RDF and Digital Libraries

Conventions for Resource Description in the Internet Commons
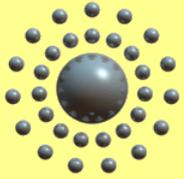
Stuart Weibel

purl.org/net/weibel
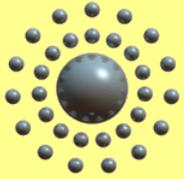
December 1998

Outline of Today's Talk

- Motivations for developing new conventions for resource description

- The Dublin Core Metadata Initiative: semantics for resource description in the Digital Library

- The Resource Description Framework: An architecture for Web metadata

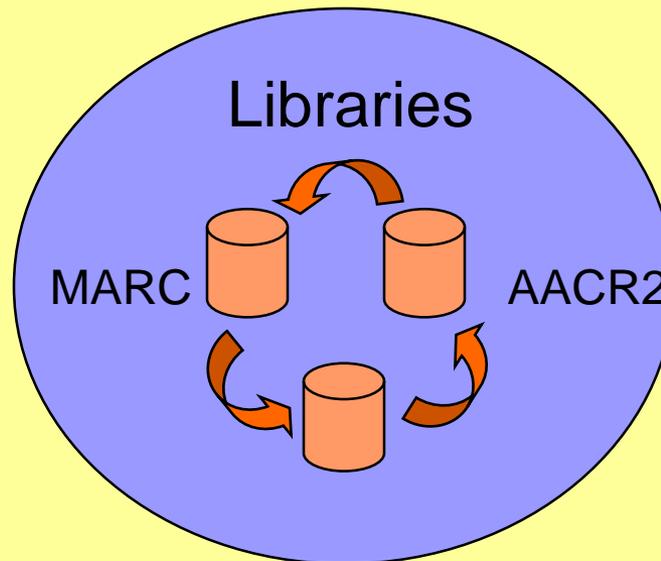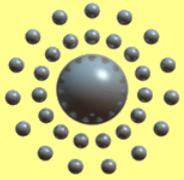# What are digital libraries?

- Institutions with articulated policies for managing information:
  - **S**election
  - **C**ollection
  - **O**rganization
  - **A**ccess
  - **P**reservation
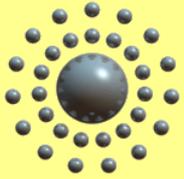
Metadata:
structured data about data

A *resource description community* is characterized by common *semantic, structural, and syntactic* conventions for exchange of resource description information

The Internet Commons embraces many formal and informal Resource Description Communities

