Open Annotation

Robert Sanderson
Herbert Van de Sompel

http://openannotation.org/

Open Annotation Collaboration
is funded by
The Andrew W. Mellon
Foundation

Open Annotation Collaboration

- Project Partners:
  - Los Alamos National Laboratory
  - University of Illinois at Urbana-Champaign
  - University of Queensland
  - University of Maryland
  - George Mason University
- Funding: Andrew W. Mellon Foundation
- Discussion Group:

  http://groups.google.com/group/oac-discuss

Current Scholarly Annotation

- Annotations stuck in silos:
  - Only consumable by client that created it
  - Many clients offline/single user only
  - Annotations not shareable beyond original environment: can not create cross system services based on (enriched & merged) annotations

- Online Annotations are Repository-centric, not Web-centric;
  - Identification in terms of local identifiers, not global URIs
  - Annotations stored in repository along with annotated content
  - Need to rethink in terms of the Web

Interoperability via OAC

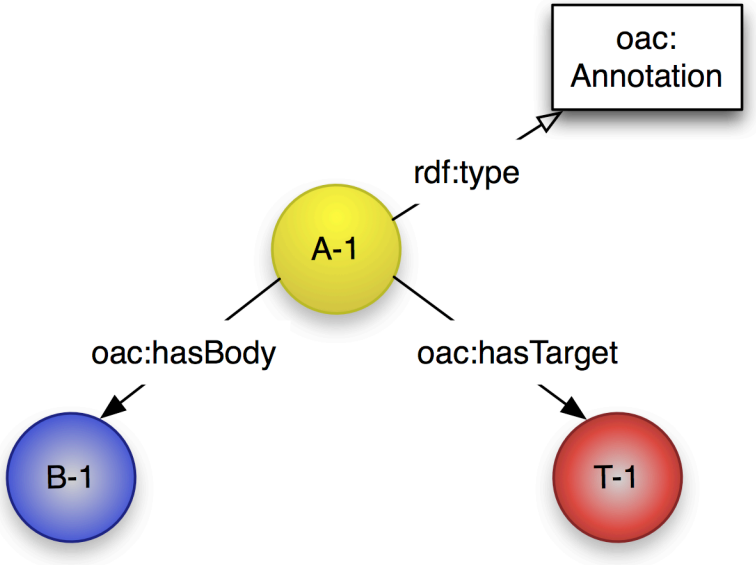
- Focus on interoperability for annotations in order to allow sharing of annotations across:
  - Annotation clients
  - Content collections
  - Services that leverage annotations

- Interoperability is at the data model level, not protocol level

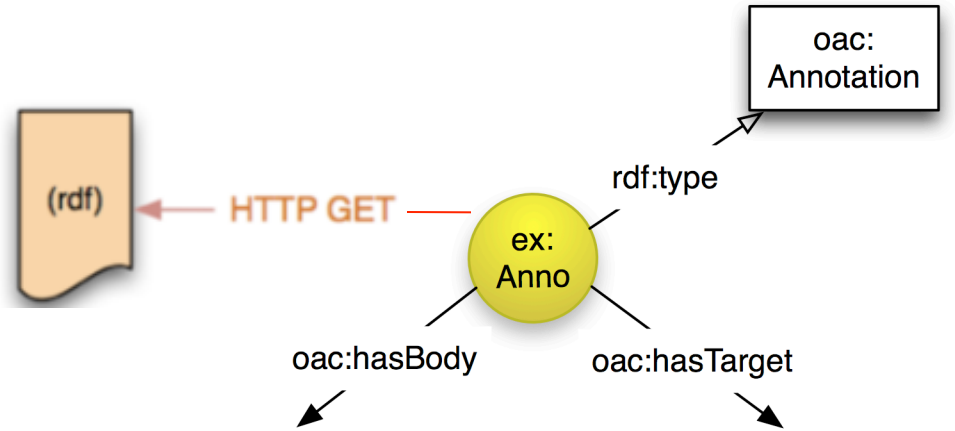
- Focus on annotation for scholarly purposes. But desire to make the OAC framework more broadly usable.
  - In order to gain adoption, we need tools, communities, integration of scholarly communication with other areas of discourse.

Basic Model

- The basic model has three resources:
  - Annotation (an RDF document)
    - Default: RDF/XML but others possible
  - Body (the content of the annotation, in any format)
  - Target (the resource the body is about, in any format)



Basic Model Example

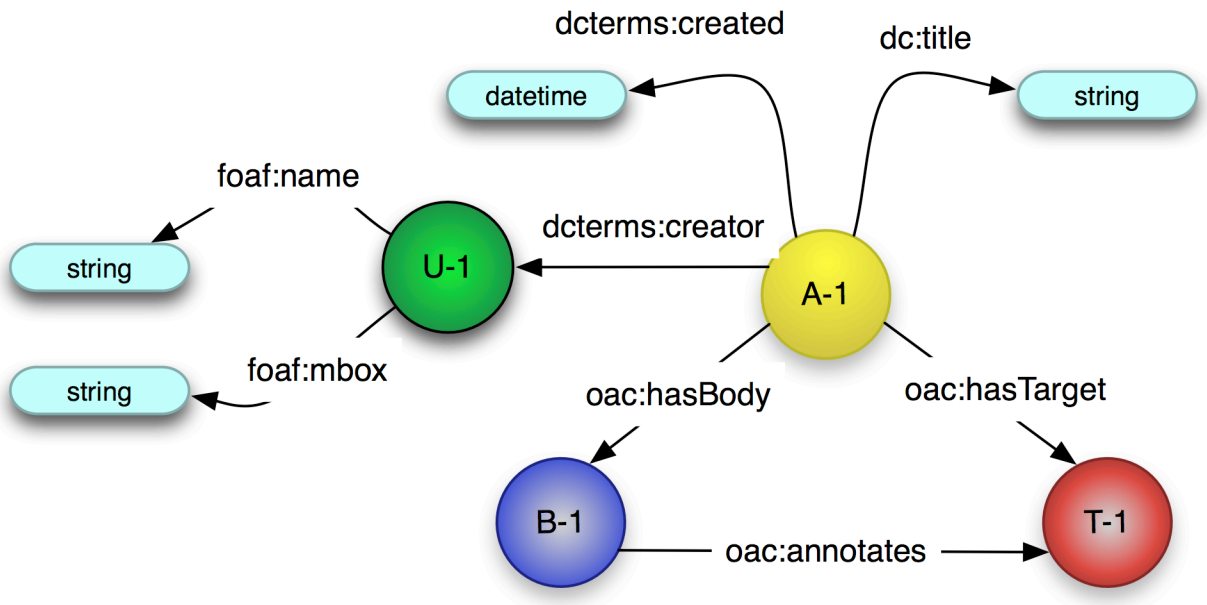


ex:HDFV

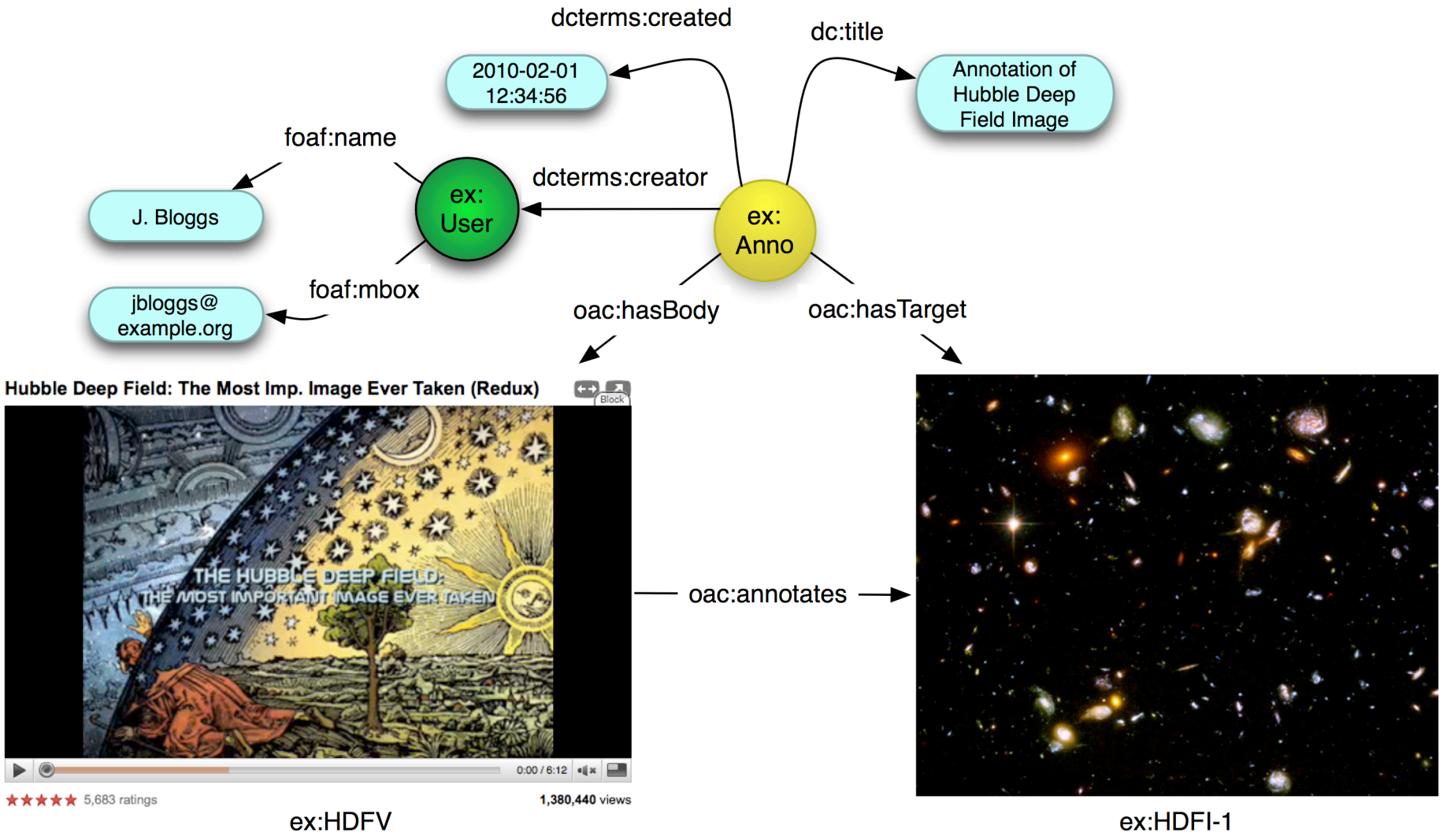
ex:HDFI-1

Additional Relationships and Properties

- Any of the resources can have additional information attached
- Links can be added to further clarify relationships

