

# *XML, Schema, Names, and Confusion*



Dan Gillman

Bureau of Labor Statistics

Washington, DC

# *Introduction*

- Problem
  - Names are not reliable identifiers, yet they are integral to data management
- Contents - Focus on XML
  - Motivation
  - Basics
  - XML and RDF
  - Registries

# *Motivation*



- Is there a need to manage individual elements?
  - Why? To know their semantics.
- More specifically - How do you find all the elements in various schemas that mean (close) to the same thing?

# Basics

- Names
  - Terminology - designations
  - Compare with Concepts
  - “A rose by any other name smells just as sweet.”
- Name reuse
  - Same name represents different concepts
- Concept reuse
  - Different names represent the same concept

- Namespaces
  - Similar to “scope” in programming languages
  - Mechanism to make names unique
  - Namespace|Name pairs
- W3C
  - Makes broad use of Namespace in XML

- Elements
  - Names are Tags
- Schemas - define syntax, almost no semantics
- Concepts
  - Unknown directly, similar to programming
  - Unique names => concepts are distinguishable within a schema
- Unique namespaces => globally unique names

- How are namespaces assured uniqueness?
  - Namespace registration
- However,
  - 1) how do we know what an element means?
  - 2) how do we know whether two elements defined in different schemas mean the same thing?

- Resource Description Framework
  - Simple ontology language
    - Model: subject - predicate - object
    - Each type is a resource
    - Resource identified by URI
  - Representation defined in XML
  - Provides meaning for elements
- So, RDF is answer to previous questions

- Or is it? Author of RDF may
  - define many elements
  - link elements with similar meaning from different schemas
- But,
  - 1) how do you find these schemas with similar elements?
  - 2) how do you determine which are similar?

# Registries

- ebXML/OASIS Registry
- UDDI - universal description, discovery, and integration for business on the web
- Need a metadata registry that
  - Manages semantics of data
  - Requires each resource be registered, for reuse
  - Provides mechanism for locating and linking similar items

# *Registries*



- Standard
  - ISO/IEC 11179

# *Conclusion*



- Questions