Interlinking the Social Web with Semantics

Uldis Bojārs, John G. Breslin, Vassilios Peristeras, Giovanni Tummarello, and Stefan Decker, *DERI, National University of Ireland, Galway*

Applying Semantic Web technologies to the Social Web can lead to a Social Semantic Web, creating a network of interlinked and semantically rich knowledge.

One of the most visible trends on the Web is the emergence of *Social Web* sites, which help people create and gather knowledge by simplifying user contributions via blogs, tagging and folksonomies, wikis, podcasts, and online social networks. The Social Web has enabled community-based knowledge acquisition with efforts such as Wikipedia, demonstrating the "wisdom of the crowd" in creating the world's largest online encyclopedia. Although defining which structures or abstractions belong to the Social Web is difficult, they all facilitate collaboration and sharing among users, although usually on single sites.

Current online-community sites are isolated from one another, like islands in a sea. Various discussions might contain complementary knowledge and discussions—parts of the answer a person is looking for—but people participating in one discussion can't readily access information about related discussions elsewhere. As more and more Social Web sites, communities, and services come online, the lack of interoperation among these data silos or "stovepipes" becomes obvious. The potential synergies among many sites, communities, and services are expensive to exploit, and their data are difficult and cumbersome to link and reuse. The main reason for this lack of interoperation is that for the most part in the Social Web, common standards still don't exist for knowledge and information exchange and interoperation.

However, the Semantic Web effort aims to provide the tools needed to define extensible, flexible standards for information exchange and interoperability. The *Scientific American* article by Tim Berners-Lee, James Hendler, and Ora Lassila defined the Semantic Web as "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."[1] During the last couple years, much effort has gone into defining standards for data interchange and interoperation. The Semantic Web technology stack is well defined, enabling the creation of metadata and associated vocabularies. The Semantic Web effort is in an ideal position to make Social Web sites interoperable. Applying Semantic Web frameworks such as SIOC (Semantically Interlinked Online Communities) and FOAF (Friend-of-a-Friend) to the Social Web can lead to a *Social Semantic Web* (see Figure 1), creating a network of interlinked and semantically rich knowledge.

What is SIOC?

The SIOC initiative aims to enable the integration of online-community information. SIOC (pronounced "shock") provides a Semantic Web ontology for representing rich data from Social Web sites in RDF.[2] It has recently achieved broad adoption in a variety of commercial and open source software applications (see the sidebar "Some Recent SIOC

1541-1672/08/$25.00 © 2008 IEEE
Published by the IEEE Computer Society

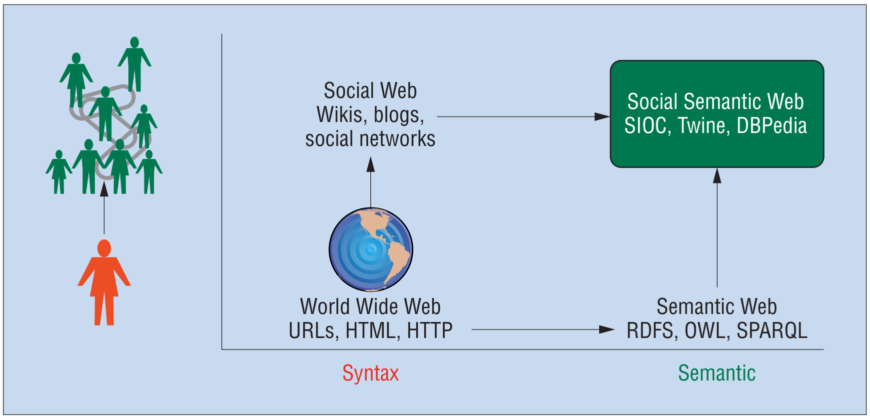

Figure 1. The Social Semantic Web. Applying Semantic Web technology to the Social Web could create a network of interlinked, semantically rich knowledge and bring the Social Web to its full potential.

Some Recent SIOC Applications

Companies are using Semantically Interlinked Online Communities (SIOC) in a variety of application domains:

- OpenLink's Data Spaces product provides access to SIOC instance data (think open social graph++) from a range of applications including blogs, wikis, aggregated feeds, shared bookmarks, discussions, photo galleries, and briefcases (for example, WebDAV file servers).
- Engage, a community information application from Talis, combines SIOC in its schematics with the Simple Knowledge Organization System (SKOS) for knowledge organization and Friend-of-a-Friend (FOAF) for person description.
- The Seesmic "video microblogging" service has also adopted the SIOC ontology as one of its open-platform formats.
- Talk Digger, a Web service from Zitgist LLC, helps people find, follow, and enter conversations on the Web. It exports all their data using SIOC.
- ImageMatters LLC recently announced gnizr, an open source social-bookmarking and mashup application that exports saved bookmarks using SIOC in combination with a tag ontology.

Many other open source applications of SIOC are coming from the Web developer community. OpenQabal, an open source social-networking and collaboration platform, supports SIOC, allowing Roller, JavaBB, and other component applications to become part of the "SIOC-o-sphere."

SIOC descriptions of forums are also being used for teaching and learning:

- Faculty Academy's Fishtank project leverages RDF's structure and searching power.
- IkeWiki, a knowledge-engineering wiki, allows discussions represented using the SIOC ontology (following a forum style with threaded views) to be attached to wiki pages.
- SWAML (Semantic Web Archive of Mailing Lists) uses SIOC as its base ontology. The project also includes Buxon, a sioc:Forum visor written in PyGTK.
- Tom Morris has created a third-party RDF exporter for the Twitter microblog service that uses SIOC (to represent all microblog entries) and FOAF (to de-

Applications"). For example, it's commonly used in conjunction with the FOAF vocabulary for expressing personal-profile and social-networking information.

By becoming a standard way to express user-generated content from such sites, SIOC enables new usage scenarios for online-community site data and lets developers build innovative semantic applications on top of the existing Social Web. In mid-2007, 16 organizations submitted the *SIOC Ontology* to the W3C, which published it (see www.w3.org/Submission/2007/02).

The SIOC ontology

The ontology consists of the SIOC Core Ontology (containing 11 classes and 53 properties, http://rdfs.org/sioc/spec) and the SIOC Services and SIOC Types modules. The SIOC Core Ontology defines the main concepts and properties required to describe semantic information from online communities. Figure 2 shows the ontology's main terms. We chose SIOC's basic concepts to be as generic as possible so that we could describe many different kinds of user-generated content.

We originally created the SIOC Core Ontology using the terms for describing Web-based discussion areas such as blogs and message boards. For example, users create posts (sioc:Post) organized in forums (sioc:Forum), which are hosted on sites (sioc:Site). In parallel with the evolution of new kinds of Social Web sites, these concepts became subclasses of higher-level concepts that were added to SIOC as it evolved: data spaces (sioc:Space, a place where data resides), containers (sioc:Container, used for grouping items together), and content items (sioc:Item). These classes let us structure the information in online-community sites and distinguish between different kinds of objects. Properties defined in SIOC let us describe relations between objects and their attributes. For example,

- the sioc:has_reply property links reply posts to the content to which they're replying,
- the sioc:has_creator and foaf:maker properties link user-generated content to additional information about its authors, and
- the sioc:topic property points to a resource describing the topic of content items (for example, their categories and tags).

