TERMINOLOGY AND RELATIONSHIPS FOR A SMOKELESS TOBACCO PRODUCT ONTOLOGY

Office of Science | Center for Tobacco Products | Authors: Vibha Kumar, Catherine Kemp, Christine Keys, Deborah Sholtes, John-William Declaris, Thomas Eads -- WiseDesign | Authors: Esther Wisdom, Christopher Chaney, Eric Fenzl, John F. "Jack" Jones

What is an Ontology?

An ontology is a portrayal of knowledge as a network of connected facts, built on the backbone of a taxonomy and precisely defined terms.

INTRODUCTION

PRODUCT REGULATION

The FDA regulates all tobacco products, including (as shown right): hookah, e-cigarettes, dissolvables, smokeless tobacco, cigarettes, all cigars, roll-your-own tobacco, pipe tobacco, and future tobacco products that meet the statutory definition of a tobacco product.

The highlighted products represent the focus of the work we're sharing with you today.
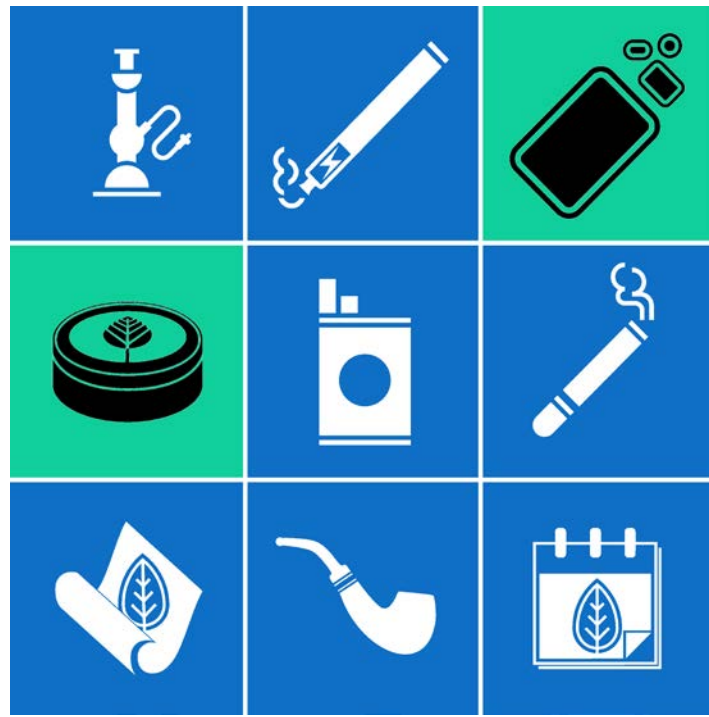
SMOKELESS TOBACCO PRODUCTS

Most smokeless tobacco use involves placing the product between the gum and the cheek or lip. Chewing tobacco, snuff, and snus are the main types of smokeless tobacco marketed in the United States:

- Chewing tobacco is cured and comes in loose leaf, plug, or twist.
- Dry snuff is loose powdered tobacco sniffed through the nostrils.
- Moist snuff and snus are cut tobacco that can be loose or pouched.

More than 30 chemicals in smokeless tobacco are linked to cancer—

Harmful chemicals are either present in the ingredients or are created during consumption of the product, or both. It is therefore important to get a good grasp of ingredients used in these products.

INFORMATION COMPLEXITY

Ingredients help determine health effects of tobacco product consumption. Information about ingredients is widely dispersed, voluminous, and variable. The ontology organizes and portrays this information in a unique way.

GOALS AND OBJECTIVES

The goal of this work is to build a smokeless tobacco product ontology to aid regulatory decision making.

The objectives are to standardize tobacco product ingredient information, organize it as a network of relationships, ensure that the relationships make sense, and preserve industry data.