Home Pages

Libraries

Geo

Commerce

Internet Commons

Scientific Databases

Museums

Whatever...
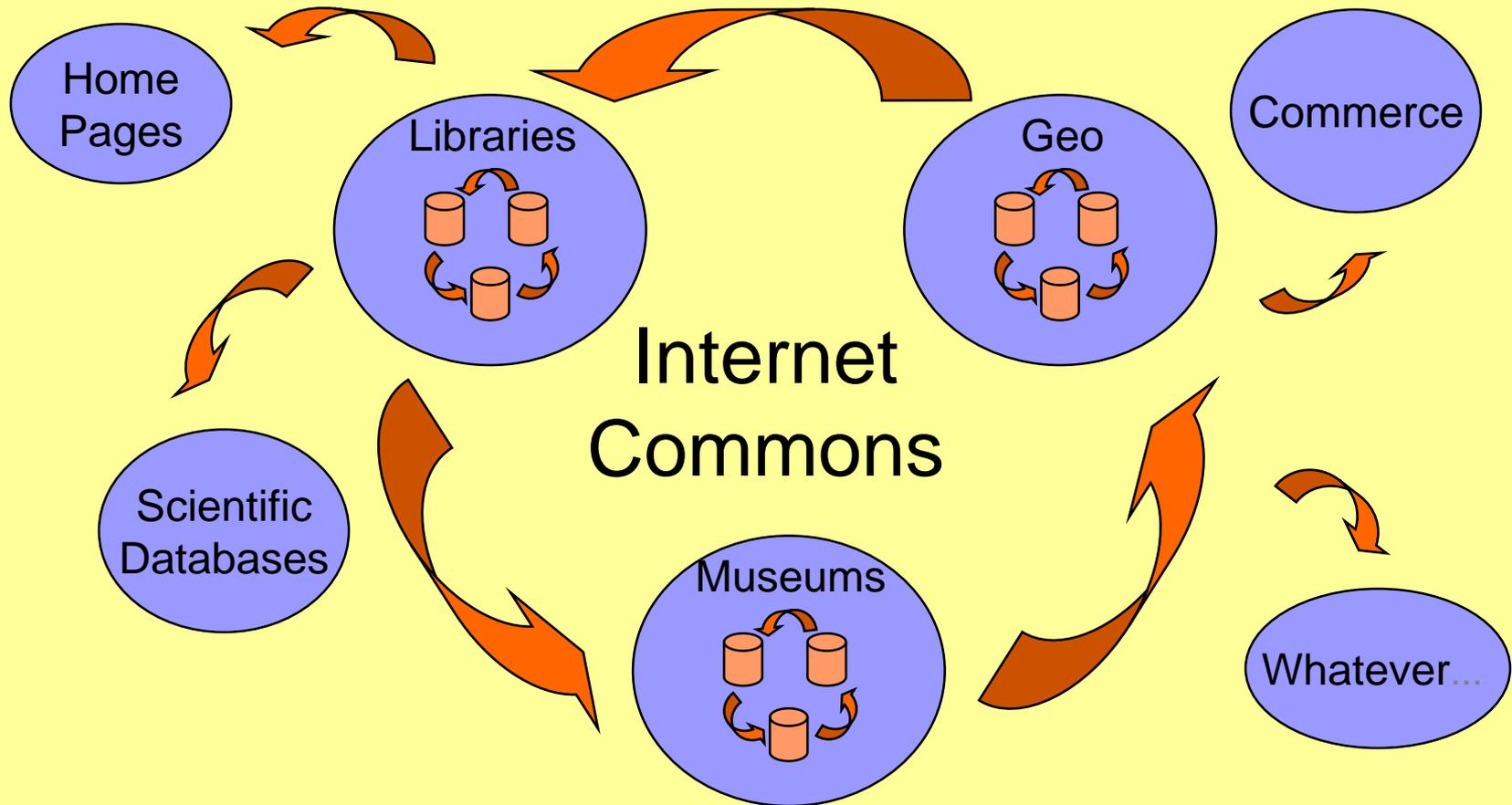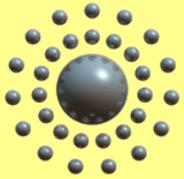
Three Levels of Interoperability
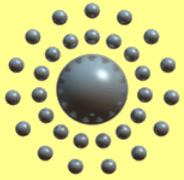
|	
Semantic Interoperability	Content Description Standard meaning (by human agreement) (DC, AACR2, TEI, FGDC…)
Structural Interoperability	RDF: A data-model for specifying semantic schemas
Syntactic Interoperability	XML: a markup idiom for structured data on the web

The Dublin Core Metadata Workshop Series

- How to improve resource discovery on the Web?
  - Agreements about resource description semantics
- Build an interdisciplinary consensus about a core element set for *resource discovery:*
  - simple and intuitive
  - cross-disciplinary
 - international
  - flexible
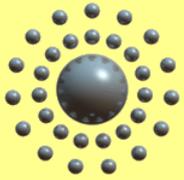
# Central Characteristics of the Dublin Core Metadata Element Set

- Descriptive metadata for resource discovery (15 elements)

- Extensible (a starting place for richer description)

- Interdisciplinary (semantic interoperability)

- International (20 languages and growing)
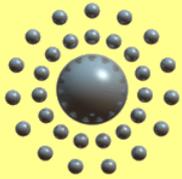
Extensibility (refined semantics)

