Consumer Health Care Taxonomy background

- Designed to support types of queries a consumer health care information service such as a website might get from a wide variety of consumers in a wide variety of care conditions.

- Project sponsor:
  - U.S. Centers for Medicare and Medicaid Services (CMS)

- Users:
  - Medicare/Medicaid beneficiary*
  - Caregiver

* Medicare is the U.S. government single payer health insurance for seniors over 65 years old. Medicaid is the U.S. jointly funded federal and State health insurance program for low-income people.
Related research: Information seeking

- Critique of social science qualitative methods (Davenport). Scientificity – consumer decision making is very different from studies of students, engineers and scientists.
- Most consumers search for health information on the Internet, usually starting with an organic search engine. The most commonly researched topics are diseases or conditions, treatments or procedures, and doctors or other health professionals. Half of online health information research is on behalf of someone else. (Pew)
Related research: Quality of care

- Evidence-based decision-making by clinicians vs. factors that patients identify as most important such as cost, qualifications and accessibility of care. (Hibbard & Sofaer)
- Patient narratives are of more interest to consumers, and easier for them to understand. (Schlesinger)
Related research: KOS development

- Most health care KOS were originally designed to support researchers, clinicians and health insurers. But they can be useful sources to build consumer-oriented health care KOS, rather than starting from scratch. (Hyvönen)
- Consumer terminology used in health care related activities can be useful to improving existing health care KOS. (Doing-Harris)
Sources: CMS health care website prototype

CareFinder prototype envisions leveraging Medicare’s massive datasets to support consumers in making better health care decisions.
Compare websites and datasets provide directory information about CMS-registered service providers and suppliers, and reported quality measures.
Sources: Inventory of data.medicare.gov datasets

Each Medicare dataset has a different structure and number of tables.
Existing quality measures were categorized by type, usefulness, availability and source.
Sources: Existing CMS and HHS taxonomies

Substance Abuse and Mental Health Services Administration (SAMHSA)

The methods and learnings from earlier projects helped inform the Consumer Health Care Taxonomy approach.
Sources: Authoritative sources, websites and query logs

- More than 100 vocabulary sources for Consumer Health Care concepts including:
  - National Library of Medicine’s Medical Subject Headings (MeSH).
  - International Classification of Diseases (ICD) used by CMS and other health insurance providers to classify diseases and conditions.
  - Unified Medical Language System (UMLS), a mapping of more than 100 vocabularies and classification systems including MeSH and ICD.
- Query logs from Physician Compare and MedLine Plus to help identify unique facets, relevant terms, and synonyms for the Taxonomy.
- Semantic relationships between Consumer Health Care Taxonomy concepts were made based on trusted sources such as:
  - Online symptom checkers from the Mayo Clinic, Cleveland Clinic, NHS UK, and HealthDirect Australia.
  - Physician Compare mappings of conditions and symptoms to medical specialties.
  - Google medical search.
Sources: Interviews with SMEs and key stakeholders

- More than 30 CMS staff working on the various Compare websites.
- CMS Compare website data contractors.
- IDEO team that built the CareFinder prototype.
- CMS user research staff.
- External health care professionals.
- Friends and family who have health care stories.
What do consumers want or need to make choices about where to get care?
Sharon is a 52 year old with ESRD who received a kidney transplant 4 years ago. The kidney has recently begun to fail and she needs hemodialysis again 3 times a week. She also recently broke her leg and needs physical therapy 2 times a week. She works as a waitress but is currently unemployed due to her accident, and does not have insurance besides Medicaid. Sharon needs to find a dialysis center close to home since a family member will need to drive her due to her leg injury. She also needs to coordinate her dialysis (3x week) with her physical therapy (2x week).

Paula is an 85 year-old woman. She was out driving in her neighborhood when all of a sudden she lost her way and she couldn’t find her way back home. This episode scared her and her family. She began to worry about her mental capacity, and wondered what kind of specialist she could see who could assess her mental acuity. Paula needs to figure out what type of specialist can help assess her mental acuity, and find a trustworthy specialist who is close to her apartment and covered by her health plan.
Consumer Health Care Taxonomy: Purpose

Needs to function as middleware that translates consumer queries into the language necessary for retrieval of data from Medicare.gov datasets and Good to Know (GTK) content.
Consumer Health Care Taxonomy: Functional requirements

- Provide enough information for any user, tool, or program to find and use content in any Medicare.gov dataset or GTK content.
- Define what vocabularies are needed to support consumer health care decision making.
- Identify authoritative vocabulary sources for each taxonomy facet.
- Provide vocabularies for each taxonomy facet that are sufficiently defined to be used to build a functional application (i.e., a CareFinder-like application).
- Be readily extensible to support new application requirements.
- Be flexible enough to accommodate additions of missing categories and changes to existing categories as needed.
- Define relationships between the vocabularies useful for searching Medicare.gov datasets and GTK content.
Consumer Health Care Taxonomy: Concept scheme

Eleven facets in the Consumer Health Care Taxonomy displayed in the PoolParty Linked Data frontend.
Consumer Health Care Taxonomy: Entry terms (skos:prefLabel)

- Entry terms have been identified by analyzing search logs for similar sites, related and curated content, popular news sources, and user research including uses cases.
- We capture both the technical version of a term and the consumer-friendly or colloquial version(s) of a term.
- Sometimes the best entry term is the technical version, and sometimes it is the consumer-friendly version.

Results of Spinal tap v. Lumbar puncture in Google Fight.
Practical consideration: Post-coordination vs. pre-coordination

- As middleware assisting consumers by reflecting their language, many multiple word concepts need to be kept together (that is, pre-coordinated) in this taxonomy.
- For example, “Hip fracture” is included in the Taxonomy as a pre-coordinated phrase in the Conditions facet.
Synonyms and Quasi-synonyms (skos:altLabel)

Variants of “End-stage renal disease”.

