

Title: USDA ARS FoodData Central Ontology Integration

Authors: Kyle A. McKillop¹, Damion Dooley², Tarini Naravane³, James M. Harnly¹, John Finley¹, Naomi K. Fukagawa¹

¹ USDA Agricultural Research Service Beltsville Human Nutrition Research Center, Beltsville , MD, USA

² Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.

³ Biological and Agricultural Engineering, University of California, Davis, CA, USA

Keywords: food composition, nutrition, diet, ingredient, database, web portal, ontology, FAIR data sharing

Abstract: The USDA Agricultural Research Service (ARS) has recently launched FoodData Central (FDC) (<https://fdc.nal.usda.gov/>), an online food nutrition data system consisting of over 300,000 foods sourced from four USDA ARS datasets, including SR Legacy (>7,000 foods), the final release of the USDA Standard Reference database reporting upwards of 150 component nutrients collected over time, and the more recent Foundation Foods that includes 113 foods representing commodity/commodity-based foods or staples with analytical data of individual samples (i.e. not composites). Foundation Foods sets a new standard for sharing ARS research data by including information on individual sample location, time, and other agricultural or market factors. The ontology integration project focuses on specifying the appropriate vocabulary capable of describing all foods. ARS has identified the OBOFoundry (<http://obofoundry.org/>) family of ontologies as an attractive platform for sourcing and managing a wide variety of food sample vocabulary needs, since it addresses the need for data interoperability with respect to industry and academic research projects and databases.

OBOFoundry ontologies are largely interoperable, meaning they share the same method of naming, defining and exposing terms for reuse in a global community of data scientists according to FAIR data sharing principles, while offering domain specific vocabularies across agricultural (AGRO), anatomical (PO, UBERON), taxonomic (NCBITaxon), chemical (CHEBI) and environmental (ENVO) domains for describing soil/plant stress conditions, crop treatment factors, food quality, and chemical nutrients. These ontologies

are in an evolutionary phase to catch up with terminology that researchers are already using, so ARS has engaged with the FoodOn curation team to help ensure that needed terms are covered within FoodOn, as well as other OBOFoundry ontologies. The process of identifying factors and defining descriptive terms within the agency, and with the external ontology community, has been a learning and communication experience that is necessary for achieving the ultimate goal of data interoperability.

The FDC project and ARS research efforts in general have pivoted from a past effort to test both single and multicomponent foods, to a new focus on more comprehensive coverage of commodity/commodity-based and staple foods via the Foundation Foods dataset. The strategy is to have ingredients well covered, thus providing the necessary data for commercial and research agencies to then model or test multi-ingredient food nutritional profiles. Currently Foundation Foods has representative beans, eggs, milk, and a selection of flours, oils, fruits, meats and vegetables. This fall we will be introducing the following Foundation foods:

- Flour, Pastry, White (Unenriched) (Unbleached)
- Flour, Rice, Brown
- Flour, Rice, Glutenous
- Flour, Soy (Defatted)
- Flour, Soy, Full Fat
- Onion, White
- Garlic
- Apples, Red Delicious
- Apples, Fuji
- Apples, Gala
- Apples, Granny Smith
- Peanut Oil
- Sunflower Oil
- Safflower Oil
- Extra Light Olive Oil

All of these food sample classes are mapped to FoodOn categories, since FoodOn specializes in naming plant and animal food sources, and single and multicomponent foods that are derived from them.

Currently under development is the next generation FDC website, which will be driven by a Neo4j property graph database model. This will be echoed as an in-house ARS database which will be augmented with OBOFoundry ontology vocabulary to describe food sample factors such as agricultural practice, and local environmental context, key variables that we have found affect the nutrient quality of commodities. ENVO and AGRO are being extended to cover these terms. The in-house graph can then be released to the live FDC website on a periodic basis. Other factors of interest include degree/method of food processing, fortifications and additives, and degree of plant maturity.