



# SIOC Ontology: Applications and Implementation Status

W3C Member Submission 12 June 2007

**This version:**

<http://www.w3.org/Submission/2007/SUBM-sioc-applications-20070612/>

**Latest version:**

<http://www.w3.org/Submission/sioc-applications/>

**Editors:**

[Uldis Bojārs](#) - [DERI, NUI Galway](#)

[John G. Breslin](#) - [DERI, NUI Galway](#)

[Alexandre Passant](#) - [LaLIC at Université Paris-Sorbonne](#)

**Authors:**

[Sergio Fernández](#) - [Fundación CTIC](#)

[Frédéric Giasson](#) - [Zitgist LLC](#)

[Kingsley Idehen](#) - [OpenLink Software Inc.](#)

Copyright © 2006-2007 by [DERI Galway](#) at the [National University of Ireland, Galway](#), Ireland.

Development of SIOC is supported by [Science Foundation Ireland](#) under grant number SFI/02/CE1/I131.

This document is available under the [W3C Document License](#).  
See the [W3C Intellectual Rights Notice and Legal Disclaimers](#) for additional information.



Regarding underlying technology, SIOC relies heavily on W3C's [RDF](#) technology, an open Web standard that can be freely used by anyone.

---

## Abstract

The SIOC (Semantically-Interlinked Online Communities) Core Ontology provides the main concepts and properties required to describe information from online communities (e.g., message boards, wikis, weblogs, etc.) on the Semantic Web. This document contains a brief overview of various SIOC implementations and applications.

## Status of this document

*This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications can be found in the [W3C technical reports index](#) at <http://www.w3.org/TR/>.*

This document is a part of the *SIOC Ontology Submission*, and is based on the [SIOC Implementations](#) page on the [SIOC wiki](#).

Authors welcome suggestions on the [SIOC Core Ontology Namespace](#) and this document. Please send comments to the [SIOC developers' mailing list](#) (SIOC-Dev), [public archives](#) are available. This document may be updated or added to based on implementation experience, but no commitment is made by the authors regarding future updates.

Please consult the [namespaces.zip](#) archive, a part of this submission, for a snapshot of the SIOC Ontology namespaces referenced in this document. Live namespace documents are located at relevant namespace URIs.

By publishing this document, W3C acknowledges that the [Submitting Members](#) have made a formal Submission request to W3C for discussion. Publication of this document by W3C indicates no endorsement of its content by W3C, nor that W3C has, is, or will be allocating any resources to the issues addressed by it. This document is not the product of a chartered W3C group, but is published as potential input to the [W3C Process](#). A [W3C Team Comment](#) has been published in conjunction with this Member Submission. Publication of acknowledged Member Submissions at the W3C site is one of the benefits of [W3C Membership](#). Please consult the requirements associated with Member Submissions of [section 3.3 of the W3C Patent Policy](#). Please consult the complete [list of acknowledged W3C Member Submissions](#).

## Table of contents

- [1. Introduction](#)
- [2. Creating SIOC data](#)
  - [2.1 SIOC APIs](#)
  - [2.2 Weblog, forum and CMS exporters](#)
  - [2.3 Other exporters](#)
  - [2.4 SPARQL endpoints](#)
- [3. Using SIOC data](#)
  - [3.1. Querying SIOC data](#)
  - [3.2. Crawling SIOC data](#)
  - [3.3. Browsing SIOC data](#)
  - [3.4 Using SIOC for new data](#)
  - [3.5 Reusing SIOC data](#)
- [4. SIOC utilities](#)
- [5. References](#)
- [6. Change log](#)

## 1. Introduction

All SIOC data uses RDF as an underlying data format, and can be created and processed as such. Various applications have been designed to use SIOC by taking some of its unique aspects into account. In this document, we will outline concrete implementations and applications that use SIOC data. (A complete state-of-the-art list of SIOC implementations is maintained at the [SIOC applications](#) page).

