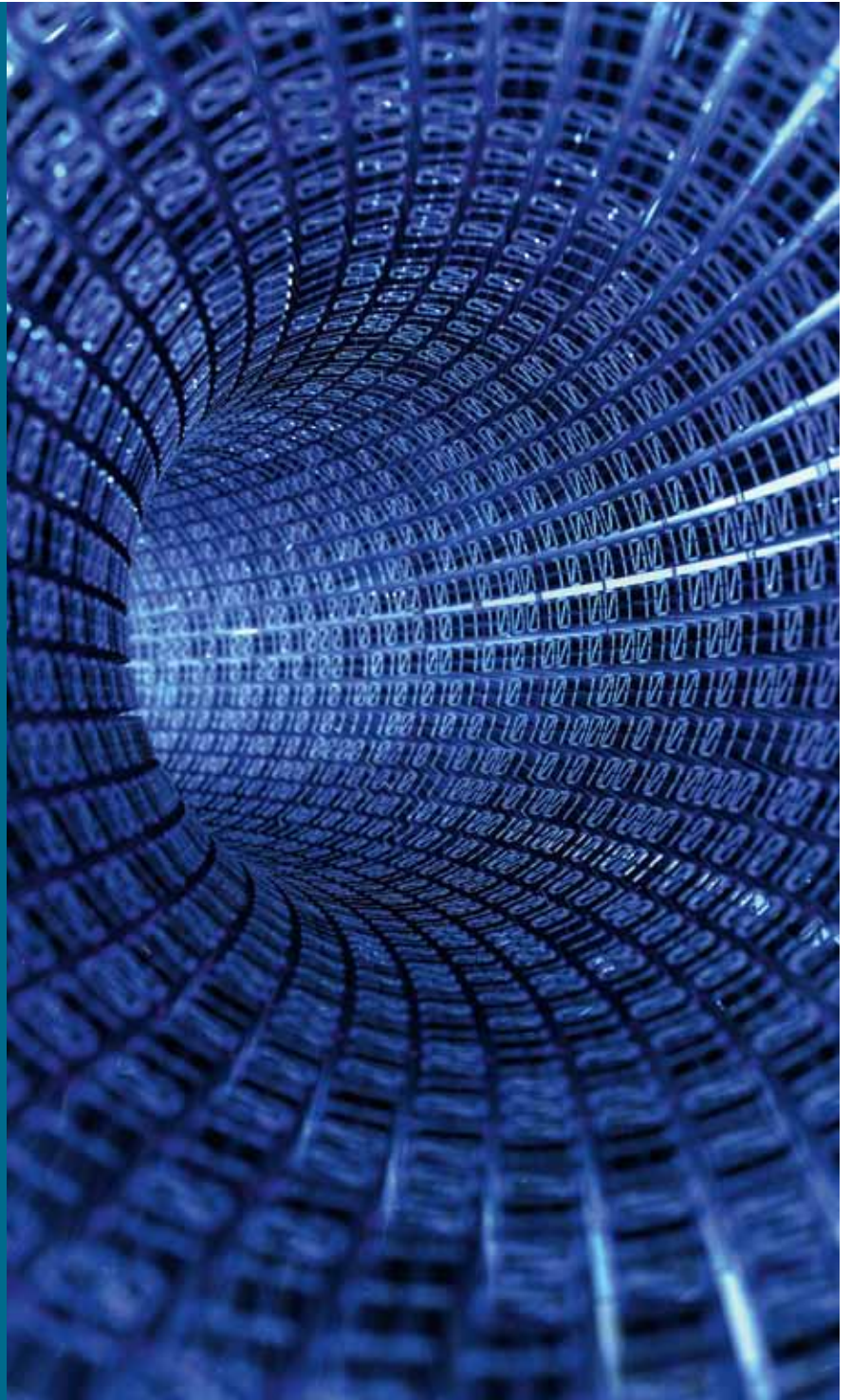
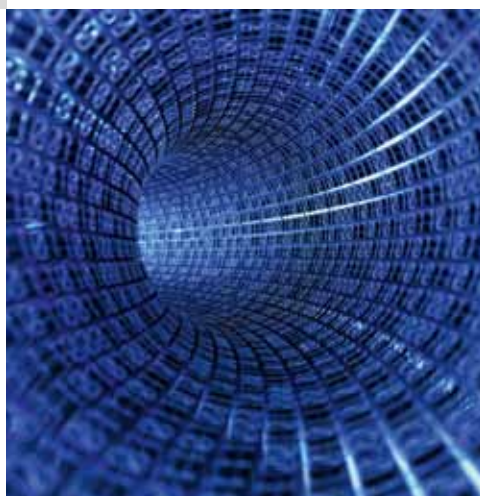