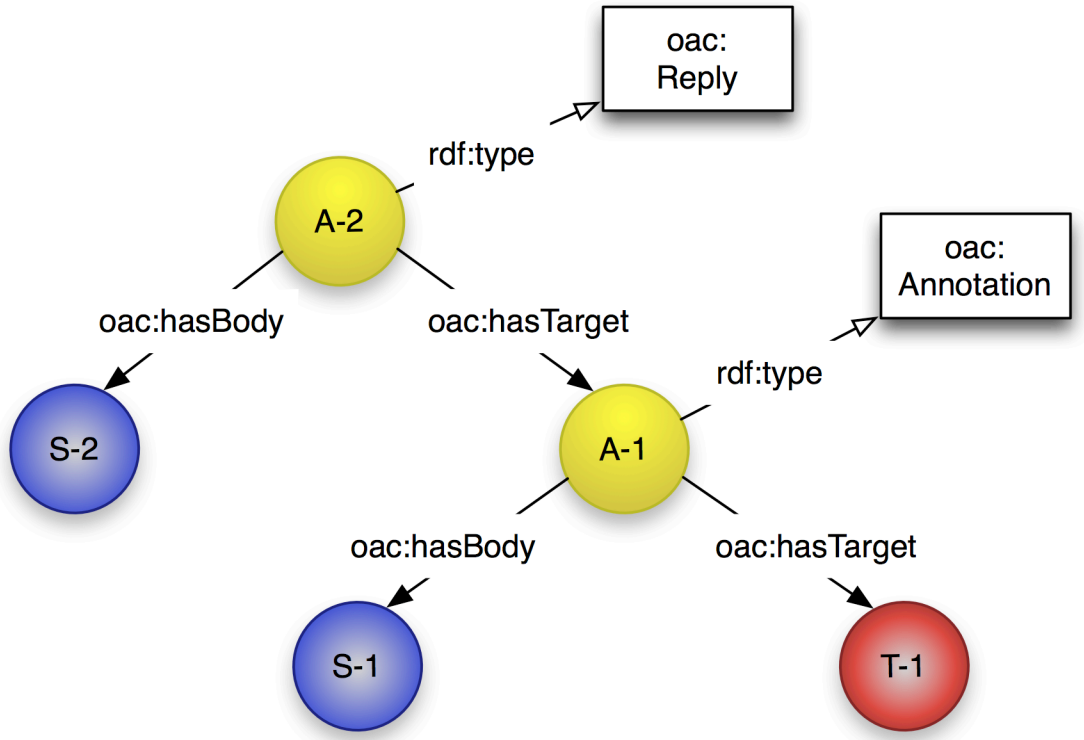


Additional Relationships and Properties Example

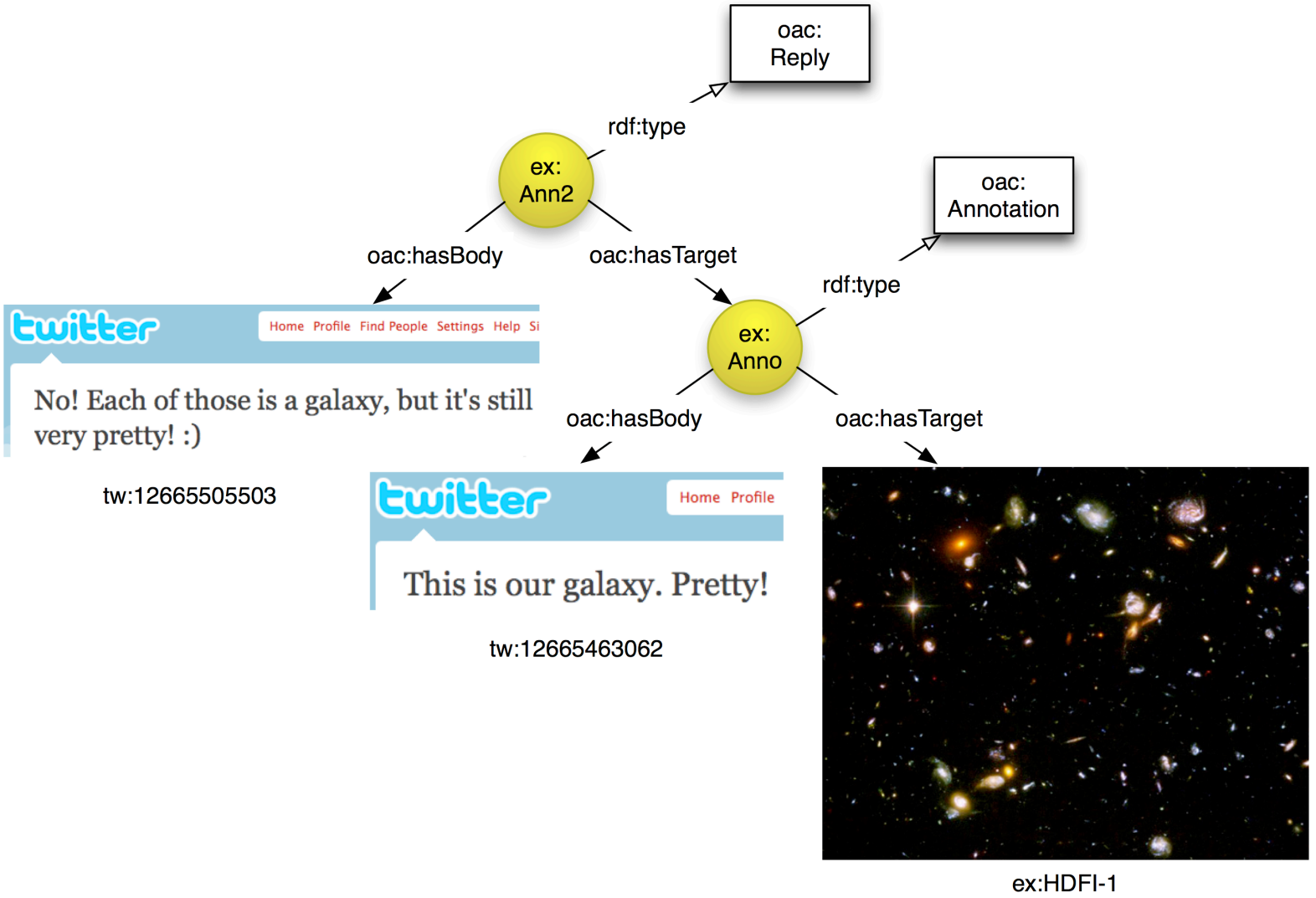


Annotation Types

- The type of the Annotation can be made more precise
- Communities can develop their own types for their requirements
- Example: Replies are Annotations on Annotations

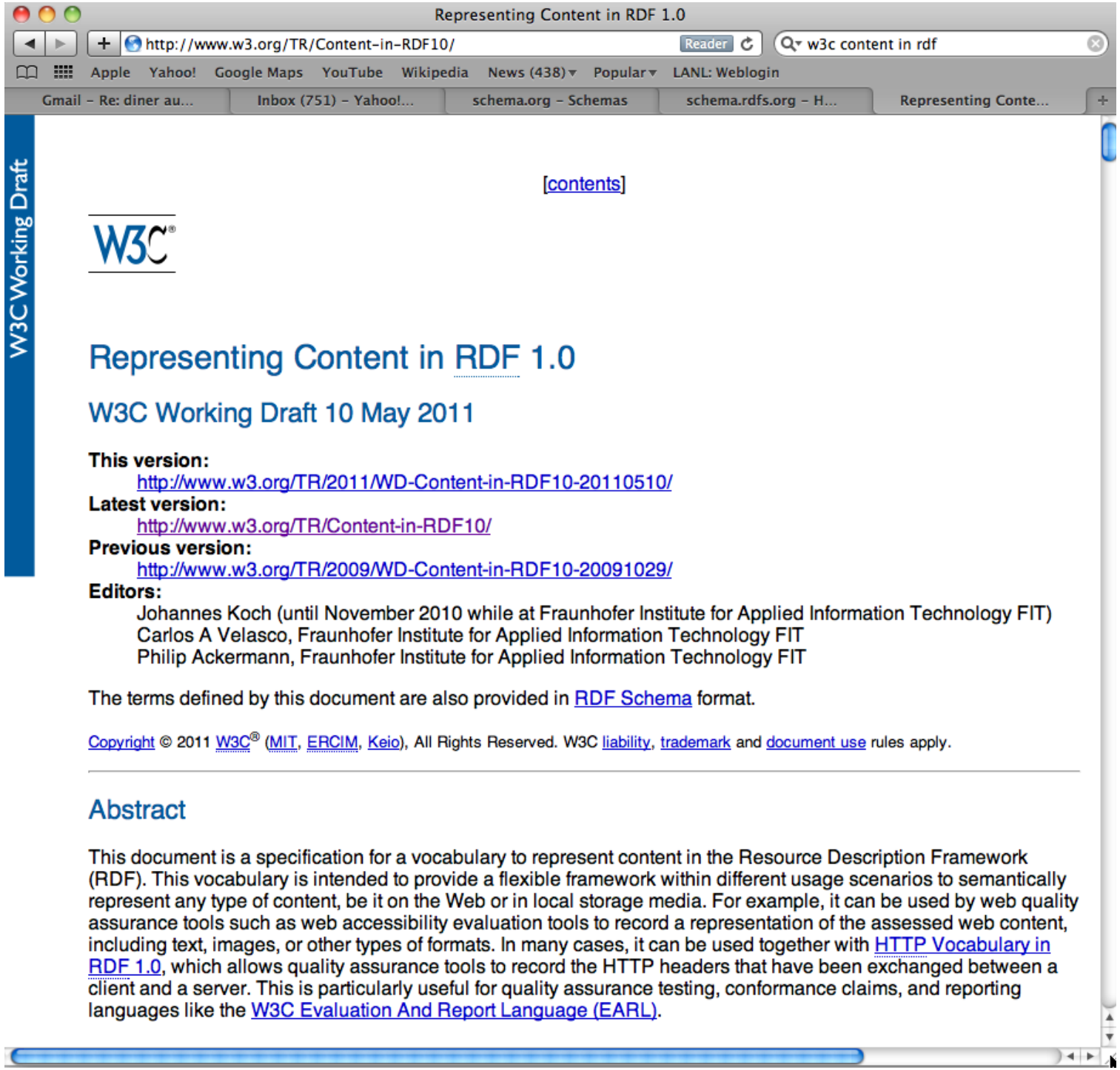


Annotation Types Example



Inline Information

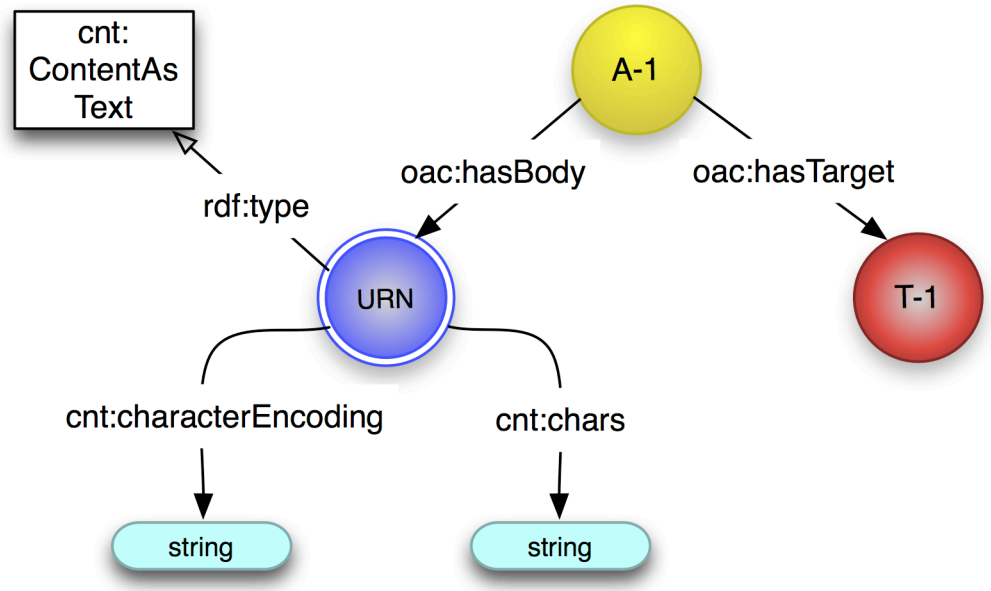
- It is important to be able to have content contained within the Annotation document.
  - Clients may be unable to mint new URIs for every resource
  - Clients may wish to transmit only a single document
  - Third parties can generate new URIs even if the client cannot
- The W3C has a Content in RDF specification for this:
  - http://www.w3.org/TR/Content-in-RDF10/


Representing Content in RDF 1.0
http://www.w3.org/TR/Content-in-RDF10/
Reader
w3c content in rdf
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (438)
Popular
LANL: Weblogin
Gmail - Re: diner au...
Inbox (751) - Yahoo!...
schema.org - Schemas
schema.rdfs.org - H...
Representing Conte...
W3C Working Draft
[contents]
Representing Content in RDF 1.0
W3C Working Draft 10 May 2011
This version:
http://www.w3.org/TR/2011/WD-Content-in-RDF10-20110510/
Latest version:
http://www.w3.org/TR/Content-in-RDF10/
Previous version:
http://www.w3.org/TR/2009/WD-Content-in-RDF10-20091029/
Editors:
Johannes Koch (until November 2010 while at Fraunhofer Institute for Applied Information Technology FIT)
Carlos A Velasco, Fraunhofer Institute for Applied Information Technology FIT
Philip Ackermann, Fraunhofer Institute for Applied Information Technology FIT
The terms defined by this document are also provided in RDF Schema format.
Copyright © 2011 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C liability, trademark and document use rules apply.
Abstract
This document is a specification for a vocabulary to represent content in the Resource Description Framework (RDF). This vocabulary is intended to provide a flexible framework within different usage scenarios to semantically represent any type of content, be it on the Web or in local storage media. For example, it can be used by web quality assurance tools such as web accessibility evaluation tools to record a representation of the assessed web content, including text, images, or other types of formats. In many cases, it can be used together with HTTP Vocabulary in RDF 1.0, which allows quality assurance tools to record the HTTP headers that have been exchanged between a client and a server. This is particularly useful for quality assurance testing, conformance claims, and reporting languages like the W3C Evaluation And Report Language (EARL).

