What is conStruct?

conStruct SCS is a distro of the Drupal content management framework. It sets new standards in data integration and as a structured content system (SCS). With conStruct, you can:

Let your data and its structure drive your applications
Easily interoperate your diverse internal information with public content on the Web, and
Leverage a platform designed from the ground up for knowledge management and collaboration.

Like Drupal, conStruct and its components are free and open source.

A Complete Data Integration Platform

conStruct combines the best of Drupal, Virtuoso for RDF and structured data, Solr for full-text and faceted search, and our own special sauces. A modular design based on a Web-oriented architecture and RESTful services assures interoperability and open substitutability.

Hosted by the Drupal CMF

conStruct is hosted by the feature-rich Drupal content management framework, with expansion planned for other CMS platforms. Drupal provides the general content development and publishing framework, user and groups management and rights, administrative functions, theming, and many other tools and capabilities. For example, more than 2500 third-party modules extend Drupal's core CMS capabilities.

Because Drupal does such a great job in general content management and in these other functions, conStruct can concentrate on its structured data focus. conStruct is designed to play nicely with the existing Drupal framework.

A Structured Data Orientation

There are two central, organizing bases for managing structured data within conStruct SCS. The first is the dataset. Datasets contain one or more data records from a single source representing the same type of instance(s). Datasets may reside on the Web as well as be stored locally. Each dataset is uniquely identified with standard metadata characterizations.

At minimum, datasets have a simple structure of attribute-value pairs for each instance record. conStruct tools operate against one or more datasets. Individual users may be assigned access to these datasets, and whether they have CRUD (*create – read – update – delete*) permissions. The combination of access rights and permissions then defines which tools and what operations are available to a given user for each dataset.

