Changing from a Hierarchical Classification to a to a Linked Data Standard: American Physical Society (APS) Taxonomy Case Study

Joseph Busch & Vivian Bliss

Outline

- Motivation and use cases
  - PACS and how it was used
  - Useful things we weren't doing with PACS
- Methodology
  - How we got to facets
  - Tools and standards
  - Building the Taxonomy
  - Validation

PACS*: 1975-2010

- Alphanumeric codes
- Strict (single-parent) hierarchy
- Incomplete and complex labels
- Some concepts are repeated many times

Semiconductors

61.43.Dq - Amorphous semiconductors, metals, and alloys

61.72.uj - III-V and II-VI semiconductors [61.72.U- - Doping …]

61.82.Fk - Semiconductors [61.82.-d - Radiation effects …]

64.70.kg - Semiconductors [64.70.K- Solid-solid transitions]

64.75.Qr - Phase separation and segregation in semiconductors

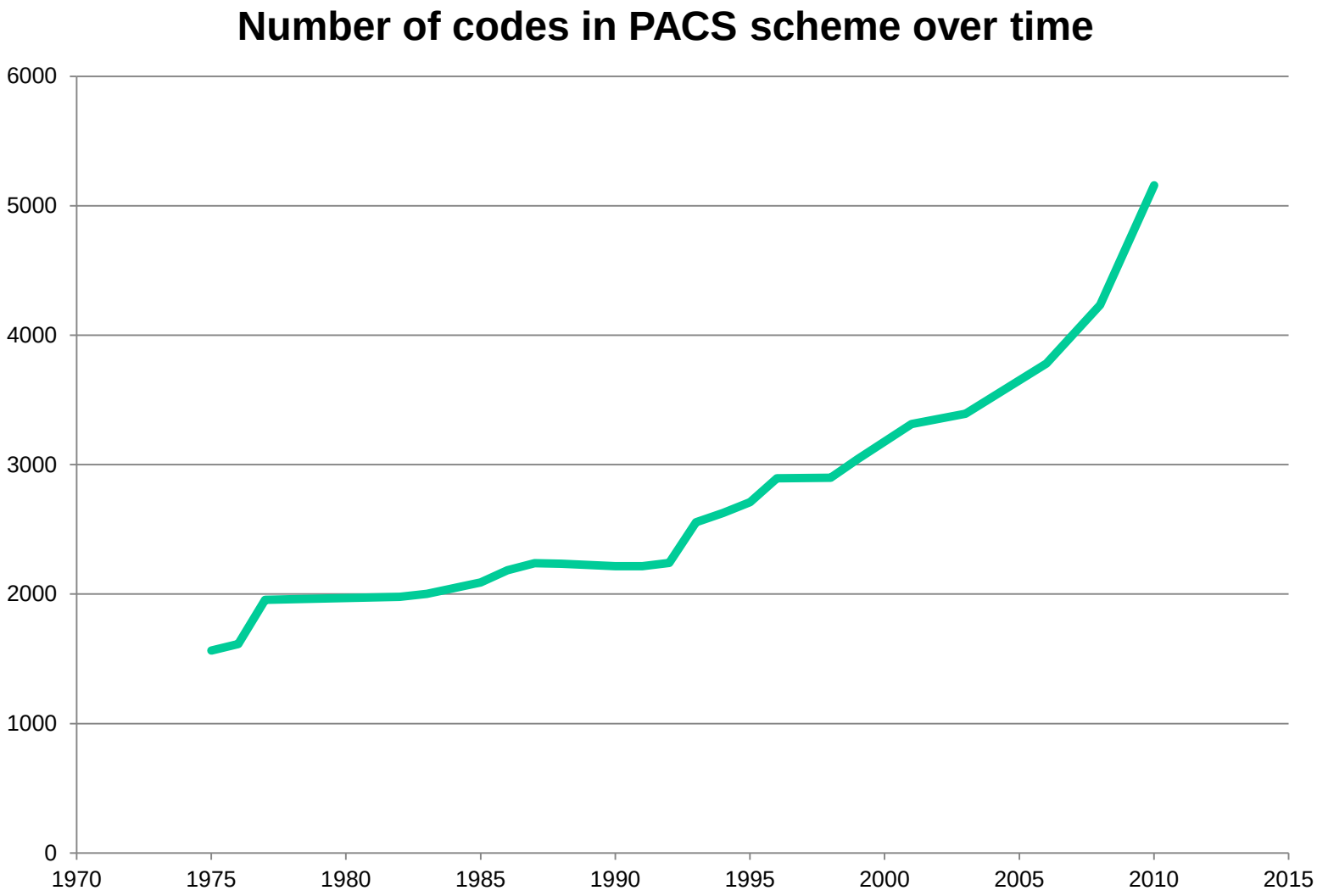
66.70.Df - Metals, alloys, and semiconductors [66.70.-f … thermal cond. …]