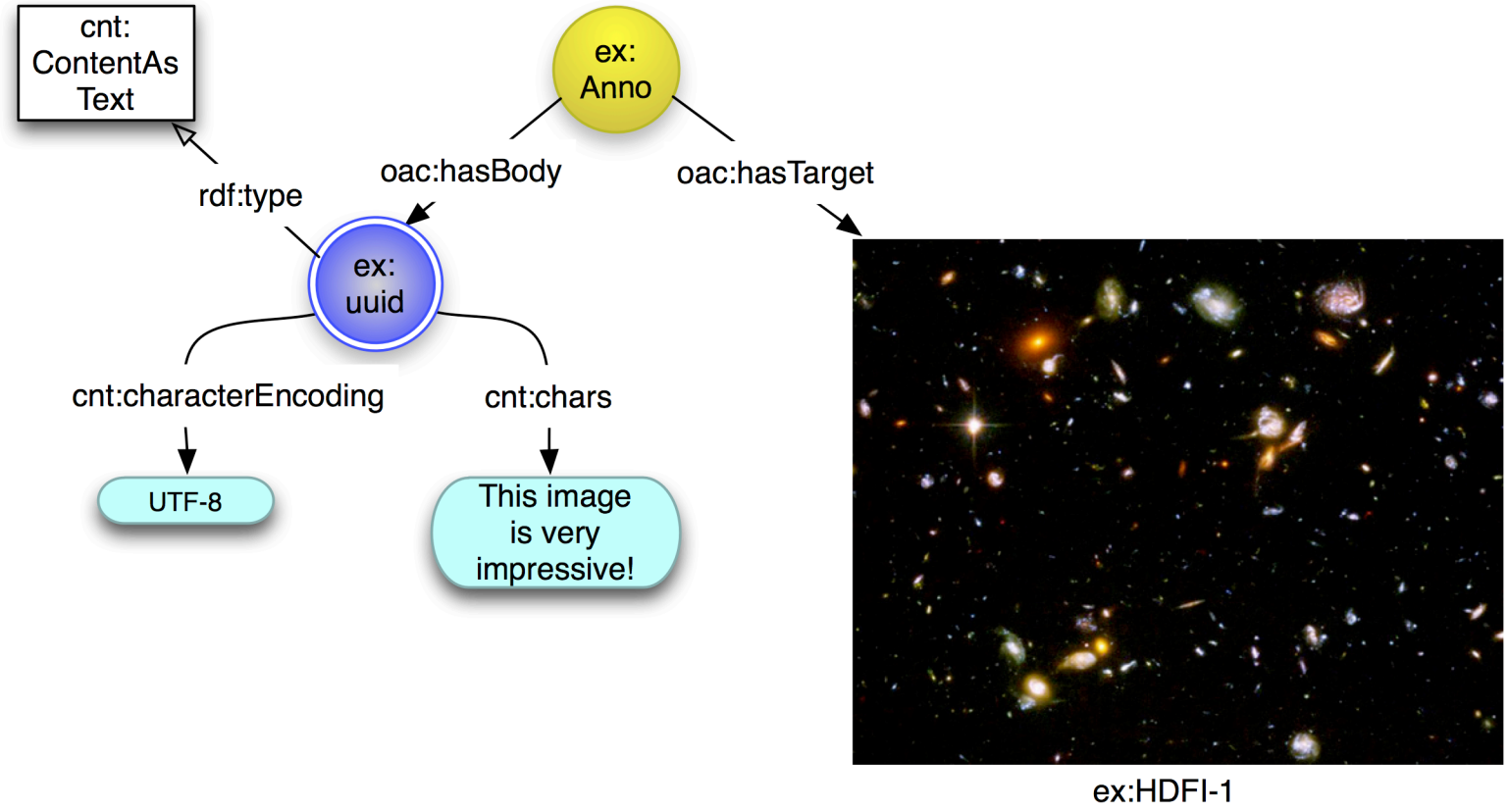
Inline Body

• We introduce a resource identified by a non resolvable URI, such as a URN:UUID, as the Body.

• The data is embedded within the Annotation document using the 'chars' property from the Content in RDF ontology.



Inline Body Example

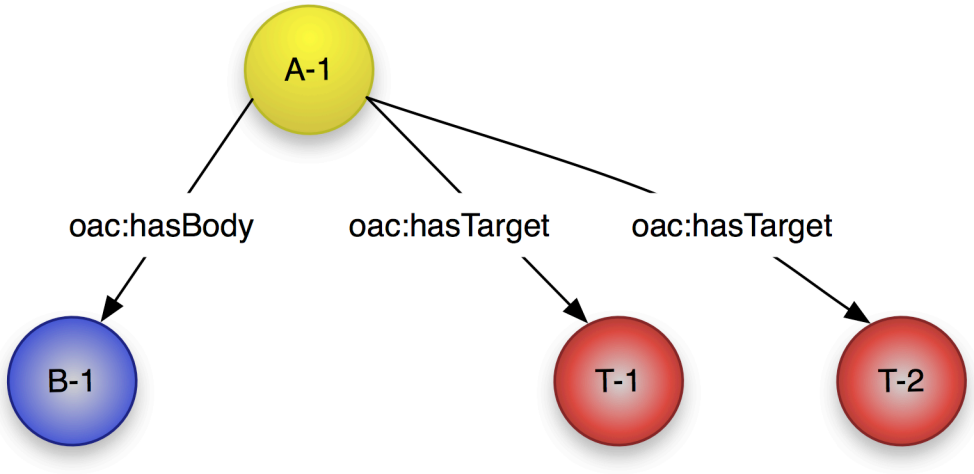


ex:HDFI-1

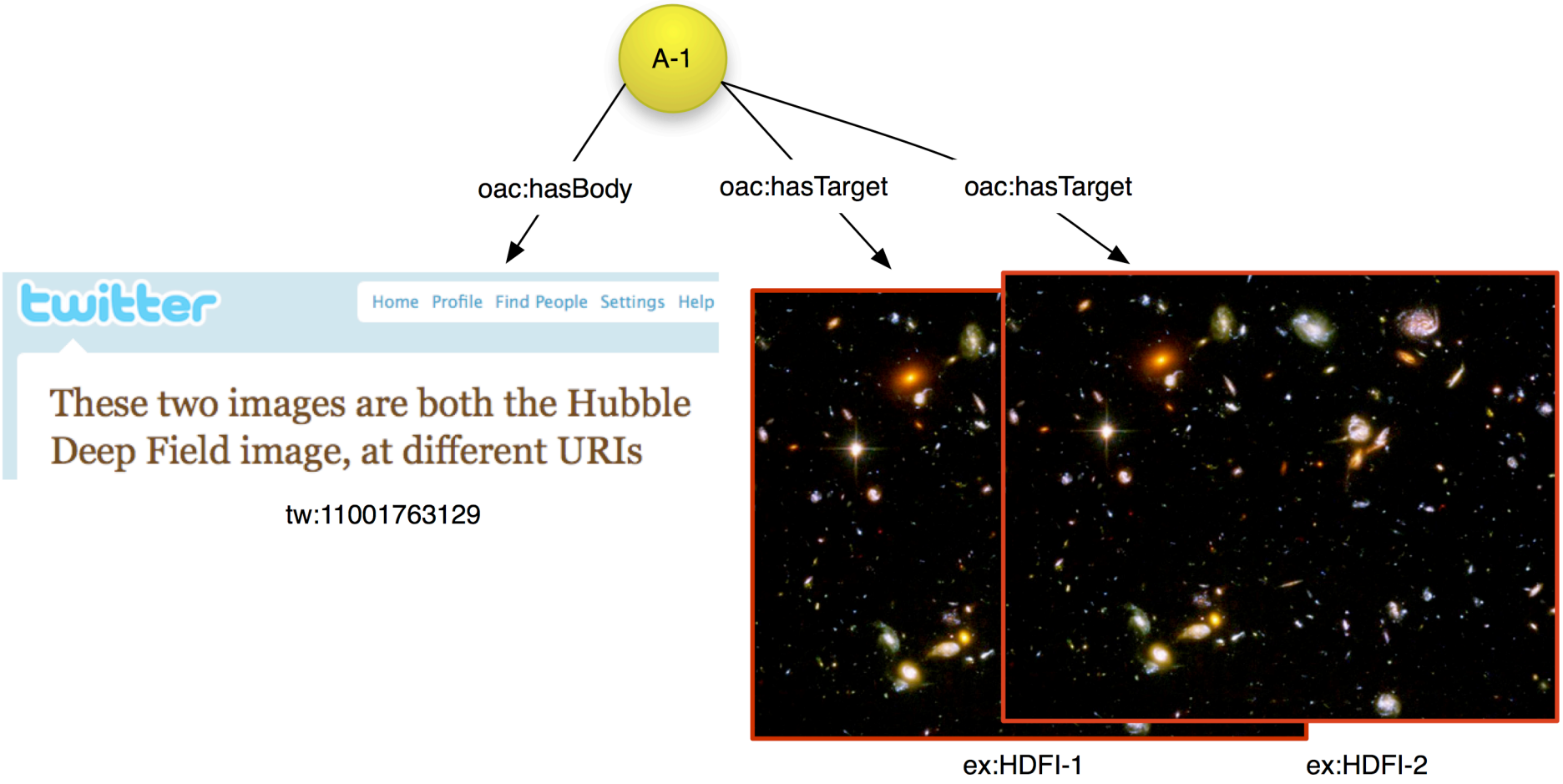
Multiple Targets

- There are many use cases for multiple targets for an Annotation:
  - Comparison between two or more resources
  - Making a statement that applies to multiple resources
  - Making a statement about multiple parts of a resource
  - …

- The OAC Data Model allows for multiple targets by simply having more than one hasTarget relationship.



Multiple Targets Example



Segments of Resources

- Most annotations are about *part* of a resource

- Different types of segment for different media types:
  - Text: paragraph, arbitrary span of words
  - Image: rectangular or arbitrary shaped area
  - Audio: start and end time points
  - Video: both area and time
  - Other: slice of a data set, volume in a 3d object, …

- We introduce a method of *constraining* resources:
  - Can be applied to either Body or Target resource
  - Use media-specific fragment identifiers; eg XPointer for XML
  - Use W3C Media Fragments for segments of image/audio/video: http://www.w3.org/TR/media-frags/
  - Introduce an approach for arbitrarily complex segments

Segments of Resources: Fragment URIs

- The Fragment part of URIs allows the creation of subsidiary URIs that identify part of the main resource
  - eg: *http://www.example.org/page.html#para1*

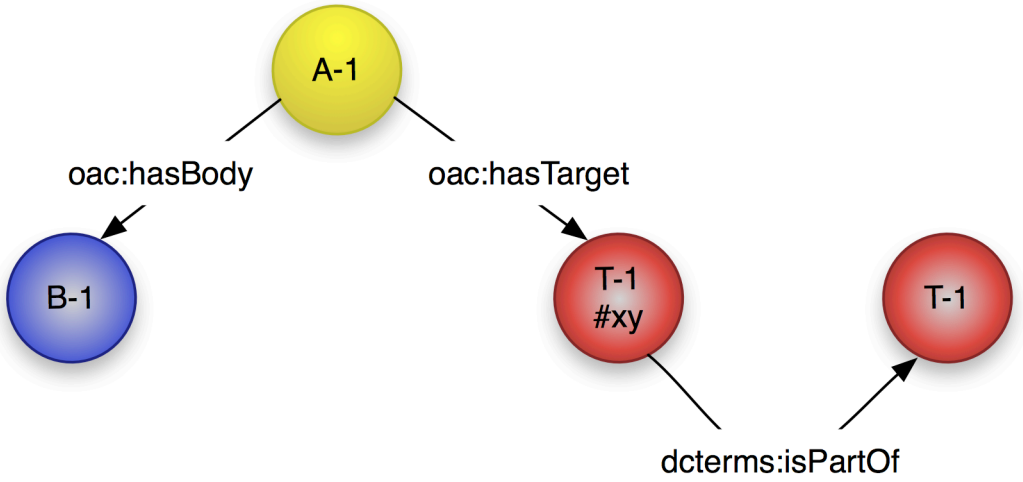
- The syntax is defined for several media types:
  - X/HTML: named anchor or identified element
  - XML: XPointer to the element
  - PDF: page number and rectangular area within
  - Plain Text: character position or line position

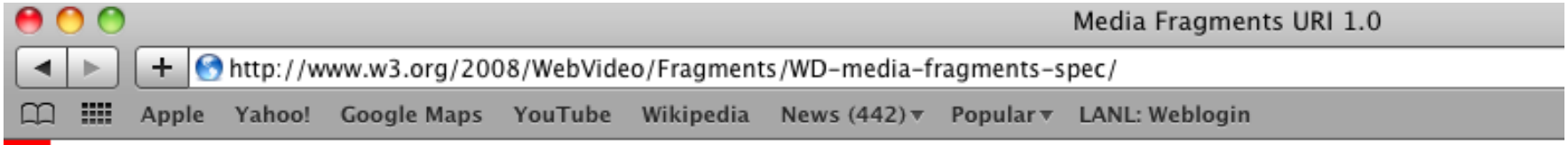
- For all types of Fragment URI, the Annotation must also create a dcterms:isPartOf link to the full resource for discovery purposes

Segments of Resources: Media Fragments

- Media Fragments allow anyone to create URIs that identify part of an image, audio or video resource.

