

Managing Data Dictionaries
using
LDAP
(**Lightweight Directory Access Protocol**)

Jim U'Ren

NASA/Jet Propulsion Laboratory

August 18, 2000

Problem

Difficult to communicate relevant information about data dictionaries between applications:

- **vocabularies** - human readable collections of terms and definitions pertaining to a domain
- **data dictionaries** - machine interpretable collections of data elements
- **schemas (information models)** - structured, machine interpretable collections of information models consisting of structured relationships between data elements

What is Needed

- A mechanism that can be used to access, publish, update, relate and integrate data dictionaries (vocabularies, data elements and schemas)
- The mechanism must be able to span domains and sub-domains e.g. engineering, science & administrative
- The mechanism must have both manual and automated interfaces
- The mechanism should follow the distributed service model (e.g. DNS (Internet Domain Name Service, x500 Directory, etc.)

A Solution

Develop a distributed data dictionary service using:

- LDAP* Internet service protocol
- ISO11179 schema

to store vocabulary, data element & schema
information

*LDAP (Lightweight Directory Access Protocol)

Why use the LDAP protocol?

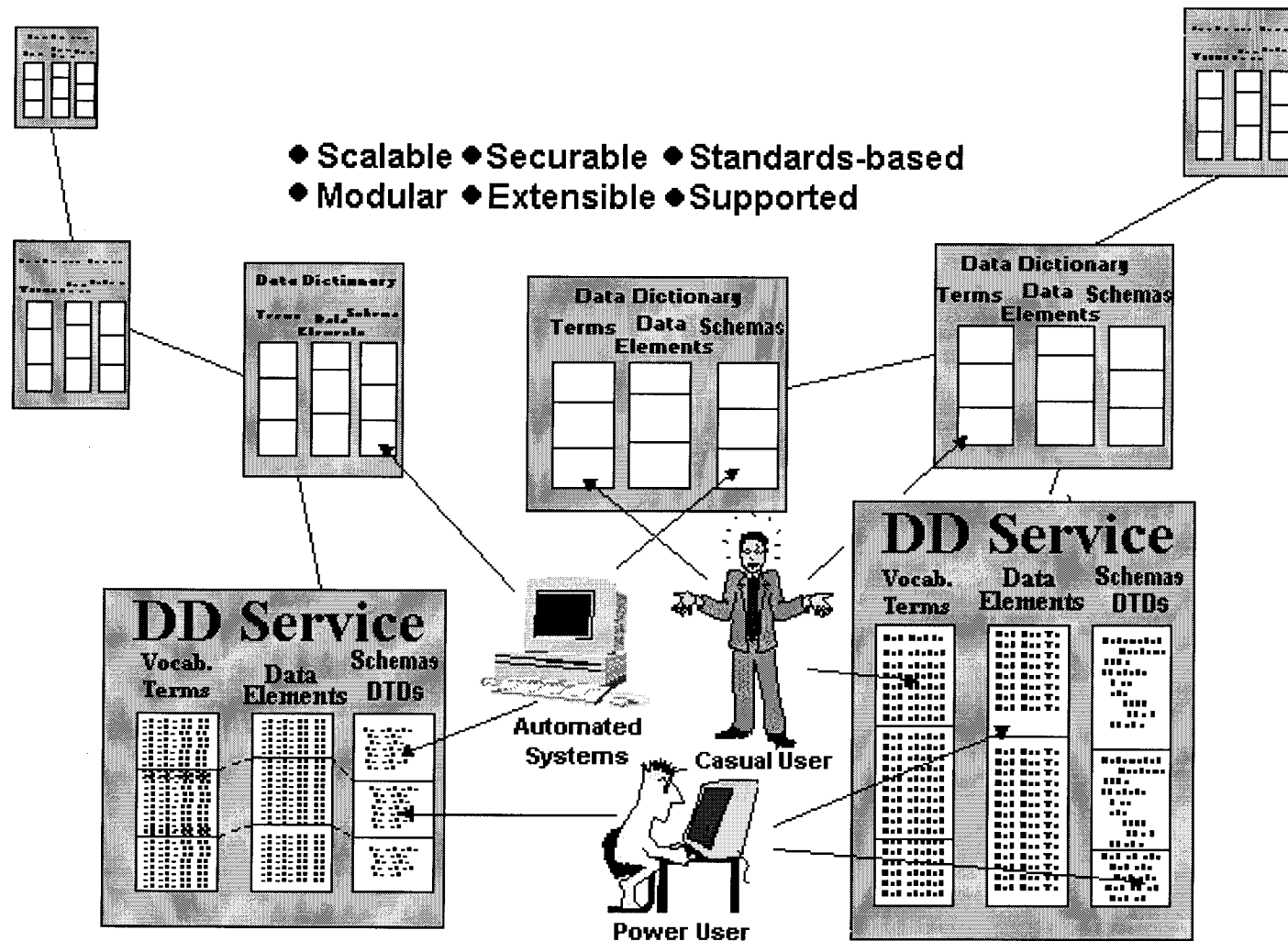
- **Universal access** - LDAP is the Internet directory standard, widely adopted and implemented by numerous vendors and open source software solutions
- **Simplicity** - LDAP is a relatively simple protocol with a straightforward API
- **Extensibility** - schemas can be easily extended and adapted
- **Access control and security** - LDAP connections can be authenticated (requiring a password or other credentials) and secured through SSL
- **Multi-platform development** - C, Perl, Java, JavaScript and other APIs are available, making LDAP directory services accessible from virtually any language, platform, or development environment