68.35.bg - Semiconductors [68.35.-p - Solid surfaces …]

68.55.ag - Semiconductors [… Nucleation & growth of thin films …]

+ about 50 more instances (71.20.Mq, 71.20.Nr, 71.55.Cn, …

* PACS – Physics & Astronomy Classification Scheme


Number of codes in PACS scheme over time
6000
5000
4000
3000
2000
1000
0
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

How APS used PACS

- Journal editorial and publishing process
  - Taxonomy term selection (indexing).
  - Authors assigning topics to their submissions.
  - Defining areas of responsibility and interest for editors.
  - Assigning articles to APS editors.
  - Describing referees areas of expertise.
  - Selecting referees to review articles.
  - Assigning articles to journal sections.
  - Generating statistical and article list reports by various subject criteria.

Use cases for the new taxonomy

- Support the publishing use cases.
- Apply the new taxonomy to journal search and browse interfaces.
- Apply it to other content
  - Search across APS sites by topic.
  - Link journal articles and meeting abstracts.
- Automated suggestion of categories (computer-assisted indexing).

Process

- Internal exploratory work (through 2013)
  - Many interested subject matter experts.
  - Unhappiness with existing alternatives.
  - But no consensus on overall structure or tools.
- November 2013: Taxonomy Strategies engaged as consultants
  - Conducted cross-organizational stakeholder interviews.
  - Identified and prioritized use cases.
  - Analyzed search queries, and existing classification.
  - Analyzed two- and three-word phrases from titles and abstracts.

Interview themes: PACS use by APS editors

- According to one editor, “Searching on PACS codes is very useful.”
- According to another editor, “I use PACS only if I’m desperate. Otherwise I use the expertise of referees.”

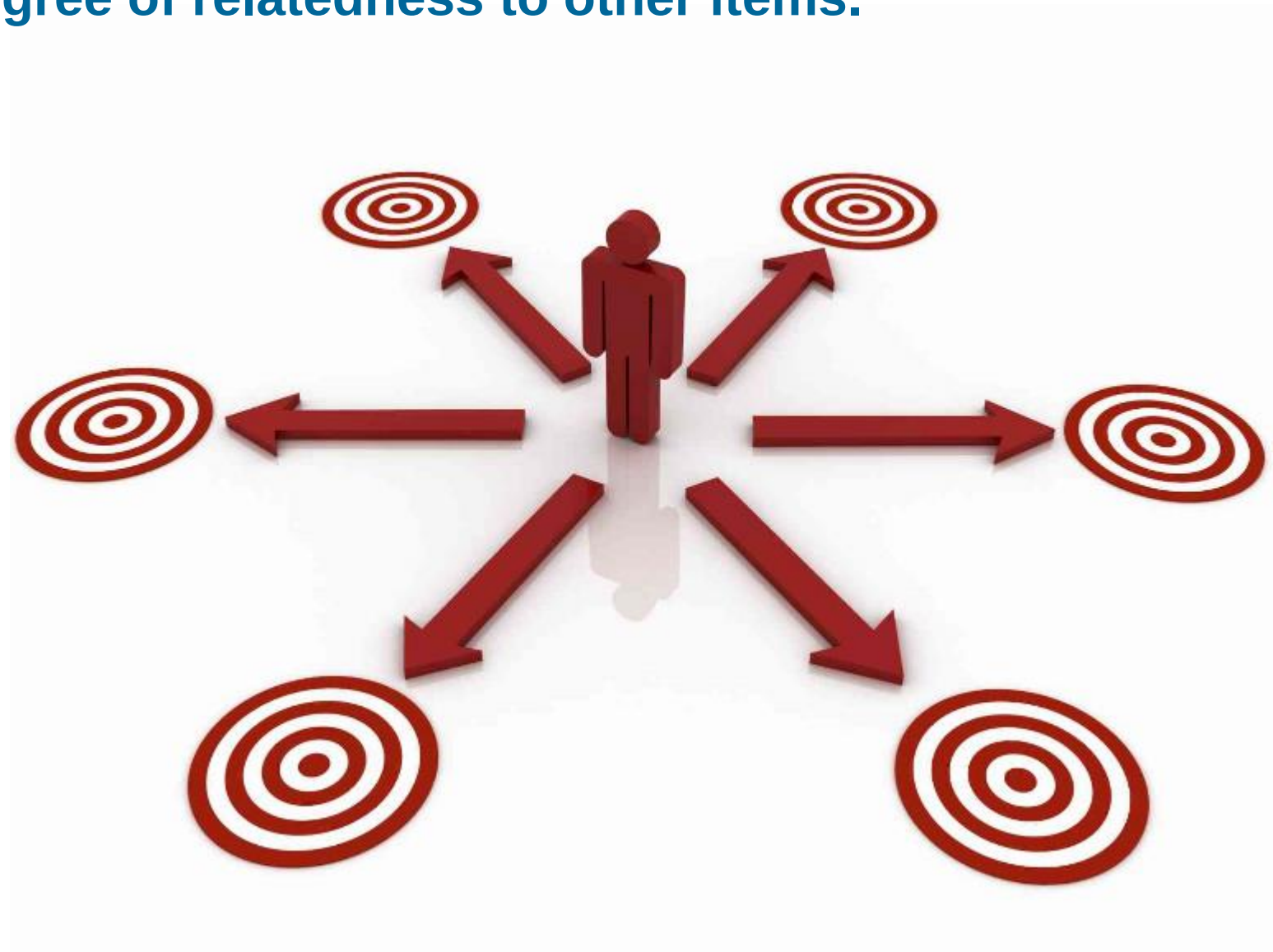
Interview themes: PACS use by physicists