- The most common use case is for rectangular areas of images:
  - http://www.example.org/image.jpg#xywh=50,100,640,480





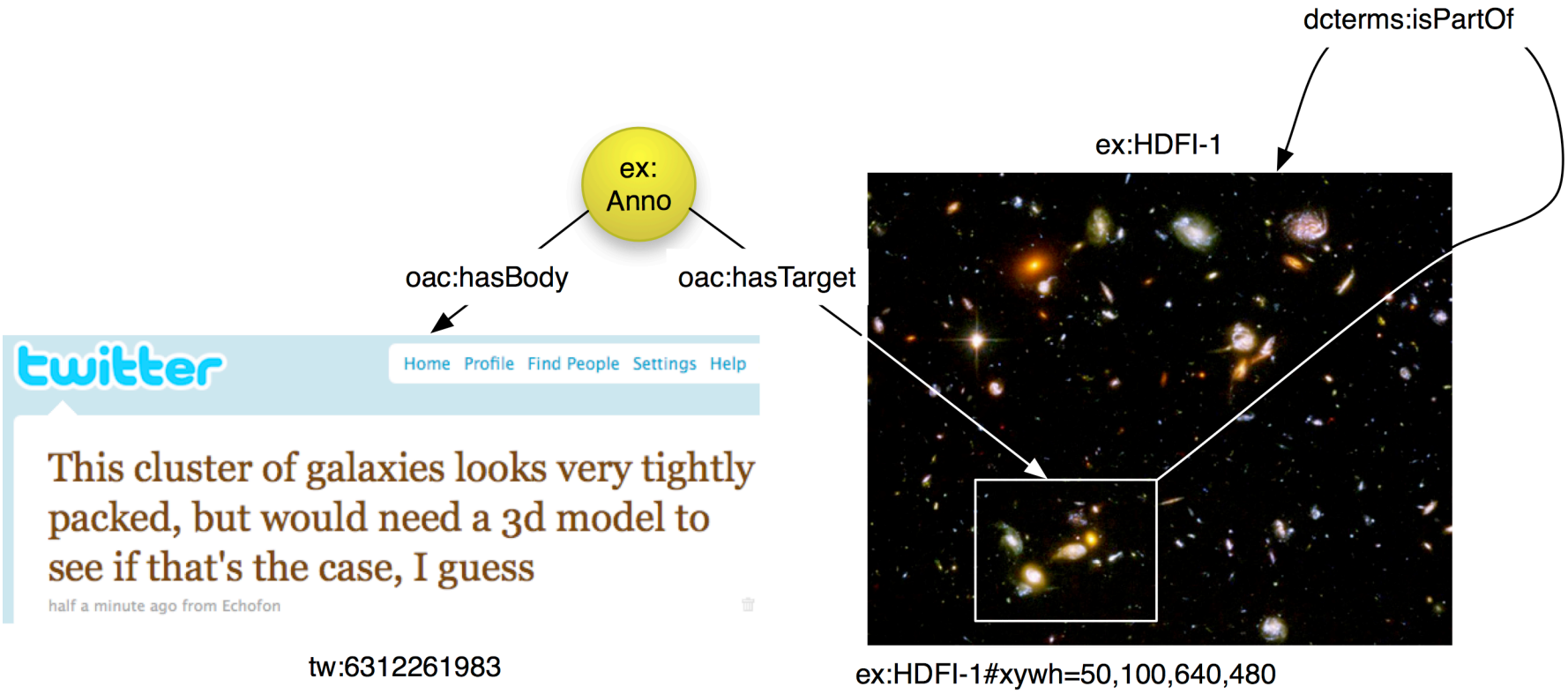
W3C Editor's Draft

Media Fragments URI 1.0

W3C Working Draft 15 March 2011

This version:
http://www.w3.org/TR/2011/WD-media-frags-20110315
Latest version:
http://www.w3.org/TR/media-frags
Previous version:
http://www.w3.org/TR/2010/WD-media-frags-20100624/
Editors:
Raphaël Troncy , EURECOM
Erik Mannens , IBBT Multimedia Lab, University of Ghent
Silvia Pfeiffer , W3C Invited Expert
Davy Van Deursen , IBBT Multimedia Lab, University of Ghent
Contributors:
Michael Hausenblas , DERI, National University of Ireland, Galway
Philip Jägenstedt , Opera Software
Jack Jansen , CWI, Centrum Wiskunde & Informatica, Amsterdam
Yves Lafon , W3C
Conrad Parker , W3C Invited Expert

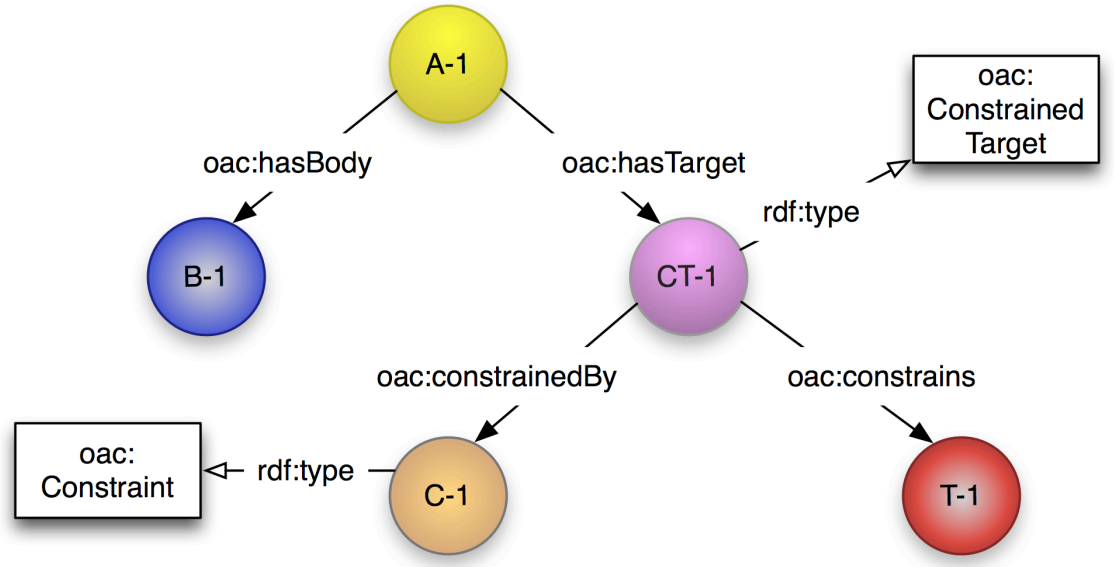
Media Fragments Example



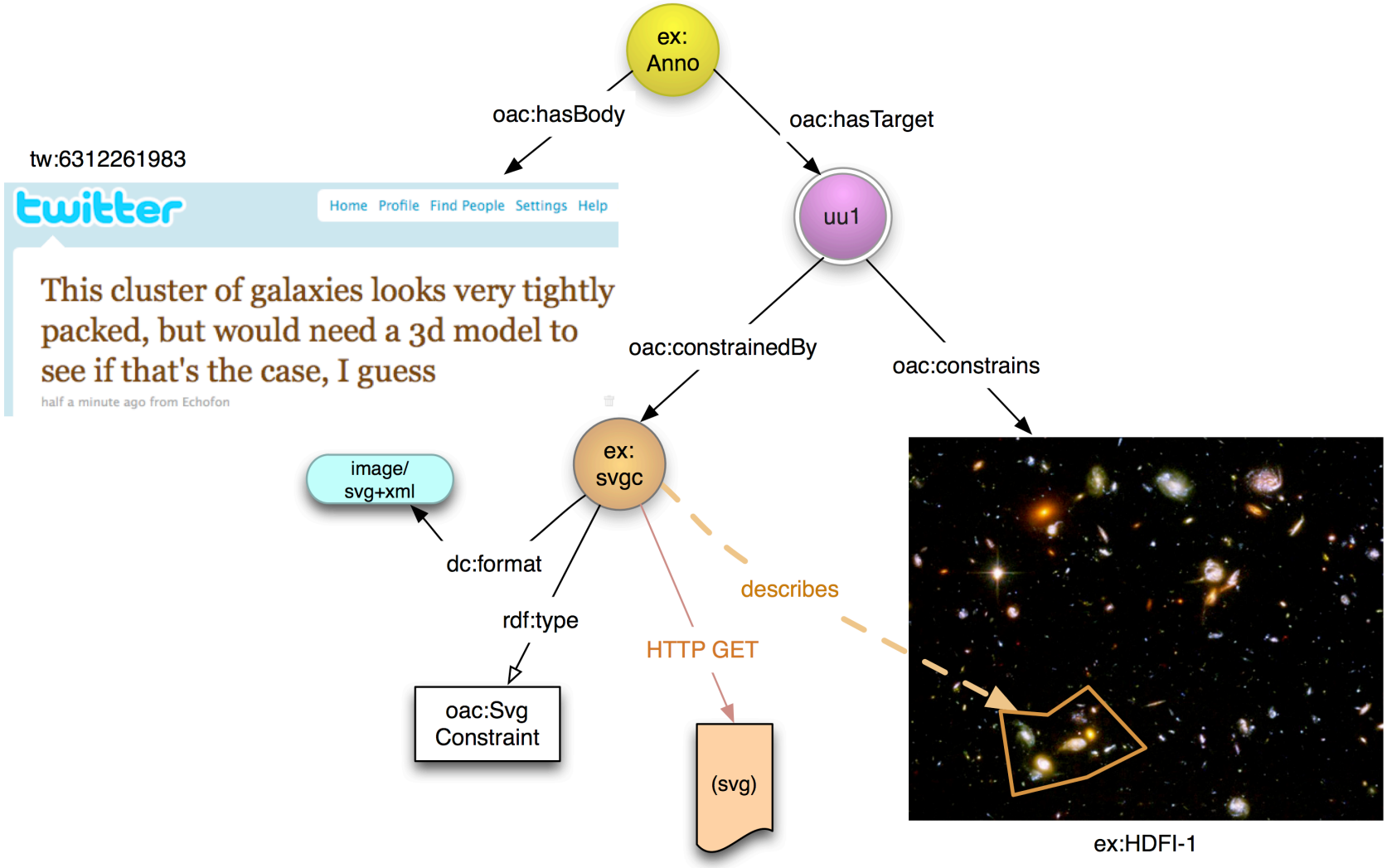
Complex Constraints

- We introduce a ConstrainedTarget resource that identifies the resource with constraints applied to it in order to fully identify the target of the Annotation
- Constraint resource describes how target resource is being used in the context of the Annotation

- The type of description is dependent on the nature of the target resource
- Different clients may support different types of description

