Modeling Complex Metadata with MarkLogic Server

The proper modeling and management of Metadata is one of the most challenging information architecture tasks in IT today. Real world metadata is hard to model. It's messy, irregular, and often unpredictable. Moreover, Metadata about people is among the hardest objects to model because of the many differences that exist between individuals (cultural, racial, geographical and historical factors to name a few provide thousands of different characteristics), but also because of the different sources and types of information that feed a complex Metadata repository sometimes called an ‘Enterprise Metadata Catalog’ – a common repository of information that needs to adapt to the differences in schema, version, semantic meaning, source format and type and ultimately the context in which that Metadata is needed since the notion of Metadata is contextual anyway (i.e. your metadata is my data). Therefore, developers need a broader awareness of these realities when modeling and implementing systems.

The challenges that developers, architects and IT managers faced when dealing with complex Metadata (as is the case with genealogical data) are several:

- **Information Access:** Disparate stores, Multiple search tools, Multiple formats, etc

- **Information Model:** Multiples or sometimes non-existing schemas, different versions of ‘the truth’, evolving Metadata models, corner cases that are difficult to fit into the standard, etc

- **Information Quality:** Poor correlation, no semantic context, limited information sharing, etc

These challenges make the typical Metadata Management approach of storing key/value pairs in a relational model inadequate to solve all possible scenarios in which Metadata can present itself. Artifacts like: Internal Markup inside the Metadata values, Hierarchy in the Metadata structure or the dreaded mixed content (both structured and unstructured), can put serious constrains on a Metadata model and implementation that is not suitable to support these variances not only in the model itself, but in the versions of that model that are inevitable to be generated as the information needs change. In addition, performance can be severely impacted when these complex and ever-changing models are not supported by the database and the business layer needs to perform the heavy lifting.

The Modeling Complex Metadata with MarkLogic Server presentation uses the example of a ‘People’s profile Database’ to describe these challenges and the techniques to improve the modeling, storage, query and analysis of Metadata that describes individuals for a global research firm (this example is loosely based from working with the intelligence community, where we're used for identifying and tracking potential terrorist suspects. One can imagine the complexities of implementing a system like this, where there are thousands of unique profile templates.)
The presentation will wrap with a high level architectural view of an ‘Enterprise Metadata Catalog’ – a platform that combines all the elements necessary for a proper Metadata strategy:

1. Flexible Modeling: supporting both simple and complex entities, properties and relationships
2. Robust storage: supporting unstructured, semi-structured and fully structured Metadata models.
3. Flexible Queries: including Fielded, Full-text, Geospatial, Paired or a combination of all.
4. High levels of performance and scalability: to manage billions of individuals in real-time.

And will conclude with a list of benefits expected of such implementation.