- According to one physicist, “I don’t think I have ever searched on the PACS numbers… Keywords and key phrases are much more universal than a numbering system.”
- One editor observed, “How often do people search on the APS website anyway. They use Web of Science, Google Scholar, etc… Even if you had a good search function there’s no reason to think people would use it.”

Interview themes: Meeting sorting categories

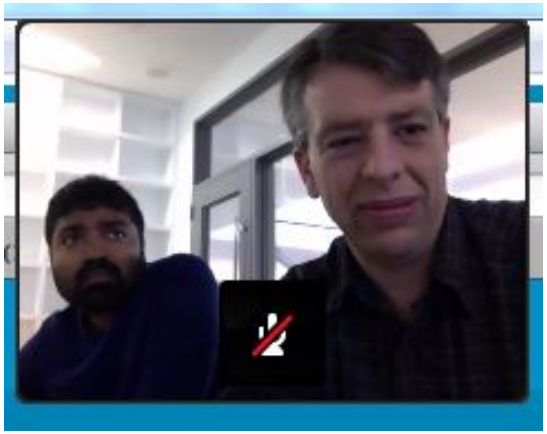
- According to one physicist, “The sorting categories are essentially a taxonomy of physics.” The discipline of Physics has two primary divisions—
  1. The physics of stuff (molecules that are bonded together to make stuff, stuff you can drop on your foot, etc.), and
  2. Theoretical physics, electromagnetics, cosmology.

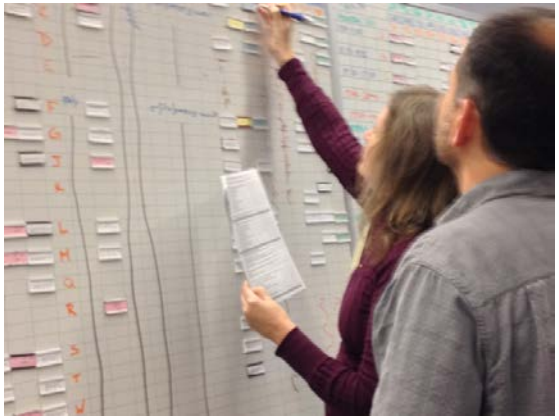
The goal of the taxonomy scheme should not just be to pigeon-hole an item in a category, but to describe the degree of relatedness to other items.



Facets emerged

- Clear need for keywords & key phrases, not numbers/codes.
- Need not just to pigeon-hole an item in a category, but to indicate relatedness to other items.
- Relationships are multi-dimensional:
  - Materials & systems - What you're studying.
  - Phenomena & properties - Why you're studying it.
  - Apparatus, theory & techniques - How you go about studying it.

