


the groups, present in a substance
quality • |'kwäləte| n. (pl. qualities)
something as measured against a
lar kind; the degree of excellence
product quality
provement in life.
quality of general excellence of standards
for connoisseurs of quality
ty beers. ■ archaic high

2008 EPA Conference on
Managing Environmental Quality Systems
April 23, 2008

Integrating the EPA System of Registries into Your Quality Program



Cindy Dickinson
U. S. EPA
Office of Environmental Information
Data Standards Branch

John Tooley
State of Washington
Department of Ecology
Applications and Data Services

Burdell Schwartz
CGI Federal
Project Manager for
System of Registries

For Conference Purposes Only





Agenda



- What is a Registry
- Types of Registries and Registry Services
- How Registries Support Quality for Data Products and Services
 - Functions
 - Internal Controls
 - Customer Services
 - Efficiency
- Using the Registries in the State of Washington
- Demonstration
- Discussion – Using Registries in EPA and Partner Quality Programs



1



What is a Registry?

- Registration
- Stewardship
- Oversight
- History
- Partnership with customers


2




Types of Data in the EPA Registries

- Registries support management and use of
 - Structured data
 - Unstructured data
- Registry content includes:
 - Controlled vocabularies (Terminology)
 - Classification schemes
 - Code sets
 - Data standards
 - Information about what is in EPA systems
 - System descriptions and classifications (where to find the system and its accessibility)
 - Data dictionaries
 - Reusable components and automated services

3



What are Registry Services?



- **Purpose**
 - Ways to integrate good quality metadata into EPA business processes and systems
 - Supports discovery, access, reuse, and understanding
- **Registry services include:**
 - People and processes to assist stewards in registering, maintaining, relating, and organizing registry content
 - Support for developers to locate existing standards, code sets, web services, XML schema, etc. (so they can be reused)
 - Automated (machine to machine services)
 - Supplying search engine content
 - Presenting EPA vocabularies on the web
 - Presenting data definitions for users of structured and unstructured data (active metadata)
 - Downloading data standards to program systems and warehouses
 - Translating code sets
 - Custom support from specialists with expertise in terminology, data concepts, and classification development and management

4





Quality for Data Products and Services



- **Registries support quality through:**
 - Terminology and data management functions
 - Internal controls
 - Customer service
 - Efficiency
- **Quality features – many!**



5



Registries Support Data Quality Functions

- **Terminology Management**
 - Processes and repository to hold, maintain, and provide access to environmental terms, taxonomies, glossaries, acronyms for reuse
 - Ability to tag terminology by concepts and specialized taxonomies (such as the Web Taxonomy) so search engines can find information
 - Ability to store definitions and “meanings” so data are understandable once it is found
 - Promotion of best practices in terminology management (to result in common, shared, controlled vocabularies)



6



Registries Support Data Quality Functions (continued)

- **Data Management**
 - Information about and access to data standards and their use
 - System inventory and data dictionaries to record information about important agency data sources
 - Ability to associate dictionary data elements to concepts and catalog them with specialized taxonomies so search engines can find where information might be located in EPA systems (future)
 - Compare tools to discover similarities and/or differences in data across systems
 - Processes and repository to hold, maintain, and access simple reusable code sets
 - Processes and repositories to hold, maintain, and access complex substance and facility standards
 - Reusable software code, XML schema, web services for developers
 - Promotion of best practices in data management



7



Registries Provide Internal Controls

- **Three tiered approach to stewardship**
 - Data stewards enter and maintain information (for systems, data dictionaries, code sets, or reusable components) – act as Subject Matter Experts
 - Managers validate information and determine appropriate level of completeness (for systems it is the IMO and IRM Branch Chiefs
 - Data Standards Branch and its contractors (with terminologists and metadata expertise) provide oversight to assure that:
 - Data are consistent
 - Concepts are correctly applied
 - Services work



8



Registries Provide Customer Services

- **Accessible data**
 - Find systems where data are stored
 - Search for information by taxonomies
 - Search for information by concept and meaning (future semantic web)
- **Data that are understood**
 - Sources identified and authentic
 - Names, definitions, concepts, meaning available
 - Data translated into user's preferred vocabulary or codes
- **Services transparent to user**