- Ukrainian Doll model
  - improve description precision with sub-structure (sub-elements and schemes)
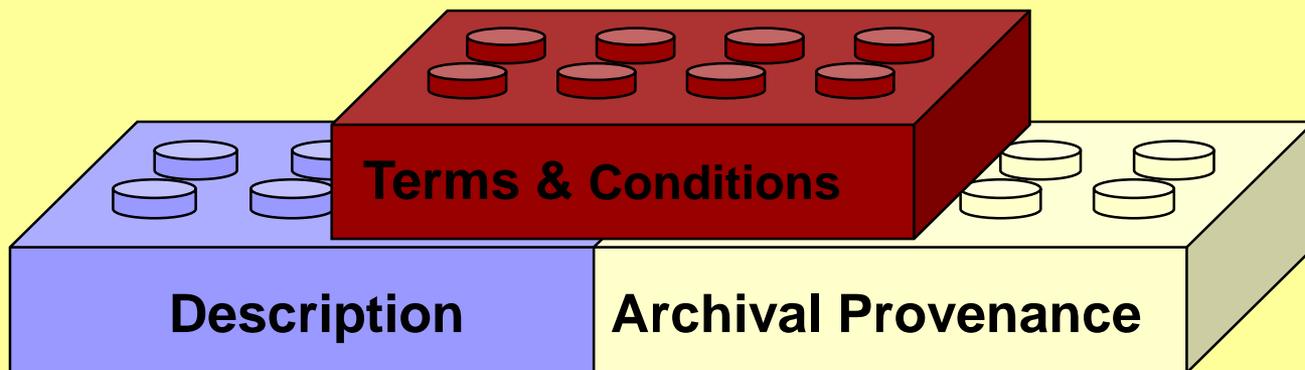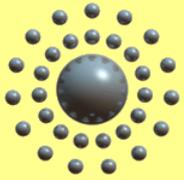 - should degrade gracefully to preserve interoperability

Creator

| Given Name	Affiliation
Surname	Contact Info
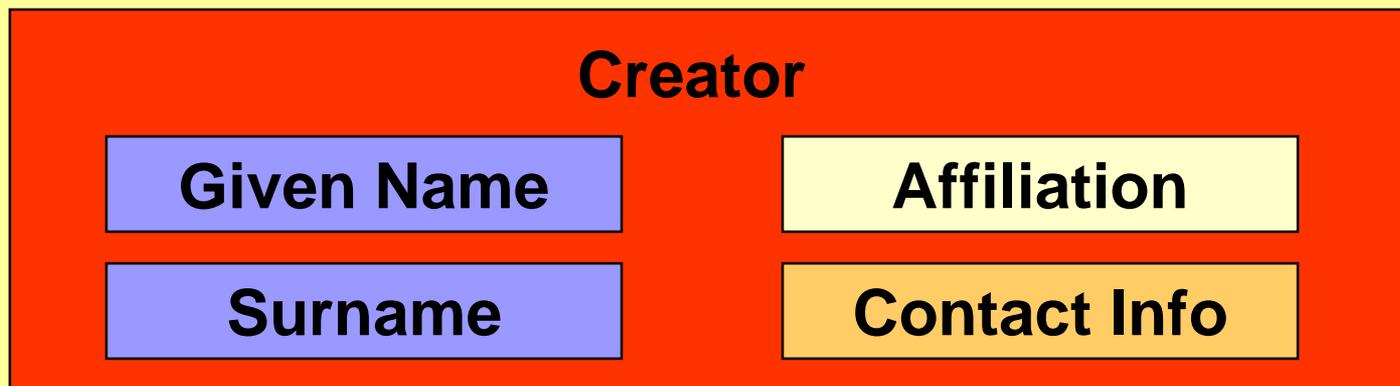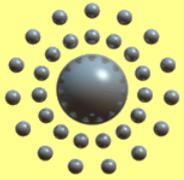
Extensibility
The Lego Metaphor

- Modular extensibility
  - additional elements to support local or discipline-specific requirements
 - complementary packages of metadata:

Terms & Conditions

Description

Archival Provenance

Dublin Core is about semantics… what about structure and syntax?

