

Porting social media contributions with SIOC

Uldis Bojars, John G. Breslin, Stefan Decker

DERI, National University of Ireland, Galway, Ireland
firstname.lastname@deri.org

Abstract. Social media sites, including social networking sites, have captured the attention of millions of users as well as billions of dollars in investment and acquisition. To better enable a user's access to multiple sites, portability between social media sites is required in terms of both (1) the personal profiles and friend networks and (2) a user's content objects expressed on each site. This requires representation mechanisms to interconnect both people and objects on the Web in an interoperable, extensible way. The Semantic Web provides the required representation mechanisms for portability between social media sites: it links people and objects to record and represent the heterogeneous ties that bind each to the other. The FOAF (Friend-of-a-Friend) initiative provides a solution to the first requirement, and this paper discusses how the SIOC (Semantically-Interlinked Online Communities) project can address the latter. By using agreed-upon Semantic Web formats like FOAF and SIOC to describe people, content objects, and the connections that bind them together, social media sites can interoperate and provide portable data by appealing to some common semantics. In this paper, we will discuss the application of Semantic Web technology to enhance current social media sites with semantics and to address issues with portability between social media sites. It has been shown that social media sites can serve as rich data sources for SIOC-based applications such as the SIOC Browser, but in the other direction, we will now show how SIOC data can be used to represent and port the diverse social media contributions (SMCs) made by users on heterogeneous sites.

1 Introduction

“Social network portability” is the term used to describe the ability to reuse one's own profile across various social networking sites. The founder of the LiveJournal blogging community, Brad Fitzpatrick, wrote an article¹ from a developer's point of view about forming a “decentralized social graph”, which discusses some ideas for social network portability and aggregating one's friends across sites. However, it is not just friends that may need to be ported across social networking sites (and across social media sites in general), but content items as well.

Soon afterwards, “A Bill of Rights for Users of the Social Web²” was authored by Smarr et al. for “social web” sites who wish to guarantee ownership and control over

¹ <http://bradfitz.com/social-graph-problem/>

² <http://opensocialweb.org/2007/09/05/bill-of-rights/>

one's own personal information. As part of this bill, the authors asserted that participating sites should provide social network portability, but that they should also guarantee users "ownership of their own personal information, including the activity stream of content they create", and also stated that "sites supporting these rights shall allow their users to syndicate their own stream of activity outside the site".

OpenSocial from Google is another related effort that has gained a lot of attention recently. While at the time of writing, OpenSocial has been mainly focusing on application portability across various social networking site, the following statement³ mentions future reuse of data across participating sites: "an OpenSocial app added to your website automatically uses your site's data. However, it is possible to use data from another social network as well, should you prefer."

To enable a person's transition and / or migration across social media sites, there are significant challenges associated with achieving such portability both in terms of the person-to-person networks and the content objects expressed on each site. As well as requiring APIs to access this data (such as SPARQL endpoints or AtomPub interfaces), representation mechanisms are needed to represent and interconnect people and objects on the Web in an interoperable, extensible way.

The Semantic Web⁴ [1] provides such representation mechanisms: it links people and objects to record and represent the heterogeneous ties that bind us to each other. By using agreed-upon Semantic Web formats to describe people, content objects, and the connections that bind them together, social media sites can interoperate by appealing to common semantics. Developers are already using Semantic Web technologies to augment the ways in which they create, reuse, and link content on social media sites. Some social networking sites, such as Facebook, are also starting to provide query interfaces to their data, which others can then reuse and link to via the Semantic Web^{5, 6}.

The Semantic Web is a useful platform for linking and for performing operations on diverse person- and object-related data gathered from heterogeneous social media sites. In the other direction, social media sites can serve as rich data sources for Semantic Web applications. As Tim Berners-Lee said in a 2005 podcast⁷, Semantic Web technologies can support online communities even as "online communities ... support Semantic Web data by being the sources of people voluntarily connecting things together". Such semantically-linked data can provide an enhanced view of individual or community activity across social media sites (for example, "show me all the content that Alice has acted on in the past three months").

