



Unlocking the Power of Data Intelligent Data Mark-up

July 2010

By Rosa Bancarotta

Data feeds Knowledge and Knowledge is Power

Collecting vast amounts of raw data can be easy; extracting valid information from that data is the challenge. Here is where the power of intelligent mark-up of the data gives the user access to information that ultimately translates into “knowledge”.

Data collection is the first step to information. Extracting information can be powerful, but, only when you can dissect the data, are you able to interpret patterns and trends that give you the knowledge to make decisions that would otherwise take hours of analysis and interpretation.

Data Mark-up

Markup of data had its beginning in the publishing and printing industry. It was developed to provide typesetters with instructions as to how things should appear on the printed page. In the electronic world, intelligent markup language is used to specify the semantic structure of data items, not their layout. The advantage of intelligent markup language is the ability to define data elements and make it easy to construct networks of interconnected information.

Technology has made it very easy to create digital content, either through original writing using tools such as MS Word, or by converting paper documents into a digital form using OCR techniques. It has also become rather easy to capture and store graphic images such as photographs, drawings, charts, graphs, etc. from multiple information sources. There are now massive volumes of information available but to make the information valuable it must be accessible.

The volume of the world’s digital content is estimated to exceed 500 billion gigabytes. Printed and bound into books, this digital content would stretch from Earth to Pluto – ten times! The speed of digital data creation is ever increasing and it is an impossible task to keep up with the mind-boggling abundance of raw data. Why not add analytics, on top of intelligent data, to draw relationships and create tools which can sort data for you?

The quest to gain knowledge from content is served best by intelligent mark-up that gives the data the structure that is necessary to define the relationships between elements. In turn, intelligent mark-up allows us to rapidly access any relevant data for specific searches or specialized applications. The more specific the

notation, the more flexibility the user has to format the data, extract it, cross reference it, evaluate it, etc.

The Power of XML

Many of us have been exposed to the data standard, XML (eXtensible Mark-up Language). XML allows us to semantically define the elements, attributes and interrelationships of data elements using tags. These tags can be used to simply define the hierarchy of the contents of a document or they can be used to define the content with greater definition and specificity, thereby providing the opportunity to leverage the content for use by a multitude of processes and presentation formats, depending on the desired results.

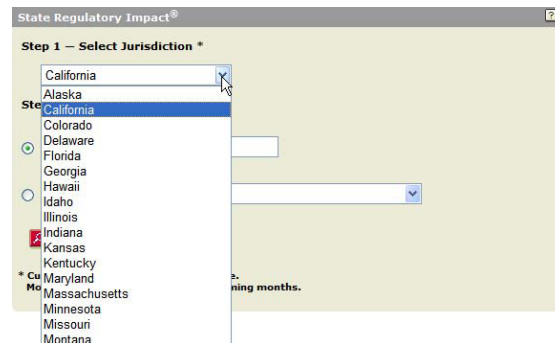
As important as the intelligence of the tagging structure are the tools used to help collect and develop, edit, and manage well-formed content. Powerful Content Management Systems (CMS) are available to help companies manage large amounts of data and these systems can be customized to provide solutions tailored to address specific needs. CMS allows the user to manage content access, editing, review, reporting, analysis, workflow and archiving.

A master repository of structured data is the base for a powerful CMS. Without data structure, effective searching, retrieval, and evaluation would be virtually impossible.

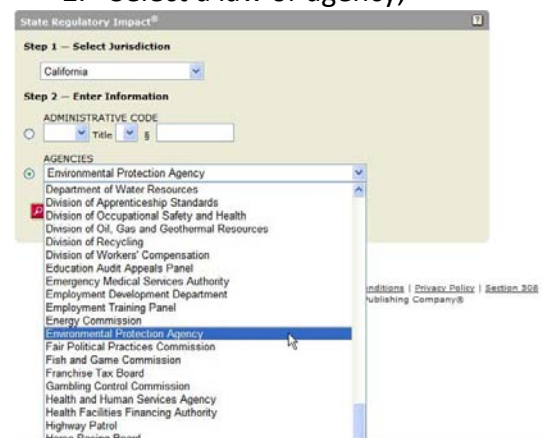
Example Analytics Tool

By applying sophisticated mark-up to bills, laws, registers and regulations, tools can be applied to help analyze and track the latest developments in the law. For example, in use today on Capitol Hill, any legislative changes impacting any area of interest are available as soon as they happen. Here is one example of the result of intelligent mark-up of data:

1. Select a jurisdiction,



2. Select a law or agency,



- Review a report with linkages to the underlying documents.

California
Environmental Protection Agency

Start Date: End Date:
Filter: 12/16/2008 through 6/16/2009 Last 6 Months [Custom](#)

241 actions found for this date range.

Tuesday, June 9, 2009

12 [2009 CA 13454](#) (3 amendments)

Reg Text	Reg Version	Version Date	Code Cite	Agency
2009 CA 13454	Proposed-Rule-Notice	June 5, 2009	13 CCR 2493	Environmental Protection Agency
2009 CA 13454	Proposed-Rule-Notice	June 5, 2009	13 CCR 2493.1	Environmental Protection Agency
2009 CA 13454	Proposed-Rule-Notice	June 5, 2009	13 CCR 2493.2	Environmental Protection Agency

Tuesday, June 2, 2009

12 [2009 CA 13387](#) (1 amendment)

Friday, May 15, 2009

12 [2009 CA 13250](#) (9 amendments)

Wednesday, May 13, 2009

12 [2009 CA 13346](#) (11 amendments)

Conclusion

By understanding the structure and uses of the data, sophisticated XML mark-up can be applied to documents, making them infinitely more valuable as an ongoing information source. This mark-up in turn allows for the creation of analytics tools like Impact to be able to draw relationships across data sets in order to mine knowledge. Data leads to knowledge. Knowledge leads to power.