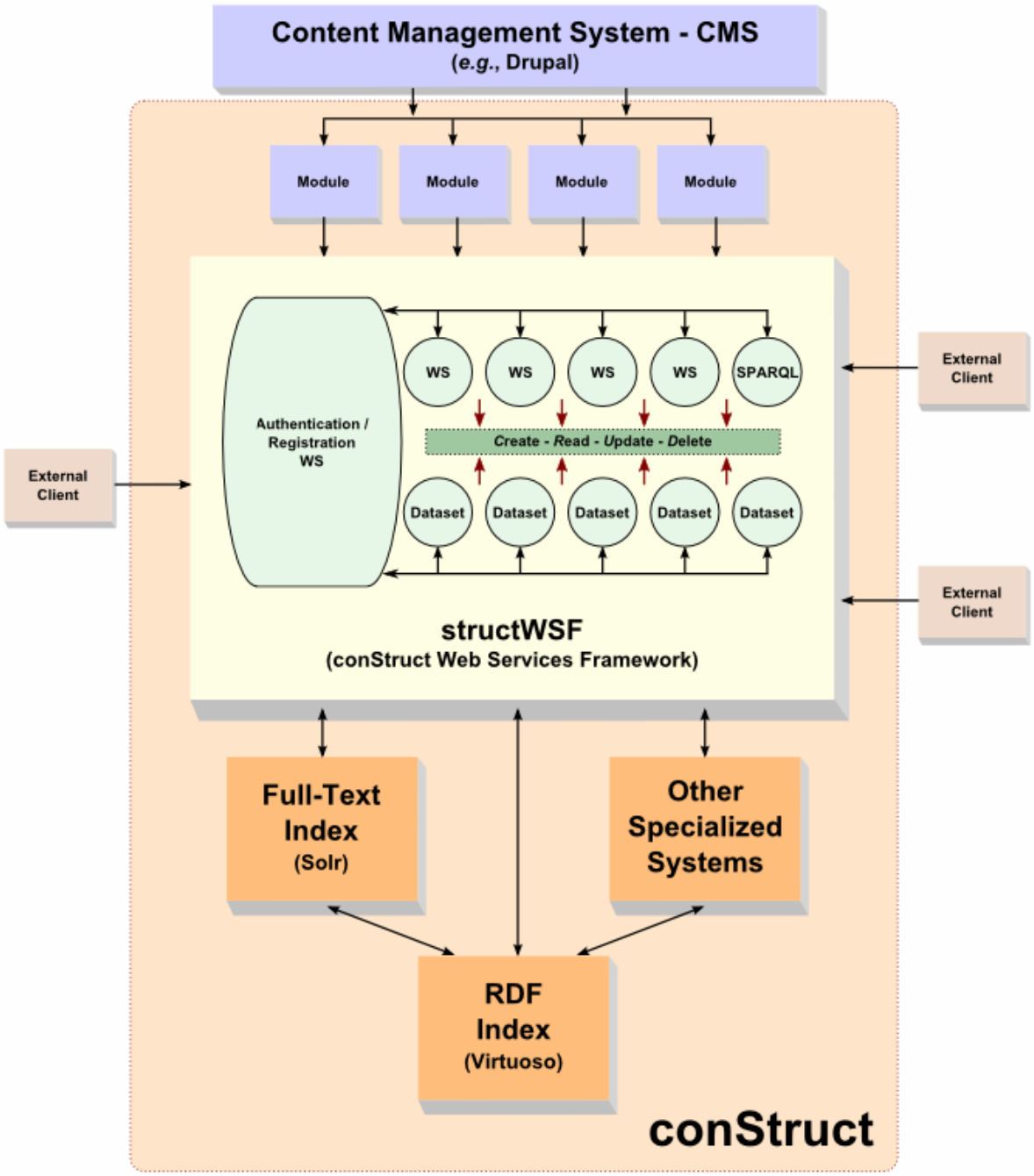

Figure 1. The conStruct Structured Content System

Structural relationships between this data – expressed as schema (ontologies) – are the second organizing basis. These ontologies describe the relationships between concepts and attributes and relate them to external schema. Actually, multiple ontologies govern the use of conStruct, some for internal and data management purposes, others as ways to describe the structural relationships in a given knowledge domain.

Only curators or site managers need worry about the ontology basis underlying conStruct. However, given this organizing importance of ontologies, conStruct provides a growing set of tools to expose these structures for enhancements or modifications.

Powerful Structured Data Management

Structured data management within conStruct is based on OpenLink's Virtuoso software. Virtuoso is a broad data storage and indexing engine. It is also a middleware platform that can integrate conventional relational database management systems (RDBMS) with Web applications and semantic Web services.

conStruct uses Virtuoso for its structured data management based on the Resource Description Framework (RDF) data model. This structured data is queried via SPARQL, as enhanced with updating and graph-management extensions. conStruct also uses Virtuoso for its native support for a variety of Web service protocols and data converters for retrievals across the network.[1]

Data and application interactions occur through Virtuoso's virtual or federated database server. This core provides internal storage and application facilities, the ability to transparently expose tables and views from external databases, and the capability to expose application logic in a homogeneous way. The variety of data sources supported by Virtuoso can be efficiently represented as RDF, which provides a coherent view of disparate data from virtually any source. Virtuoso can also provide the RDF data to drive other specialized systems, such as analysis or inference engines.

Full-text, Faceted Search

Many structured data systems lack good performing full-text or faceted search. Also, structured data based on linked data RDF often substitutes Web identifiers for literal text values. While good for linking and tracking purposes, this practice can indirectly shield indexed text from standard search engines, leading to apparently incomplete results sets when searching.

To address these issues, conStruct adds the textual information of the resources being referenced for each indexed resource. It does this by issuing a series of queries to the RDF triple store in order to retrieve the desired text descriptions for the structured data, which is then indexed by the Solr faceted text-search engine. Solr is an open source enterprise search server based on the Lucene Java search library, with faceted search, caching, and many more features.

These SPARQL queries trace the identifiers and then extract the full-text information to which they reference. Each conStruct record thus contains a full-text representation of all objects and properties. Among other benefits from the Solr search engine, this approach removes the conventional full-text limitations of RDF while also gaining contextual faceting.

Built on the Independent structWSF Architecture

conStruct is built around a middle layer – structWSF – of simple and compound Web services. These Web services (WS) expose the underlying structured data. They have uniform interfaces and conventions and share the error codes and standard functions of HTTP, consistent with

[1] Internal Drupal content is still maintained by the MySQL database. In addition, the basic design and architecture of conStruct's independent structWSF Web services layer allows the substitution of other RDF triple stores (so long as they have update and named graph capabilities).

RESTful design. conStruct further exposes its structured data as RESTful linked data, useful for sharing and integrating with external services.

This structWSF layer is a separate open source component under Apache 2 license. It can be queried and used independently of the broader conStruct or Drupal. It is especially designed to provide an interoperable data layer that resides independently of a specific portal or CMS system.

structWSF Web services and endpoints lend themselves to connecting multiple conStruct installations into an open, platform-independent and distributed data collaboration framework. The dataset structure noted above is also designed for such networked interoperability.

Basic and Extended Tools Suite

All conStruct tools are basically the packaging of these Web services with user interfaces (UIs) to expose them at the Drupal CMS level. Each tool package is called a '*module*' in the Drupal context. conStruct modules are therefore very lightweight wrappers that conform to the registry and hooks of Drupal and tie into one or more Web services at the structWSF level.

For ease of discovery in listings and such, the name of each conStruct Drupal module is prefixed by '*struct*'. Most of the current modules are part of the base conStruct distro, though we anticipate more modules (and hopefully some from third parties) over time.

Because of the data structure-driven nature of conStruct, many of its tools are able to present context-specific dropdown list choices and context-specific auto-completion of entered values in the user interface, all driven solely by the underlying structure of the governing ontologies.

conStruct Core

This is the essential conStruct module that bridges the existing Drupal system and its hooks with structured data and RDF. Multiple pre-configured RDF ontologies drive this module, which automatically brings faceted browsing and searching and structured data organization to Drupal.

conStruct Core is the bridge between Drupal, Virtuoso and its ontologies and structured data. It also provides basic user interface (UI) capabilities for CRUD (*create, read, update, delete*) activities.

structDisplay

Various views of structured data and associated facets and attributes are provided via structDisplays created with this module. The structDisplay module provides a WYSIWYG editing of template HTML and CSS with variable assignments within the template. The module combines modifications to the Drupal TinyMCE rich text editor (RTE) and the Smarty templating system, all linked to conStruct Core and ontology variables. structDisplays can be assigned to Drupal blocks for further presentation flexibility.

structOntology

The structural relationships that drive conStruct are provided via a series of structured vocabularies, or ontologies. Ontologies define hierarchical concept relations, relatedness, and synonyms for key concepts. Ontologies are also the basis for inferencing and other reasoning tasks done by the Virtuoso system on the structured data, as well as for labels and synonyms used within the UI.

There are some standard, pre-configured ontologies provided with the package. New domains or new structural considerations are provided by supplementary ontologies.

The structOntology module provides the facility for constructing and managing these additional ontologies. It also provides import support for externally specified representation of ontologies in text format (via, as one option, Excel spreadsheets).

The conStruct Distro

conStruct SCS may be obtained and built in one of three ways:

1. As a set of standard Drupal module downloads, which requires all non-module components to be separately downloaded and installed (instructions provided):
   - conStruct Core
   - structDisplay
   - structOntology
2. As a single distribution package, which includes the modules listed above, plus these additional components with installation and build instructions:
   - structWSF Web services framework
   - Smarty templating system
   - Virtuoso open source (VOS) edition, and the
   - Solr faceted, full-text search engine.
3. As an Amazon EC2 AMI image that can be copied into a new instance with a single click.

Community and Open Source Commitments

OpenStructs.org is the distribution point for the full conStruct distro and for the separate structWSF Web services framework. OpenStructs.org is also a knowledge base and clearinghouse for other modules and open source templates, harvesters and extractors developed by others for the system.

All conStruct modules from Structured Dynamics are being developed in strict accordance with best Drupal practices and hooks under GPL license. It is the intent that all contributed SD modules interoperate seamlessly with Drupal core and key other contributed modules.

conStruct may also evolve over time to support other CMS and RDF triple stores.