Social media sites should be able to collect a person's relevant content items and objects of interest and provide some limited data portability (at the very least, for their most highly used or rated items). We will refer to these items as one's social media contributions, or SMCs. Through such portability, the interactions and actions of a person with other users and objects (on systems they are already using) can be used to create new person or content associations when they register for a new social media site.

³ <http://code.google.com/apis/opensocial/container.html>

⁴ <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

⁵ <http://www.openlinksw.com/blog/~kidehen/?id=1237>

⁶ <http://www.dcs.shef.ac.uk/~mrowe/foafgenerator.html>

⁷ <http://esw.w3.org/topic/IswcPodcast>

In [2], it was shown that social media sites can serve as rich data sources for SIOC-based applications such as the SIOC Browser. In the other direction, we will demonstrate in this paper how SIOC data can be used to represent and port the diverse SMCs made by users on heterogeneous sites.

2 Getting content items using SIOC

The SIOC initiative [3] was initially established to describe and link discussion posts taking place on online community forums such as blogs, message boards, and mailing lists. As discussions begin to move beyond simple text-based conversations to include audio and video content, SIOC has evolved to describe not only conventional discussion platforms but also new Web-based communication and content-sharing mechanisms. In combination with the FOAF vocabulary for describing people and their friends, and the Simple Knowledge Organization Systems (SKOS) model for organising knowledge, SIOC lets developers link posted content items to other related items, to people (via their associated user accounts), and to topics (using specific “tags” or hierarchical categories).

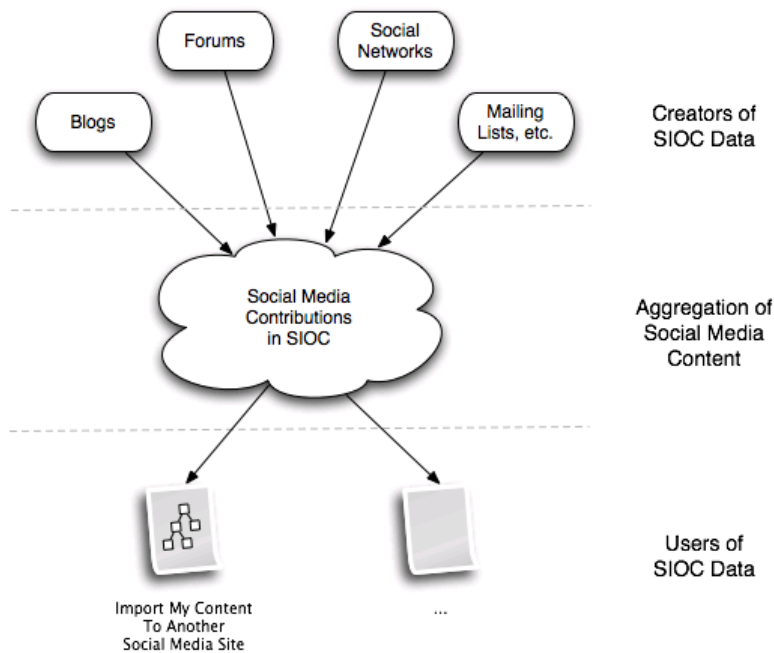


Fig. 1. Porting social media contributions from data providers to import services

Various tools, exporters and services have been created to expose SIOC data from existing online communities. These include APIs for PHP, Java and Ruby, data exporters systems like WordPress, Drupal and phpBB, data producers for RFC 4155

mailboxes and SIOC converters for Web 2.0 services like Twitter and Jaiku, and commercial products like Talis Engage and OpenLink Virtuoso. A full set of applications that create SIOC data is available online⁸.