SIOC data can also be processed and used by many generic Semantic Web applications, capable of using RDF. A full list of these applications is outside the scope of this document. For more information about Semantic Web applications and libraries please see "[Where do I find tools for Semantic Web development?](#)" section of the [Semantic Web FAQ](#).

## 2. Creating SIOC data

SIOC is designed to export information about the content and structure of online community websites in a machine-readable form. Thus, various tools, exporters and services have been created to expose SIOC data from existing online communities.

### 2.1 SIOC APIs

**SIOC Export API for PHP.** In order to help people to write SIOC exporters, a [SIOC Export API for PHP](#) has been designed, offering an easy way to manipulate SIOC data through PHP objects and methods, and rendering content in an RDF/XML file. The API creates and exports SIOC concepts about the authors (sio:User plus foaf:Person), posts and comments (sio:Post and sio\_t:Comment), and the structure of the website (sio:Site and sio:Forum).

**SIOC API for Java.** A [SIOC API for Java](#) has been created, based on [semweb4j](#). For each object in the SIOC ontology, this API generates classes with links between the objects realised as Java properties.

### 2.2 Weblog, forum and CMS exporters

Different SIOC exporters have been written for a number of popular weblogs, forums and content management systems (CMS). All of these exporters feature RDF auto-discovery links for SIOC data, and are available via open-source licences.

**WordPress SIOC Exporter.** [WordPress](#) is a popular blogging platform based on PHP/MySQL. The [WordPress SIOC Exporter](#) allows the production of SIOC metadata from WordPress-based blogs, by simply installing two plugin files in the plugins folder and enabling the SIOC plugin from the WordPress control panel. This plugin is the most widely used SIOC exporter.

**Dotclear SIOC Exporter.** [Dotclear](#) is a widely-used French blogging

platform. The [Dotclear SIOC Exporter](#) produces SIOC metadata using the SIOC export API for PHP, and exports information about the blog itself, the blog users, posts and comments.

**b2evolution SIOC Exporter.** [b2evolution](#) is a multi-blog platform that evolved from the same roots as WordPress (from b2/cafeblog). An early version of a [b2evolution SIOC Exporter](#) has been built upon the SIOC export API for PHP.

**Drupal SIOC Exporter.** There is also a [Drupal SIOC Exporter](#), which can be used to export SIOC data from [Drupal](#) sites, including blogs and forums. As Drupal can be used as a multi-user blogging platform, the plugin will export all blogs and all user accounts, so that each post can be clearly identified by its users.

**phpBB SIOC exporter.** [phpBB](#) is one of the most used open-source message board platforms. A [phpBB SIOC Exporter](#) has been written that produces SIOC metadata about forums, posts and the users that created them.

## 2.3 Other exporters

**Talk Digger.** [Talk Digger](#) is a web service that helps people to find, follow and enter conversations on the Web, in order to see who is linking to a specific web page. Users can create a personal profile, define their interests, make new friends, track conversations, leave comments in conversations, etc. All data from this service is [exported in RDF/XML using SIOC](#).

**SWAML.** [SWAML](#) is an exporter for mailing list content in Semantic Web format. SWAML reads a collection of e-mail messages stored in a mailbox (from a mailing list compatible with RFC 4155) and generates an RDF description of it. It is written in Python, using SIOC as the main ontology to represent a mailing list in RDF. SWAML is also available as a Debian package (in testing).

**Mailing List Archives.** A Java-based application for [generating SIOC data from mailing list archives](#) has been developed, leveraging RSS and Atom feeds from web-based message archives. The [source code](#) uses the [RDFReactor](#) library for creating RDF APIs, and some [sample SIOC output data](#) is also available.

**Twitter2RDF.** An [RDF exporter for Twitter microblogs has been created](#) that uses SIOC (for the microblog entries) and FOAF (for describing the people). For example, here are representations of [Twitter](#) microblogs for two users: [captsolo](#) and [johnbreslin](#).

**IRC2RDF.** An [RDF converter for IRC](#) has been created that exports metadata in [Turtle](#) format, and SIOC is being used as one of the main representation formats.