9



Registries Support Efficiency

- **Enable use of data standards**
 - Locate international, national, federal and EPA standards
 - Provide guidance for implementation
- **Promote reuse of data as appropriate**
 - Find and reuse system data elements
 - Find and reuse code sets
 - Find and reuse taxonomies
 - Find and reuse terms and glossaries
- **Promote quality development**
 - Reuse web services, system components, data models, and XML schema that work with our data



10



System of Registries Development Schedule

- **Release 1 – July 2008**
 - New terminology services user front end (supporting retrieval of terms and vocabularies) with link to COTS software
 - New substance registry services user front end (supporting retrieval and stewardship of substance information)
 - New data standards web page (one stop shopping for EPA developers)
 - New data registry COTS software (supporting standards, data dictionaries, and code sets)
 - New SOR home page
 - New subscription service for customers
 - New training registration module for customers
 - Incorporates existing EPA System Inventory (READ)
 - Incorporates existing links to Envirofacts to access the Facility Registry System
 - Incorporates existing links to Exchange Network XML Registry
 - Collaboration software available to each registry service



11



System of Registries Development Schedule

- **Release 2 – November 2008**
 - New data dictionary user front end (supporting retrieval and stewardship of dictionaries)
 - New code set module user front end (supporting retrieval and stewardship of code sets)
 - New custom glossary creation capability
 - New presentation web services to allow glossaries in the terminology services tool to appear on Agency web pages
 - Substance web services re-engineered to meet Exchange Network Node 2.0 specifications
 - Active link between data standards web page and data registry to improve access to data standards
 - Active link between system inventory (READ) and system data dictionaries in data registry

12



System of Registries Development Schedule

- **Release 3 – February 2009**
 - New compare tool for the data registry (to compare agency data elements from one system to another and from systems to data standards)
 - Improved tracking of data standards conformance in the data registry
 - Launch of Reusable Component Services
 - Web services catalog including existing UDDI (Universal Description, Discovery, and Integration) capability
 - XML Registry broadened to support any EPA XML schema
 - Possible schema builder workbench tool
 - May be additional services
 - READ enhancements

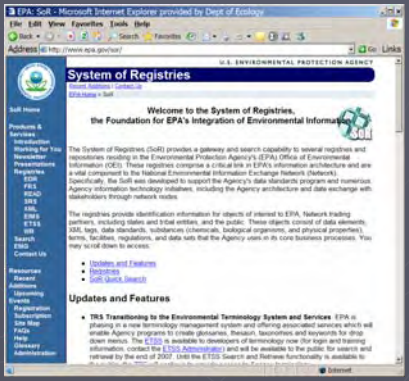
13

Customer Perspective - Using the Registries

System of Registries Washington State Perspective

14

System of Registries (SoR) Washington State Perspective



The screenshot shows the EPA SoR website with the following content:

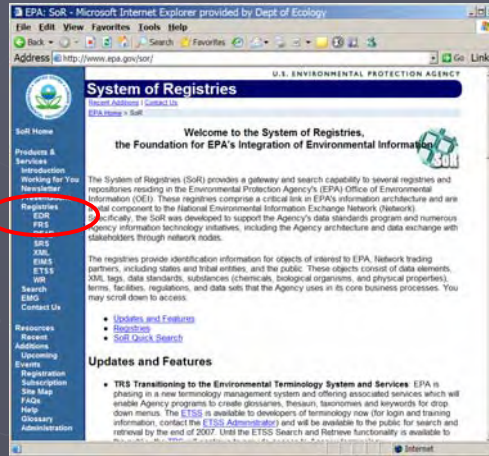
- System of Registries**
- Welcome to the System of Registries.**
- the Foundation for EPA's Integration of Environmental Information**
- Updates and Features**
- This Transitioning to the Environmental Terminology System and Services**