- HTML META tags jump-started deployment
- HTML 4.0 provided additional structural richness, but suffers from deficiencies
- Embedded metadata only (other important models are not supported)
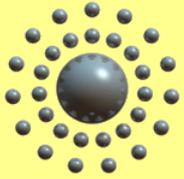
Resource Description Format:
## An Architecture for Metadata on the Web

- Conventions to support interoperability among applications that exchange metadata
- RDF provides a data model and structural conventions that comprise an architecture for Web metadata
- XML is the encoding syntax
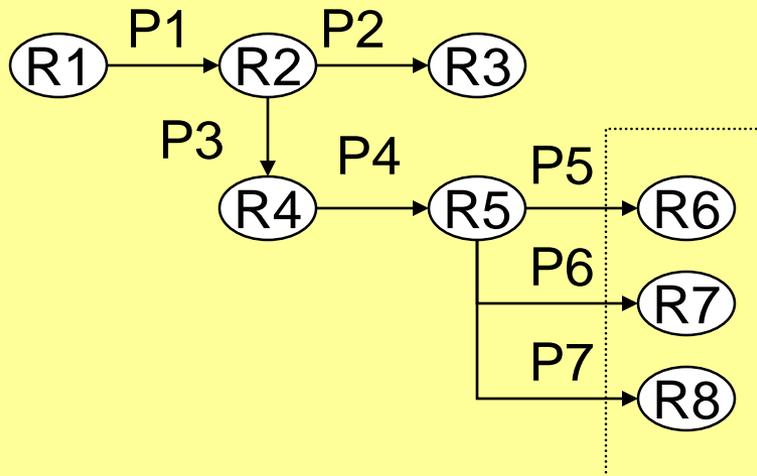- Semantics defined by *stakeholders* (not software developers)
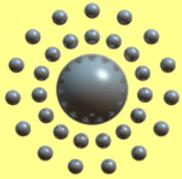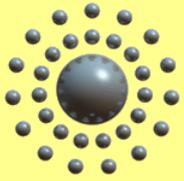
The RDF Data Model
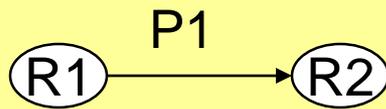
Nodes are **resources** connected by **named properties**
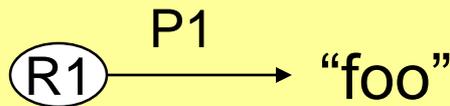
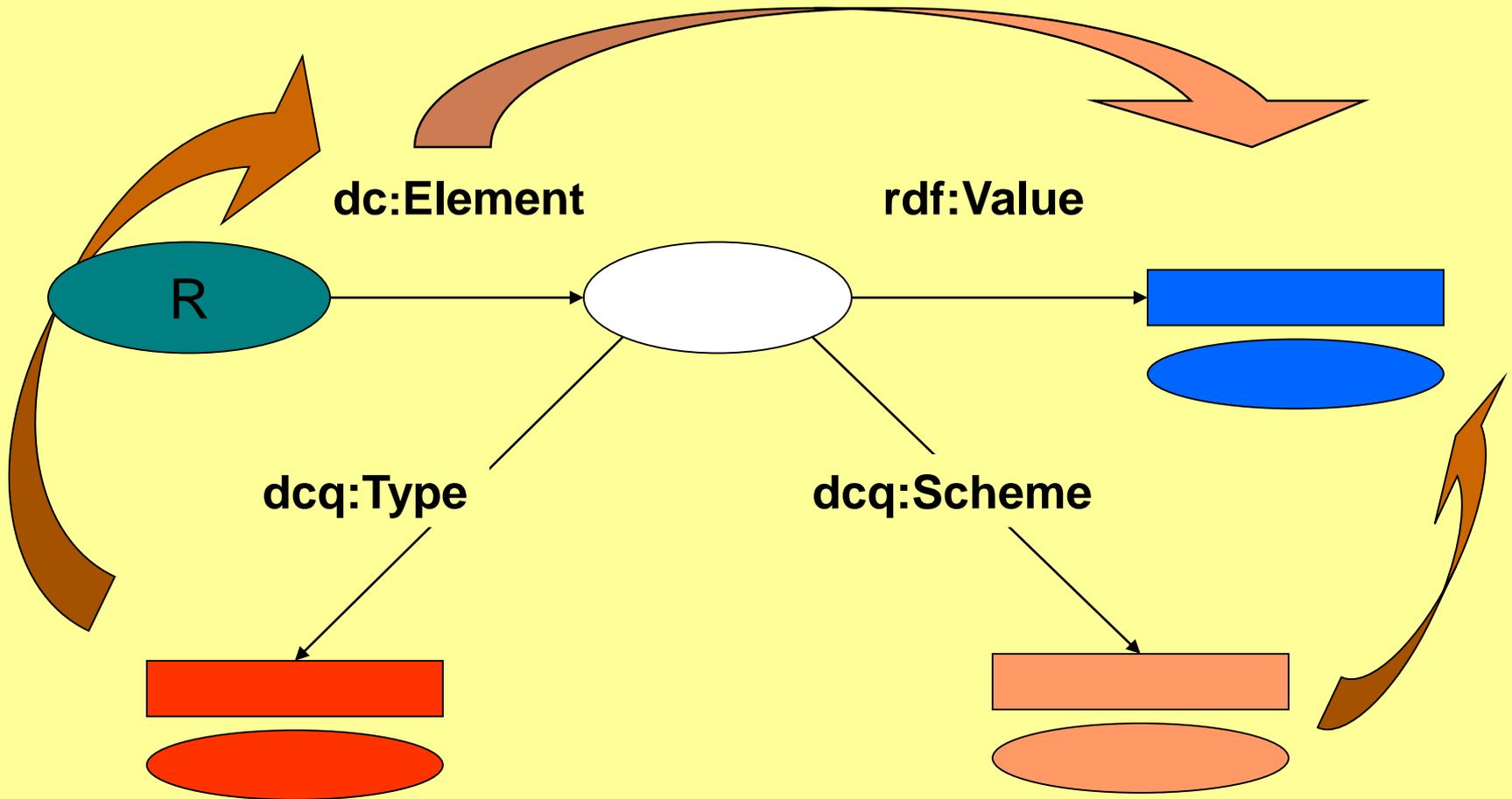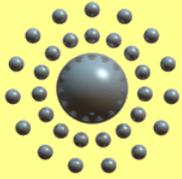The degenerate case is an arc terminating in a fixed value

