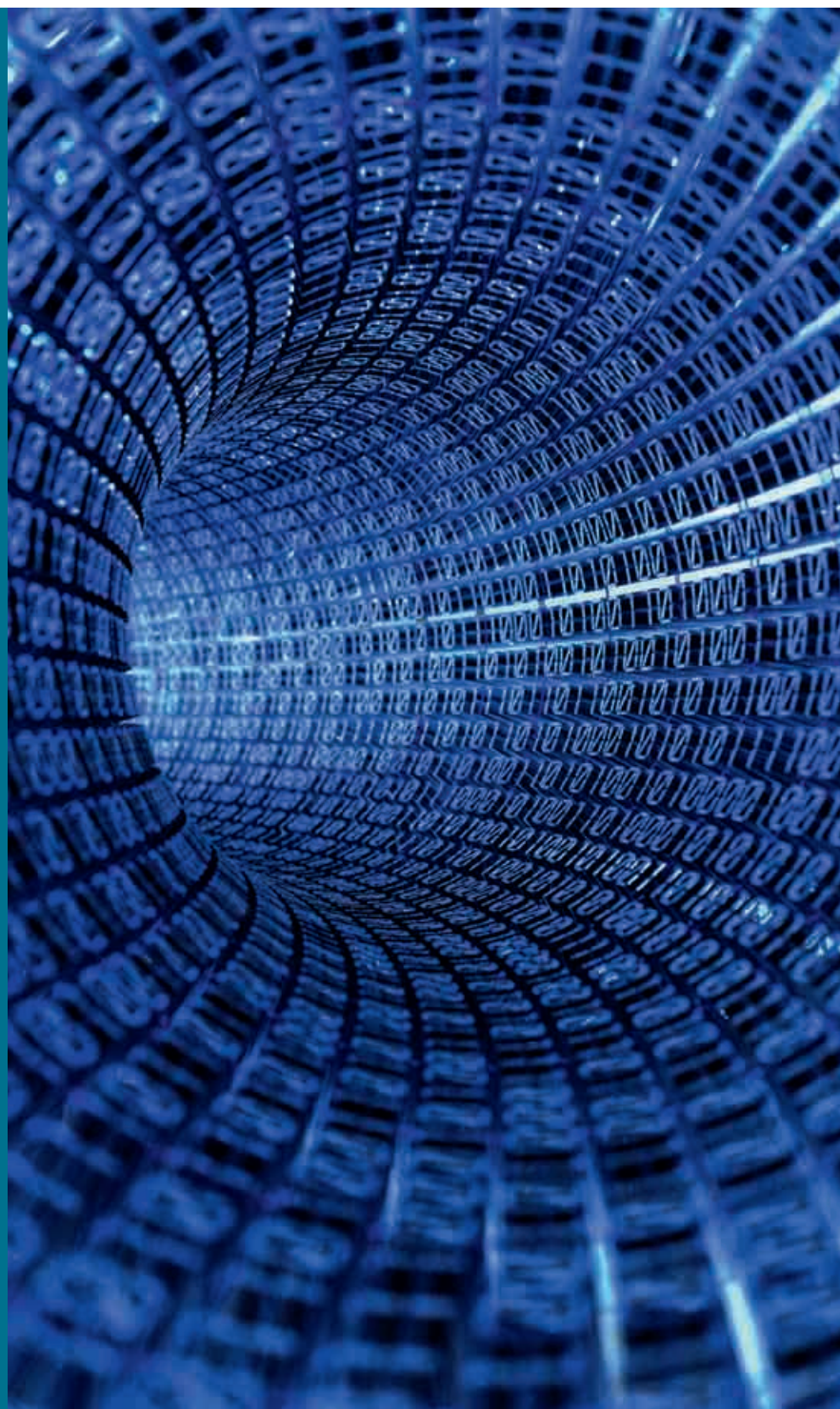


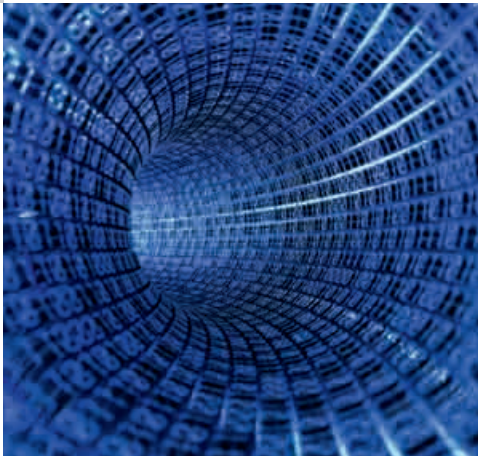
Opening up government data

TSO's solutions to improve
transparency and re-use



November 2010

www.tso.co.uk/opendata



TSO has been at the forefront of working with public sector clients to open up public data. Our experts help to create, structure, capture, transform and deliver some of the most important government information. We create tools and processes to allow data to be created in a structured way; enrich data using text engineering techniques; convert data into formats to publish as linked data on the web and provide and host web environments that allow both people and machines to access the data.

Today's government information needs to be published in an open and re-usable way

The Government has committed to open up government data to improve transparency and has set deadlines for publishing data. Prime Minister David Cameron has specified that 'from July 2010, government departments and agencies should ensure that any information published includes the underlying data in an open, standardised format'.