John Tooley
Washington State Department of Ecology

15

System of Registries - EDR

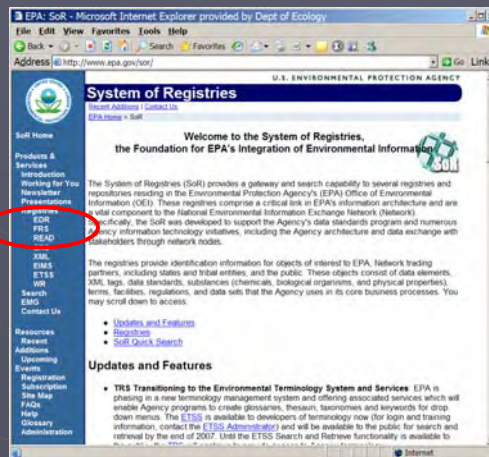
EDR
Environmental
Data
Registry



16

System of Registries - FRS

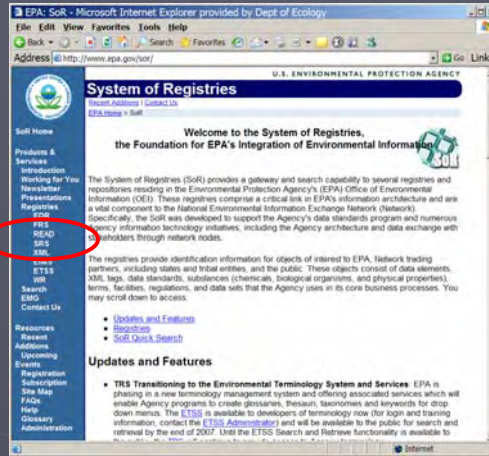
FRS
Facility
Registry
System



17

System of Registries - SRS

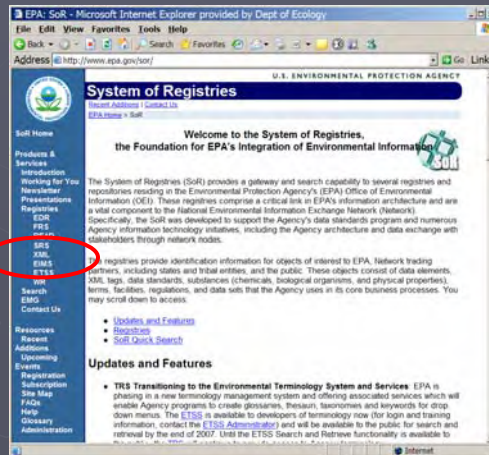
SRS
Substance
Registry
System



18

System of Registries - XML

XML
Registry
System



19

Environmental Data Registry



20

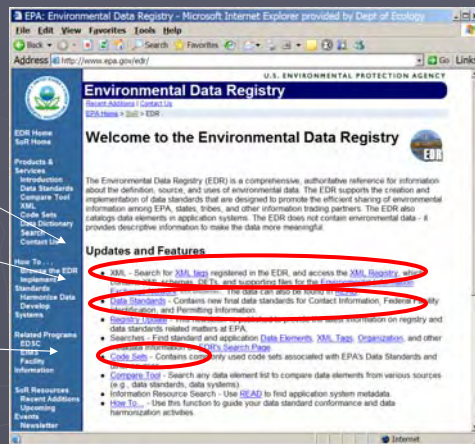
Environmental Data Registry



XML Tags
XML Registry

Data Standards

Code Sets



21

XML Tags; XML Registry



The screenshot displays two web browser windows. The left window is titled 'Search - Microsoft Internet Explorer provided by Dept of Ecology' and shows the 'Pacific Northwest Water Quality Exchange' search page. It includes a search bar and several filter sections: 'Data Source' (with checkboxes for Alaska, West US, Oregon, and Washington), 'Data Provider', 'Project Organization', 'Project Data Range', 'Responsible Organization', 'Station Name', 'Station Type', 'Station Location', and 'Location Description'. The right window is titled 'Exchange Network Node Log of Activities - Microsoft Internet Explorer provided by Dept of Ecology' and shows a table of activity logs. The table has columns for 'Primitive Name', 'Service Name', 'Service Requestor', and 'Service Date'. Below the table, there is a 'Status Detail' section with XML data.

Primitive Name	Service Name	Service Requestor	Service Date
Get Status Detail	Schedule	GetMeasures	system
Get Status Detail	Schedule	GetMeasures	4/28/2008 8:09:43 PM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 8:04:52 PM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 4:03:41 PM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 9:11:17 AM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 9:44:24 AM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 12:13:18 PM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 12:13:14 PM
Get Status Detail	Query	GetMeasurements	gpm@ecapp@deq.state.or.us
Get Status Detail	Query	GetMeasurements	4/28/2008 10:11:01 AM

Data Standards



Ecology uses, or considering, the following for adoption:

- Biological Taxonomy (Ver 2)
- Chemical Identification (Ver 2)
- Environmental Sampling, Analysis and Results: Monitoring Location (Ver 1)
- Environmental Sampling, Analysis and Results: Project (Ver 1)
- Facility Site Identification (Ver 2)
- Latitude/Longitude (Ver 2)
- Measure (Ver 1)
- Method (Ver 1)
- SIC/NAICS (Ver 2)
- Tribal Identifier (Ver 2)
- Well Information (Ver 1)

The screenshot shows the EPA Environmental Data Registry website. The page title is 'Environmental Data Registry' and the sub-page is 'Data Standards'. It includes a navigation menu on the left with links like 'Home', 'Products & Services', 'Introduction', 'Data Standards', 'Compare Tool', 'XML', 'Code Sets', 'Data Dictionary', and 'Search'. The main content area has a heading 'Data Standards' and a list of links: 'EPA Final Standards', 'EPA Standards Under Development', 'EPA Sponsored Standards', 'Program Specific Data Standards', 'Data Standards Process', and 'Data Standards Information'. Below this is a paragraph explaining the vision of the EPA data standards program and the role of the Environmental Data Standards Council (EDSC). At the bottom, there is a section titled 'EPA Final Standards' with a paragraph of text.

Codes Sets

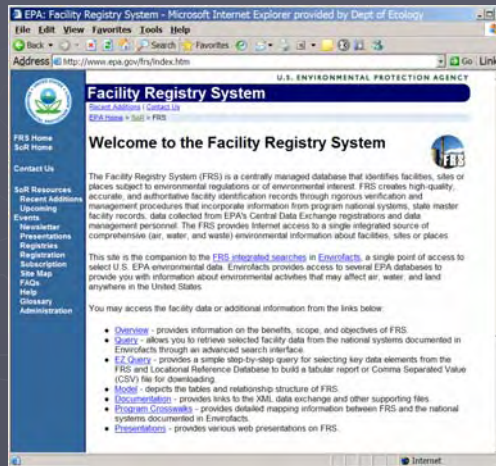


Ecology Uses:

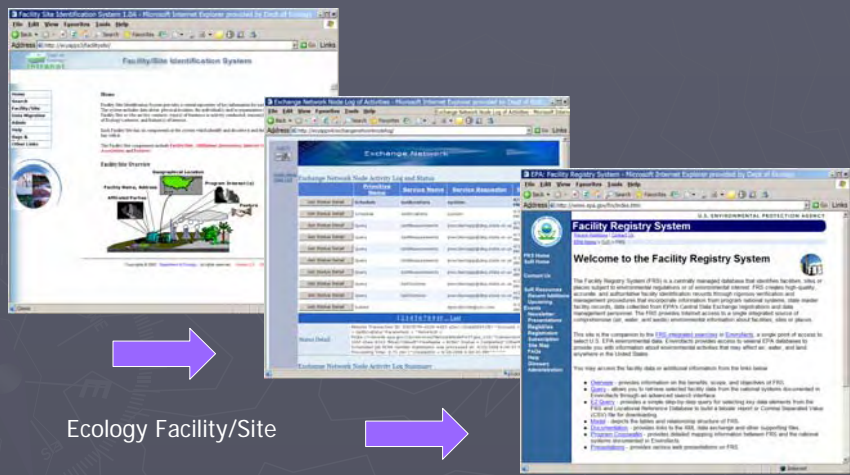
- Coordinate Data Source Codes
- Coordinate Data Source Names
- Facility Types Names
- Geometric Type Codes
- Geometric Type Names
- Horizontal Collection Method Codes
- Horizontal Collection Method Names
- Horizontal Reference Datum Codes
- Horizontal Reference Datum Names
- Reference Point Codes
- Reference Point Names
- Tribal Land Names
- Tribal Names
- Verification Method Codes
- Verification Method Names
- Vertical Collection Method Codes
- Vertical Collection Method Names
- Vertical Reference Datum Codes
- Vertical Reference Datum Names
- Geographic Location Collection Methods