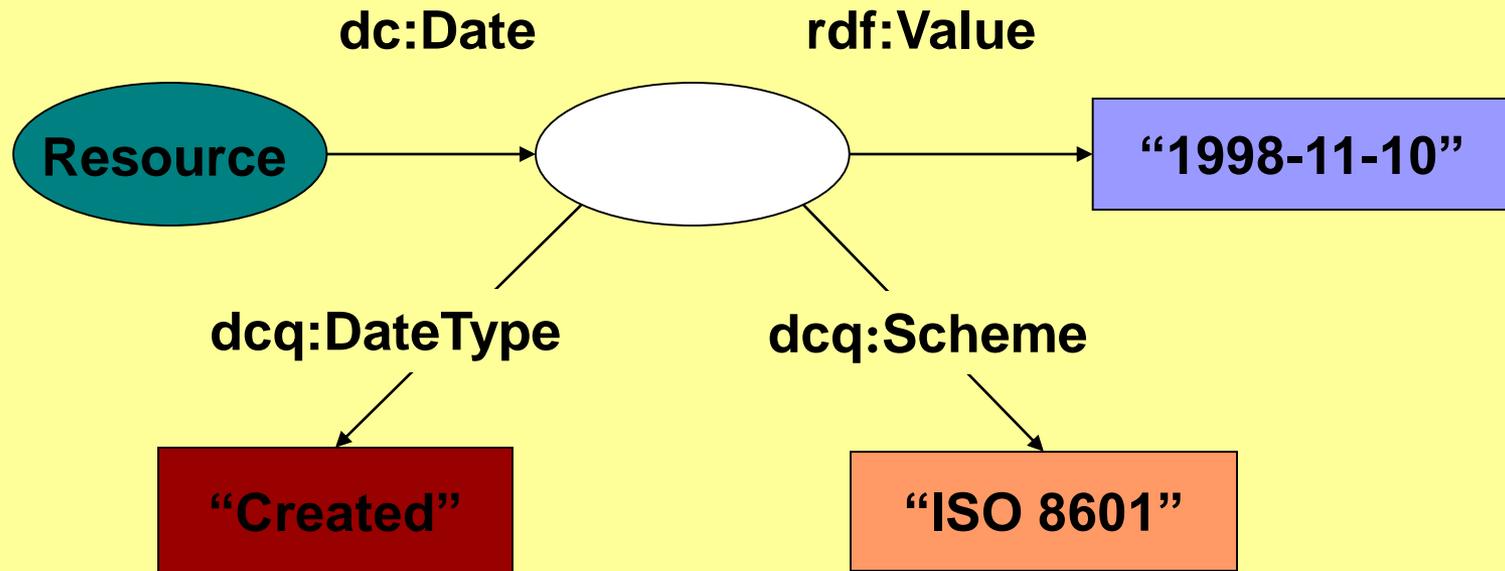An **RDF description** consists of a directed graph of **arbitrary complexity**