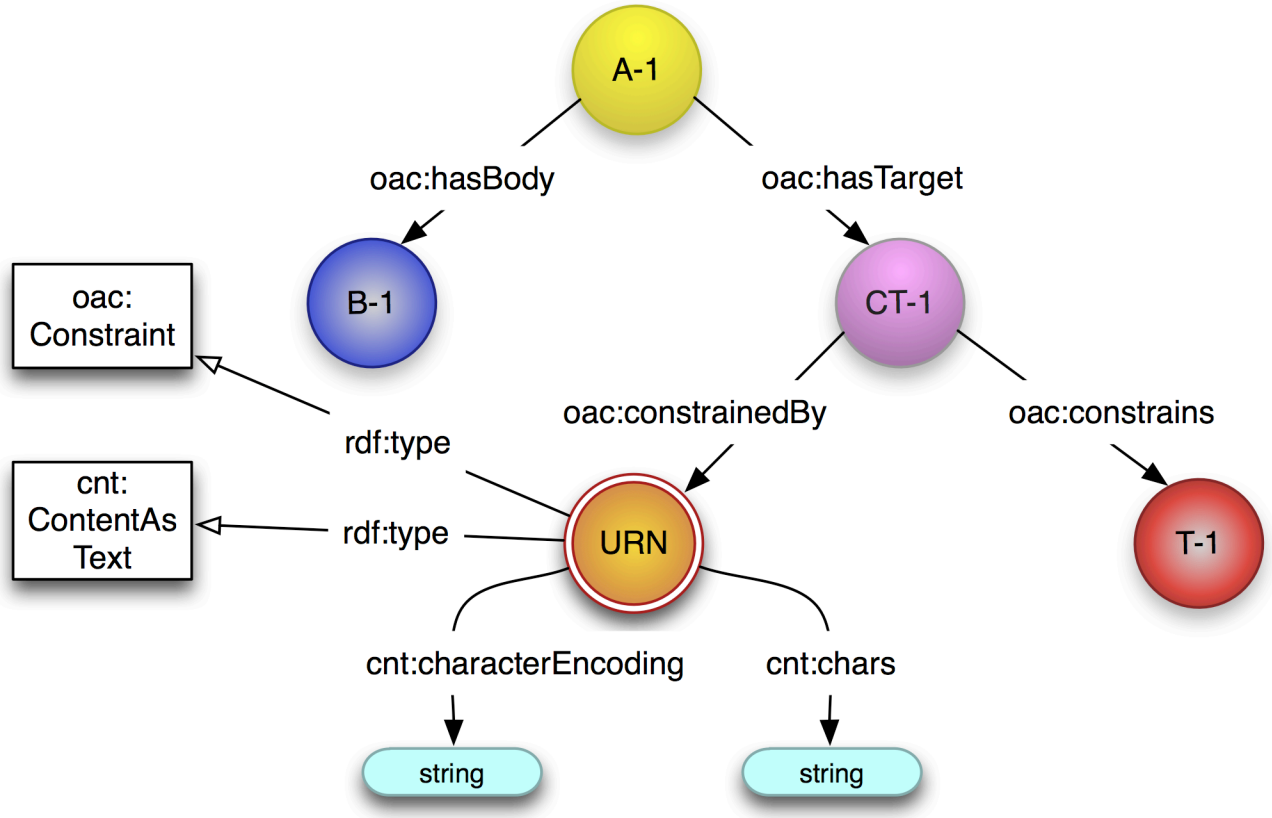


Complex Constraints Example



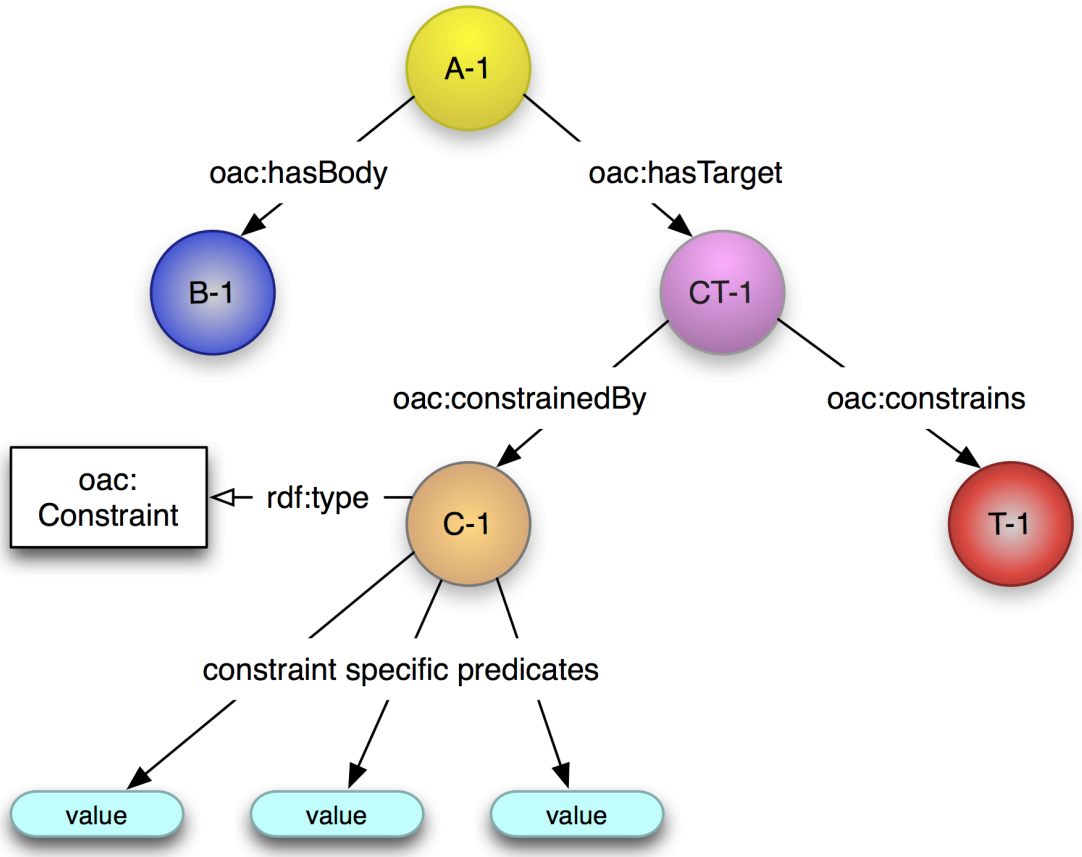
Inline Constraints

• We can also use inline information in the same way as for the Body resource to include the Constraint data.

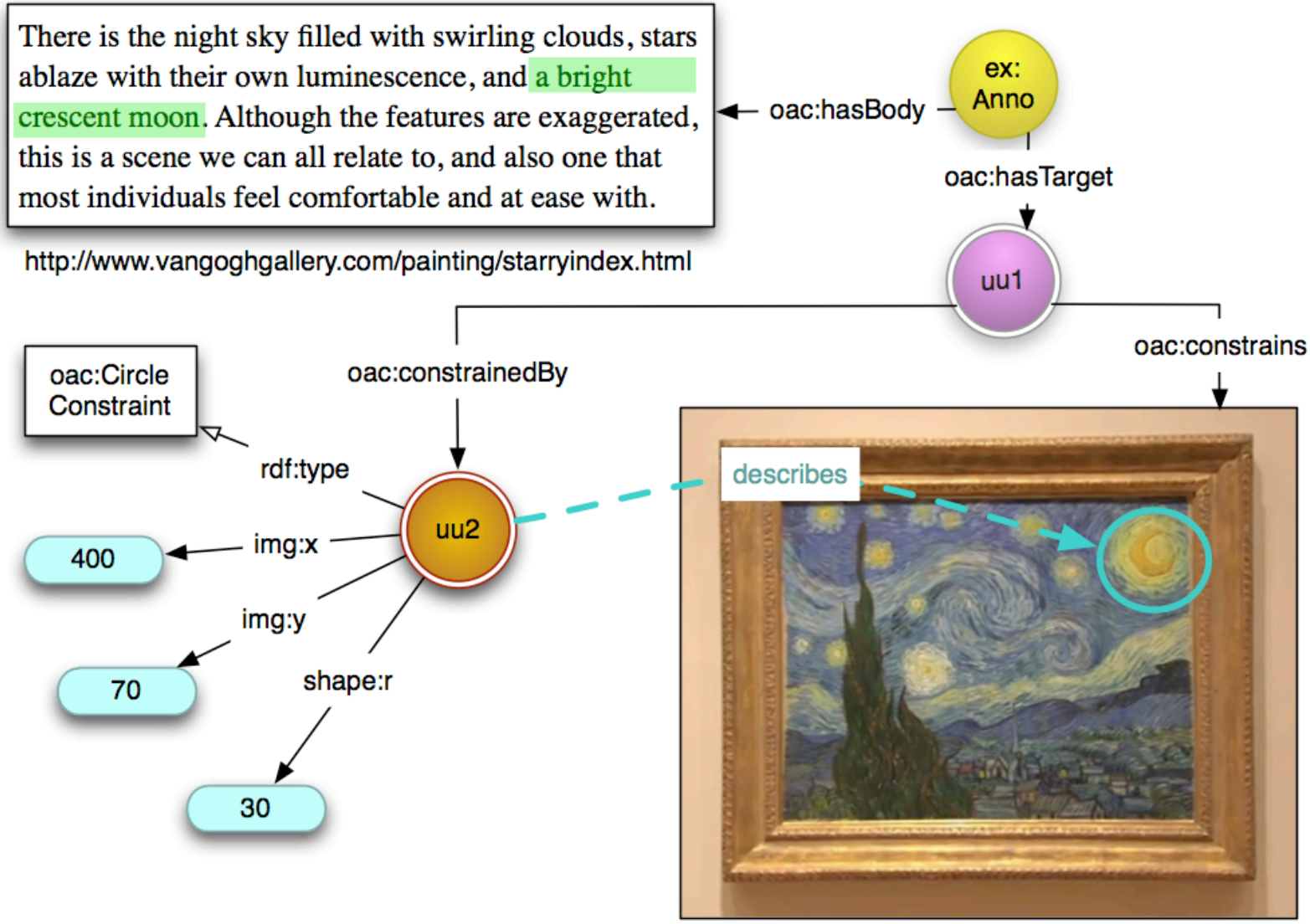


RDF Constraints

• The information could instead be linked to the Constraint, within the Annotation document.

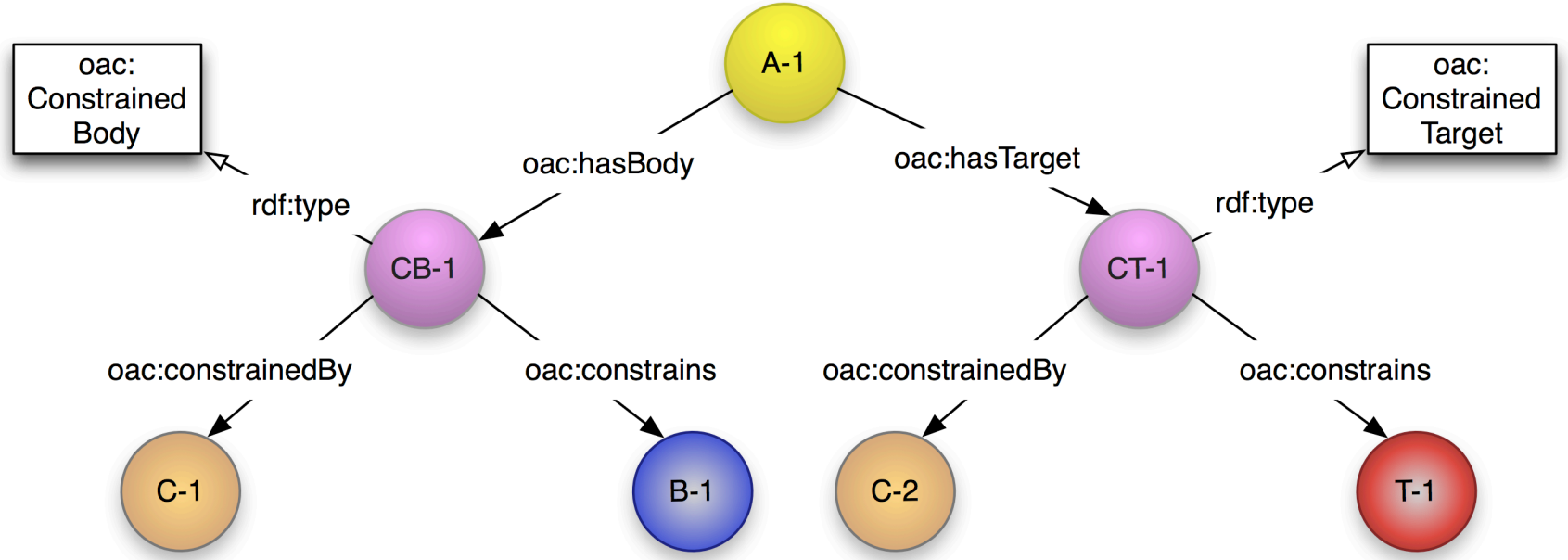


RDF Constraint Example



Constrained Body

- The Body may also be constrained in the same way as Targets



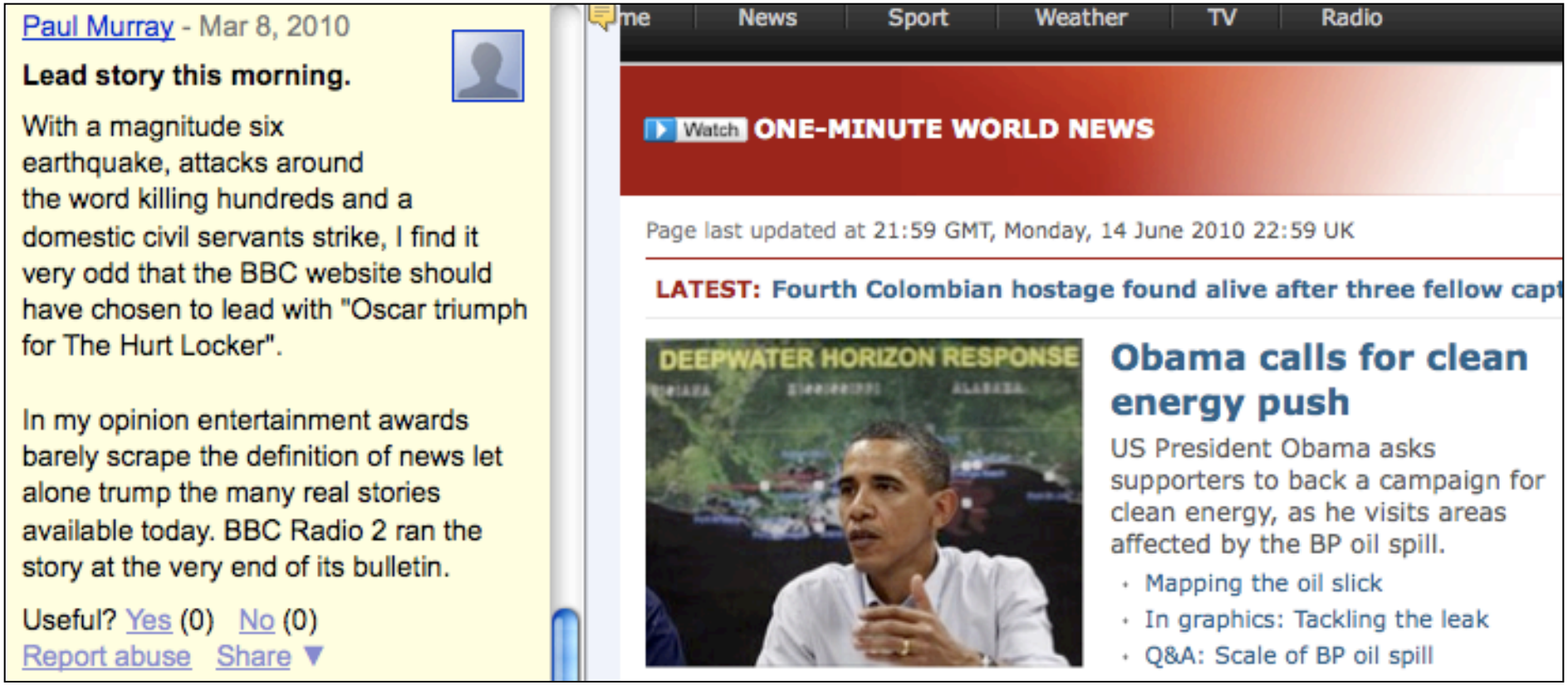
Open Annotation: Dealing with Web Time

- As regular Web resources, Body and Target of an Annotation have representations that can change over time.

- Body and Target can change independently of each other.