Facility Registry System




Facility Registry System



Ecology Facility/Site Exchange Network FRS Flow to EPA

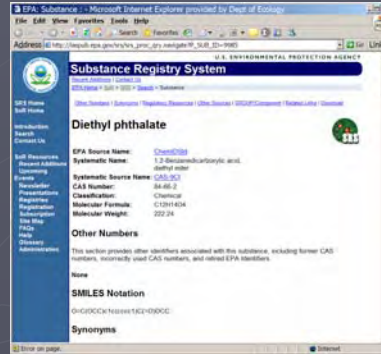
26

Substance Registry System



27

Ecology's Environmental Information Search (EIM)



Hyperlink to SRS



28

Query SRS Web Service Custom EIM Substance Page



XML

SRS Web Service

XML

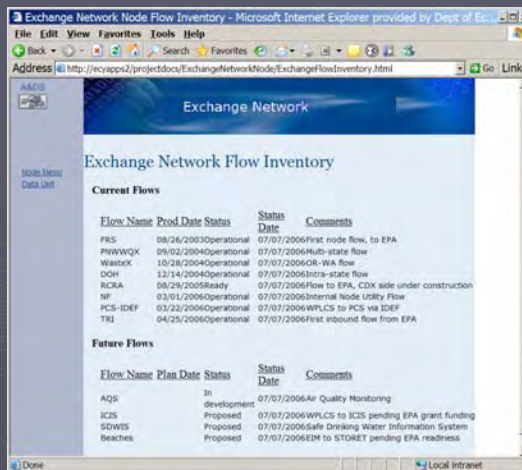
29

XML Registry





30

Exchange Network Data Flows using XML Registry





31



Demonstration and Discussion



- **Demonstration**
 - System of Registries
 - Home page and overall concept
 - Terminology services
 - System inventory services
 - Data registry services
 - Substance registry services
 - Facility registry system
 - Reusable component services
 - Data standards web page
- **Discussion**
 - Using the System of Registries in EPA and partner quality programs

34



Appendix: Data Registry Services



35



Data Registry Contents (Current Project)

- Data standards
- Data dictionaries (for EPA's most important systems)
 - Link to information about the system (the source)
 - For each data element in the dictionary:
 - Name
 - Definition
 - Meaning (when it isn't the same as a definition)
 - Concept – to allow the data element to be group with like data elements
 - Link to the standard element (if applicable)
 - May have formats, example representations (think 04/23/08 or April 23, 2008 for date)
- Code sets or valid value lists (when important or reusable)

36



Data Registry Services (Current Project)

- Compare tool to search from a data standard to find all systems containing that standard element
- Compare tool to search for like terms across EPA systems by name but also by concept and meaning
- Services to allow for automated translation of one code set to another
- Reports about usage of data standards in EPA systems

37

Data Element Display (for Facility ID Standard)

Item Type	Preferred Name	Preferred Definition	Context	Registration Status
Data Element	Facility Site Environmental Interest	The environmental permits and regulatory programs that apply to the facility site.	Facility Site Identification	STANDARD
Data Element	Environmental Interest Type Text	The environmental permit or regulatory program that applies to the facility site.	Facility Site Identification	STANDARD
Data Element	Environmental Interest Start Date	Date that US EPA, Tribe or State became interested in the facility site for a particular environmental interest type.	Facility Site Identification	STANDARD
Data Element	Environmental Interest End Date	Date that US EPA, Tribe or State ceased to be interested in the facility site for a particular environmental interest type.	Facility Site Identification	STANDARD
Data Element	Environmental Interest Start Date Qualifier Text	The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest start date.	Facility Site Identification	STANDARD
Data Element	Environmental Interest End Date Qualifier Text	The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest end date.	Facility Site Identification	STANDARD
Data Element	Environmental Information System Identification Number	The identification number, such as the permit number, assigned by an information management system that represents a facility site, waste site, operable unit, or other feature tracked by that Environmental Information System.	Facility Site Identification	STANDARD
Data Element	Facility Site Standard Industrial Classification	The Standard Industrial Classification (SIC), or type of business activity occurring at the facility site.	Facility Site Identification	STANDARD
Data Element	SIC Primary Indicator	The name that indicates whether the associated SIC Code represents the primary activity occurring at the facility site.	Facility Site Identification	STANDARD
Data Element	Facility Site North American Industry Classification	The North American Industry Classification System (NAICS) code, or type of industrial activity occurring at the facility site.	Facility Site Identification	STANDARD
Data Element	NAICS Primary Indicator	The name that indicates whether the associated NAICS Code represents the primary activity occurring at the facility site.	Facility Site Identification	STANDARD
Data Element	Federal Facility Ownership and Operation Information	Provides information about the legal right of possession or proprietorship.	Facility Site Identification	STANDARD
Data Element	Facility Ownership Type Code	The code of the type of ownership of a facility.	Facility Site Identification	STANDARD
Data Element	Facility Operation Type Code	The code that represents the type of party that operates a facility.	Facility Site Identification	STANDARD

