

Sustainable semantic publishing

Using semantic
technology to unlock the
value in your data



September 2011

<http://openup.tso.co.uk>



All organisations can benefit from using semantic technology to unlock the value in their data. To gain the biggest benefits you need a sustainable process – one that automates the capture, transformation and publishing of linked data in an efficient, seamless and reliable way. At TSO, we publish daily updates in multiple data formats, including XML and RDF, and have developed a collection of scalable and resilient integrated services which help to capture, transform and deliver data.

For more information about TSO's OpenUp® services visit <http://openup.tso.co.uk/openup-platform>



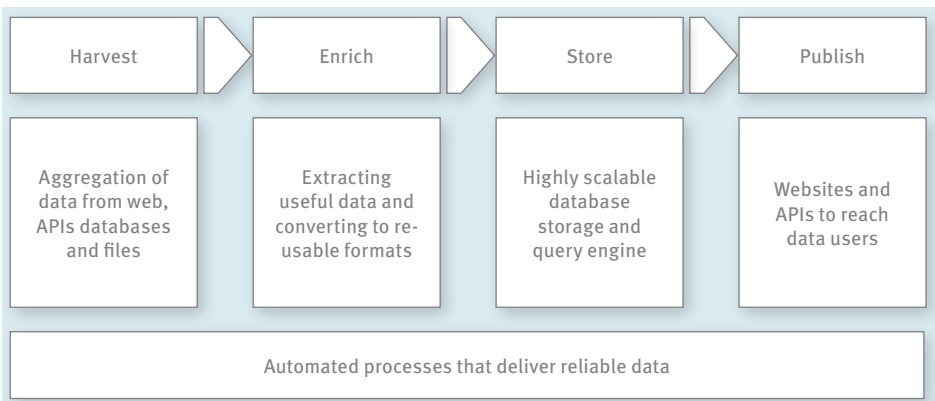
What is sustainable data publishing?

Using semantic technology is a great way of enhancing and getting more value out of your business data, but to make it effective you need an automated way of creating and disseminating data.

For every organisation it will be different, but the key steps in developing a sustainable publishing solution will usually include:

- Data harvesting – aggregation of data by pulling (or pushing) it from the web, APIs, databases or files
- Data enrichment – taking text, extracting what's useful and converting it into the required format for re-use
- Storing and querying RDF – providing a highly scalable database storage and query engine
- Publishing data – creating websites and APIs to access your information in the most effective way

The TSO OpenUp® platform is a collection of integrated services available as Software as a Service (SaaS), allowing fast deployment at a reasonable cost. Together, the services provide a highly scalable and resilient platform that allows you to store, query and enrich your data.



TSO's OpenUp® sustainable data publishing platform

Data harvesting

Our harvesting system can pull data from the web, APIs, databases or from supplied files, and provides an increasingly sophisticated means of controlling the acquisition of data. It also exposes an API allowing data to be pushed into the system.

For example, it is possible to schedule data collection at predetermined times each day, throttling the level of requests so that a website system is not overloaded, whilst allowing for retries for any failures after configured periods to ensure all data is collected. We can work with you to create a harvesting system that works for your data. The harvesting system allows configurable pipelines to be put in place to control the entire acquisition and manipulation process.

Try the Data Enrichment Service

<http://openup.tso.co.uk/developer/des>



Data enrichment

If you are handling text that you know has information hidden away inside it, then our Data Enrichment Service (DES) will help you to get at that information. The DES is a web-based scalable information extraction system built on top of the GATE framework. It allows bespoke processing of data, but can also be used in conjunction with the core DES system, which provides continually improving identification and extraction of entities, such as people's names, organisations or places, with a focus on UK-based data.

The extracted information can be made available in a variety of formats to suit the requirements of the user. One of the formats that is available is RDF/XML. This information can then be stored in TSO's RDF store, making it available for querying. The combination of extraction and RDF store provides a powerful approach to maximising the information potential of documents.

TSO's core Data Enrichment Service is free to use. For data publishers needing to enrich more than 10,000 documents per day with an SLA or with specific data extraction requirements, we can offer the Professional version of the Data Enrichment Service for a monthly fee.

Storing and querying RDF

To handle the linked data aspects of OpenUp®, we provide an RDF database storage and query engine. Our datastore is built around 5Store, a highly scalable, clustered, commercial database storage and query engine that hosts RDF data. 5Store was designed and developed by Garlik, leaders in semantic technology. TSO's RDF store is designed to scale to well over a trillion triples.

We can provide a number of approaches to access data stored on the platform, including SPARQL endpoints and Linked Data APIs.

Publishing data

Having transformed your data, we can also help to develop and host the resulting web applications. We can create websites and host them with our website hosting service, certified to ISO 27001 Information Security Standard. We have delivered a range of projects for our clients including the creation of APIs to access linked data.

To find out more about any of these services or how they can be combined to create a managed data publishing service please email us at opendata@tso.co.uk

Scalability and resilience

At the heart of the platform is the ability to scale – and to do so in a resilient manner. The platform can be expanded at all levels, so if a client has particular needs then we can expand that part of the system. If required, we can provide dedicated hardware for you for those parts of the system where you need that extra level of performance.

In addition, the platform will provide geographic failover so that, in the event of disaster at one data centre, there is the reassurance that downtime will be minimal.

Professional services

We have built a great team of talented software developers with a semantic web skill base, in addition to our experienced C++, C# and XML developers. The team has been involved in a number of key semantic projects including:

- Data Enrichment Service – an automated annotation tool
- Queryable RDF store with ready-to-use data sets
- XBRL to RDF conversion for public sector accounts (to be released via data.gov.uk)
- Government Department names ontology (accessible via data.gov.uk)
- RDF glossary of terms for Best Management Practice
- Harvesting feeds of live data into the RDF store from london-gazette.co.uk and legislation.gov.uk
- RDFa annotation in the London Gazette
- Creating an implementation of the linked data API

Want to find out more?

We are always looking for opportunities to use semantic technologies to further enhance information use. If you would like to know more, have ideas you would like to share or think you could add value by partnering with us, then please get in touch with opendata@tso.co.uk

To find out more about the OpenUp® platform visit <http://openup.tso.co.uk>

To keep up to date with the latest information from TSO, follow us on Twitter www.twitter.com/TSOTechnology

www.tso.co.uk

75% recycled
This is printed on 75% recycled paper



CONFIDENTIALITY STATEMENT

The contents of this document together with all other information, data, materials, specifications or other related documents provided by Williams Lea ("WL") (together "materials") shall be treated at all times by the recipient as the confidential and proprietary information of WL. The recipient shall not disclose any such materials to any third parties without the express, prior written approval of WL. Where such express approval is granted by WL, the recipient shall ensure that all third parties to whom disclosure is made shall keep any such materials confidential and shall not disclose them or any part of them to any other person. All intellectual property rights in the materials shall remain the property of WL, or its third party licensors, and are protected by copyright.

© 2008-09 Williams Lea Group

DISCLAIMER

This document may be incomplete without reference to any oral briefing provided by WL, reflects current conditions and WL's views as of this date and is subject to correction or change at any time. Although the information contained in this document is believed to be accurate in all material respects, neither WL nor any of WL's advisers, agents, officers or employees accepts responsibility or liability for or makes any promise, representation, statement or expression of opinion or warranty, express or implied, with respect to the accuracy or completeness of the content of this document (to the extent permissible by law) unless and save to the extent that such promise, representation, statement or expression of opinion or warranty is later expressly incorporated into a legally binding contract.



Part of the Williams Lea Group