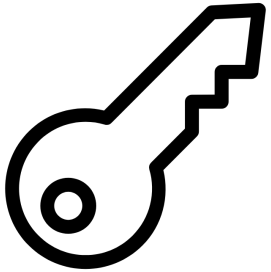
Big Data and Controlled Vocabularies

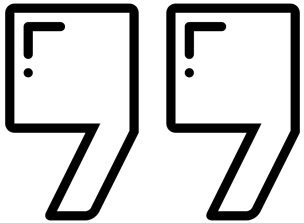
Camille Mathieu
Information Science Specialist
17 June 2019

Jet Propulsion Laboratory
California Institute of Technology

Quick-Take Overview



Key Concepts



Examples from JPL

Taking Action

#SLAtaxo

Big Data? Controlled Vocabularies?

Credit: JustAGirlAndHerBlog.com

Stuff vs **Strategy**

What organizational schemas can you impose on existing and future content to make it easier to find, aggregate, manage, and disposition?

#SLAtaxo

Key Concepts: the Stuff

- Big Data → “an accumulation of data that is too large and complex for processing by traditional database management tools”
- The 5 Vs: volume, velocity, variety, veracity, and value

#SLAtaxo

Key Concepts: the Stuff

Credit: LifeOnACanadianIsland.Blogspot.com

- Big Data → the digital information reality of many organizations

#SLAtaxo

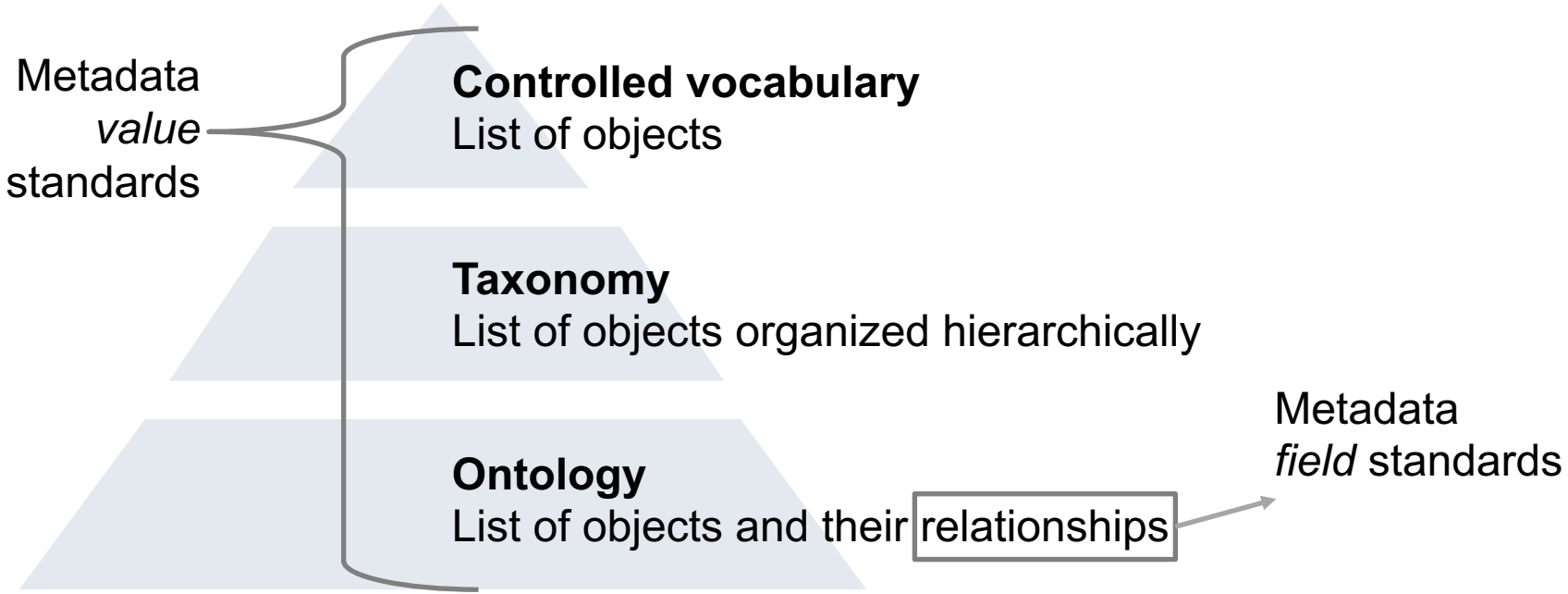
Key Concepts: the Strategies

Credit: LifeOnACanadianIsland.Blogspot.com

- Controlled vocabulary
- Taxonomy
- Ontology
- Metadata standards

#SLAtaxo

Key Concepts: the Strategies

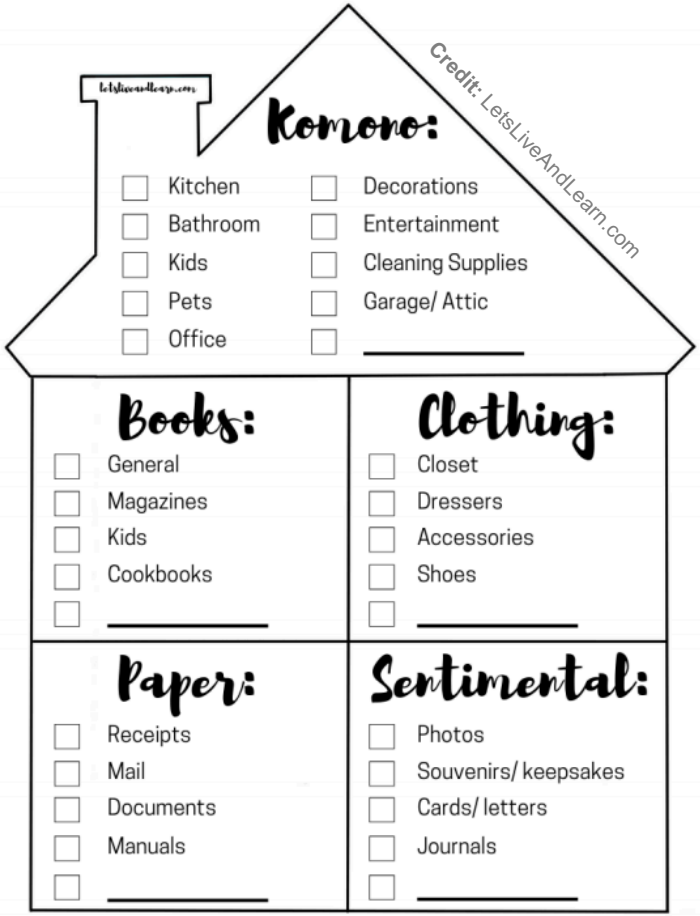


#SLAtaxo

Key Concepts: the Strategies

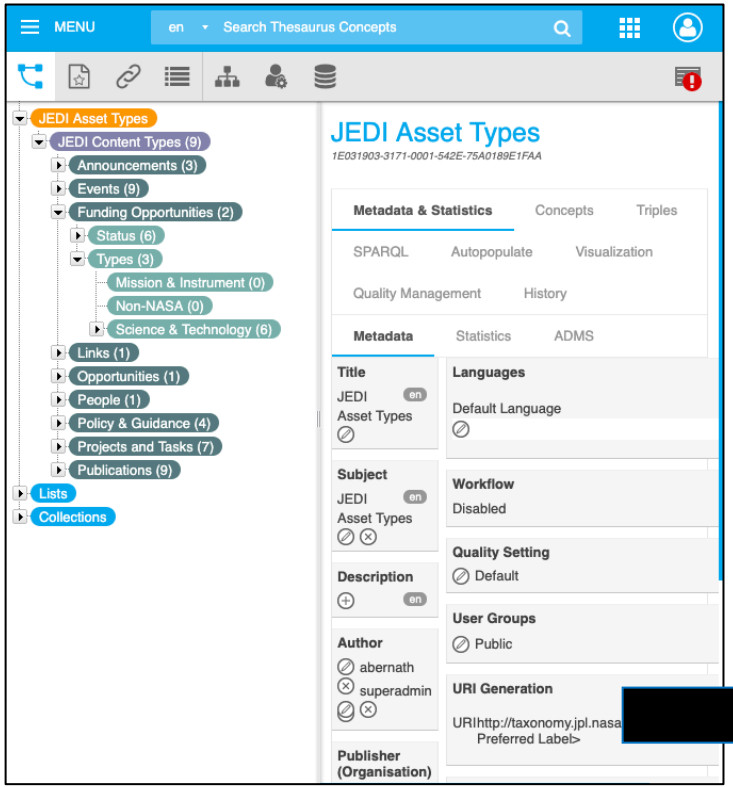
- Controlled vocabulary
- Taxonomy
- Ontology
- Metadata standards