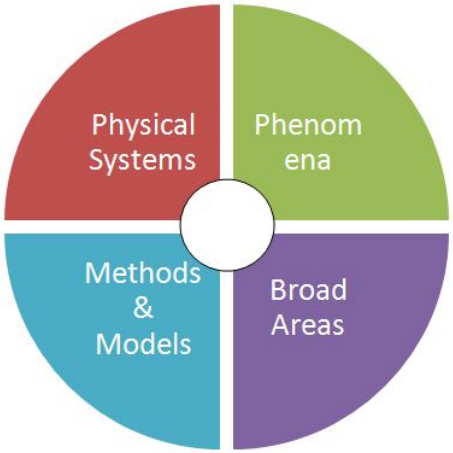




Reviewed tools & standards

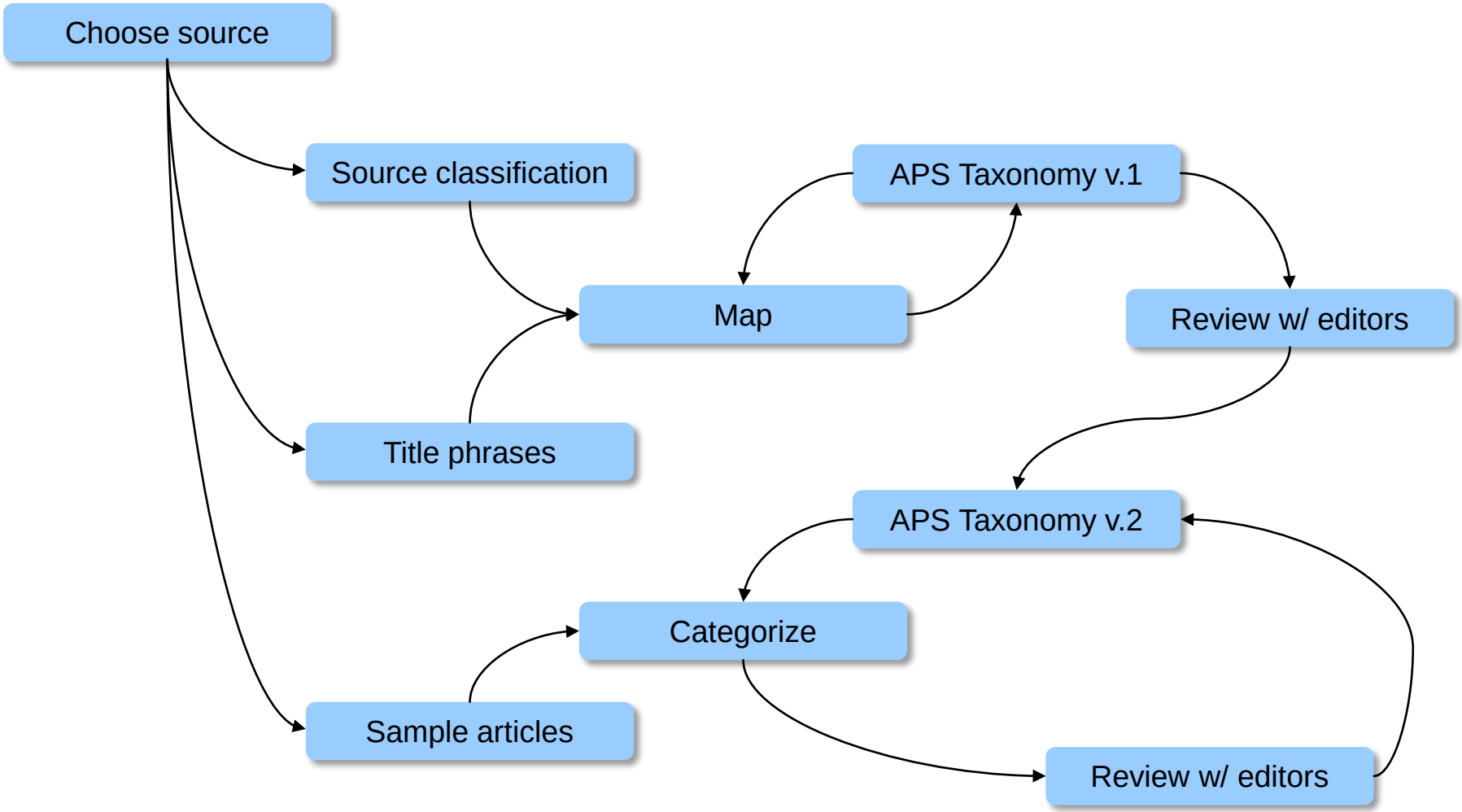
- Basic taxonomy/thesaurus capability:
  - Concepts, relationships and alternate terms.
  - Search/indexing and query tools.
  - Mapping of schemes to each other.
- Support multiple users via a web based service.
- Keep things open, as simple as possible (encourage reuse via Linked Open Data)
- Content indexing, term suggestions.