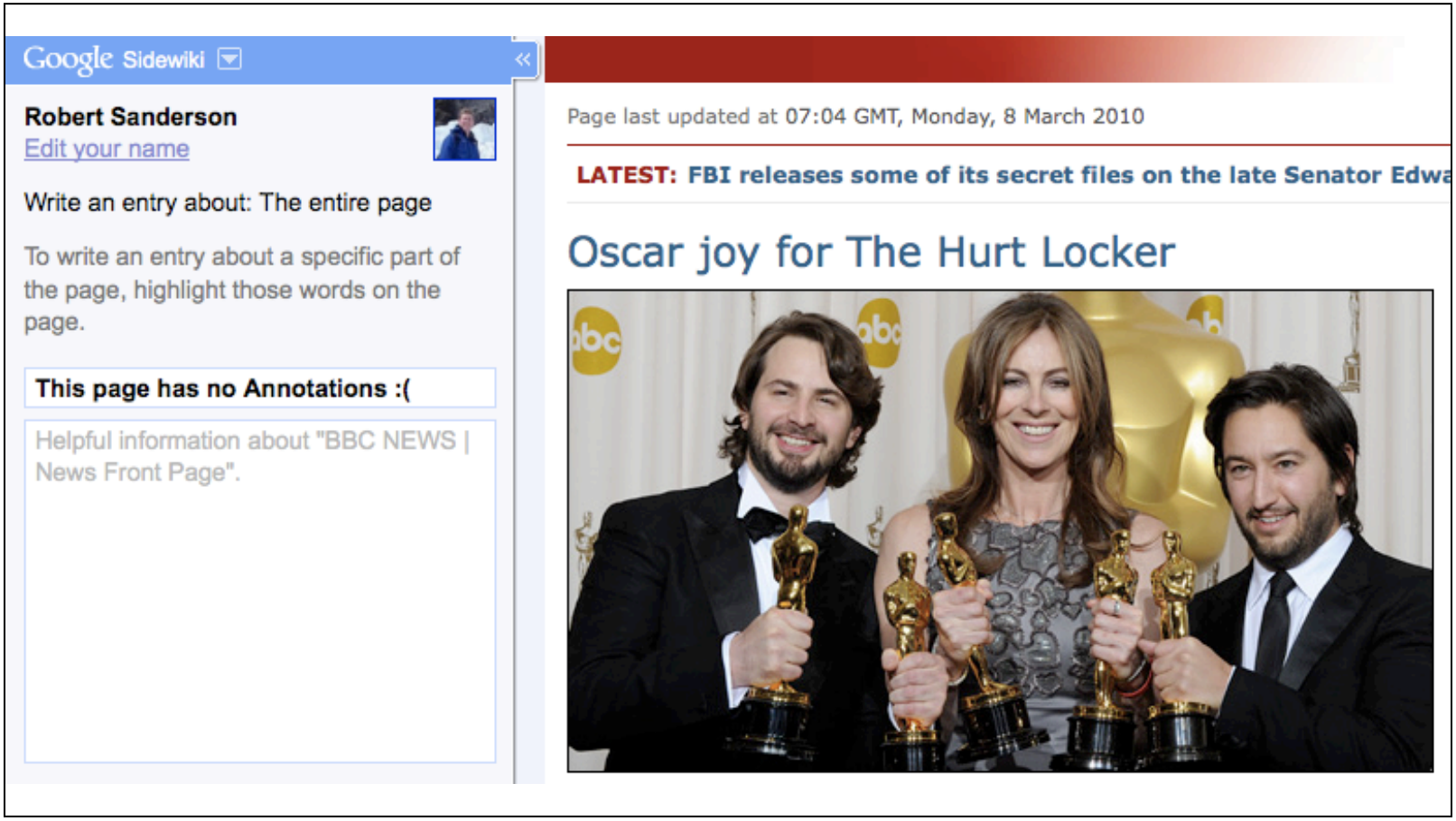
- If an Annotation involves resources as they existed at a particular point in time, this needs to be recorded.

Web-Centric Annotation: No Persistence



Google Sidewiki Annotation on *http://news.bbc.co.uk/* as of 2010-06-14

Web-Centric Annotation: No Persistence



Archived page from:
http://www.dracos.co.uk/work/bbc-news-archive/2010/03/08/07.05.html

Web-Centric Annotation: Desired Persistence

Paul Murray - Mar 8, 2010

Lead story this morning.

With a magnitude six earthquake, attacks around the word killing hundreds and a domestic civil servants strike, I find it very odd that the BBC website should have chosen to lead with "Oscar triumph for The Hurt Locker".

In my opinion entertainment awards barely scrape the definition of news let alone trump the many real stories available today. BBC Radio 2 ran the story at the very end of its bulletin.

Useful? Yes (0) No (0)
Report abuse Share ▼

Page last updated at 07:04 GMT, Monday, 8 March 2010

LATEST: FBI releases some of its secret files on the late Senator

Oscar joy for The Hurt Locker

Open Annotation: Dealing with Web Time

- As regular Web resources, Body and Target of an Annotation have representations that can change over time.

- Body and Target can change independently of each other.

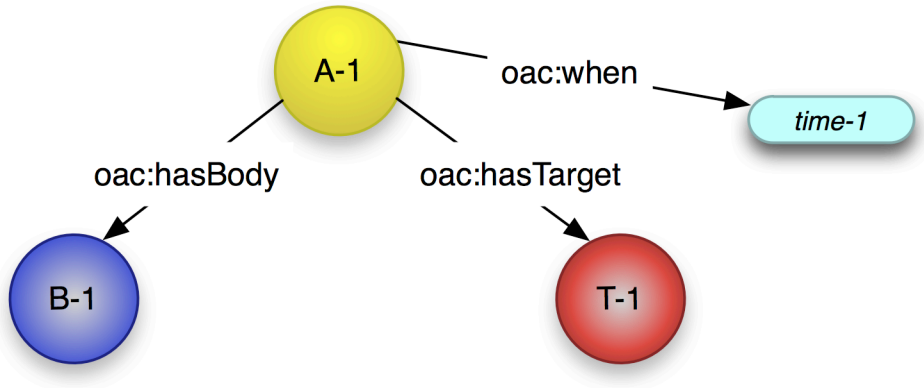
- If an Annotation involves resources as they existed at a particular point in time, this needs to be recorded.

- The OAC model provides hooks for doing so:
  - Timeless Annotations;
  - Uniform Time Annotations;
  - Varied Time Annotations.

Time: Uniform Annotations

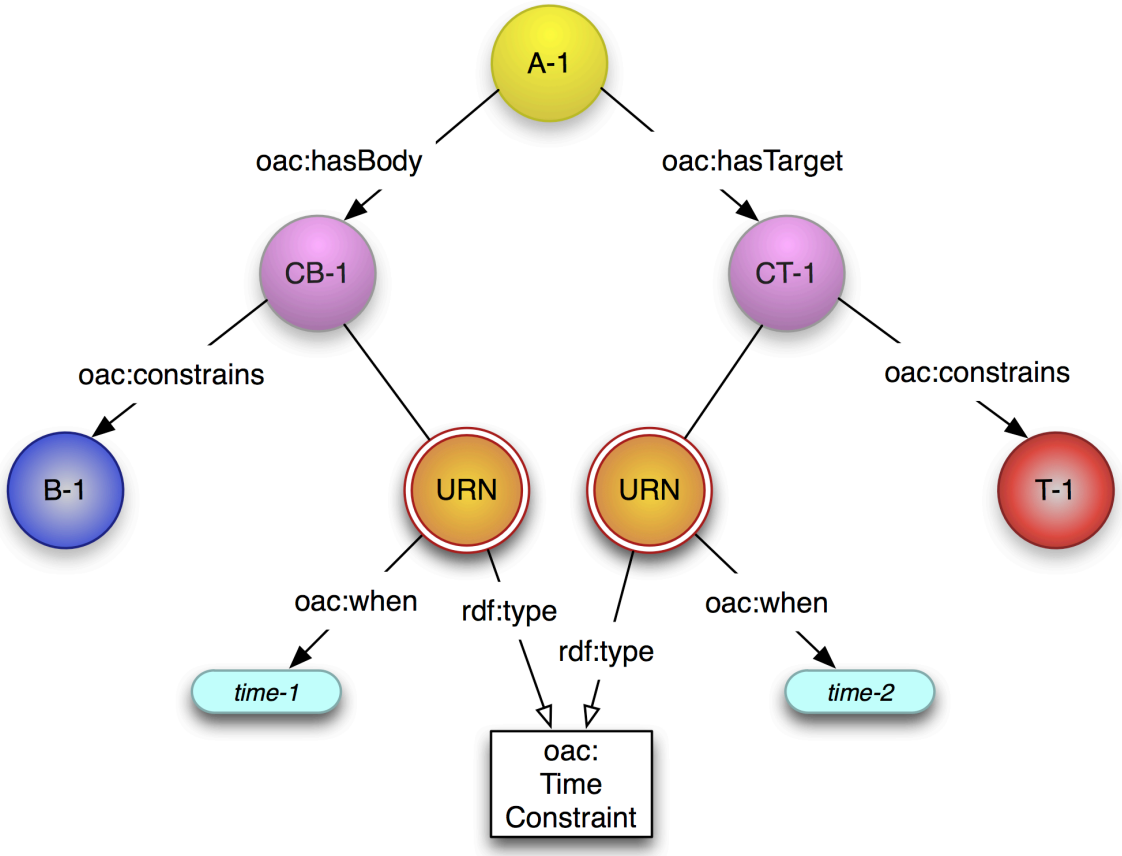
• If a single point in time is applicable to all resources, we attach it to the Annotation using the oac:when predicate

• This timestamp can be used to discover appropriate, archived copies of the resources



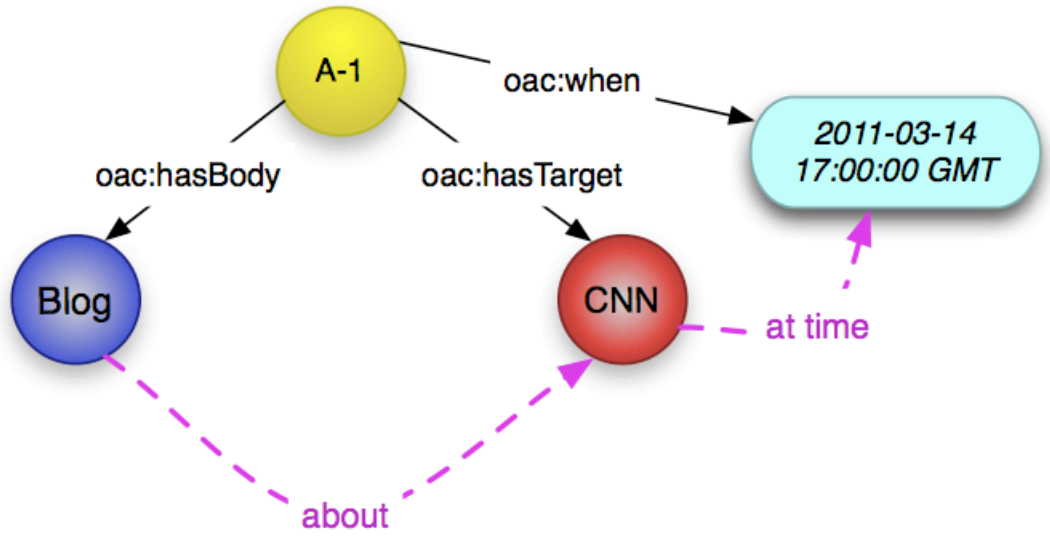
Time: Varied Annotations

• If different timestamps are required for each resource, we use oac:when from an oac:TimeConstraint.