The high-level concepts sioc:Space, sioc:Container, and sioc:Item are at the top of the SIOC class hierarchy; most other SIOC classes are subclasses of these. A data space (sioc:Space) is a place where data resides, such as a Web site, personal desktop, or shared file space. It can be the location for a set of Container(s) with content Item(s). Subclasses of Container can be used to further specify typed groupings of Item(s) in online communities. The class sioc:Item is a high-level concept for content items; it describes user-created content.

We usually use these high-level concepts as abstract classes from which we can derive other SIOC classes. We need them to ensure that SIOC can evolve and be applied to specific domain areas where definitions of the original SIOC classes such as sioc:Post or sioc:Forum can be too narrow. For example, an address book, which describes a collection of social and professional contacts, is a type of sioc:Container, but it's not the same as a discussion forum.

Let's look at some subclasses of these high-level SIOC concepts from the SIOC Types module.

SIOC modules

The SIOC Types module defines more specific subclasses of the SIOC Core concepts that we can use to describe the structure and types of content found in Social Web sites. This module defines subtypes of SIOC objects needed to more precisely represent various elements of online-community sites (for example, sioc_t:MessageBoard is a subclass of sioc:Forum). It also introduces new subclasses for describing different kinds of Social Web objects in SIOC. In addition, the module points to existing ontologies suitable for describing these objects' details (for instance, a sioc_t:ReviewArea might contain Review(s), described in detail using the Review Vocabulary). Examples of SIOC Core Ontology classes and their corresponding SIOC Types module subclasses include

- sioc:Container: AddressBook, AnnotationSet, AudioChannel, BookmarkFolder, Briefcase, EventCalendar;
- sioc:Forum: ChatChannel, MailingList, MessageBoard, Weblog; and
- sioc:Post: BlogPost, BoardPost, Comment, InstantMessage, MailMessage, WikiArticle.

Community sites typically publish Web service interfaces for programmatic search and content management services (typically SOAP or REST [Representational State Transfer]). These services can be generic (with standardized signatures covering input and output message formats) or service specific (where service signatures are unique to specific functions performed, as in current Web 2.0 API usage patterns). The SIOC Services module lets us indicate that a Web service is associated with (located on) a sioc:Site or part of the sioc:Site—for example, a sioc:Forum. This module provides a simple way to tell others about a Web service and shouldn't be confused with detailed Web service definitions. The sioc_s:service_definition property relates a sioc:Service to its full Web service definition (for example, in WSDL).

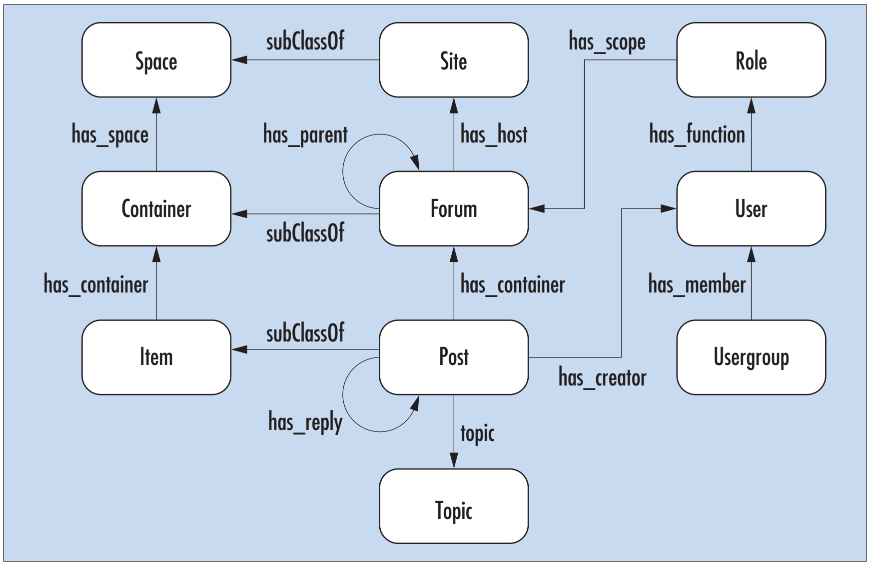

Figure 2. The main classes and properties in the Semantically Interlinked Online Communities (SIOC) ontology. The ontology describes common terms corresponding to the application data found in online communities.

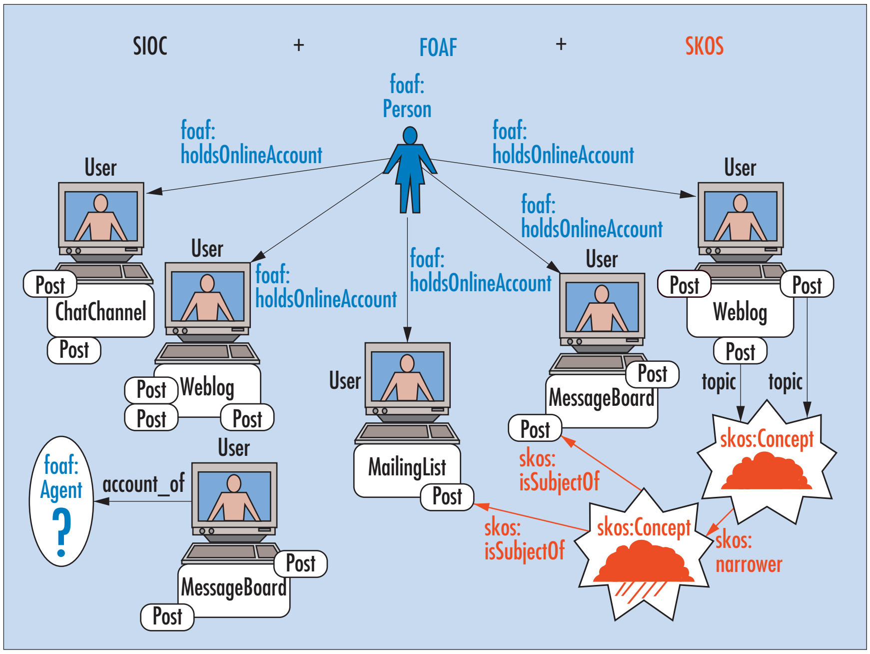

Figure 3. Relationships between the SIOC, Friend-of-a-Friend (FOAF), and Simple Knowledge Organization System (SKOS) ontologies. This example shows how SIOC uses FOAF terms to describe person-centric data and uses a SKOS schema to represent topic hierarchies and relationships.

Relationships between SIOC and other ontologies

One of the Semantic Web best practices is the reuse of existing ontologies and vocabularies, leading to better data interoperability. The SIOC ontology follows this practice by reusing the FOAF vocabulary to describe person-centric data and the Dublin Core (DC) vocabulary to describe properties of SIOC content items. Figure 3 shows some of the relationships between SIOC and other vocabularies. A person (described

Useful SIOC Resources

These resources will be useful in finding out more about the Semantically Interlinked Online Communities (SIOC) initiative:

The SIOC project page:
http://sioc-project.org
The SIOC W3C member submission:
www.w3.org/Submission/2007/02
A SIOC developer mailing list:
http://groups.google.com/group/sioc-dev
An Internet relay chat channel about SIOC:
Channel #sioc on IRC server irc.freenode.net
A comprehensive list of SIOC applications:
http://rdfs.org/sioc/applications
The SIOC Browser prototype:
http://sparql.captsolo.net/browser
The Semantic Radar extension for Firefox:
https://addons.mozilla.org/en-US/firefox/addon/3886
PingTheSemanticWeb.com's namespace statistics page:
http://pingthesemanticweb.com/stats/namespaces.php
SIOC is listed here as the fourth-most-used namespace out of more than 500. According to this service, more than 100,000 RDF documents use SIOC on the Web.