By early 2014



- High level design
  - Facets rather than single hierarchy
  - Allow poly-hierarchy
- SKOS (Simple Knowledge Organization System) –
  - An open W3C standard
- PoolParty server
  - Web based tool from *Semantic Web Company*
  - Enforces SKOS rules
  - Allows additional schemas
- Develop custom content indexing interface to integrate with existing services.

Taxonomy build-out process

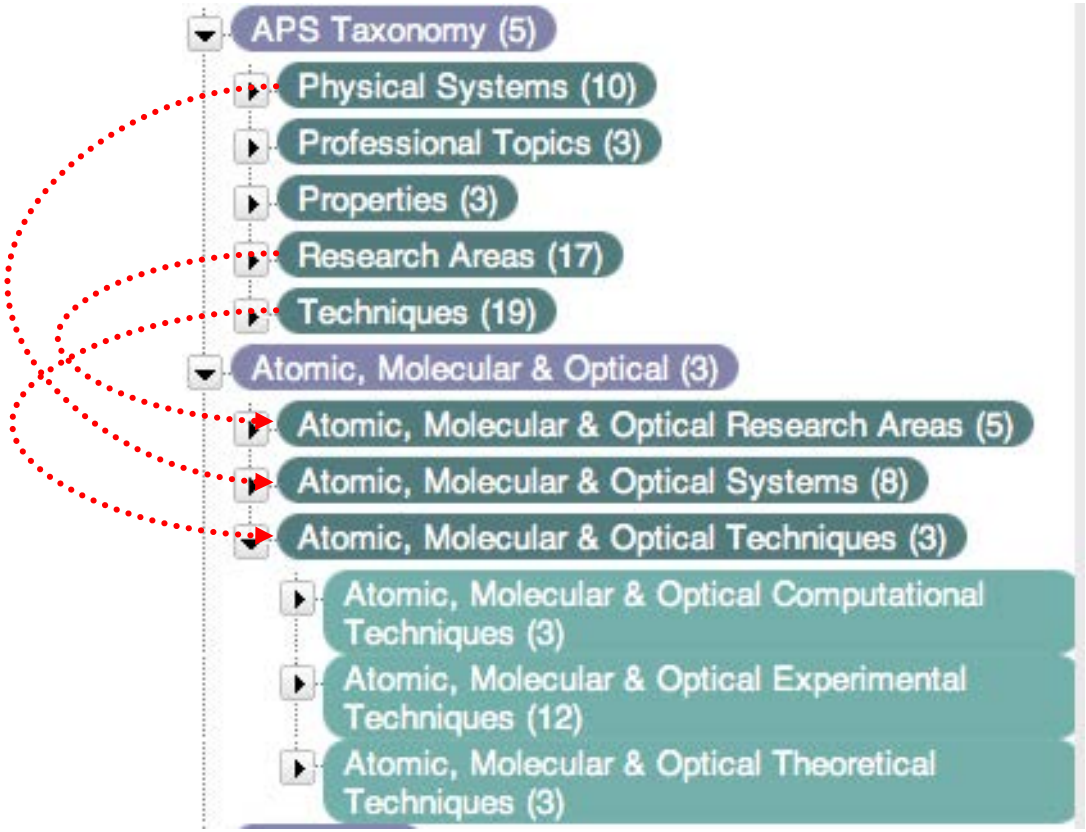


Building out the taxonomy

- Worked with subject matter experts:
  - Entering detailed terms was easy.
  - But there was strong disagreement on overall structure.
- Concern: Desire to hide terms not relevant for a discipline.
- Taxonomy in the tool provided a great visual (much better than a spreadsheet)
- Leveraged Skype with screen sharing for collaboration.

