



Opening up data for transparency and re-use

Information Sheet

Data Enrichment Service (DES)



What is the Data Enrichment Service?

TSO's Data Enrichment Service (DES) is information extraction infrastructure for processing plain documents. It enables you to extract entities from text, XML or XHTML and output them in RDF/XML, JSON, XML or XHTML+RDFa with the last two adding mark-up to the original document.

The DES is built upon the GATE system from the University of Sheffield (<http://gate.ac.uk>). It can potentially run any GATE application and we can scale out to provide higher document throughput if required.

In addition to the main infrastructure, the DES Starter is an application that can bootstrap extraction processes by marking-up entities such as names, places, postcodes, streets and organisations using mainly UK data, with much of the work on data.gov.uk being utilised. DES Starter can easily be added to the start of your bespoke application.

TSO can also develop bespoke applications for the Data Enrichment Service. For example, we have developed an application for legislation which extracts, marks up and makes available the effects appearing within legislation so that it can be amended by approved parties. The goal of this project is to make significant savings to government and value-add for all classes of legislation user. Bespoke applications are accessed in a simple way, via the web, using the standard DES API.

The DES can output to a wide range of formats. As part of the OpenUp® platform, RDF/XML and XHTML+RDFa outputs can be stored in the 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.

How can the DES help me and my organisation?

The DES takes textual information from sources and adds value to the data to make it useful to other computer systems. The DES can benefit many different organisations that need to unlock valuable information from text. The purpose of the DES is two fold:

- To identify key elements within data and introduce machine-readable mark-up
- To relate the mark-up to other data sources and create linked data

How can I access the DES?

You can access the DES through our OpenUp Platform at <http://openup.tso.co.uk/developer/des>

The service is free to use within a certain limit. For data publishers needing to enrich more than 10,000 documents per day with an SLA, we can offer the Professional version of the Data Enrichment Service for a monthly fee.

Find out more

To find out more about how TSO can help to create and enrich data, visit <http://openup.tso.co.uk>

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

More information on the DES – Data sets used

Sources of data that drive the DES Starter include:

- Data.gov.uk
- Legislation.gov.uk
- Ordnance Survey
- Wikipedia
- Geonames
- DBpedia

Provenance

The provenance of information can be an important factor for certain users. To that end the DES is starting to implement the Open Provenance Model (OPMV) as is being created with the ongoing work being done by data.gov.uk. More details on OPMV can be found at SourceForge.

Currently the DES is adding in provenance information for RDF/XML serialisations of documents. As the OPMV work is still ongoing the information generated in this respect is subject to change.