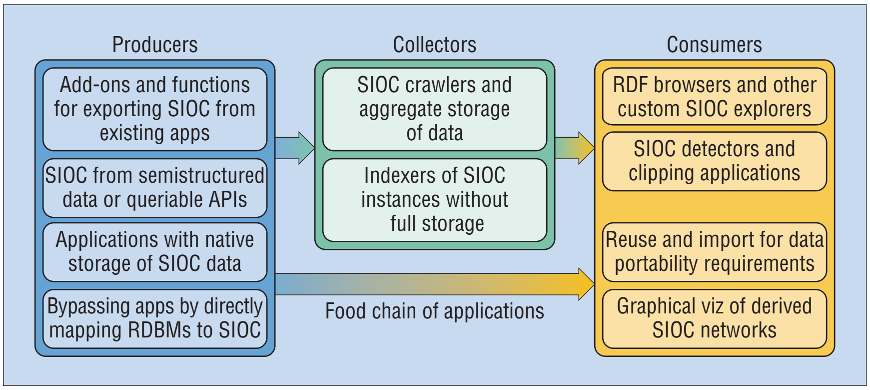

Figure 4. The SIOC food chain. Many types of applications produce, collect, and consume SIOC data.

by foaf:Person) will usually have a number of online accounts (sioc:User) on different online-community sites. They use these accounts to create content (sioc:Item or sioc:Post). The class sioc:User is a subclass of foaf:OnlineAccount, and the foaf:holdsOnlineAccount property links a person to his or her online accounts. SIOC content items (shown as Post objects in Figure 3) are described using properties from SIOC, FOAF, and DC. By using FOAF to point to multiple social-media site accounts registered to a user and using SIOC to express user-generated content on these sites, we can aggregate the content created by a person all across the Social Web.

Topics are usually present on the Social Web as categories and tags assigned to content items and are represented in SIOC using a sioc:topic property. SIOC doesn't require the values of sioc:topic to be in a particular ontology (apart from the value being a URI). It's left up to information system architects to choose the most appropriate ontology to represent topics in each case. One approach, illustrated in Figure 3, uses the Simple Knowledge Organization System (SKOS) schema to represent topic hierarchies and their relationships.

The W3C Member Submission documents for SIOC contain more detailed information on the relationships between SIOC and other RDF ontologies. (For more information on SIOC, see the sidebar "Useful SIOC Resources.")

What SIOC is not

SIOC tries to represent the information of online-community sites and human communication. However, unlike other representation approaches (for example, OWL Time[3]), SIOC isn't aiming at a domain's axiomatization. We deliberately designed the SIOC ontology to capture existing information that's mostly present in current Web information systems and to provide a low entry barrier for users and developers. SIOC's design lets us export information relatively simply from many sources and thus generate a critical mass of information. This leads quickly to the Web community adopting SIOC, providing additional exporters, and writing applications. SIOC aims at a minimal consensus of a given domain rather than a complete specification.

SIOC-enabled applications and tools

SIOC gives different online-community sites a common format for expressing their data in a rich, interlinked form. This interconnection of Social Web content using Semantic Web technologies can lead to many interesting possibilities on the individual and community level. Taking advantage of the information represented in SIOC, developers can build various browsers and applications on top of SIOC.

The SIOC food chain

Figure 4 shows some application types in the SIOC food chain, illustrating where SIOC data are being produced, collected, and consumed. This isn't exhaustive but does cover most available SIOC application types (http://rdfs.org/sioc/applications).

SIOC data producers include

- application add-ons and modules that reuse and build on existing internal functions to export SIOC data in RDF format (for example, the SIOC exporters for WordPress or Drupal);
- sources of semistructured data or queriable APIs that output data in formats that can be converted to SIOC and augmented with other RDF data (such as SWAML [Semantic Web Archive of Mailing Lists] for mailing list mailboxes, the Flickr2RDF tool, and Sioku, which leverages the Jaiku microblogging API);
- applications that store semantic data natively and use SIOC as one of their repre-

sentation formats (for example, Talis Engage); and

- schema-mapping tools that directly access MySQL or other relational database stores—bypassing default application access methods—to generate SIOC metadata (for example, OpenLink Data Spaces).

To help produce SIOC data from new applications, developers have also created reusable data-export APIs for PHP, Ruby on Rails, and Java.

The SIOC data these applications produce can be discovered, collected, stored, or indexed in an intermediary step or consumed directly by end-user applications.[4] Collectors can include Semantic Web spiders ("scutters") or crawlers that are specifically tailored to gather SIOC data. These crawlers can store SIOC instances from multiple sources in a single data store (such as SWSE [Semantic Web Search Engine], Swoogle, Zitgist, and the SIOC Crawler). Other collectors include indexers that store URLs of sites where developers can access SIOC data instances (for example, Sindice and PingTheSemanticWeb.com).

SIOC data consumers include

- generic browsers for RDF data (such as Tabulator and Disco);
- browsers customized to display SIOC data (for example, SIOC Explorer and Buxon);
- extensions for Web browsers that detect the presence of SIOC and other RDF data;
- extensions for pinging data-indexing services or reusing portions of community contributions elsewhere (for example, the Semantic Radar);
- applications that can import SIOC data—for example, for portability of data contributions between user accounts on different systems or for migrating a community from one site to another (for example, the SIOC importer for WordPress); and
- applications with visualizations oriented toward SIOC-linked data graphs (for example, Alexandre Passant's "SIOCal network" browser [http://apassant.net/home/2006/06/sioc-browser/index.php/network], which displays SIOC comment structures in terms of a network of their creators).

Generating the data

SIOC export tools are a class of application that produces SIOC RDF data from online-community sites, often implemented as plug-ins to a site's specific content-management system. By enriching social-community sites with SIOC RDF exporters, we can automatically create high-quality data without screen-scraping or reconstructing relationships from the information visible on Web pages.

SIOC export plug-ins are available for several common content creation platforms. These SIOC exporters represent information about every content item on a site in RDF, making all the main information at a site available in RDF and ready for reuse.

The SIOC Explorer aggregates data from various online-community sites and lets users browse and explore all the disparate information in an integrated manner.

Most SIOC exporters also use RDF autodiscovery links (that is, a link to an RDF document that's inserted in the HTML HEAD element of Web pages on the site) that enable tools such as the Semantic Radar plug-in for Firefox (https://addons.mozilla.org/en-US/firefox/addon/3886) to automatically discover this content.

The WordPress SIOC export plug-in, developed for a popular blogging platform, is one of the most popular SIOC export tools. A number of different online-community sites have similar export tools (http://rdfs.org/sioc/applications/#creating-others). We developed SIOC APIs to lower the barrier for entry and to help people who are unfamiliar with the Semantic Web to write SIOC exporters for their community sites. One such library, the SIOC API for the PHP scripting language, lets users easily manipulate SIOC data through PHP objects and methods, and renders the data into RDF/XML. The API creates and exports SIOC concepts about authors (sioc:User and foaf:Person), posts and comments (sioc:Post and sioc_t:Comment), and the Web site structure (sioc:Site and sioc:Forum).