Quasi-synonyms of “Canes”.

Taxonomy Strategies  The business of organized information
Hierarchical relationships (skos:broader, skos:narrower)

Part of “Hips, legs and teet”.

Type of “Fracture”.

Taxonomy Strategies  The business of organized information
Relationships to entry terms in other facets are a custom schema

<table>
<thead>
<tr>
<th>Semantic Relation</th>
<th>Inverse Semantic Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Facet Class</strong></td>
<td><strong>Facet Class</strong></td>
</tr>
<tr>
<td>Body Locations and Systems</td>
<td>is affected by</td>
</tr>
<tr>
<td>Kidneys</td>
<td>is affected by</td>
</tr>
<tr>
<td>Conditions</td>
<td>has treatment of</td>
</tr>
<tr>
<td>End-stage renal disease</td>
<td>has treatment of</td>
</tr>
<tr>
<td>Conditions</td>
<td>is concern of</td>
</tr>
<tr>
<td>End-stage renal disease</td>
<td>is concern of</td>
</tr>
<tr>
<td>Conditions</td>
<td>needs medical supply</td>
</tr>
<tr>
<td>End-stage renal disease</td>
<td>needs medical supply</td>
</tr>
<tr>
<td>Care Setting</td>
<td>is location for treatment</td>
</tr>
<tr>
<td>Dialysis Facilities</td>
<td>Is location for treatment</td>
</tr>
</tbody>
</table>
Relationships to entry terms in other facets are a custom schema (2)

<table>
<thead>
<tr>
<th>Semantic Relation</th>
<th>Facet Class</th>
<th>Facet Class</th>
<th>Inverse Semantic Relation</th>
<th>Facet Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Care Settings</td>
<td>Specialty Areas</td>
<td>is specialty of</td>
<td>Care Settings</td>
</tr>
<tr>
<td></td>
<td>Dialysis Facilities</td>
<td>Dialysis Services</td>
<td>is specialty of</td>
<td>Dialysis Facilities</td>
</tr>
<tr>
<td></td>
<td>Medical Supplies &amp;</td>
<td>Tests &amp; Treatments</td>
<td>uses medical supply</td>
<td>Medical Supplies &amp;</td>
</tr>
<tr>
<td>Equipment</td>
<td></td>
<td></td>
<td></td>
<td>Equipment</td>
</tr>
<tr>
<td></td>
<td>Dialysis Equipment</td>
<td>Dialysis</td>
<td>uses medical supply</td>
<td>Dialysis Equipment</td>
</tr>
<tr>
<td>&amp; Supplies</td>
<td></td>
<td></td>
<td></td>
<td>&amp; Supplies</td>
</tr>
<tr>
<td></td>
<td>Specialty Areas</td>
<td>Tests &amp; Treatments</td>
<td>is part of practice area</td>
<td>Specialty Areas</td>
</tr>
<tr>
<td></td>
<td>Nephrology</td>
<td>Dialysis</td>
<td>is part of practice area</td>
<td>Nephrology</td>
</tr>
<tr>
<td></td>
<td>Care Settings</td>
<td>Conditions</td>
<td>is focused on in setting</td>
<td>Care Settings</td>
</tr>
<tr>
<td></td>
<td>Dialysis Facilities</td>
<td>End-stage renal disease</td>
<td>is focused on in setting</td>
<td>Dialysis Facilities</td>
</tr>
<tr>
<td></td>
<td>Body Locations &amp;</td>
<td>Tests &amp; Treatments</td>
<td>treats body location</td>
<td>Body Locations &amp;</td>
</tr>
<tr>
<td>Systems</td>
<td></td>
<td></td>
<td></td>
<td>Systems</td>
</tr>
<tr>
<td></td>
<td>Kidneys</td>
<td>Dialysis</td>
<td>treats body location</td>
<td>Kidneys</td>
</tr>
</tbody>
</table>
Semantic relationships diagram

All relationships

ESRD relationships
Mapping to Medicare.gov dataset values (skos:relatedMatch)

Example of mapping a Specialty Area to Medicare.gov data set. (Initial mappings are narrow in scope).
Project observations

- Consumer healthcare related decision-making behavior is different from clinicians.
  - Focus on the problem to be solved: Translate consumer queries into the language necessary for retrieval of data from Medicare.gov datasets and Good to Know web content.
  - Exhaustivity is not a requirement.

- While there are many healthcare-related technical KOS available, consumer-friendly terminology is generally not available from authoritative sources.
  - A lot of work is required to compile a useful KOS from many sources.
  - Documentation of editorial guidelines supports this activity and helps to make it scalable.
A small set of extensible taxonomies and custom semantic relationships are sufficient to develop the domain model.

- A concise set of subject predicate object relationships, e.g., **Condition is_concern_of Specialty Area**.
- SKOS is not intended for encoding more complex ontologies beyond thesaurus relationships (hierarchy, equivalent and generic associative), so a custom schema was developed for specific associative relationships.

A strategy to setup separate concept schemes for the Consumer Health Care Taxonomy and the Medicare.gov datasets controlled vocabularies provided flexibility and extensibility.

- SKOS relatedMatch was used to map across the concept schemes.

KOS management tools are immature in their capacity to accurately and efficiently batch import and export KOS, interim taxonomies and semantic relationships.


Resources: Center for Medicare and Medicaid Services (CMS) websites

Resources: Symptom checkers

Questions?

Joseph Busch (Washington, DC, USA)
jbusch@taxonomystrategies.com
skype: jbusch94110
mobile: 1-415-377-7912