Why use the ISO 11179 schema?

- an established international standard
- widely supported by organizations & government agencies - US Census Bureau, NIST, Defense Information Systems Agency, Environmental Security, DoE, DoJ, Bureau of Labor Statistics, DoT, EPA , etc.
- flexible use of elements within the schema
- can be easily implemented in an LDAP directory service - flexible and easily configured LDAP servers are well suited to flexible 11179 schema

Distributed Data Dictionary Service

using the LDAP Protocol and ISO 11179 schema



Terminology Lookup Scenarios

- An end user needing to clarify use and meaning of a word used within a specific context, performs a multi-domain vocabulary lookup across multiple DD services looking for the published vocabulary of the referenced domain
- An end user, confronted with a number of new acronyms used in a presentation, accesses a local DD Service to lookup the acronyms based within probable domains eliminating the alternative meanings e.g.. STEP standards work vs the JPL STEP project
- As a search engine scans through a document, it discovers a keyword list and finds a "reserved word"; the document includes a reference to a domain specific vocabulary list in a DD service; the search engine uses this vocabulary to be certain it is indexing the keywords in the right context
- An engineer or scientist, writing a technical paper, needs to include a glossary of relevant terms in the paper; by performing a multi-service search, terms and definitions that relate to the topic of the paper are quickly found and inserted into the paper with the corresponding attributions.

Validation Scenarios

- A system integrator receives an MCAD geometry model (e.g. STEP AP203 Part 21 file) of a component to be integrated into any assembly; automatically, a standard validation routine is performed that checks for use of the standard units of measure as called for in the contract
- As a STEP model is checked into a PDM system, an automated validation routine checks the model using the schema (located in the DD Service) that is identified in the Part 21 data file
- An MCAD geometry model is sent from design to thermal analysis and validation is performed using the correct schema version as referenced in the model; validation is an automated process that occurs before any work is done with the model as it is transferred between domains

Other Scenarios

- Data modeling - a data modeler, charged with developing a information model for a new application, uses data elements published in several DD Services (much like a parts library) ensuring that the new information model will have compatible interfaces with data sets that share the same data elements or collection of elements.
- Creating a TDP (technical data package) - an applications performs a schema check against objects about to be wrapped into a TDP (e.g. STEP AP232 or PDM Schema TDP) to ensure their correct structure and meta-data content
- Data integration - an analyst, charged with integrating data from two or more data sets, accesses the “correct” version of each schema as referenced in the data set from the “DD service space” allowing them to identify / map interfaces between the data sets e.g. MCAD-ECAD-Cost data
- Extending a schema - to solve a local problem, a data modeler uses data elements from a published collection of data items to extend an existing “official” schema; the new schema is published in the DD Service with traces/links back to the “official” schema

What's next

- Begin populating the prototype LDAP DD server(s) with real vocabulary, data element and schema sets
- Test various LDAP domain naming conventions
- Test client interface(s) e.g. Perl, JAVA, etc.
- Link LDAP DD servers (include GSFC - Waterbury)
- Test cross domain and cross server searching
- Test “wrapping” existing DDs with LDAP (e.g. PDS, DMIE)
- Publish whitepaper on LDAP DD Service discussing architecture and prototype implementation (ETA - Fall '00)