The Dublin Core Data Model
(an application of the RDF data model)
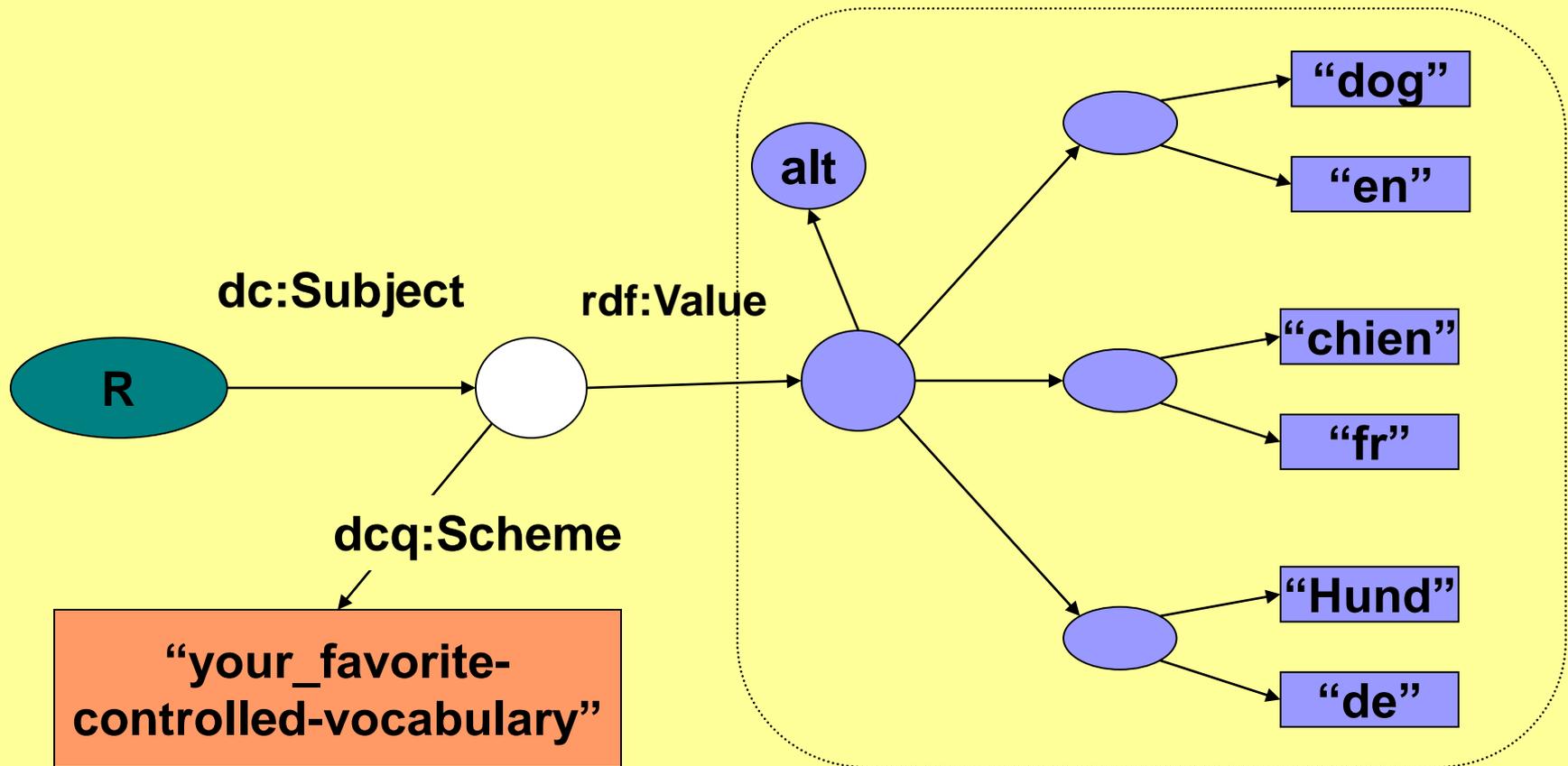
dc:Date