Reusing SIOC data

As the Social Web begins to generate more SIOC data, SIOC data consumers can reuse this information—for example, to provide better tools for finding related information across online-community sites or to transfer rich information about content items between those sites. The pilot implementation of a SIOC import plug-in for WordPress (http://wiki.sioc-project.org/w/SIOC_Import_Plugin) is one application for reusing SIOC data between social-media sites.

Regular bloggers can reuse SIOC data by installing this plug-in to the WordPress blog engine. A user begins by supplying the URL of some SIOC data. The plug-in then retrieves the data from this URL, extracts all the content items, and posts them to the blog. While the pilot implementation processes only one SIOC data file at a time, it can be easily extended to mirror all content from a blog or forum site.

Tools such as this SIOC import plug-in let us transfer content items between not only the same type of sites (for instance, blogs) but also different types of social-media systems (for example, between a mailing list, expressed in SIOC using SWAML, and a blog). By combining SIOC data import and export tools, we enable various scenarios for data portability of social-media contributions (described later), letting people save their contributions and move them between different online-community sites.

Faceted exploration of SIOC data

The SIOC Explorer (https://launchpad.net/sioc-ex) aggregates data from various online-community sites and lets users browse and explore all the disparate information in an integrated manner.[5] This application's core is BrowseRDF—a domain-independent, faceted navigation system for RDF data that provides a generic view of all RDF data associated with SIOC concepts (for example, the foaf:maker relation between authors and posts). The application aggregates SIOC content into a local RDF store and provides various ways to view the content and associated data. The start screen lets users choose from a list of available sioc:Forums, with data coming from SIOC-enabled systems such as online-community forums, blogs, and mailing lists.

Figure 5. Faceted exploration of SIOC RDF data. The SIOC explorer lets users browse a broad view of disparate information in an integrated way.

When viewing posts from an individual forum or a group of forums, the user sees a list of posts in reverse chronological order. Each post is summarized (see Figure 5), but the user can expand it to read the full content. Clicking on a post's creator shows all posts (including comments and replies) written by this person across all forums; clicking on a topic shows all posts tagged with this topic, again across all forums. In contrast to ordinary feed readers, such lateral browsing works across all different types of community forums that can be described in SIOC. For instance, clicking on the user "Elias Torres" will show not only his blog posts but also his emails and contributions to Internet relay chats.

Finally, Figure 5's left side shows a generic faceted navigation interface, including relevant facets that aren't already shown as part of the default browsing interface. The application builds facets dynamically at view time, showing the properties and values derived from actual data as well as the properties that might not be known at system design time. Some facets (like the year) contain only "simple" values, whereas complex facets such as maker or topic can be further expanded to show subsequent subfacets (see the bottom left of Figure 5). Application developers can customize the facet navigation to their needs and can exclude or include certain facets or certain advanced operators such as the inverse join or the existential join.

The SIOC Explorer is built on the Ruby on Rails framework for Web application development. It uses three components for consuming and processing Semantic Web data. The first is ActiveRDF for mapping RDF data onto programmatic objects. The second is BrowseRDF, a faceted browsing engine that enables navigation of large Semantic Web data sets without domain-specific navigation knowledge. The third component is the SIOC RDF crawler, which crawls, extracts, normalizes, and integrates SIOC data from various community sites.

Exploring people's social connections via objects

Object-centered sociality (www.zengestrom.com/blog/2005/04/why_some_social.html) refers to the hypothesis that people are connected through the objects they create and collaborate on. On the Social Web, people are related through user-generated content and annotations. Figure 6 shows a conceptual illustration of this idea using a model of content "circles" created by a person via multiple online accounts.[6] Connections are formed between these circles by people creating similar content or using similar annotations.

For example, Bob and Carol are connected via bookmarked URLs that they both have annotated and through events that they're both attending, and Alice and Bob use similar tags and subscribe to the same blogs. SIOC and FOAF can be used together (www.mkbergman.com/index.php?p=372) to describe the objects in this social network of users, with FOAF describing information about the people and SIOC describing user accounts and their user-generated content. All this information, integrated across the Social Web, lets us build a picture of all the objects that a user has created, interacted with, and commented on across different social-media sites, from which the links between the users themselves emerge.

The Social SIOC Explorer[7] is an extension of SIOC Explorer that lets us see and explore social relations on the Social Web as manifested via user-generated content. The rich data structure, including links between an original post and its replies and links between a post and additional information about its author, is collected from SIOC data producers. The Social SIOC Explorer can use this information to mine relations between people on the Web—for example, to find a set of people who have participated in a discussion with or commented on the content created by a certain user (see Figure 7 on page 36).

In addition, this application contains a component that analyzes social networks using information about the content created by users, extracts social-context information from the SIOC data, and allows visualization of this information in the user interface.

SWPop: Live, embeddable SIOC aggregation

SWPop is a framework for creating semantic-data-powered plug-ins that can be embedded in Web applications. It demonstrates a synergy of applications for collecting and consuming SIOC data. Using SWPop, developers have created SIOC-powered widgets for the WordPress blog engine, which enhances the blogging platform with information derived from harvested SIOC data. When the SWPop plug-in is activated, clicking on highlighted links in the Web application will activate pop-ups that show a summary of information and discussions from the entire "SIOC-o-sphere" (all available sources of SIOC data).

Let's look at an example of a forum dedicated to work opportunities for Web-technology experts. Suppose the Human Resources Department is interested in hiring a Semantic Web consultant and posts a