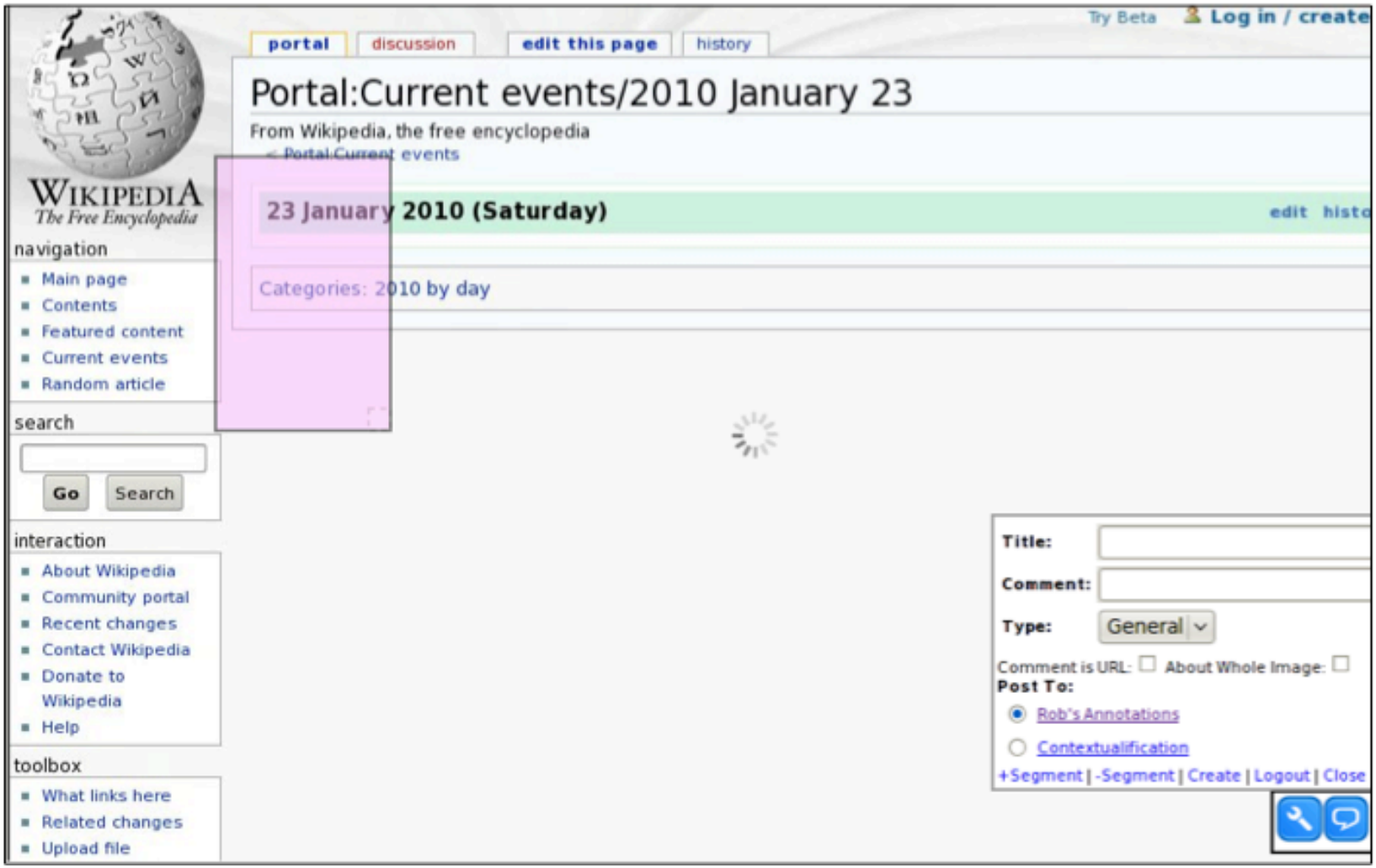


Memento + Open Annotation: Persistent Annotations

• In order to reconstruct the Annotation as intended: Use Memento to obtain an archived representation of B and T as they existed at the `oac:when` datetime.

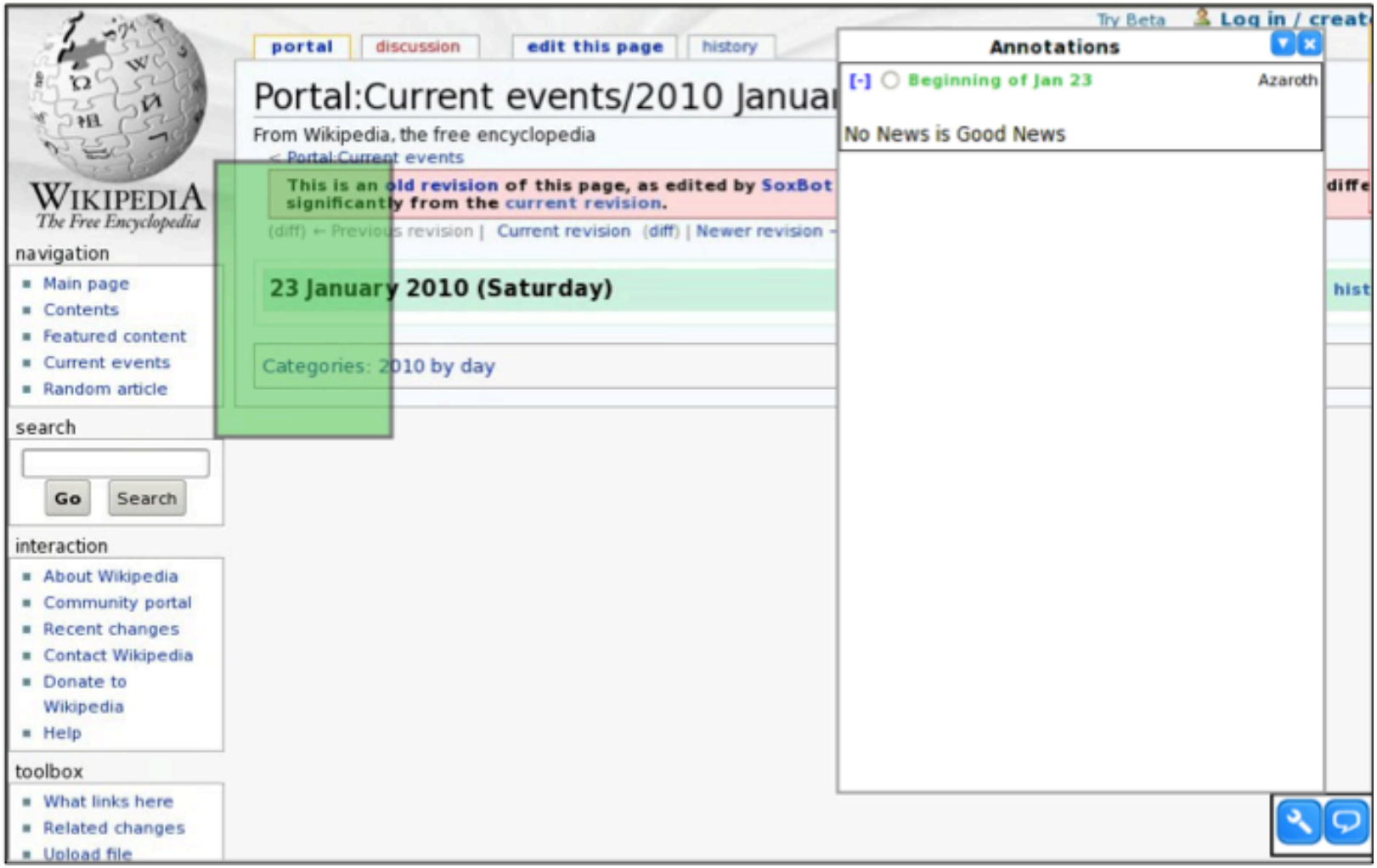


Create an Annotation



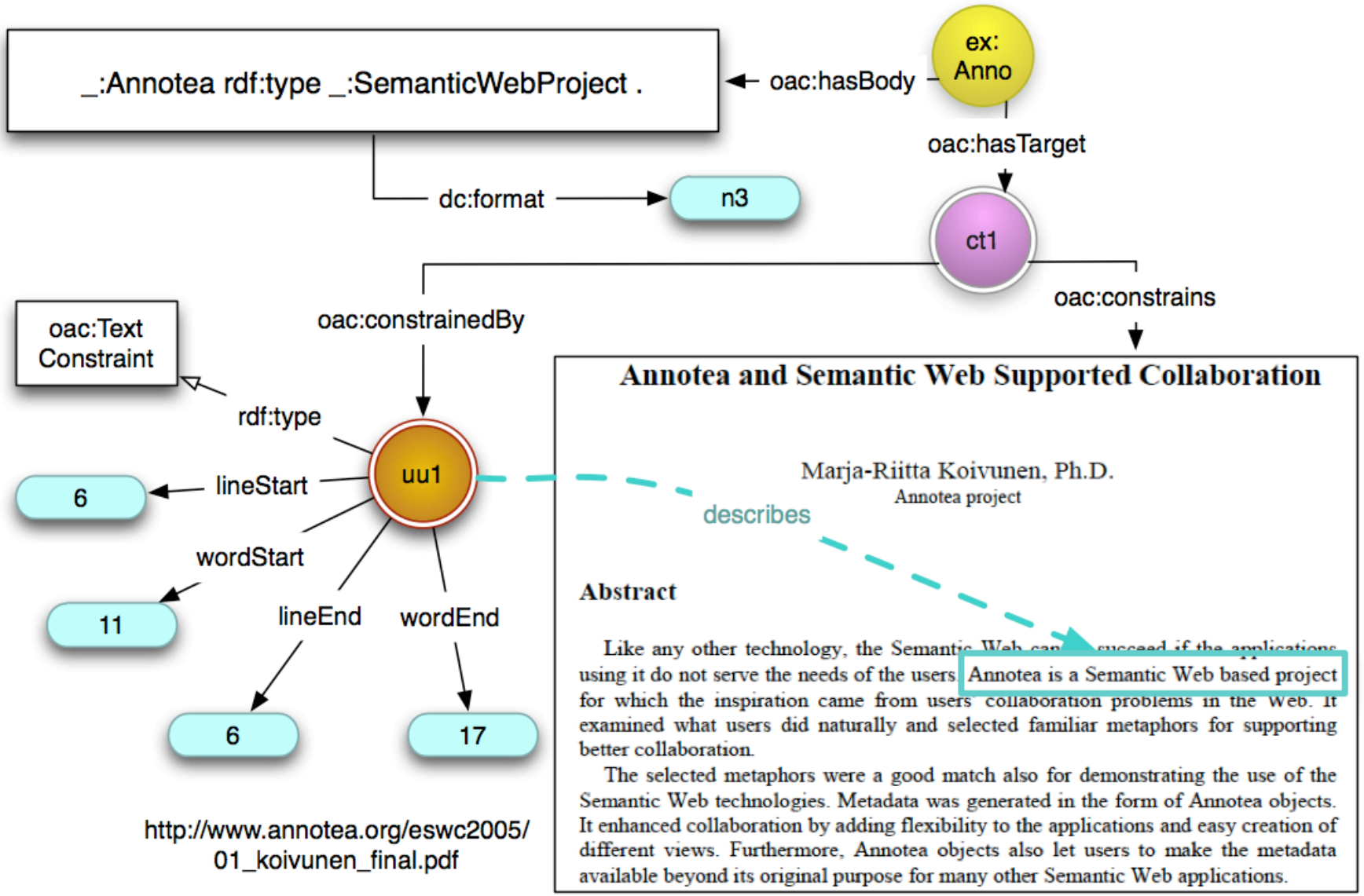
Reconstruct the Annotation without Memento