dc:Date   **rdf:Value**

Resource → ( ) → **"1998-11-10"**

dcq:DateType   **dcq:Scheme**
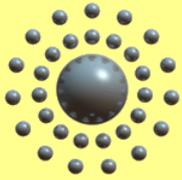
"Created"   **"ISO 8601"**
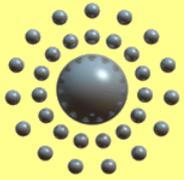
dc:Subject
(with multi-lingual substructure)

CORC
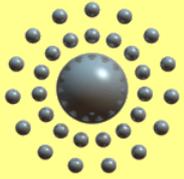## Cooperative Online Resource Catalog

- An OCLC research project exploring the cooperative resource selection, and creation and sharing of metadata by libraries
  - Cooperative cataloging of web resources
  - Accommodating both local and shared metadata (electronic and physical resources)
  - Multi-standard support
   - Dublin Core, MARC, XML, RDF
  - http://purl.org/corc

CORC System Components

- Web resource harvesting and selection
- Automatic classification and subject assignment (Dewey Decimal Classification)
- Metadata extraction, editing and authority control
- Resource Identifier maintenance (PURL)
  - http://purl.org
- Flexible delivery mechanisms (databases, Web portals)
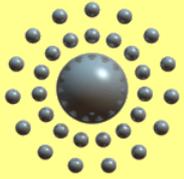
The Reggie Metadata Editor

- Metadata Editor supporting multiple content standards

- Early RDF prototype

- Developed at Distributed Systems Technology Centre, University of Queensland, Brisbane

- http://dstc.edu.au
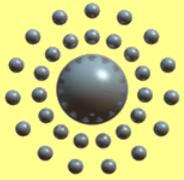
Summary

- Dublin Core semantics and RDF structure are co-evolutionary

- HTML META tags have jump started deployment, but are limited

- RDF provides an enabling architecture and a data model for Web metadata

- Tools are beginning to be appear (and evolve)

- Cross disciplinary resource description is within reach
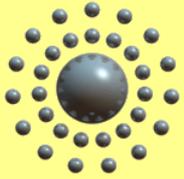
Additional Information
Dublin Core and RDF

- Dublin Core Homepage

  http://purl.org/dc

- RDF Home Page

  http://www.w3.org/RDF

- PURL home page

  http://purl.org