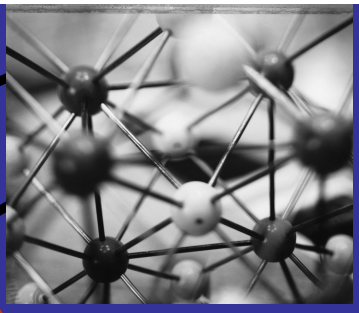
Using Standardized XML Models to Enable Semantic Warehousing



Introduction

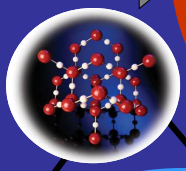
The use of Extensible Markup Language (XML) in healthcare and life sciences (HCLS) is rapidly spreading. The expressive power of XML is crucial to describe the complex phenomena in HCLS. For purposes of biomedical information representation and exchange there is a need to standardize the XML content to support semantic persistency, computation, and interoperability.

- ✓ Information standards can be used to create an underlying data model for semantic warehousing, in addition to their common use for information exchange.
- ✓ Inbound data can be persisted into an XML-based warehouse in its native XML format, as mandated by some HCLS standards.
- ✓ This makes it easier to preserve the full richness of the source information being integrated in a warehouse, while revealing the similarities found in data sets received from multiple data sources.

The Challenge

While XML is very expressive it is also very flexible thus allowing slightly different representation for the same phenomenon. This a challenge in a clinical environment that requires algorithms for identifying semantically isomorphic XML structures.

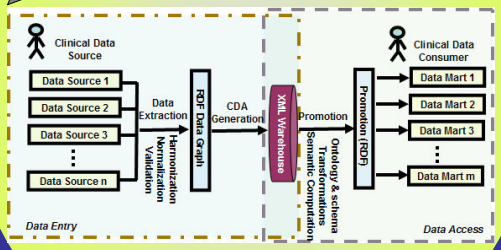
In research, XML can be used to preserve the slight differences between disparate data sets that need to be integrated, while revealing the similarities among the source data sets.



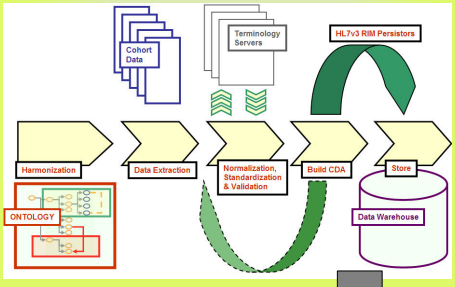
Our Goal

To create a Biomedical Information Infrastructure (BII) that has at its core an XML warehouse, based on internationally approved HCLS standards (implemented in XML). The BII will be equipped with direct access mechanisms, along with services to export subsets of data to user-defined marts.

Overview of XML Warehousing Equipped with Data Entry and Access Services



Data Entry Work Flow



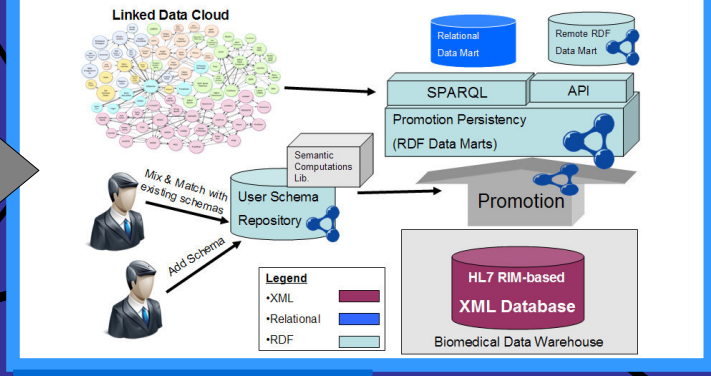
The data and knowledge entry processes heavily rely on internationally controlled terminologies, along with a project-specific ontology that defines the conceptual organization of terms used in that project.

A standard data instance is generated from the template model using a common instance generation mechanism

Clinical Instance

```
<component contextConductionInd="true" typeCode="COMP">
    <section classCode="DOCSECT" moodCode="EVN">
        <code code="29762-2" codeSystem="2.16.840.1.113883.6.1"
            codeSystemName="LOINC" displayName="Social History"/>
        <entry contextConductionInd="true" typeCode="COMP">
            <observation classCode="OBS" moodCode="EVN" negationInd="true">
                <templateId root="2.16.840.1.113883.3.18.99.1.1.4.1"/>
                <code code="77176002" codeSystem="2.16.840.1.113883.6.96"
                    codeSystemName="SNOMED CT" displayName="Smoker"/>
            </observation>
        </entry>
    </section>
</component>
```

From a Warehouse to Multiple Marts



Using 'promotion' processes, it is possible to promote certain XML contents to variables within a user-defined mart schema.

Contact information:
Carmel Kent
IBM Haifa Research Lab
E-mail: carmelk@il.ibm.com