The establishment of the Public Sector Transparency Board, with Sir Tim Berners-Lee and Professor Nigel Shadbolt as board members, and the continued growth of www.data.gov.uk will drive forward the Government's transparency agenda. The Public Sector Transparency Board's principles state that public data should be published in re-usable, machine-readable form. Since launch more than 4,000 datasets have been listed on www.data.gov.uk, but this is only the starting point. The real challenge is to publish those datasets as open, linked data to enable relationships to be made with other data – and then to keep that data up-to-date. Enriching content and transforming it into open, linked formats such as RDF, XML and JSON; using URIs effectively; and publishing data through a RESTful API enables that data to be re-used in more innovative ways.

At TSO we have been working with public sector organisations to open up their information and publish it as linked data. We are particularly expert in the publication of regularly updated, fine grained data and our established, reliable web publishing processes enable simultaneous publication in many different formats. Examples of regularly updated websites built on open linked data principles include www.legislation.gov.uk and www.london-gazette.co.uk which enable the information to be re-used. We also host www.data.gov.uk, the website created to provide a way in to the wealth of government data.



www.data.gov.uk was launched in January 2010 to provide a way in to government data.



www.govspark.org.uk, partially funded and hosted by TSO, is an example of how government open data has been re-used to create a more useful application.

The Statutory Instruments template includes metadata to create structured data and enable publishing in multiple formats.



TSO's Data Enrichment Service can identify and extract useful pieces of information such as names, dates, organisations or places from unstructured text and convert that information into linked data.



Legislation.gov.uk, developed by TSO, was built on open data principles to provide full access to the statute book as open data.

"a major contribution to the Government's transparency agenda"
– Lord McNally, Minister of State and Deputy Leader of the House of Lords

Three steps to opening up data

1. Creating and capturing your data in a structured way

To enable your data to be re-used it must be created in a structured way. Our experts will first work with you to understand the requirements for your data. Once these are established, we will create tools and processes that will capture content in the most appropriate and efficient way, depending on the nature of your data, the number of users capturing it and their skill set. We have created tools and processes using MS Office templates, XML authoring tools and web portals for content validation and submission.

TSO developed an MS Word based drafting tool for The National Archives and relevant government departments to create secondary legislation. The template contains all elements in the development of a Statutory Instrument, including typesetting styles and metadata, enabling online validation, reducing errors and allowing a quicker publishing process. The template is used to create structured XML and print ready PDFs, with website content generated automatically from source XML.

2. Enriching your data to enhance and transform

To make your data open, linked and re-usable, it must be unlocked from the usual print and web formats and converted into linked, machine-readable formats. Our experts will use text analysis frameworks such as GATE to automatically enrich your content and extract information from it, enabling it to be converted into open formats, including RDF (the recommended format for linked data) XML, XHTML+RDFa and ATOM. The approach can be integrated with templates to improve the automation of your data publishing.

The data in more than 250,000 notices on the London Gazette website has been enriched using GATE to apply RDF and create machine-readable data. The information in the London Gazette is now available in a range of formats including print, XHTML, XML and RDF and the data is versatile enough to re-use in combination with other data.

3. Delivering open, re-usable data to engage citizens

The ultimate aim of opening up data and creating it as machine-readable, linked data is to enable the creation of new, more useful data applications for the public. TSO provides a scalable and secure environment for hosting your RDF data, making it easy for publishers and developers to build the next generation of semantic web applications. Our platform is built on 5Store, a highly scalable, clustered, commercial database storage and query engine that holds RDF data, designed and developed by Garlik who are leaders in semantic technology. Several APIs (Application Programme Interfaces) are available to extract the data, including a SPARQL endpoint.

Legislation.gov.uk was built on open data principles to enable information to be published as both human readable and machine-readable content. Users are able to browse the content online in accessible HTML format or download in accessible PDF format. The underlying data is also available in re-usable XML, RDF and ATOM formats through a published RESTful API.

Our open data publishing solutions:

However you want to open up your data, TSO can help to transform and deliver in the most effective way.

- Data hosting – a highly scalable and secure triplestore environment for hosting RDF datasets
- Data transformation – taking human readable content and transforming it into machine-readable formats
- Professional data services – helping to create URI schemes and APIs to make data re-usable
- Managed data publishing service – an end-to-end approach to managing your data publishing ensuring streamlined processes that enable simultaneous web publication in many different data formats.

About TSO

TSO (The Stationery Office) is the leading provider of information management and publishing solutions to the public sector. We are the largest publisher by volume in the UK, publishing more than 8,000 titles a year in print and digital formats. Our experts help to create, structure, capture, transform and deliver some of the most important government information. TSO provides services, consultancy and infrastructure to deliver all aspects of the information lifecycle to the highest standards for our clients.

TSO is part of the Williams Lea Group, the leading global provider of Corporate Information Solutions.

Find out more

To discuss your requirements with one of our experts please email opendata@tso.co.uk

To find out more about how TSO can help to create, enrich and engage citizens with your data visit www.tso.co.uk/opendata and to try TSO's open data tools and services visit <http://openup.tso.co.uk>

To find out about OpenUp, TSO's open data challenge, which asks data users to let us know what public data they need and how they would like it presented, visit www.tso.co.uk/openup

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



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.

