

### Market Need

Unstructured and semi-structured data within an enterprise can be many folds larger than structured data. But this data is often difficult to leverage to new business opportunities due to a lack of domain models for the data.



Figure 1. Untamed Assets: Unstructured and Semi-Structured Data

Identifying concepts held in existing data and the relationship between these concepts is a crucial step in opening up new product potential from existing data. The OntoCore Project helps to achieve this goal.

### Technology solution

**OntoCore** - a tool-kit for the semantic enrichment of unstructured and semi-structured data with relevant concepts and relations extracted from publicly available Linked Open Data (LOD) repositories.

### Technology Challenge

While augmentation of data with an existing external data source is technically straightforward, the technical challenge in this project is to select the 'most relevant' information for augmenting the input data.

Our research has focused on the mechanisms for selecting:

- The correct senses of concepts and relations found in input text; and
- The identification of a limited set of most relevant LOD concepts for a given input document.

### Work Flow

- Textual data is fed in to OntoCore for analysis
- Key semantic entities in the data are identified
- Relevant Linked Open Data is selected for augmenting the input data stream
- Original text data is augmented with most relevant conceptual and relational data through document meta-data

### Research Team

**Dr. Eoghan O'Shea**, Dublin Institute of Technology

**Dr. Robert Ross**, Dublin Institute of Technology

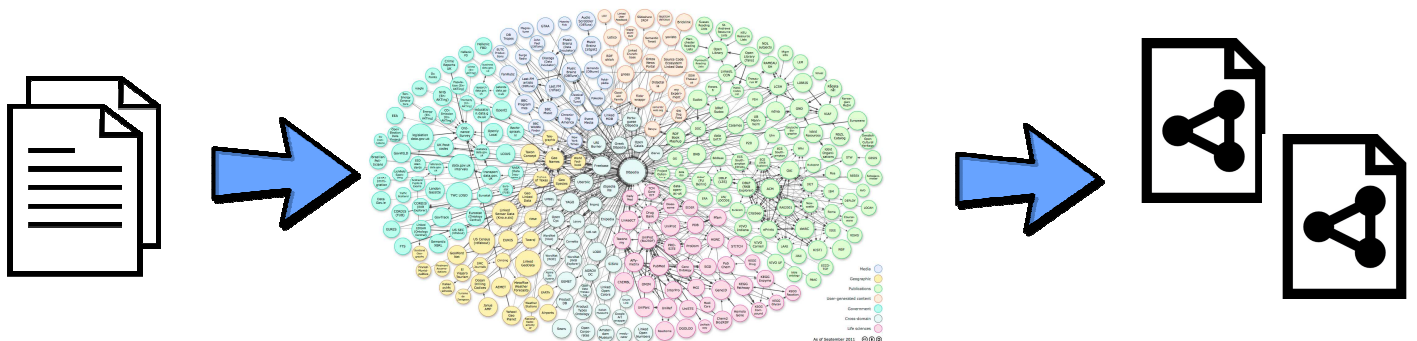


Figure 2: Document Transformation – Raw Unstructured documents augmented through reference to Linked Open Data Repositories

Follow CeADAR Ireland:

