Content Interoperability

Donatella Castelli
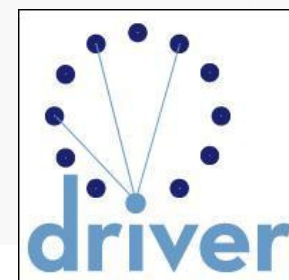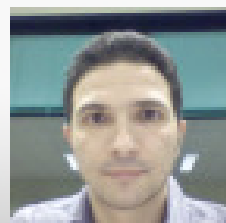
CNR-ISTI

(on behalf of the DL.org Content WG)

Outline

- Context

- Content Interoperability Framework

- Scope
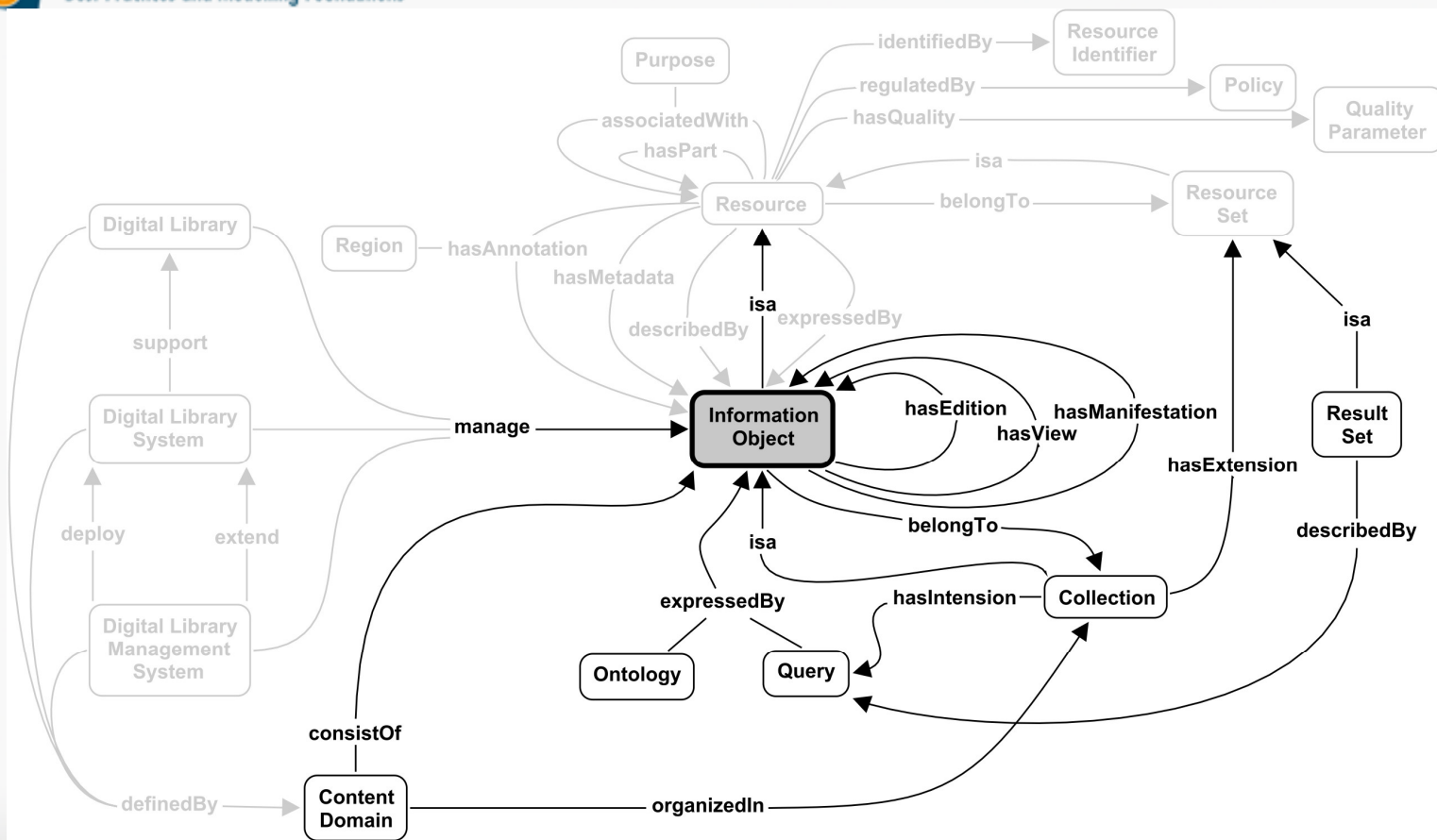
- Next Steps

Content WG members

WG Mission & Scope

Mission

- Survey & analyse existing approaches to content interoperability

- Identify possible solutions to mitigate/resolve interoperability issues
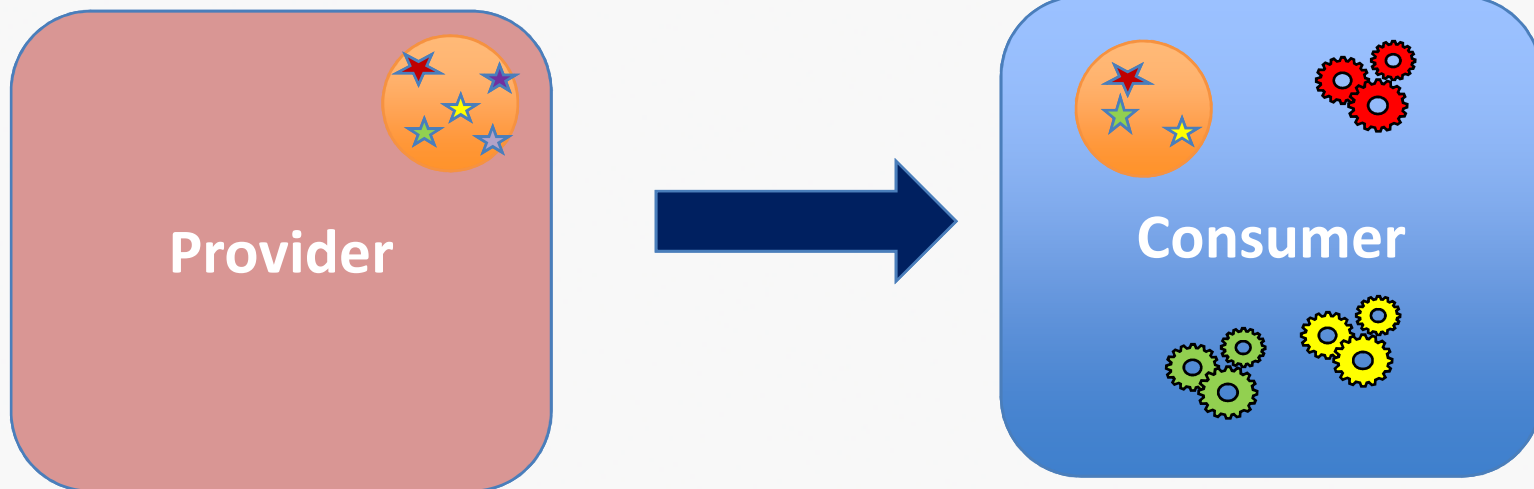
- Proposing effective patterns

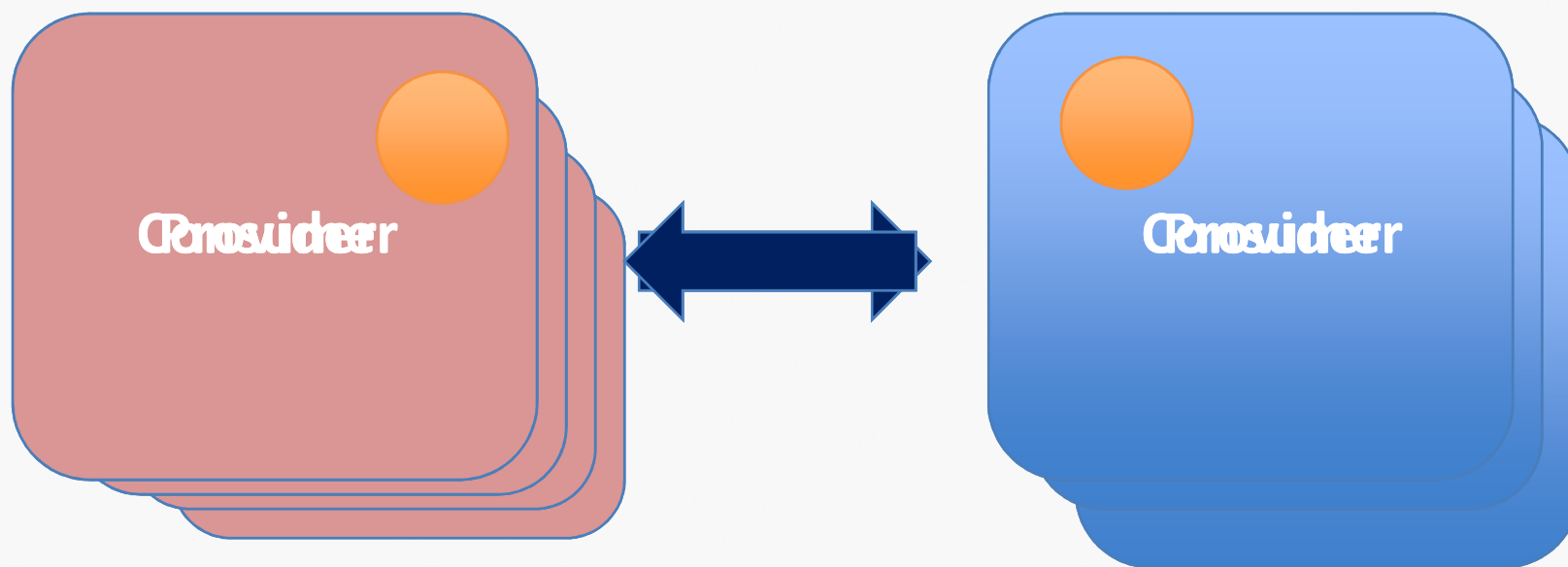# The Content Domain (cont.)

Scope
Focus on a selected subset of content aspects

The problem

The problem

Protocol

- Interoperability subsumes **communications** between the provider and the consumer

- The protocol captures all the **steps needed to reach a common understanding** on the resource between the provider and the consumer and might include the steps governing the actual consumption

Interoperability problem (1)

- **Which information object facet is involved?**

  i.e. interoperability with respect to the IO structure guarantees that the consumer can access and process the information object parts using appropriate type-specific functions

Interoperability problem (2)

- **Which level of abstraction is involved?**
    - *syntactic*: representation of the information object facet
    - *semantic*: meaning of the information object aspect facet
    - ......
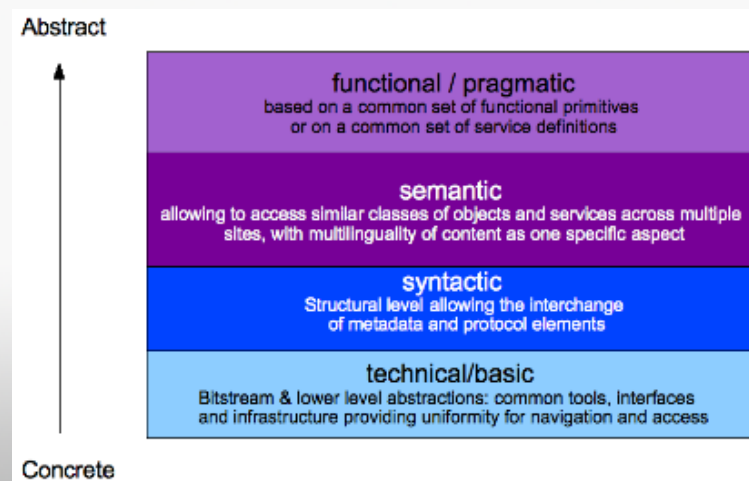
Interoperability solution (P)

- **Interaction Model/Approach**,

   How the specific level of interoperability is achieved

   e.g. cooperative model: common metadata schema

   uncooperative model: metadata mapping

- **Quality**

   A measure of the quality of the approach

- **Time**

   The time-related feature characterizing the interoperability scenario

Content Interoperability Framework

Resource Facet

Abstraction

Approach

Quality

Time

WG Scope

- Subset of the Information Object facets
  - **Structure**
  - **Metadata/Attributes**
  - **Context**
  - **Provenance**
  - **Identifier**

Structure

- The structural properties of the objects (document model, "data type")

  Interoperability enables safe and/or efficient execution of operations over an Information Object based on the structural "assumptions" declared by the associated Information Object structure

- *Static data models*
  - Dspace Item Object Model
  - Fedora Document model

- *Dynamic data models*
  - OpenDLib DomDL

Metadata (Attributes)

- "metadata" enriching the information object for various management purposes

- **Element**
  - Dumbing Down
  - Mapping

- **Value set**
  - Mapping

Context

- A specific kind of metadata devised to characterize the circumstances that form the setting for the Information Object e.g. **relations with other entities** like persons, places, moments in time or abstract ideas that complement the object semantics

  – Key-value Models

  – Markup Scheme Models

  – …….

  – Ontology Based Models

Provenance

- Provenance of an Information Object is **a record** of the processes applied to that object

  – Identifier of the source repository of the object

  – Precise description of all the transformations applied to the object

  – Static (e.g. OPM), provenance record given to some inference engines that deduces causality dependency

  – Executable, i.e. provenance as a saved task

Identifier

- The Identifier is a token bound to the Information Object distinguishing it from the others within a certain scope.

- **Persistent identifier**
  - identifiers have to remain unchanged over time
  - identifiers must remain resolvable (corresponding locations must be known).
  - it must be possible to find identified objects at the locations indicated by the resolver
  - identifier must permanently identify the same object through time

Abstraction

- **Syntactic**, e.g. XML

- **Semantic**, e.g. Fedora datastream, Fedora behavior, DSpace item

- **Interoperability levels for DC metadata**
  - Level 1: Shared Term Definitions
  - Level 2: Formal Semantic Interoperability
  - Level 3: Description Set Syntactic Interoperability
  - Level 4: Description Set Profile Interoperability

State-of-the-Art Survey

- **Literature**

- **Projects**

  Europeana, Europeana Connect, D4Science, STERNA, DRIVER, EFG, Video Active, Papyrus, EUProvenance, CATCH, Fedora, Dspace, … and many others

Assessment

Facet: Structure	Approach_1	Approach_2	Approach_3	Approach_4
static	XX	XXX		
dynamic		YY		
parts- only-rel			YYYY	
....		WW		WWW

Next steps

- Framework definition and stabilization

- State-of-the-Art enrichment

- Identification of patterns

- Feedback to the Reference Model

workinggroups.wiki.dlrg.eu/index.php/Content_Working_Group

content@dlorg.eu

QUESTIONS?