# Opening up government data

Williams Lea's solutions to  
improve data transparency  
and re-use





**Williams Lea is a Business Process Outsourcing company specialising in information management and customer communications. As part of our solution, our publishing 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 humans and machines to access the data.**

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

On 31 May 2010, Prime Minister David Cameron wrote to government departments outlining the new Government's commitment to opening up government data to improve transparency.

The Public Sector Transparency Board will drive forward the Government's transparency agenda, making it a core part of all government business and ensuring that all Whitehall departments meet the new tight deadlines set for releasing key public datasets.

In addition it is responsible for setting open data standards across the whole public sector, listening to what the public wants and then driving through the opening up of the most needed data sets.

Following the launch of data.gov.uk, the Public Sector Transparency Board has published draft data principles which include the need to publish public data in machine-readable form, unlocking it from PDFs and other unprocessable formats and making it re-usable. The principles recommend that data is re-published in linked data forms, such as RDF, which allow the most powerful and easiest re-use of data.

Public bodies should also actively encourage the re-use of their public data, working with businesses to grow new, innovative uses of data such as mobile applications.

### **Meeting the requirements of the Public Sector Transparency Board**

Williams Lea has already been helping public sector clients to publish data in re-usable linked formats, publishing using open standards and following the relevant recommendations of the World Wide Web Consortium.

We are particularly expert in publishing regularly updated, fine grained data.

Our established web publishing processes enable simultaneous publication in many different formats and ensure that the initial timely release is an accurate version of the published content in all formats.

## Three steps to opening up data

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

### 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.

Our publishing division 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.

We have enriched the data in more than 250,000 London Gazette notices to create machine-readable data.

### 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. An example data mash up, combining data from different sources, can be found on the London Gazette website: [www.london-gazette.co.uk/demo](http://www.london-gazette.co.uk/demo)

Legislation.gov.uk was built on open data principles to enable information to be published as both human readable and machine-readable content.

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

The ultimate aim of opening up access to public data and creating it as machine-readable linked data is to enable the creation of new, more useful data applications for the public. We provide 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:

- 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 Williams Lea

Williams Lea is the leading global provider of Corporate Information Solutions. As a BPO organisation, we combine consultancy, unique expertise and the latest integrated technology to develop solutions that transform the way organisations communicate and exchange their printed and electronic information. Williams Lea employs more than 9,000 staff across Europe, the US and Asia Pacific with key centres in London, Dublin, Paris, Frankfurt, Bonn, New York, Chicago, Los Angeles, Tokyo, Sydney, Hong Kong, Beijing and Chennai. We work with many world leading organisations in sectors such as financial services, investment banking, legal and professional services, retail, telecoms, automotive, energy, utilities and pharmaceuticals. Since 2006, Deutsche Post DHL, the world's leading logistics group, has held a majority stake in the company. With the long-term backing of Deutsche Post DHL, Williams Lea will continue to invest in and evolve its Corporate Information Solutions proposition that represents the future of the BPO space.

TSO, the leading provider of publishing services to the public sector, became part of the Williams Lea Group in 2007. TSO has been at the forefront of working with public sector clients to open up published data. For more information please visit [www.tso.co.uk/opendata](http://www.tso.co.uk/opendata)

## Find out more

Our experts would be delighted to talk to you about how our services can help to open up your organisation's data. Please contact us at [publicsector@williamslea.com](mailto:publicsector@williamslea.com)

**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.

### 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.



**WILLIAMS LEA**  
corporate information solutions