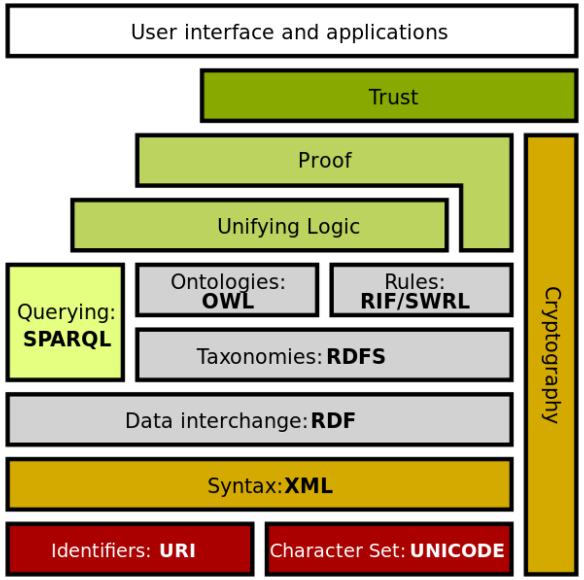
Introduction

- NPL generates large volumes of scientific and engineering data. These data are either directly related to the science under study (e.g. variables measured for a given scientific phenomenon), or are used to describe the experiment process (meta data).
    - Example: In a laboratory that studies the stress-strain relationship for a particular material, where the physical geometry, type of material and stress and strain values are direct values that are relevant to the material under study. Additionally, the temperature, pressure and humidity of the lab environment may also be important factors that affect the data collected, and consequently the physics of stress-strain relationship for the material.
- Current data curation processes at NPL vary considerably across the laboratory. Some laboratories such as Mass Spectrometry deploy automated data collection, analysis and storage, while others collect data in notebooks and transcribe it using various vehicles such as Excel, Matlab, LabView, Python, C# etc.
- The processed data are then typically used in an output product e.g. calibration certificate, scientific paper, software, measurement device etc. Beyond this point no further use is made of the data. Two factors contribute to this: 1) the data cannot easily be discovered (no recovery), and 2) the data generation process is not documented (no reproducibility). Even if the data are found, there are major limitations to the ability to reuse this data due to lack of the data curation process. In addition to the loss of data curation information, the data provenance is also in general, lost.
- A PoC development system is currently being developed for the newly renovated Stress-Strain laboratory where Advanced Materials will be tested. An advanced material ontology will form a core component of this system.

Solution: Linked Data Graphs



Unlike traditional database schemas that are ‘fixed’ between phases of development, linked data graphs stored as Subject-Predicate-Object triples are ‘unfixed’. This provides a storage system that evolves dynamically and reflects changes in data structures without the need to re-define the database schema.

However, the applications and human actors that interact with the triplestore need to be intelligent enough to ‘understand’ what is meant by the data within the system – the **Semantics**. To do this requires an ‘**Ontology**’ – a dictionary of terms and relations between them.

In short, what does this technology stack give us? A searchable, flexible storage environment that can grow as new knowledge is gained without the need to have to re-design the system from scratch every time the data structure changes.

Other Relevant Activities

1. NPL Materials Centre staff are on many international standards committee and frequently involved in vocabulary for new emerging areas. For example, Prof. Mark Gee took a leading role in the preparation of the vocabulary standard EN ISO 80,004 (several parts) for nano-technologies.
2. Prof. Graham Sims as International Chair for the G15 + EC, Versailles Project on Advanced Materials and Standards (VAMAS) pre-normalisation was the Steering Committee member responsible for Technical Working Area (TWA) 35 on “Interoperability of Materials Databases”.
3. Prof. Graham Sims has led the NPL input into the Bureau International des Poids et Mesures (BIPM) initiative to generate an International Metrology Resource Registry (IRMM) for high quality data, where a widely supported ontology and metadata are important aspects of the success of this initiative supported by National Metrology Institutes worldwide.
4. Prof. Graham Sims has been an invited plenary speaker at two conferences in the Asian Materials Database series, including in 2012 speaking on “Material Databases - their role in eco-design and sustainability”.
5. Prof. Graham Sims chairs a working group on the World Materials Research Institute Forum (WMRIF) on “Promotion of Global Databases”.