All of these data sources provide accurate structured descriptions of social media contributions (SMCs), that can be aggregated from different sites (e.g., by person via their user accounts, by co-occurring topics, etc.). Figure 1 shows the process of porting SIOC data from various sources to SIOC import mechanisms for WordPress and future applications. We will now describe the SIOC import plugin for WordPress.

3 Import SIOC data, with a WordPress example

The SIOC import plugin for WordPress⁹ is an initial demonstrator for social media portability using SIOC. This plugin creates a screen (see Figure 2) in the WordPress administration user interface which allows one to import user-created content in the form of SIOC data.

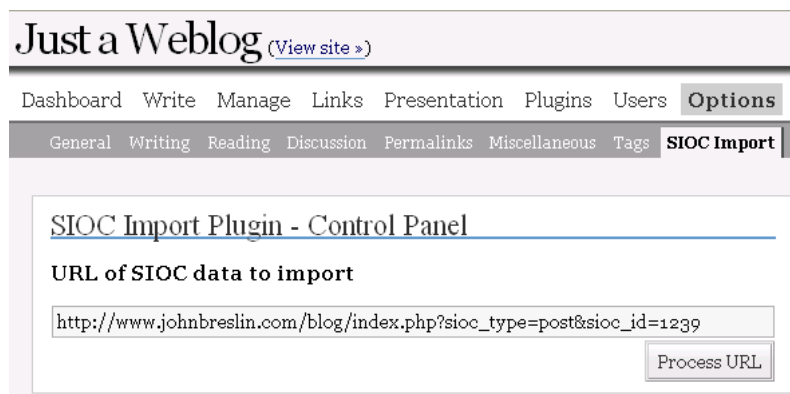


Fig. 2. SIOC RDF import into WordPress

Data to be imported can be created from a number of different social media sites using SIOC export tools (as described above). For example, a SIOC exporter plugin for a blog engine would create a SIOC RDF representation of every blog post and comment, including information about:

- The content of a post
- The author
- The creation / update date
- Tags and categories
- All comments on the post
- Information about the container blog

⁸ <http://rdfs.org/sioc/applications/#creating>

⁹ http://wiki.sioc-project.org/w/SIOC_Import_Plugin

The data representation used (RDF) enables us to easily extend this data model with new properties when they become necessary. The import process implemented by the WordPress SIOC import plugin is the following:

- Parse RDF data (using a generic RDF parser called ARC)
- Find all posts - `sioc:Post(s)` - which exhibit all of the properties required by the target site
- For each post found:
 - Create a new post using WordPress API calls

The pilot implementation currently works with a single SIOC file and imports all the posts contained within it. Figure 3 shows an example post imported into WordPress:

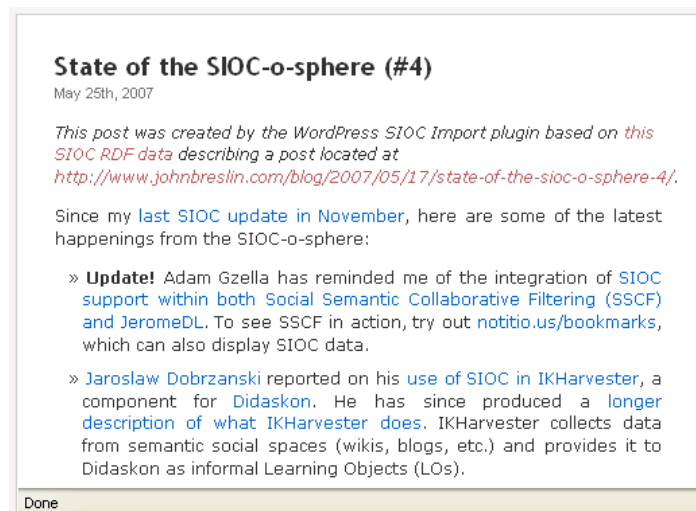


Fig. 3. Imported post in WordPress

Since SIOC is a universal data format, and not specific to any particular site, this pilot implementation already allows us to move content between different blog engines or even between different kinds of social media sites. However, the import of a single file shown here is useful for demonstration purposes.

We will now describe how a SIOC import tool can be extended to port all user-created content from one social media site to another. By starting from a site's main SIOC profile, we can retrieve machine-readable information about all the content of this site - starting with the forums hosted therein, and then retrieving the contained posts, comments, and associated users. This extended SIOC import tool needs to retrieve all SIOC data pages (possibly limited by some user-defined filters) and to re-create all the data found in this SIOC page on the target social media site.

This will result in a replica of the original site, including links between objects (e.g., between posts and their comments). Often, a part of the content that a user wants to port is not publicly available. SIOC exporters can also be used in this case, but the user will first need to authenticate at the source site and ensure that they have enough privileges to access all the data that need to be migrated.

Another step in social media portability is keeping two sites synchronised (if required): having the same set of users, posts, comments, category hierarchies, etc. In principle, this can be achieved by importing a full SIOC dataset and then monitoring SIOC data feeds for new items added (some SIOC export tools may need to be extended to do this). Implementing this in practice will undoubtedly unfold some interesting challenges.

Another example for using a complete site import would be for platform migration. For example, this could occur if a person has been using a mailing list for a particular community, and they then decide that the extended functionality offered to them by a Web-based message board platform is required.

It is not just discussion-type content items that can be ported. Using the SIOC Types module¹⁰, various content types can be collected in `sioc:Container(s)` and ported in the same way (Sounds, MovingImages, Events, Bookmarks, etc.)

4 The role of SKOS and FOAF

SIOC allows us to describe most user-created content, but it can also be combined with other RDF vocabularies such as Dublin Core (DC), Friend-of-a-Friend (FOAF) and Simple Knowledge Organisation Systems (SKOS). These vocabularies can be used when there is a need to migrate some additional data specific for a particular site.

DC provides a basic set of properties and types for annotating documents and resources. DC's "Type" vocabulary also defines various document types such as MovingImages, Sound, etc., that can be used to describe media elements from social media sites.

SKOS is designed for describing taxonomies such as category hierarchies. By exposing categories in SKOS we ensure the portability of this information to other social media sites.

Finally, FOAF is designed for describing information about people and their social relations. This vocabulary is already used together with SIOC to describe information about users, and additional properties from FOAF (e.g., `foaf:knows`) can be used to describe users' social networks. This can be useful when porting data from a social networking site.

5 Conclusions and Future Work

In this paper we have shown how SIOC data can be used to represent and port the diverse social media contributions being made by users on various sites. We began by describing the need and requirements for such portability, then talked about sources of data including various SIOC data producers, and next we described how such SIOC data can be imported into a system such as WordPress. We finally talked about how this data can be augmented using other vocabularies such as FOAF. For future work, we mentioned the issue of who should be allowed to reuse certain data in other sites

¹⁰ <http://rdfs.org/sioc/types>

(as spam blogs are often duplicating other people's content without authorisation for SEO purposes). As well as collecting a person's relevant content objects, social media sites may need to verify that a person is allowed to reuse data / metadata from these objects in external systems. This could be achieved by using SIOC as a representation format, aggregating a person's created items (through their user accounts) from various site containers, and combining this with some authentication mechanisms to verify that these items can be reused by the authenticated individual on whatever new sites they choose.

References

1. T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", *Scientific American*, vol. 284, no. 5, pp. 35–43, May 2001.
2. J.G. Breslin, A. Harth, U. Bojars, S. Decker, "Towards Semantically-Interlinked Online Communities", *Proceedings of the 2nd European Semantic Web Conference (ESWC '05)*, LNCS vol. 3532, pp. 500-514, Heraklion, Greece, May 2005.
3. U. Bojars, J.G. Breslin, A. Passant, "SIOC Browser - Towards a Richer Blog Browsing Experience", *Blogtalk Reloaded: Social Software - Research and Cases*, ISBN 3833496142, Vienna, Austria, October 2006.