Reconstruct the Annotation with Memento

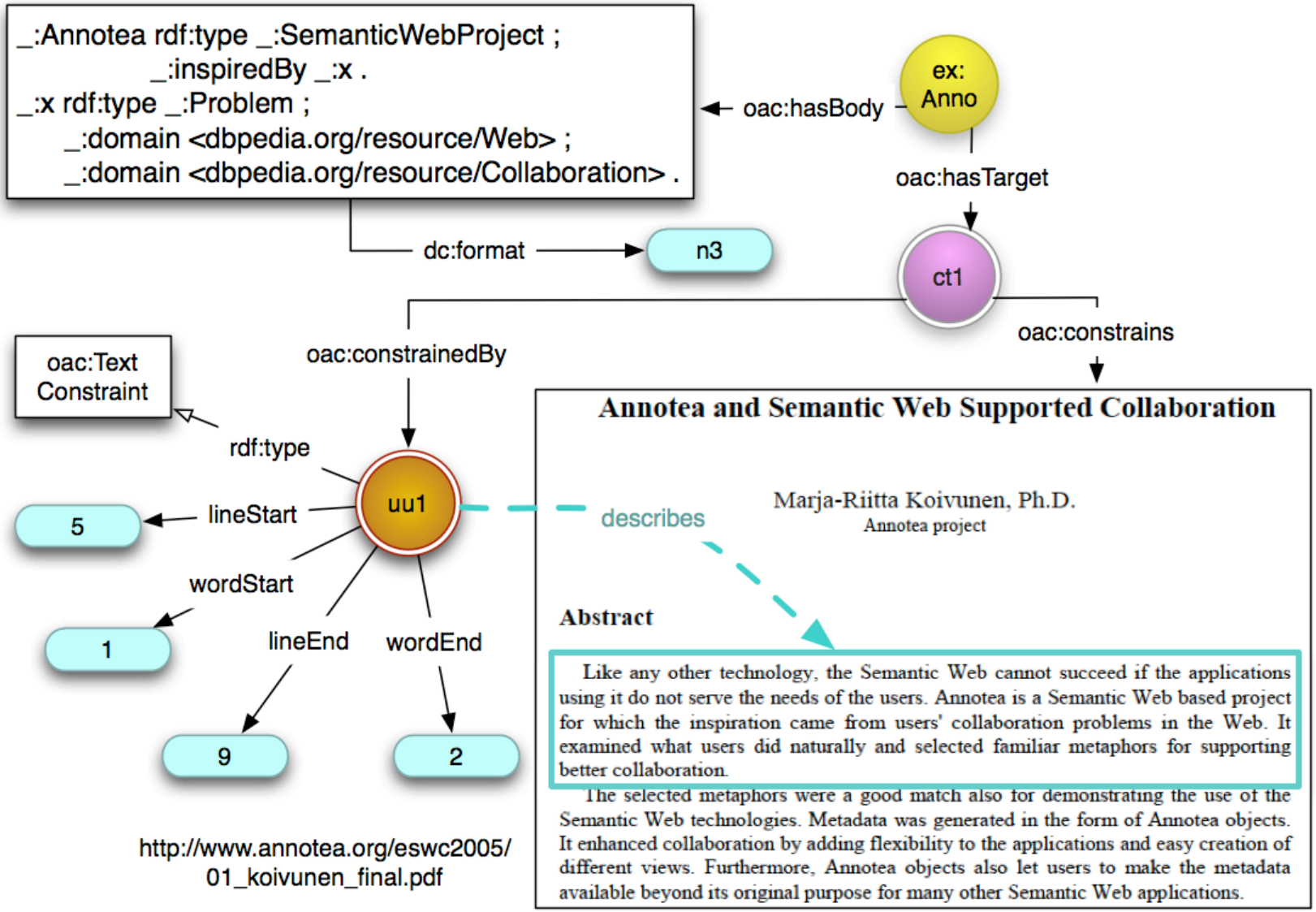


paper at *http://arxiv.org/abs/1003.2643*

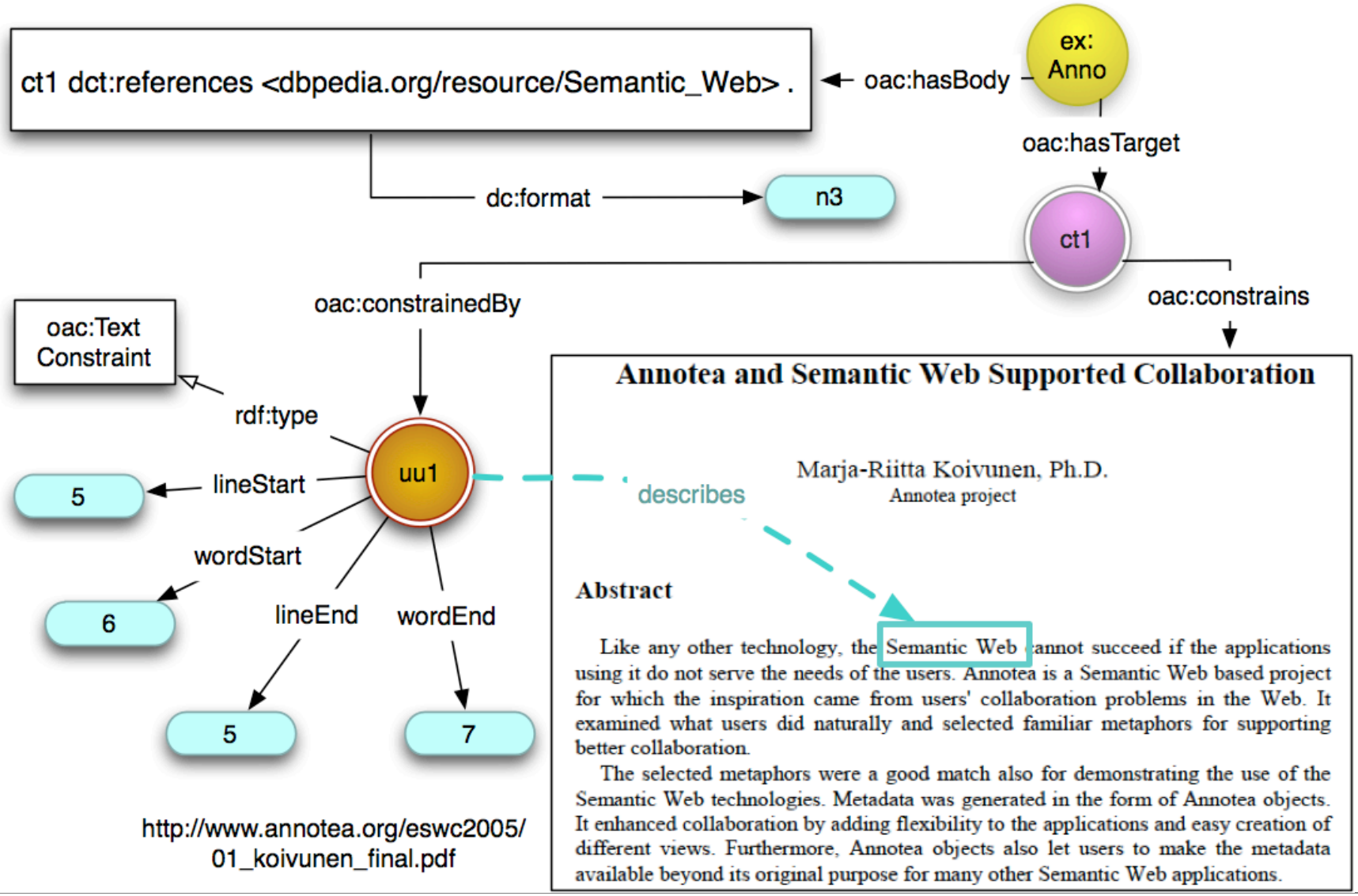
Structured Annotations (1)



Structured Annotations (2)



Structured Annotations (3)

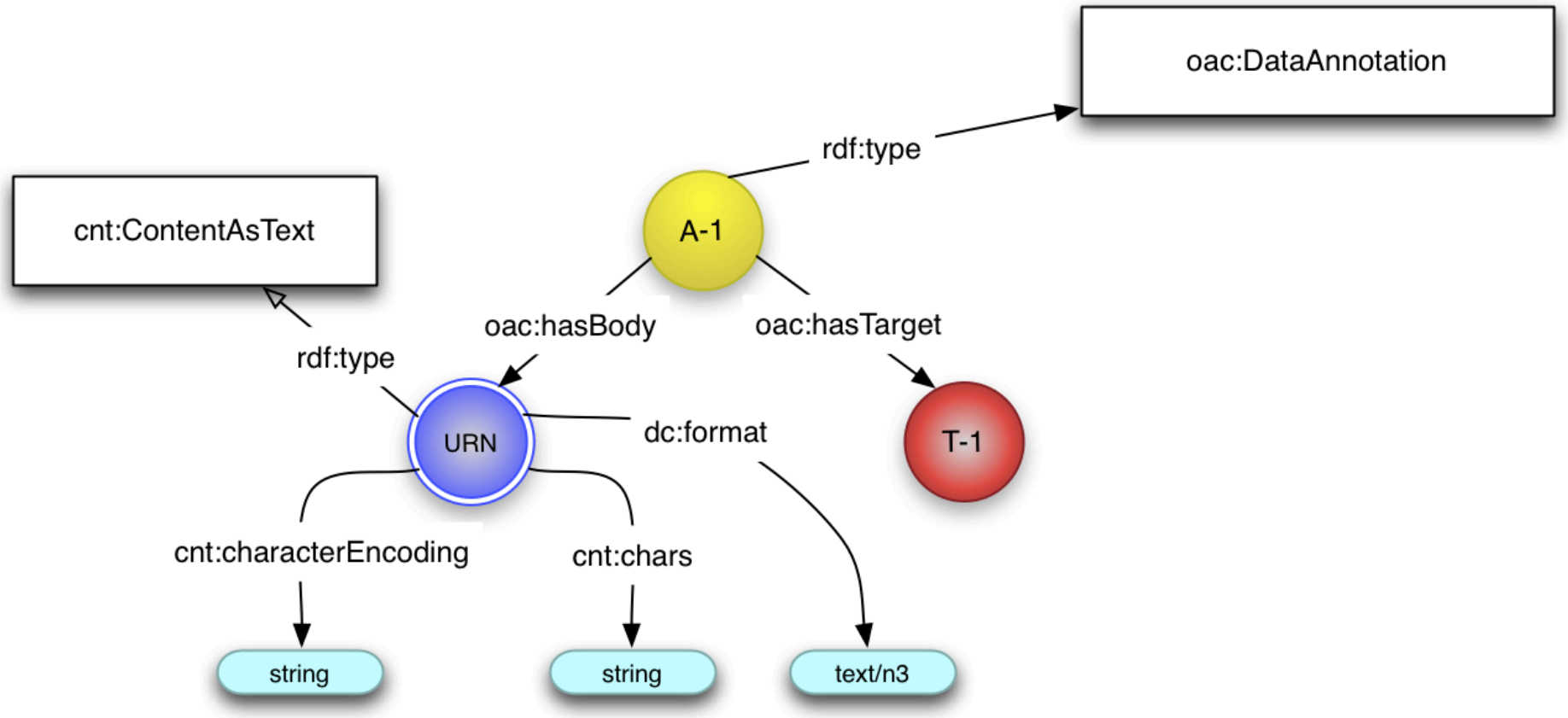


Approach: Inline Structured Body

- Typing the Annotation, e.g. rdf:type oac:DataAnnotation

- Inline structured body:
  - URN-identified node as Body
  - Use rdf:type “rdfg:Graph” to indicate node is a named graph – http://www.w3.org/2004/03/trix/rdfg-1/
  - Use W3C Content in RDF approach to embed structured Body: http://www.w3.org/TR/Content-in-RDF10/
    - Can be as XML or as Text
    - Use dc:format to express MIME type, application/rdf+xml ; text/n3 ; text/turtle

Approach: Inline Structured Body

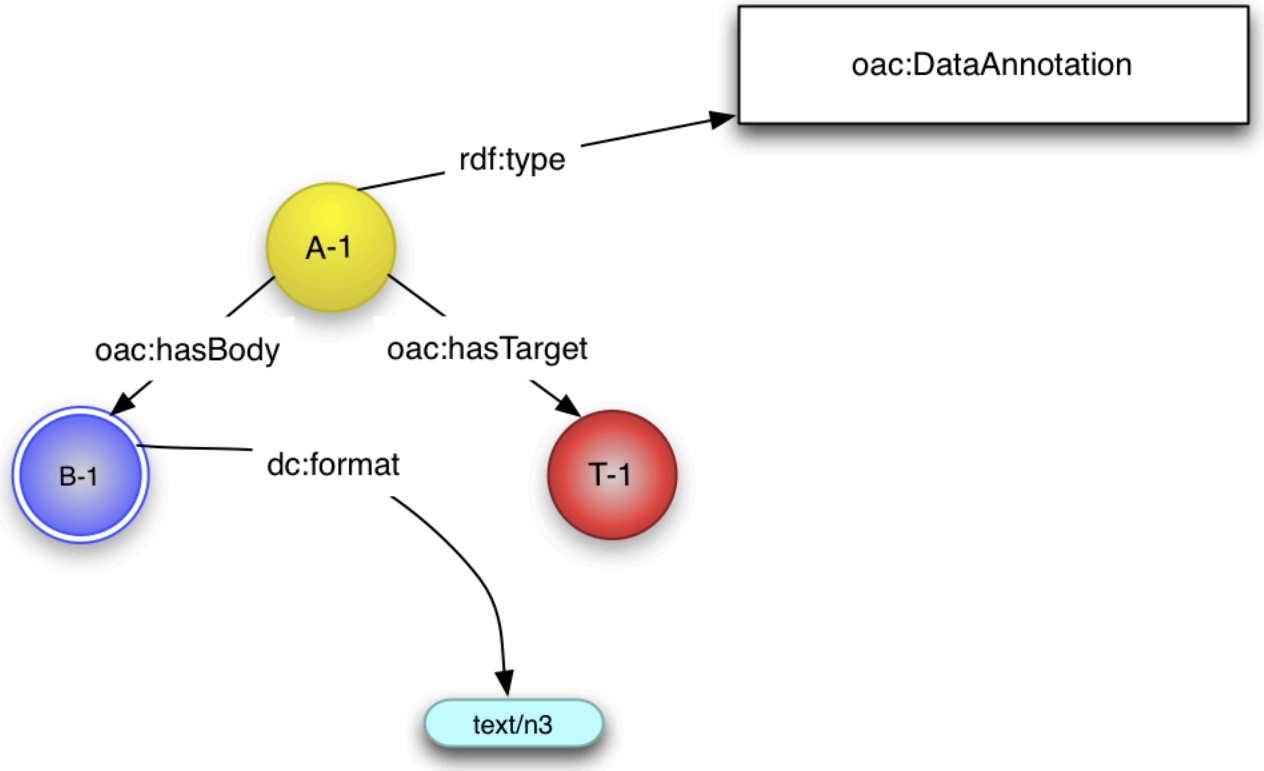


Approach: Out of Bound Structured Body

- Typing the Annotation, e.g. rdf:type oac:DataAnnotation

- Out of Bound structured body:
  - HTTP URI for Body
  - Body is a Web-based document
    - Use dc:format to express MIME type, application/rdf+xml ; text/n3 ; text/turtle

Approach: Out of Bound Structured Body



References – Open Annotation & SharedCanvas

- Sanderson, R., and Van de Sompel, H. (2010) Making Web Annotations Persistent over Time. Proceedings of the 10th ACM/IEEE-CS Joint Conference on Digital libraries. http://arxiv.org/abs/1003.2643
- Sanderson, R., Albritton, N., Schwemmer, R., Van de Sompel, H. (2011) SharedCanvas: A Collaborative Model for Medieval Manuscript Layout Dissemination http://arxiv.org/abs/1104.2925
- Sanderson, R., Van de Sompel, H. (2011) Open Annotation Alpha3 Data Model Guide. http://www.openannotation.org/spec/alpha3/