**Sioku.** [Jaiku](#) is another microblogging site for which the [Sioku Jaiku2RDF service](#) has been created using Ruby on Rails. SIOC and FOAF are used as the main vocabularies for representing streams of microblog entries and for describing people and their contacts respectively.

**Custom exports.** Some sites have developed custom SIOC exports for their own applications. For example, here is some [SIOC forum data](#) produced from a [Dutch community forum](#). Some other custom blog sites are producing SIOC data in RDFa or eRDF ([1](#), [2](#)). A custom [SIOC exporter for a blog aggregator](#) has also been produced.

## 2.4 SPARQL endpoints

**OpenLink Data Spaces.** [OpenLink Data Spaces](#) (ODS) SPARQL endpoints provide access to SIOC instance data from a range of ODS application instances. The [ODS SIOC reference](#) wiki page describes the SIOC data available from these applications via ODS, including blogs, wikis, aggregated feeds (RSS 1.0, 2.0 and Atom), shared bookmarks, discussions (i.e. comment threads), photo galleries, briefcases (e.g. WebDAV file servers), etc. The live [ODS demo server](#) and [MyOpenLink.net \(alpha\) service](#) are examples of ODS instances that can expose SIOC instance data to SPARQL query service clients, also in the form of real and virtual RDF graphs.

## 3. Using SIOC data

### 3.1. Querying SIOC data

All SIOC data can be queried using SPARQL, once the SIOC Core Ontology and Module Namespaces are defined in the SPARQL query.

**OpenLink Data Spaces.** As mentioned in [section 2.4](#), [ODS](#) exposes all its data as real or virtual RDF graphs via its [Virtuoso](#)-based quad store. The [ODS SIOC reference](#) wiki page describes how various application realms are mapped to SIOC, along with an extensive collection of SPARQL query examples and live demonstration links for interacting with the SIOC instance data.

**#B4mad.Net.** The [#B4mad.Net SPARQL endpoint](#) has been set up to query SIOC data from [PlanetRDF](#) and the [SIOC-Dev](#) mailing list at Google Groups. This service uses [ARC](#) and the [XMLArmyKnife SPARQL AJAX library](#). Some [demos of SIOC queries](#) are given.

### 3.2. Crawling SIOC data

**SIOC Crawler.** SIOC data can be collected by a crawler that traverses the Web and retrieves any SIOC data it finds. The crawler starts with a list of "seed" SIOC URLs and follows `rdfs:seeAlso` links used to point to more SIOC and RDF data. This is a generic principle for crawling RDF

documents, so a generic RDF crawler could be used. The [SIOC Crawler](#), however, has additional knowledge about the structure of SIOC data which allowed the enhancement of this crawler with advanced functionality, e.g., incremental retrieval of new SIOC data in threads.

### 3.3. Browsing SIOC data

**SIOC Browser.** The [SIOC Browser](#) allows people to browse and receive additional information from SIOC data sources or data stores. Browsers can work in two modes - on-the-fly mode and crawler mode - or can use a combination of both ([Bojars et al., 2006](#)). The [on-the-fly or live browser](#) is a simple and effective way to explore community information available in SIOC. It gives a user-friendly look at the internal structure of the data without requiring the viewers to dive into a more complex RDF/XML syntax. A triple-store interface - that can be plugged onto any triple store that offers a SPARQL endpoint - has also been written for [browsing crawled SIOC data](#), providing methods to visualise this data in both textual and graphical ways.

**Buxon.** [Buxon](#), a sioc:Forum browser, was released as a part of SWAML 0.0.3 and is now available as an independent package. Written in PyGTK, it reads sioc:Forum information from RDF files and shows it as a tree of message threads. See this [Buxon screenshot](#) from the application. It is available as a Debian package.

**SIOC Explorer.** The [SIOC Explorer](#) is a web application which can aggregate posts from community web sites publishing SIOC data. The SIOC Explorer allows you to view and navigate based on all exported RDF data, not just SIOC, by utilising a domain-independent faceted-browsing approach. It has been implemented in Ruby on Rails and the [ActiveRDF](#) / [SWORD](#) Semantic Web application framework for Rails.

