

APS Taxonomy Project

Arthur Smith, American Physical Society

April 2014

History (APS subjects perspective)

- 1965 – “Analytic Subject Index”, PRL reorder
- 1970 – Phys Rev split into A-D
- 1975 – PACS (from ICSU, AIP, IEE + others)
- 1977 – revised PACS, internal database
- 1997+ – journals go online, indexes obsolete?
- 2010 – last PACS update from AIP
- 2011 – PRX topical browsing, not PACS

Project Timeline

- Initiated in early 2013
 - Discussions with APS stakeholders
 - No consensus on data structure, organization
 - Significant software development needed
- November 2013: consultant engaged
 - Taxonomy Strategies (Washington DC)
- By March 2014: some decisions on structure, software:
 - Facets rather than single hierarchy
 - SKOS/RDF
 - Acquired PoolParty server to create and manage

Taxonomy project goals

- new capability for topical browsing of online journals, expanding the limited interface used now for some APS journals.
- Easy to use for authors to index their submitted journal articles.
- PACS mappable.
- Assists editorial workflow, e.g., assigning articles to journal sections or particular editors, finding referees with the right expertise, etc.
- Applicable to all APS content, e.g., meeting sessions and legacy content.

SKOS/RDF

- Several standard approaches to taxonomy/thesaurus creation:
 - ISO Thesaurus standard (see ISO 25964, 2012/2013)
 - SKOS (“Simple Knowledge Organization System”)
 - <http://www.w3.org/2004/02/skos/>
- SKOS based on W3C “Resource Description Framework” (RDF) standard, compatible with “Linked Data”
- Basic components of SKOS:
 - “concept” = “unit of thought”, identified by a URL, given preferred and alternate “labels”
 - “label” = string representing concept in a given language (term)
 - “relationships” = links between “concepts”. Hierarchy (broader/narrower), simple associations, or across concept schemes (“match” relationships)
 - “concept scheme” = grouping of concepts into a vocabulary/taxonomy (labels should be unique within a given concept scheme)
- SKOS is extensible through RDF to add custom properties and relationships

Indexing articles with SKOS

Example:

<<http://dx.doi.org/10.1103/PhysRevD.89.072001>>

dcterms:subject

<http://link.aps.org/cs/APSTaxonomy/top_quark> .

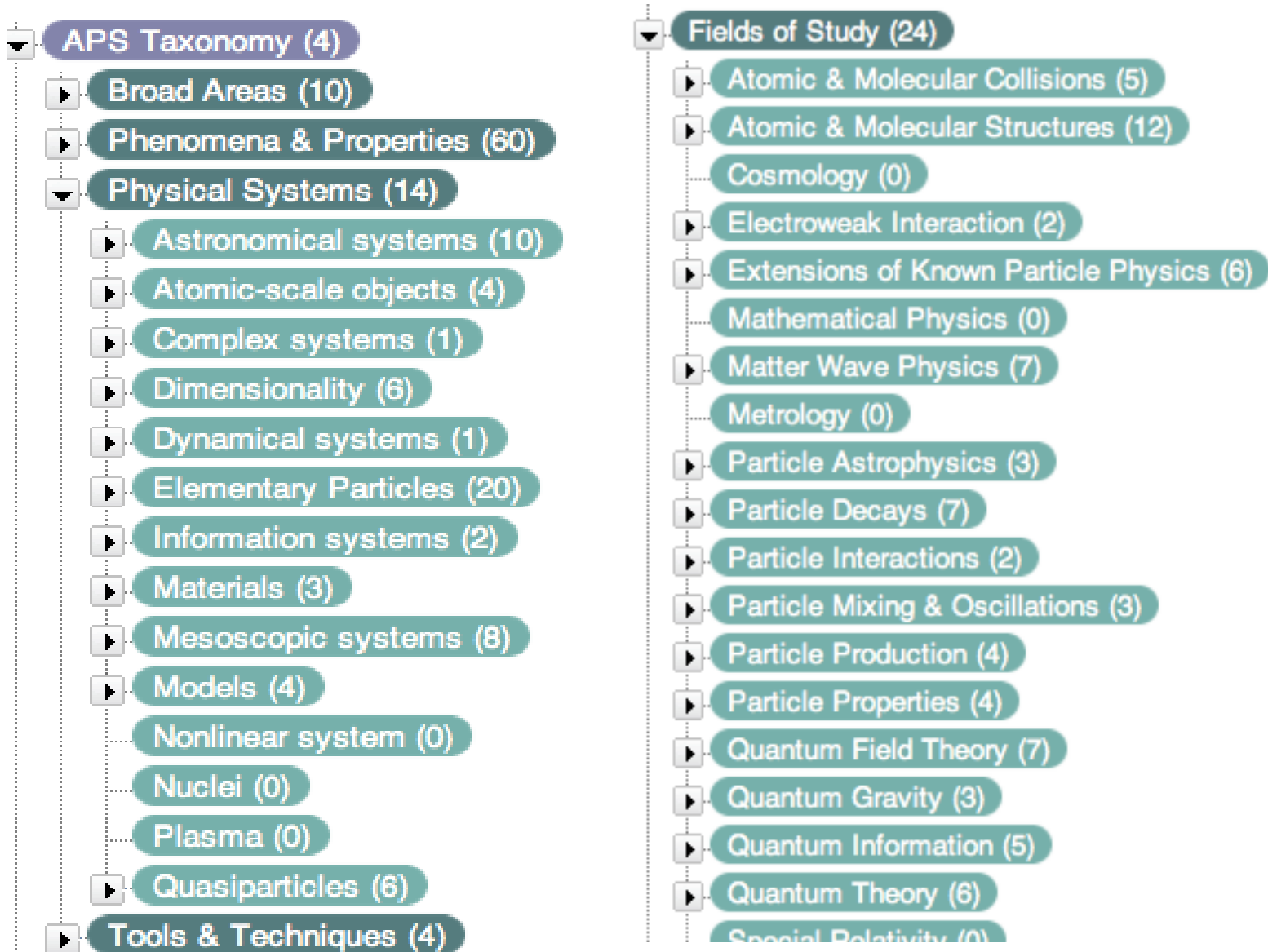
We may allow more detailed indexing using existing vocabularies (HEP ontology?):

<<http://dx.doi.org/10.1103/PhysRevD.89.072001>>

dcterms:subject

<<http://cern.ch/thesauri/HEPontology.rdf#Fermilab>> .

What does it look like (early draft)?



What's next?

- Finalize organization and rules for completion
- Work with physicists (mostly editors) to fill out and reconcile the facets
- User interfaces for browsing, adding terms, editorial assignment
- Publish taxonomy as linked data and use it – expected later this year.

Questions

- Arthur Smith, Lead Data Analyst, APS Journal Information Systems, apsmith@aps.org
- Joseph A Busch, Principal, Taxonomy Strategies, jbusch@taxonomystrategies.com