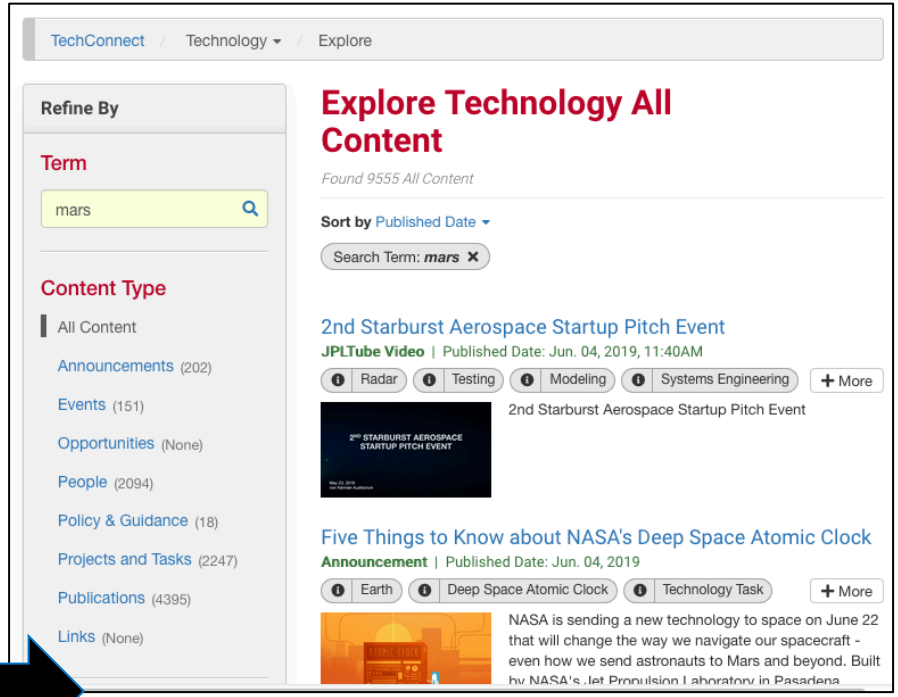
=



#SLAtaxo

Example #1: PoolParty Taxonomy Management





#SLAtaxo

Example #2: Metadata Standardization

Field Name	Definition	Justification
Creator	An entity primarily responsible for making the resource.	Facilitates discovery (search); allows content to be associated with JPLer.
Created	Date of creation of the resource.	Vital for effective ERM and disposition.
Identifier	An unambiguous reference to the resource within a given context.	Vital for effective ERM and disposition; facilitates "dashboard" use and reuse of content.
Title	A name given to the resource.	Facilitates discovery (search).
Format	The file format, physical medium, or dimensions of the resource.	Vital for effective ERM and disposition.
Type	The nature or genre of the resource.	Facilitates discovery (search); vital for effective ERM and disposition; facilitates "dashboard" use and reuse of content.
Rights	Information about rights held in and over the resource.	Facilitates "dashboard" use and reuse of content.
Keywords	The topic of the resource per enterprise keywords.	Facilitates discovery (search).

- *Field standards* (schema) drive *value standards* (taxonomy)

#SLAtaxo

Example #3: Crowdsourced Folksonomies

Labels

Labels for Subject Terms

The following table shows the preferred labels for subject terms related to Spacecraft Mechanisms. An article may be labeled with many subject terms to allow for faceted searching.

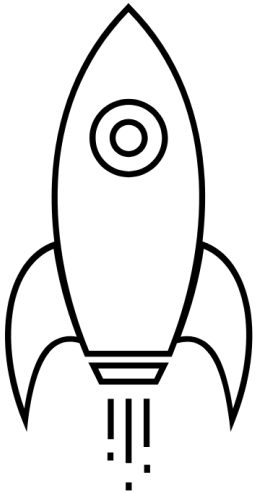
Label
actuators
bearing
damper
feedback-device
flexure
gear
gearbox
hinge
latch
locking
motor
release-device
seal
spring
valve

Component	Summary	Related pages
Damper	*Dampers is used in a broad sense in this category. This can include traditional fluid or eddy current dampers, but it also includes end of travel stops like crushable honeycomb that dissipates energy.* **Label** damper	Click here to view related content Cardboard Crushable Honeycomb Energy Absorber Crushable Honeycomb Energy Absorber DEB Viscous Rotary Dampers Eddy Current Dampers JPL Damper

#SLAtaxo

Taking Action

- Getting started means understanding both the *stuff* you need to manage, and the *strategies* at your disposal
  - Implementation requires evaluation of tools
- Designing taxonomies and other standards should be done with findability/reusability goals in mind
- Source controlled vocabularies from published taxonomies, and (more importantly) from subject matter experts
- Want linked data? Taxonomy is the first step to ontology



#SLAtaxo

The Life-Changing Magic of Taxonomied Data

Big Data

Credit: DreamsTime.com

Big Data **+ Strategy**

#SLAtaxo

thank you!

Jet Propulsion Laboratory
California Institute of Technology

Camille Mathieu
cmathieu@jpl.nasa.gov
@searchbrarian
/camillemathieu

Thoughts? **#SLAtaxo**