SRI Work on Science Taxonomies

Jeffrey Alexander, Ph.D.
Senior Science & Technology Policy Analyst
Center for Science, Technology & Economic Development
August 2011
Who We Are

SRI is a world-leading independent R&D organization

- Founded by Stanford University in 1946
  - A nonprofit corporation
  - Independent in 1970; changed name from Stanford Research Institute to SRI International in 1977
- Sarnoff Corporation acquired in 1987
  (formerly RCA Laboratories) will become fully integrated into SRI effective January 1, 2011
- 2,000 staff members combined
  - 800 with advanced degrees
  - More than 20 locations worldwide
- Consolidated 2010 revenues: $500+ million
Center for Science, Technology & Economic Development
Sample Activities

- Science, technology, and innovation measurement and indicators (NSF, Saudi Arabia)
- State/regional economic and innovation strategies (Florida, Virginia)
- Research institute planning and evaluation (Japan, Saudi Arabia)
- University strategic planning and economic development (KAUST, Imam U, Princess Noura U)
- Strategic technologies planning (Saudi Arabia, Korea)
- Evaluation of science, technology, and innovation programs: National Science Foundation, National Institutes of Standards and Technology, Ohio
- Training workshops in R&D management, program evaluation, and international collaboration (Korea, Saudi Arabia)
- Design and implementation of technology development organizations – national labs, incubators, innovation centers (Saudi Arabia)
Taxonomy Support for NCSES

• SRI is a long-time support contractor for the National Center for Science & Engineering Statistics at the NSF (formerly the Division of Science Resource Statistics)

• Engaged in 2009 for yet another project on classifications of fields of science
  • Five (5) taxonomy reports since 2000
  • One (1) SRS taxonomy team (2001/2002)
  • Two workshops (2006, 2008)

• Dealing with perennial challenges of interdisciplinary research and taxonomy harmonization across NCSES surveys and publications

• Working with Patrick Lambe of Straits Knowledge, Singapore & led by Jeri Mulrow, NCSES
Some Observations So Far

• Differences in taxonomies driven by data collection needs versus data analysis needs

• What is it that we really want to classify?
  – Fields of research (what is going on in the lab)
  – Fields of study (NOT the same as disciplines)
    • Especially significant when studying Ph.D.’s and post-docs
  – Fields of education (roughly analogous to disciplines?)
  – Departments (organizational units)
  – Occupations
  – Fields of application (technologies?)
    • Often tied to markets/industries
One Illustration of the Problem
Courtesy of Diana Hicks, Georgia Tech
One Approach: Faceted Taxonomies

- Framework to classify in multiple dimensions
  - Each dimension may be a separate tree or list
  - Dimensions are orthogonal – mutually exclusive
Example: ANZSRC

• Type of Activity
  – Pure basic research
  – Strategic basic research
  – Applied research
  – Experimental Research

• Field of Research – categorizes the methodology
  – 3 hierarchical levels
  – 22 Divisions, 157 Groups, 1238 Fields

• Socio-Economic Objective – purpose or outcome
  – 5 Sectors; 17 Divisions, 119 Groups, 187 Objectives

• Minor revisions every 5 years; major every 10
Current NCSES Activities

**TAXONOMY MANAGEMENT POLICY**
- Provides a formal structure to think about taxonomy improvement and principles everyone can follow
- Defines roles and responsibilities, clarifies expectations
- Guides decisionmaking on taxonomy improvement towards greater harmonisation and currency
- Provides for feedback mechanism

**TAXONOMY MANAGEMENT SYSTEM**
- Provides a common environment to store, share, consult, link and manage classifications and vocabularies
- Creates greater opportunity for visibility into what we have
- Makes harmonisation less manual and arduous
- Makes it easier to show the ramifications of change
- Enables making reasoning, principles, meanings of terms, reasons for term exclusion explicit
- Retains local control where needed

**SCOPING R&D TAXONOMY**
- Helps us figure out what’s involved in developing a taxonomy to be used as inter agency standard and whether it’s worth the effort
- Helps us clarify our objectives and understand our options
- Gives us a more confident and complete feel for what’s already out there
- Helps us define the taxonomy strategy eg hierarchy, facets, thesaurus
Operationalizing Taxonomy Management

External Authority acts as external reference & data exchange standard

SRS Standard acts as internal reference & data exchange standard

CIP supplies core of SRS Taxonomy of Disciplines (ToD) to must resolve to

FLAG TERMS AND DOCUMENT: variations in meaning from CIP additional terms variations in structure

Taxonomy Mgmt System tracks variations, different usage of terms (e.g. FoSt, FoR)

BRDIS FoR/S&E
Facilities FoR/S&E
FF FoR/S&E
GSS FoSt
HERD FoR/S&E
NSCG FoSt NSCG FTER
NSRCG FoSt NSRCG FTER
PubAtt FoST
SDR FoSt SDR FTER
SED FoSt SED FoR/S&E

Provides starting vocabulary and structure for ToD

FLAG TERMS AND DOCUMENT: variations in meaning from ToD additional terms variations in structure

SESTAT FoSt SESTAT FTER
WebCASPAR FoSt WebCASPAR FoR/S&E
SEI FoSt SEI FTER SEI FoR/S&E
W&B M FoSt

Determine relationship with ToD after survey linkages established
Other Relevant Work

• Exploratory project with STAR Metrics to develop taxonomies to work with topic modeling efforts in CHEM Directorate

• Discussions of user needs for a multi-agency system for the Classification of Research & Development Activities (CORDA)

• IARPA FUSE Project
Prospective Areas of Inquiry

• Do we really know what IS a science?

• Technical approaches can eliminate SOME issues
  – Triangulation to identify and correct errors
  – Better enforcement of consistent reporting policies
  – Unified format standards and data architectures
  – Increased use of machine analysis

• Abandon the goal of a unified taxonomy
  – Create an environment which accommodates multiple competing taxonomies
  – Leverage new technologies in text analysis, concept inference, Semantic Web
  – Make taxonomies self-organizing and self-correcting
  – Requires computing power, intensive design effort, and RESOURCES
Thank You

*Headquarters: Silicon Valley*

**SRI International**
333 Ravenswood Avenue  
Menlo Park, CA 94025-3493  
650.859.2000

*Washington, D.C.*

**SRI International**
1100 Wilson Blvd., Suite 2800  
Arlington, VA 22209-3915  
703.524.2053

*Princeton, New Jersey*

**SRI International Sarnoff**
201 Washington Road  
Princeton, NJ 08540  
609.734.2553

*Additional U.S. and international locations*

[www.sri.com](http://www.sri.com)