38

Data Dictionary – Table Display (for READ System)

System/Application	Table Name	Table Description
READ	SYS_USER_LOGIN_LOG	user login events to the READ system
READ	SYS_XML_EXPORT	specification of XML exports supported by the READ application
READ	TELEPHONE_COMMUNICATION	telephone connections
READ	REF_TECHNOLOGY	technology
READ	REF_TECHNOLOGY_CATEGORY	technology categories
READ	REF_TELEPHONE_NUMBER_TYPE	types of telephone numbers
READ	REF_TRIBAL_LOCATION	American Indian tribal or Alaskan Native entity lands
READ	REF_VALIDITY_RATING	validity ratings applicable to READ records
READ	REPT_PARAMETER	parameters used in READ reports
READ	REPT_SAVED_PARAMETER	report parameters saved by READ system users
READ	SYS_GROUP	security groups
READ	SYS_GROUP_INFO_RESOURCE_XREF	associations denoting information resources used by system groups
READ	SYS_GROUP_SYS_USER_XREF	association denoting information resources and READ system users for application security purposes
READ	SYS_OPTION	options for the operation of the READ system
READ	SYS_PORTLET	portlets contained within the READ application
READ	SYS_PORTLET_FIELD_DEFINITION	explanatory definitions for READ system portlet fields
READ	SYS_PORTLET_HELP	guidance in understanding READ system portlets
READ	SYS_PORTLET_INSTRUCTION	guidance instructions in operating READ system portlets
READ	SYS_USER	users of the READ system

39

Data Dictionary – Tables with Columns Display (for READ System)

System Tables	System Columns	Table	Column Name	Column Description	Column Data Type	Nullability
INFO_RESOURCE_CONTACT	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	CONTACT_ROLE_CODE	code designating a contact role	VARCHAR2	
	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	INDIVIDUAL_IDENTIFIER	READ system unique identifier used to uniquely designate an individual within a context	NUMBER	
	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	INFO_RESOURCE_PK_ID	READ system-generated unique identifier that is uniquely assigned to designate an information resource	NUMBER	
	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	CREATED_TMSP	date/time that the instance record in the entity was originally created	DATE	
	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	CREATED_USER_ID	user identifier of the individual that originally created the instance record in the entity	VARCHAR2	
	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	USER_ID_LAST_UPDATED	user identifier of the individual that last updated the instance record in the entity	VARCHAR2	
	<input type="checkbox"/>	INFO_RESOURCE_CONTACT	TMSP_LAST_UPDATED	date/time that the instance record in the entity was last updated	DATE	

40

Code Set Display (for USPS State Codes)

Administrative Item	Permissible Value	Permissible Value Meaning	Effective Date	Expiration Date	Legacy Code
US Postal Service State Abbreviations	LA	LOUISIANA	01/09/2009		
Alternate Names	MA	MASSACHUSETTS	01/09/2009		
Audit Information	MD	MARYLAND	01/09/2009		
Comments	ME	MAINE	01/09/2009		
Examples	MH	MARSHALL ISLANDS	01/09/2009		
Item Classified Under	MI	MICHIGAN	01/09/2009		
Reference Documents	MN	MINNESOTA	01/09/2009		
Data Elements	MO	MISSOURI	01/09/2009		
Permissible Values	MQ	NORTHERN MARIANA ISLANDS	01/09/2009		
Related Value Domains	MS	MISSISSIPPI	01/09/2009		
	MT	MONTANA	01/09/2009		
	NC	NORTH CAROLINA	01/09/2009		
	ND	NORTH DAKOTA	01/09/2009		
	NE	NEBRASKA	01/09/2009		
	NH	NEW HAMPSHIRE	01/09/2009		
	NJ	NEW JERSEY	01/09/2009		
	NM	NEW MEXICO	01/09/2009		
	NV	NEVADA	01/09/2009		
	NY	NEW YORK	01/09/2009		

41