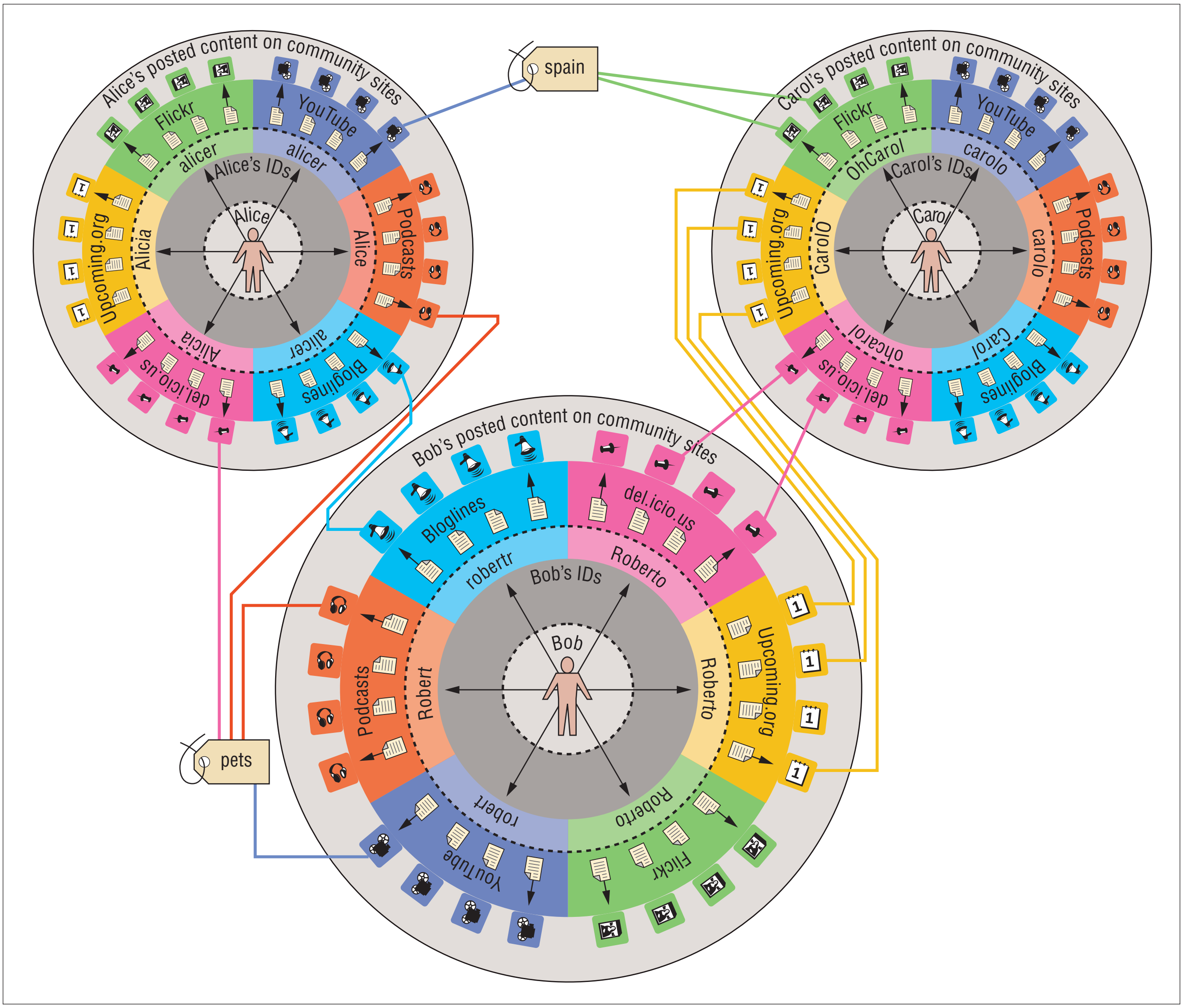

Figure 6. Object-centered sociality. The Social Web connects users through user-generated content and annotations.

request for this position. Many people reply with questions and proposals. With the SWPop plug-in, a user clicks on an author's name (for instance, "John") and sees a popup window with topics and posts John has recently created anywhere on SIOC-enabled online-community sites (see Figure 8 on page 36). The user can then navigate across his posts and restrict results to a given topic, if needed. Similar cross-site navigation is possible starting from links that are mentioned in a post or from topics (keywords). This not only is interesting for casual background checks but also introduces the concept of reputation portability by enabling others to quickly assess the value of John's statements given his previous distributed online activity.

Technically, SWPop plug-ins for Web applications operate as follows. The site owner adds SIOC RDF data export functionality to the site. When a new post is created on the site, WordPress pings the Sindice Semantic Web search engine. SWPop enriches the rendered Web page by adding a microformat containing a SHA1 hash of the author's email and the SWPop JavaScript (JS) code, which will skin the microformat with an SWPop information popup.

When the page is loaded, the SWPop JS detects the microformat and performs an Ajax call to the back-end SWPop application. The back-end application is based on the Sindice Semantic Web search engine (see "Indexers of SIOC Instances without Full Storage" in Figure 4, www.sindice.com), which keeps a frequently updated index of all the RDF resources on the Semantic Web.[8] Sindice's public API provides various ways to locate RDF sources—for example, by text keywords, URI, or inverse functional properties, such as the FOAF SHA1 hash of an email address in this case (mbox_sha1sum).

The SWPop back end uses these APIs to locate RDF information related to the post's author on the basis of the SHA1 hash of his email address. The information returned might include articles and comments made by the poster across the SIOC-o-sphere (with data about posts such as topic, date, and subject), details about the poster (such as photos) from other sources (for example, FOAF profile files), and social-network information. SWPop collects all

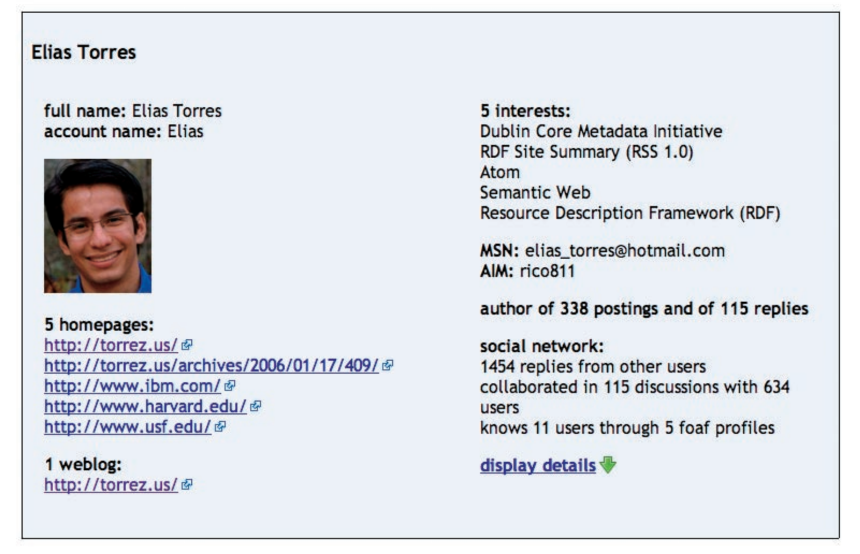

Figure 7. The Social SIOC Explorer's aggregate view of information about a specific person. The screen shows the total number of replies to that person's posts calculated across multiple sites on which he or she participates.

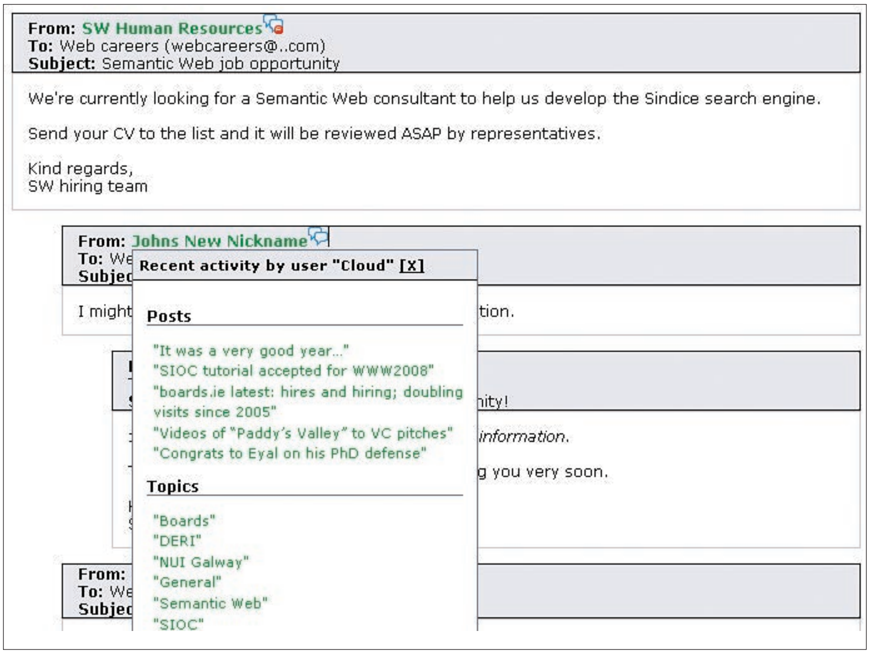

Figure 8. The SWPop framework. The back end locates RDF information about John, including recent activity on SIOC-enabled online-community sites.

this information into a named graph repository on top of which Sindice performs analysis and queries.