**Other browsers.** SIOC data can also be browsed using generic tools, such as [Tabulator](#), [Disco](#) or [Timeline](#), directly using SIOC data in RDF/XML or by translating it into a specific data type.

### 3.4 Using SIOC for new data

**Fishtank.** [SIOC descriptions of fora for teaching and learning](#) demonstrate another [use for SIOC data](#) in the [Fishtank](#) application for the [Faculty Academy](#). This application also aims to use the structure and searching power of RDF to fully utilise tags and feeds on blogs, by combining people's RSS feeds with SIOC data using RAP and [Triplr](#).

**BAETLE.** [BAETLE](#) (Bug And Enhancement Tracking Language) aims to create a software bug ontology that can be used by various repositories to enable people to query for bugs across these repositories. SIOC is being used to define some of the required terms.

**RDFa on Rails.** [RDFa on Rails](#) is a library of helper methods to help

Ruby on Rails developers with producing RDFa data. SIOC terms are used to describe blog posts in this library.

### 3.5 Reusing SIOC data

**IKHarvester.** [IKHarvester](#), a component for the [Didaskon](#) curriculum assembly framework, collects data from semantic social spaces (wikis, blogs, etc.) and provides it to Didaskon as informal learning objects (LOs). SIOC data exported from blogs and wikis is gathered and mapped to learning object metadata (LOM) with IKHarvester.

**notitio.us and JeromeDL.** [notitio.us](#), a social bookmarking and knowledge harvesting system, provides [SIOC metadata support through SSCF](#) (social semantic collaborative filtering). The SSCF functionality can be seen in action at [notitio.us/bookmarks](#), which can also display the associated SIOC data from bookmarked sites, forums and posts. This functionality is also implemented in the [JeromeDL](#) semantic digital library system.

## 4. SIOC utilities

**Semantic Radar.** To facilitate end-user access to SIOC data, the [Semantic Radar](#) - a Firefox browser extension - detects the presence of SIOC, FOAF and DOAP data in a web page, and alerts a user who then has the possibility to browse the data in an online SIOC browser.

**PingTheSemanticWeb.** The Semantic Radar application can also ping the [PingTheSemanticWeb](#) (PTSW) website, an online service that collects, stores and distributes links to RDF documents for every ping, and this is an efficient way to find and index SIOC data over the Web ([Bojars et al., 2007](#)). Through this index, external services such as [doap:store](#) or [Sindice](#) can use the PTSW service to find data.

**SpecGen4.** The [SIOC Core Ontology Specification](#) is generated using the [SpecGen4](#) Python-based ontology specification generator for RDFS/OWL. This utility identifies SIOC class and property terms from the [SIOC Core Ontology Namespace in RDFS/OWL](#), and generates a customised HTML specification file using these terms in combination with a template and some per-term definition files.

## 5. References

- [RFC 4155] E. Hall, "The application/mbox Media Type," RFC 4155, The Internet Society, September 2005, <http://www.ietf.org/rfc/rfc4155.txt>.
- [Bojars et al., 2006] U. Bojars, J.G. Breslin, A. Passant, "[SIOC Browser - Towards a Richer Blog Browsing Experience](#)", The 4th Blogtalk Conference (Blogtalk Reloaded), Vienna, Austria, October 2006.
- [Bojars et al., 2007] U. Bojars, A. Passant, F. Giasson, J.G. Breslin, "An Architecture to Discover and Query Decentralised RDF Data", The 3rd

Workshop on Scripting for the Semantic Web, The 4th European Semantic Web Conference (ESWC '07), Innsbruck, Austria, June 2007.

## 6. Change Log

- 2006-10-24: Initial version of this document.
- 2007-05-09: Revised document, added information about Buxon, SIOC explorer.
- 2007-05-30: Revisions for member submission.

[uldis.bojars@deri.org](mailto:uldis.bojars@deri.org)

[john.breslin@deri.org](mailto:john.breslin@deri.org)