of the notification service is gathering reported event descriptions and passing them along in a
predictable way, nothing more. This focus allows SHARE to build a service that raises awareness of new resources from sources to consumers as quickly as possible.

SHARE does understand that some processing of these events will be helpful to real-world use cases. Part of SHARE's long-term strategy is to perform this kind of processing at our registry layer, where we plan to build a registry of resources based on the stream of event notifications. This registry layer would be tasked with determining the relationships between various events and massaging the appropriate event data. This document will not consider the challenges of this registry work.

At each stage of SHARE’s architecture we invite consumers to make direct use of our services. However, it will be important for consumers to have a reasonable set of expectations about what they would be receiving. The notifications from our notification service will be quite raw, but very fast. This will be most helpful to consumers who are interested in a comprehensive idea of what is emerging in the research ecosystem and not too troubled by the fact that they may hear about the same thing more than once or receive a varied set of metadata with each notification. This may be useful, for example, to repository managers trying to get a more comprehensive idea of what may be available to collect or university sponsored research offices trying to learn what kind of research output is flowing from their institutions. It may be less useful to someone trying to follow a specific researcher’s work, unless they are prepared to absorb a fair degree of duplication and ambiguity with regard to the researcher’s name, for example.

**Event reports**

When a source becomes aware of a new publication, dataset, presentation, or other results of scholarly research, it generates a report of that research release event for the SHARE notification service. This report is, in essence, a bundle of metadata to be shipped to SHARE. In order to keep things as simple as possible for the source, SHARE does not presume this metadata will be available in any ongoing way at the source. In other words, SHARE does not require any state be preserved at the source.
When an event is reported to SHARE, two pieces of contextual information are implicit in the act of the report itself: the source of the report and the time of the report (which may be different than the source and creation time of the resource being described). This information is considered part of the report even though it is not explicitly packaged into the payload of the report.

The report payload consists of two metadata packages. Event metadata will use existing standards wherever possible, but must follow specifications developed for the notification service. Resource metadata may include any metadata about the resource being described that the source wishes to pass along, though SHARE will express preferences for certain kinds of metadata.

Details of the event metadata have yet to be determined, but it will, at a minimum, include an identifier for the resource being described by the event (a DOI, Handle, or URI) so that consumers of the notification service can return to the resource for more information. The event metadata will likely also include as appropriate: a bibliographic citation, type, and availability of the resource; the names, identifiers, and institutions of the creators of the resource; and sponsor and grant award information related to the resource.

The provision for resource metadata in the event report allows a source to include full Dublin Core or more detailed metadata so that fewer return trips are required by future consumers of event notifications. Such resource metadata won’t be required or specified, since SHARE wants to allow the community to determine what would be of value more organically.

**Event notifications**

The notifications provided by the service to consumers will be almost identical to the event reports it receives from sources. The only substantial change is that the implicit contextual information will become explicit report metadata in those notifications.

The report metadata will include the source and time of the event report. The event and resource metadata will remain the same as they were in the report.
Relationship to sources

In order to report research release events to the notification service, sources will have to register and authenticate with the service.

The notification service will develop a simple way for sources to push reports about individual research release events to the service. But knowing that some sources won’t have the wherewithal to add these push reports to their platforms, the notification service will also be capable of harvesting reports from sources. In the case of harvests, the service will use existing standards (OAI-PMH and ResourceSync are both being considered at this time) to harvest metadata from the source and will itself assemble event reports from that metadata.

At the time a source registers with the notification service it will specify which method of reporting it prefers and supply relevant connection information. Sources will authenticate with the notification service when sending reports.

Relationship to consumers

Consumers of the notification service will either be able to subscribe to certain pre-defined channels of event notifications or execute searches against the digest of notifications stored by the service.

Channels will be defined by the service with input from its consumers. For example, there may be channels based on the source of event reports, the institution related to an event, or the sponsor or grant award. Channels will provide notifications to subscribers via a push model; in other words, the consumer will get updates on a regular basis without making any further effort to retrieve the notifications. The number of channels will be limited due to the processing cost of the continual internal queries required to produce their feeds. Consumers will have to register with the service in order to subscribe to notification channels.

The precise push models for notification channels have not been determined. It is likely that the service will provide a simple subscription model. We are also considering making channels available via ResourceSync.

Searches against the digest of notifications will only query the report and event metadata related to an event. The results of searches will include the full event notification, though the query will not retrieve from the resource metadata. Searches will allow the consumer to define their own criteria for retrieval, beyond the bounds of the preset channels. Consumers will have to register with the service to save searches for repeated use, but otherwise will be allowed to search the notification service without registration.
An API will also be made available to allow notification channels and search queries to be used directly by other applications.

**Relationship to other layers of SHARE**

The notification service is only the first step along the road to SHARE’s long-term vision of a robust repository ecosystem for research output. The notification service provides essential raw material for a distributed registry layer that can accommodate not just publications but research data. This registry in turn feeds a discovery layer to help interested parties find research outputs across repositories. The notification service can also aid with the creation of a content aggregation layer that moves beyond curation and discovery to facilitate data and text mining of large corpora of content, as well as other community-driven value-added services.

As a direct consumer of SHARE notification service notifications, the SHARE registry will also serve as a test of the usefulness of these notifications.

**A note about shifting terminology**

Those who have been paying attention to SHARE over the past few months will notice that some of our terminology is shifting.

We want to be very clear that what SHARE is building in this instance is a service for others to build upon rather than a complete system that stands on its own. Though we have called this a “Notification System” in the past, we will refer to this as a “Notification Service” going forward.

In some documents we referred to “inbound” and “outbound” metadata and transport for the notification service. That language made no sense from any perspective other than inside SHARE, so we have shifted to the “reporting” and “notifying” terms present in this document to describe that distinction.