On the basis of analysis results, SWPop adds a speech balloon with posts and other information about the person, collected from across the Social Web—for example, next to the poster's name on the Web page. When a visitor clicks the balloon, the server will create an aggregate of post titles, sorted by date, along with recent topics about which the poster has been writing from all sources of SIOC data. It will then send this data to the SWPop JS using JavaScript Object Notation (http://swm.deri.org:8080/swpop/samples/thread2.html).

A benefit of applications such as SWPop is that little effort is required to install them and to tap into the existing vault of information described in SIOC and other Semantic Web vocabularies. The application's client-side part is lightweight; the SWPop back end and large RDF indexing applications such as Sindice do all the "heavy lifting."

SIOC-related initiatives

Various working-group initiatives are promoting SIOC and tailoring it for their community domains, including social-network portability, collaborative work spaces, and life-science discourse. The initiatives described here have found value in SIOC and are finding it useful in their projects.

Data portability using SIOC

The DataPortability working group (www.dataportability.org) was recently established to look at ways to port data from one social-media service to another. It aims to document the best practices for integrating existing open standards and protocols to enable end-to-end data portability between online tools, vendors, and services. Another of its goals is to let users move, share, and control their identity, photos, videos, and all other forms of personal data.

One sample data portability scenario involves using the Yadis communications protocol to discover a particular person's identity. The protocol returns a Yadis XRDS (Extensible Resource Descriptor Sequence) document indicating which identities that person prefers to use and what services those identities are held on. Then, we can use the WRFS (Web Relational File System) abstraction model to find out what containers the returned identities hold on those services (http://dataportability.pbwiki.com/WRFS%20Prototype%20Workspace).

SIOC is an ideal representation method for describing all the content items a person has created (via his or her user accounts) on various social-media sites and the structure contained therein. A combination of the FOAF vocabulary and SIOC is well suited to describe the relationship between a person and multiple social-media site accounts this person owns.

In Figure 9, Bob holds user accounts on various social Web sites (two shown for clarity, but there could be many more). Via those accounts, he creates content items on those sites (usually in containers of some sort, such as a bookmark folder, personal blog, message board, or image gallery). He should be

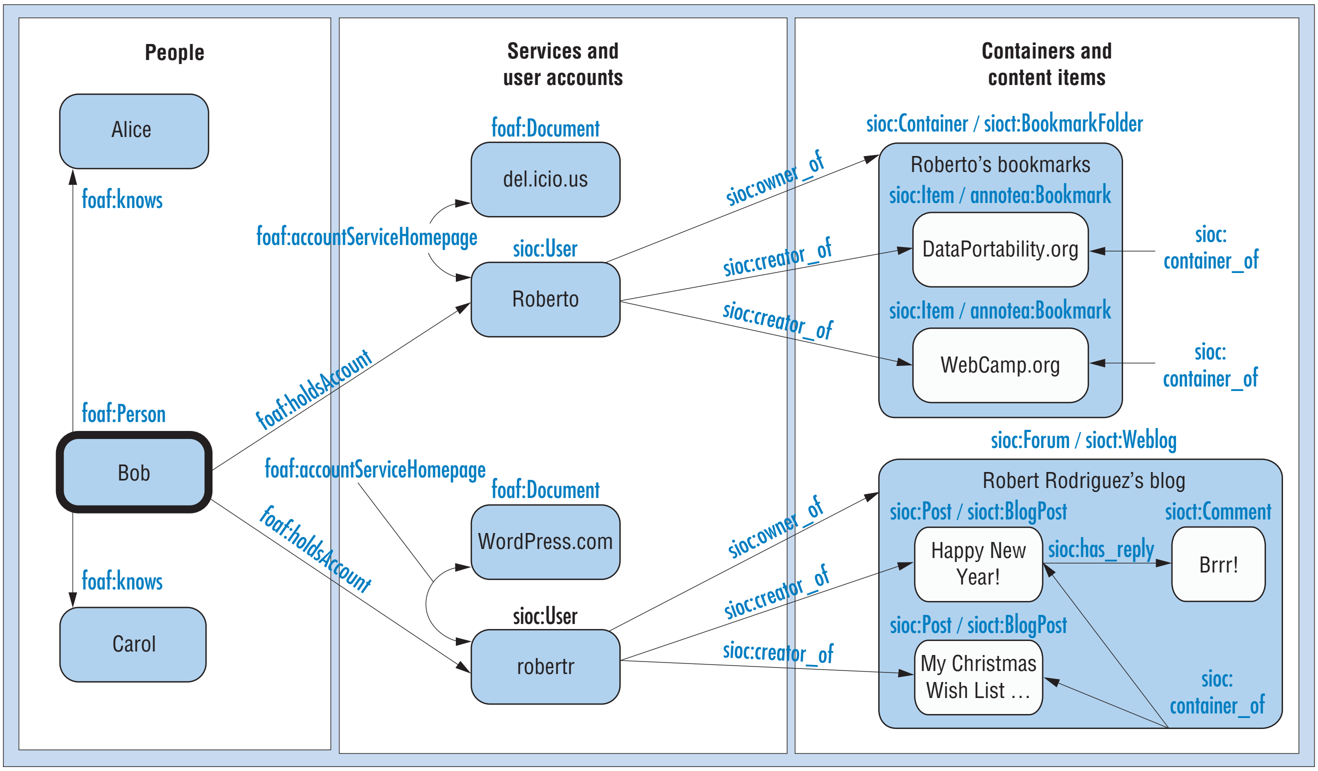

Figure 9. Bob's user accounts on two object-centered social networks. Via these and other Social Web sites, Bob creates content items in various containers.

able to port not only his social graph (in this case, his connections to Alice and Carol) but also his personal containers or sets of content items and perhaps even associated comment replies. The vocabulary terms (foaf:knows, sioc:User, and so on) are shown in blue.

SIOC is more that just a way to represent personal data containers. The DataPortability working group and others are exploring methods for porting not just small user-centric data sets but whole sets of community data. We created SIOC to provide a way to describe the content from online communities (mailing lists, message boards, and so on). While people soon used it for blogs and more recently for other personal sets of Web 2.0-type content items, it has the concepts needed to describe the structure and contents of a community site as a whole. If someone runs a community site and decides to port a group from one place to another, he or she can use SIOC to describe the existing community site's structure and content to recreate it on a different information system.

One challenge for DataPortability is the mapping of different data standards. This initiative can use related but different data representation methods for various scenarios (for instance, XFN/hCard and FOAF), which raises the issue of mapping between these formats. We hope that existing work from the Semantic Web domain—such as ontology mapping and GRDDL (a technique for *g*leaning *r*esource *d*escriptions from *d*ialects of *l*anguages)—will significantly address this issue.

For this initiative, SIOC is an open standard for describing user-created content and thus enabling data portability. Because it's based on Semantic Web technologies such as RDF, SIOC data can be easily mixed with other RDF data formats (such as DOAP [Description of a Project], Review, and so on) that more specialized community sites might require.

Semantically interlinking CWEs

Collaborative-work-environment (CWE) platforms such as Lotus Notes, Microsoft SharePoint, and BSCW (Be Smart—Cooperate Worldwide) provide integrated work spaces to support knowledge workers inside the enterprise. During the last 20 years, the main driver for developing such platforms has been the need for interoperability among numerous applications that were initially built separately to support specific tasks, such as tools for project and document management, calendars, forums, and mailing lists. These CWE platforms integrate CWE tools and related information, enhancing intraorganizational knowledge storage and processing.

With the introduction of these CWE platforms, the interoperability problems were adequately addressed internally. However, at the interorganizational level, problems remain. As legacy and isolated systems, these platforms remain informational silos that can't communicate with the rest of the world. From an organizational perspective, this creates insurmountable barriers to information exchange with peer organizations. From the single user's perspective, it creates a situation in which information that might exist in separate systems can't be automatically combined and must be collected and processed manually.

The case of e-professionals (for example, engineers, lawyers, and researchers) who work concurrently on multiple projects that are supported by different CWE platforms is quite interesting. How can they find a comprehensive list and links to all documents (and posts) that have been uploaded (or edited) during the last week in all the projects (and platforms) in which they might be involved? And how can they

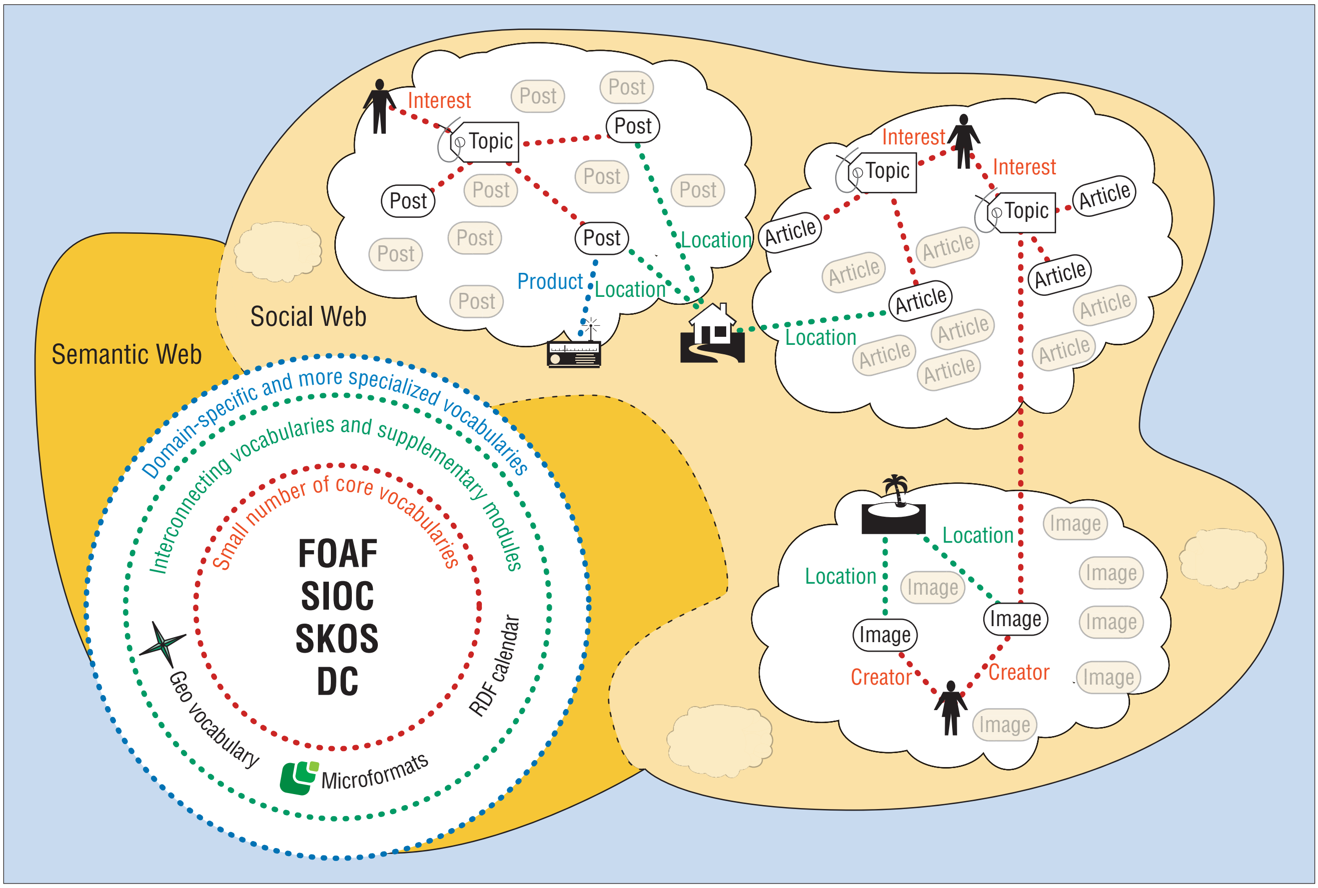

Figure 10. A possible "crystallization point" for domain-specific ontologies. SIOC will bring multiple domain ontologies together to better describe user content.

identify links to resources categorized—for example, as "architecture"—that might be stored in all these different platforms?

Currently, the Ecospace Integrated Project is addressing these types of problems. The project has adopted the SIOC ontology to provide the basis for much-needed multiplatform integration and to allow cross-project querying and access to this semantically interlinked information.[9] This has occurred in three stages:

1. Concepts (such as document, folder, user, or post) that exist in the CWE domain and appear in the involved platforms—namely, BSCW and Business Collaborator (BC)—were mapped to the SIOC ontology. This was feasible because we developed SIOC to cover online communities, a domain quite related to CWE. In this way, SIOC provides a metalanguage for the CWE domain.
2. The consortium developed SIOC exporters for BSCW and BC. These tools, based on the conceptual mappings created in the previous stage, annotate the internal data and export them as SIOC RDF data. Because these tools use the same SIOC ontology, the internally kept data acquires the same semantics and can be understood and processed by any RDF-based application.
3. Our research institute developed a specialized SIOC4CWE Explorer for navigating and querying aggregated SIOC data from heterogeneous shared workspaces in a unified way. This gives the user a single interface to query all the projects he or she participates in regardless of whether they're stored in BSCW, BC, or any other platform.

So, queries such as those presented earlier for e-professionals can be answered. The user can directly access documents, posts, calendars, contacts, and any other kind of resource that were previously "locked" inside each separate platform. Following this approach, any CWE platform can become interoperable with others and accessible by the SIOC4CWE Explorer from the moment it exports SIOC-annotated data. Developing a SIOC exporter is a comparatively easy task that could give a huge return on investment for favorite CWE platforms, because it provides a semantic bridge to the outer world. The Ecospace Integrated Project is now looking at the development of a SIOC exporter for SAP NetWeaver.

Bio-zen and the art of scientific community maintenance

The bio-zen initiative (http://neuroscientific.net/index.php?id=43) is attempting to represent on the Semantic Web data, information, and knowledge from research in all facets of the life sciences. It aims to unify information that's now scattered through a multitude of different data structures, exchange formats, and databases. As part of this, SISC (Semantically Interlinked Scientific Communities, http://neuroscientific.net/sisc/sisc_introduction.htm) aims to improve how we currently represent and communicate scientific data and knowledge.

Bio-zen and SISC employ the SIOC, FOAF, DC, Creative and Science Commons, Open Biomedical Ontologies, and Health Care and Life Sciences ontologies and technologies. The initiative's creators adapted SIOC to represent basic scientific discourse in scientific publications and on the Web. According to Matthias Samwald, one of the creators, they chose SIOC as a base ontology because it provides "an excellent tool to describe scientific discourse in a practical, Web-centric manner."