APS Taxonomy Version 2

- Universal “facets” with discipline-specific sub-schemes,
- Subject coverage now – about 80%



Validating the taxonomy

- Initial validation round with spreadsheet
  - Small sub-scheme (Education Research)
  - Results (10 users) took some effort to compile.
  - Several new term suggestions, overall consistent.
- Web application for subsequent rounds
  - Easily handles much larger concept schemes.
  - Automate reporting.
  - But still a work in progress…

Condensed Matter - Sample # 35

Title: Spin-chain description of fractional quantum Hall states in the Jain series

Abstract: We discuss relationship between fractional quantum Hall (FQH) states at filling factor n = p/(2p+1) and quantum spin chains. This series corresponds to the Jain's states n = p/(2mp+1) with m=1 where the composite fermion picture is realized. We show that the FQH states with toroidal boundary conditions beyond the thin-torus (TT) limit, can be mapped to effective quantum spin S=1 chains with p spins in each unit cell. We calculate energy gaps and the correlation functions for both the FQH systems and the corresponding effective spin chains, using exact diagonalization and infinite time-evolving block decimation (iTEBD) algorithm. We confirm that the mass gaps of these effective spin chains are decreased as p is increased which is similar to S=p integer Heisenberg chains. These results shed new light on a link between the hierarchy of FQH states and the Haldane conjecture for quantum spin chains.

Wrapper PDF

Tag this article

Select one or more tags from each dropdown/search list that indicates the primary research focus of the article. If no tag seems to fit you can leave the box empty.

Condensed Matter Research Areas

Fractional quantum Hall effect

Condensed Matter Systems

Select one or two terms from this facet

Condensed Matter Techniques

Heisenberg model

- **Suggestions:** Any additional tags you think would be helpful to classify or find this article? If more than one suggestion, separate terms with a semicolon.
- **Difficulty:**
  It was easy to tag this article.

Summary report for Condensed Matter

10 tagging sessions have been recorded for this scheme, involving 50 content samples.

Sample #	Condensed Matter Research Areas	Condensed Matter Systems	Condensed Matter Techniques
1	3 Domain walls 2 Spin-orbit coupling 1 Magnetic domains 1 Spin torque 1 Magnetic texture 1 Transport properties	3 Magnetic systems 1 One-dimensional systems	2 Landau-Lifschitz-Gilbert equation
2	1 Ferromagnetism	2 Ferroelectrics 1 Heterostructures 1 Thin films	1 Landau-Ginzburg theory
3	1 Integer quantum Hall effect 1 Quantum Hall effect 1 Optical & microwave phenomena	2 Chalcogenides 1 Thin films 1 Two-dimensional electron gas	1 Magneto-optical Kerr effect 1 Theories of electronic structure
4	4 Spin torque 3 Spin waves 3 Spin Hall effect 2 Spin current	1 Thin films 1 Magnetic multilayers 1 Nanowires 1 Magnetic systems	1 Transport measurements 1 Scanning electron microscopy
5	3 Transport properties	2 Nanowires 2 Nanostructures 1 Nanoclusters	2 First-principles calculations 2 Nonequilibrium Green's function 1 DFT methods

Status - November 2014

- Continued input from subject-matter experts in many of the sub-schemes.
- Validation testing, revisions expected.
- Concept mapping under way (from PACS, journals, meeting categories)
- Next steps:
  - IT work: content-concept relationships, user interfaces for indexing etc. (publishing use cases)
  - Versioning and release schedule.

Questions?

- Joseph Busch, jbusch@taxonomystrategies.com, @joebusch
- Vivian Bliss, vbliss@taxonomystrategies.com

Abstract

Since 1975, APS *Physical Review* article submission, editorial assignments and journal tables of contents have been based on the Physics and Astronomy Classification Scheme (PACS) developed by the American Institute of Physics (AIP), a 5 to 6-level hierarchical classification system. In 2013 APS decided to develop a new taxonomy. AIP is no longer maintaining PACS, and APS wanted a single scheme for all types of content (journal articles, conference papers and website content), and to develop new capabilities for topic-based online services such as facetted navigation, targeted alerts, personalized subscriptions, etc. The new taxonomy includes several discrete sets of categories (facets) allowing content to be categorized along distinct dimensions rather than trying to use a single concept as with PACS. APS also wanted to base the new taxonomy on Linked Data standards, in particular the SKOS knowledge organization standard, and is using the PoolParty thesaurus management tool to build, maintain and publish the new facetted taxonomy. This presentation will discuss the process that has been developed to build, maintain and ultimately publish the new APS Taxonomy so that is, as far as possible, backward compatible with the legacy content categorized using PACS, and is extensible and scalable to support new information services.