Future "shock"

Following on from the SIOC ontology's successful dissemination and implementation in over 40 modules and applications, we plan to extend our SIOC research by looking at more specific application domains. The SIOC project has a stable ontology and a good foundation of tools using it, but we need to figure out what the requirements are in particular domains, whether they involve interlinking real-estate communities or extracting personal skills from user-generated content. We are now targeting a number of use cases to see how to augment SIOC with domain-specific terms. Some domains of interest include facilitating customer health support groups (patient support from peers on the Web), providing structured representations of professional scientific discourse, and modeling or navigating interactions in prelegislative consultation.

An interesting dimension is the application of SIOC in CWE platforms. SIOC's initial purpose was to link open online communities on the Web. CWEs are using SIOC in a fundamentally different environment: linking legacy, closed, and proprietary systems. We intend to reuse this experience for applying SIOC-based solutions in other similar areas. The more general area of enterprise application integration seems to be a challenging candidate where interoperability issues are critical.

We expect SIOC to act as a "crystallization point" along with FOAF, DC, and SKOS (see Figure 10) in bringing various domain ontologies together to better describe user-generated content both on the Social Web and in enterprise environments. SIOC provides a framework to which we can attach further details about content items. If additional information (for example, embedded in the item or retrieved using natural-language-processing techniques) is available about a content item, we can describe this information in RDF and attach it to existing SIOC data about this item.

We'll also investigate Social Semantic Web solutions for enhancing discussion topics on social software systems such as forums, blogs, wikis, and podcasts with semantics, to yield what we might term "semantic dialectic topics." This research will build on SIOC, SALT (Semantically Annotated LaTeX, http://salt.semanticauthoring.org), and other efforts to represent argumentative discussions such as IBIS (Issue-Based Information Systems), Diligent, and Compendium to create a structure for defining the topics and argumentative nature of future discussions. It will also provide methods to classify existing topic discussions. The application domains of scientific discourse (as in SISC) and legislative debate are of particular interest here.

Systematic access to knowledge is critical for solving today's problems on an individual, organizational, and global level. To develop new solutions and enable innovation, networks are necessary—networks of people and organizations having access to networks of knowledge. Although knowledge is inherently strongly interconnected, current information fabrics don't reflect this interconnectedness, so they're not optimal for supporting the development of solutions and innovation. The lack of interconnectedness hampers basic information management and problem-solving capabilities such as finding, creating, and deploying the right knowledge at the right time for problem solving and collaboration.

We can direct the combination of Semantic Web and Social Web collaboration technologies toward a universal collaboration and networked knowledge infrastructure, enabling knowledge management capabilities as expressed by visionaries such as Vannevar Bush[10] and Doug Engelbart.[11] For the most part, their ideas remained just a vision for far too long because the necessary foundational technologies hadn't been invented. Figuratively speaking, these ideas were proposing jet planes when the rest of the world had just invented bicycles. Starting with the Semantic Web and the Social Web, the

The Authors

Uldis Bojārs is a researcher and PhD student in the Digital Enterprise Research Institute at the National University of Ireland, Galway. His work focuses on the SIOC (Semantically Interlinked Online Communities) project. His research interests include the Semantic Web, online communities, and social software. Bojārs received his master's in computer science from the University of Latvia and is a member of the ACM. Contact him at uldis.bojars@deri.org.

John G. Breslin is a researcher and adjunct lecturer in the Digital Enterprise Research Institute at the National University of Ireland, Galway. His research interests include social software, online communities, and the Semantic Web. Breslin received his PhD in electronic engineering from the National University of Ireland and is a member of the IEEE. Contact him at john.breslin@deri.org.

Vassilios Peristeras is a research fellow and adjunct lecturer in the Digital Enterprise Research Institute at the National University of Ireland, Galway. His research interests include e-government and e-participation, collaborative work environments, social software, and the Semantic Web. He received his PhD in electronic government from the University of Macedonia. Contact him at vassilios.peristeras@deri.org.

Giovanni Tummarello is a research fellow and adjunct lecturer in the Digital Enterprise Research Institute at the National University of Ireland, Galway, where he leads the Data Intensive Infrastructures research unit. His research has focused on computer science, computational intelligence, and multimedia semantics. He leads or led projects such as DBin, Sindice, and Semantic Web Pipes and started the Semantic Web Applications and Perspectives conference series. Contact him at giovanni.tummarello@deri.org.

Stefan Decker is a professor at and director of the Digital Enterprise Research Institute at the National University of Ireland, Galway. His research interests include the Semantic Web, digital libraries, and the social semantic desktop. Decker received his PhD in computer science from the University of Karlsruhe. Contact him at stefan.decker@deri.org.

foundations to make these visions a reality now exist.

Acknowledgments

We acknowledge the contributions of Fabio Corneti and Adam Westerski on SWPop and of Eyal Oren and Benjamin Heitmann on the SIOC (Semantically Interlinked Online Communities) Explorer. Science Foundation Ireland funded this work under grant SFI/02/CE1/I131, as did Ecospace under grant FP6-IST-5 35208.

References

1. T. Berners-Lee, J.A. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, no. 5, 2001, pp. 34–43.
2. J.G. Breslin et al., "Towards Semantically Interlinked Online Communities," *Proc. European Semantic Web Conf.* (ESWC 05), LNCS 3532, Springer, 2005, pp. 500–514.
3. J.R. Hobbs and F. Pan, "An Ontology of Time for the Semantic Web," *ACM Trans. Asian Language Processing*, vol. 3, no. 1, 2004, pp. 66–85.
4. U. Bojārs et al., "An Architecture to Discover and Query Decentralized RDF Data," *Proc. Scripting for the Semantic Web Workshop at ESWC*, CEUR Workshop Proc., 2007, http://CEUR-WS.org/Vol-248/paper11.pdf.
5. B. Heitmann and E. Oren, "Leveraging Existing Web Frameworks for a SIOC Explorer to Browse Online Social Communities," *Proc. Scripting for the Semantic Web Workshop at ESWC*, CEUR Workshop Proc., 2007, pp. 52–61, http://CEUR-WS.org/Vol-248/paper6.pdf.
6. J.G. Breslin and S. Decker, "The Future of Social Networks on the Internet: The Need for Semantics," *IEEE Internet Computing*, vol. 11, no. 6, 2007, pp. 86–90.
7. U. Bojārs, B. Heitmann, and E. Oren, "A Prototype to Explore Content and Context on Social Community Sites," *Proc. 1st Conf. Social Semantic Web* (CSSW), LNI P-113, 2007, pp. 47–58.
8. G. Tummarello, R. Delbru, and E. Oren, "Sindice.com: Weaving Open Linked Data," *Proc. 6th Int'l Semantic Web Conf.* (ISWC), LNCS 4825, Springer, 2007.
9. K. Ning et al., "A SIOC-Enabled Explorer of Shared Workspaces," *CSCW/Web 2.0 Workshop 10th European Conf.* (Ecscw 07), 2007.
10. V. Bush, "As We May Think," *Atlantic Monthly*, vol. 176, no. 1, 1945, pp. 101–108.
11. D.C. Engelbart, *Augmenting Human Intellect: A Conceptual Framework*, tech. report AFOSR-3233, Stanford Research Inst., 1962.