The main challenge was resolving thorny issues in standardizing tobacco product ingredient information.
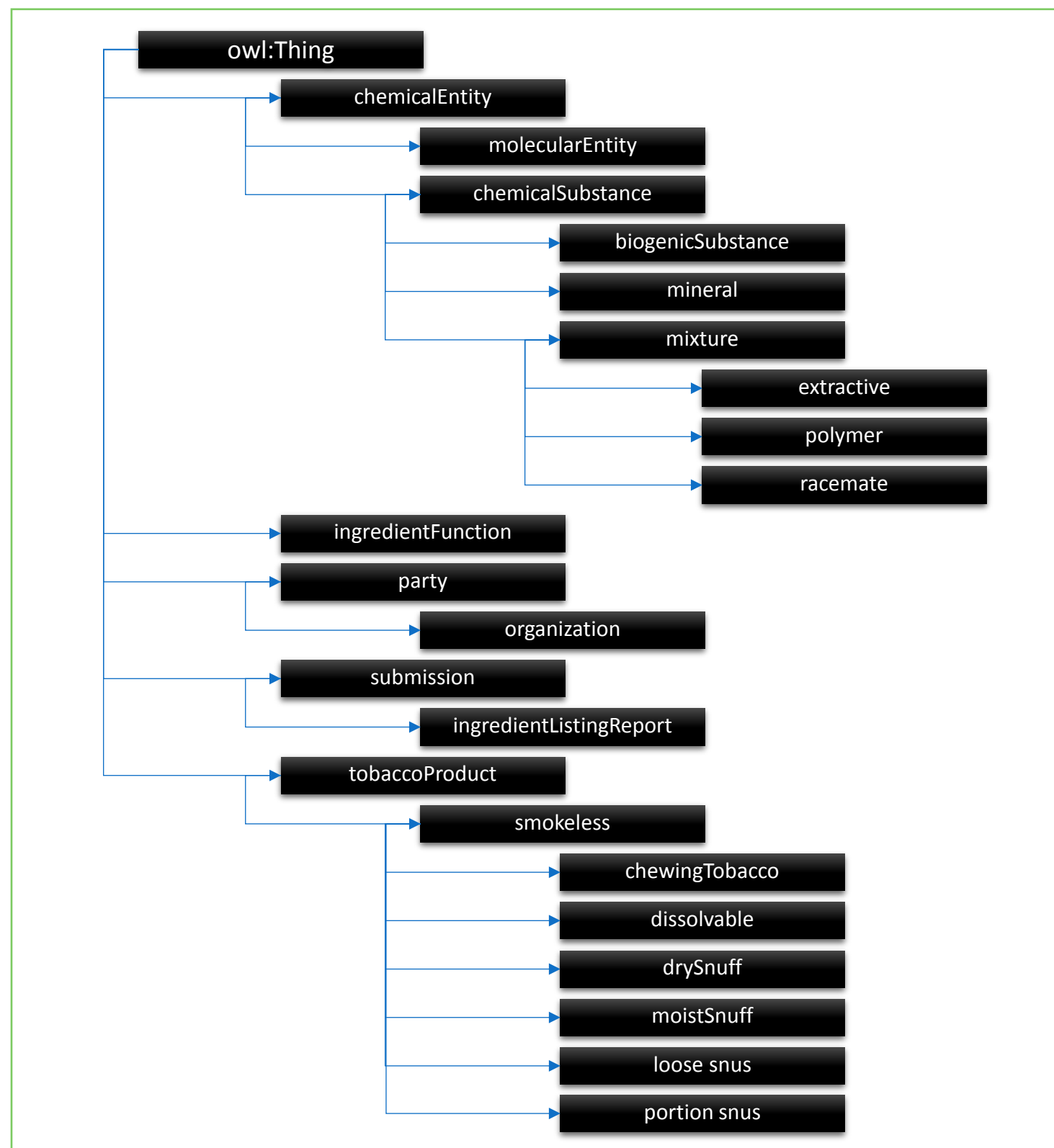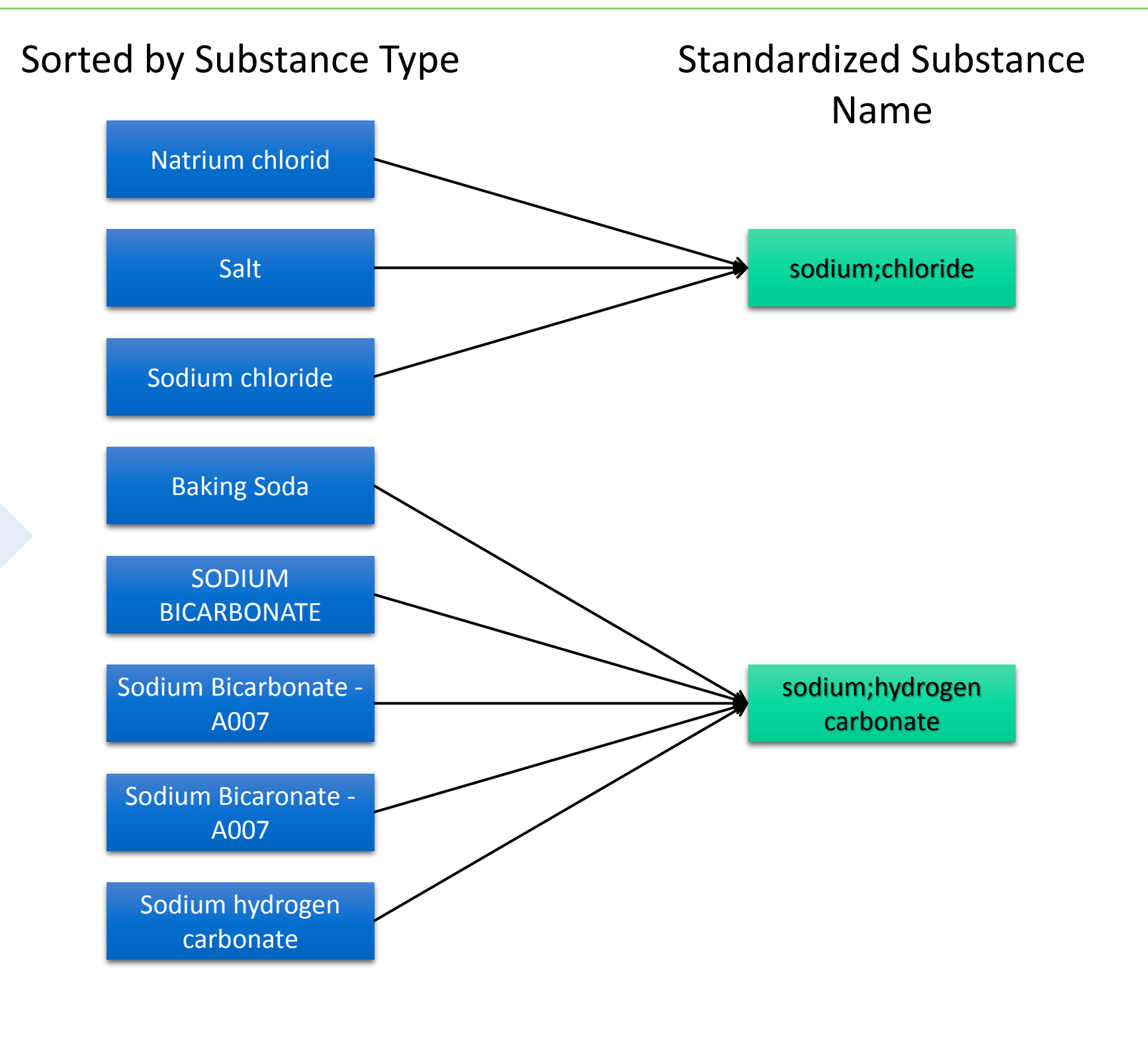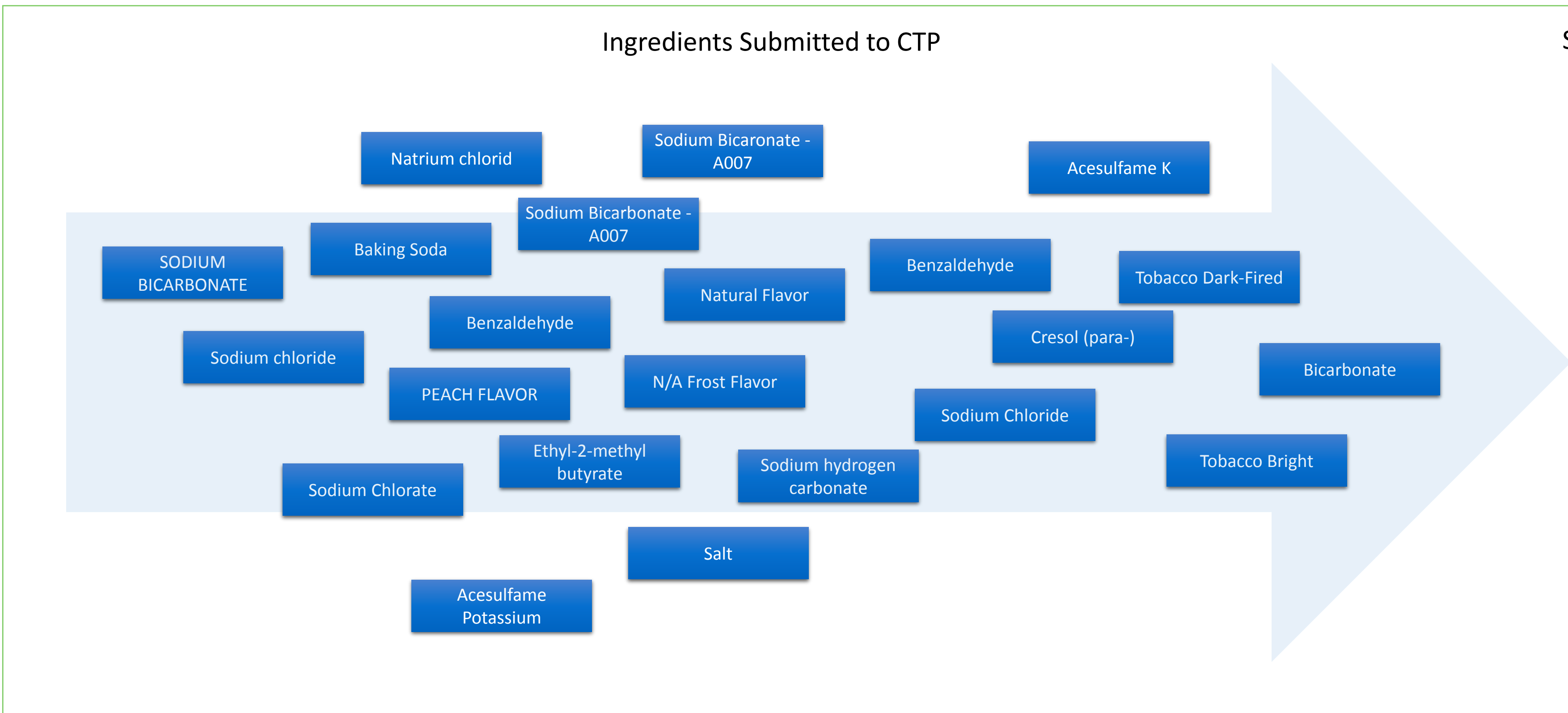
METHODS

Curate Ingredient Data
- Aggregate data
- Correct misspelling, punctuation, spacing
- Develop ingredient naming rules
- Remove duplicates
- Preserve unique attributes
- Apply substance classification rules

Construct Prototype Ontology
- Describe products, ingredients and their functions
- Build a data model
- Encode in web ontology language (OWL2)
- Review and vet all content with the CTP-OS Division of Product Science

RESULTS

- The ontology displays relationships in multidimensional graphs.
- Curation deduplicated manufacturers, tobacco product names, and ingredient names.
- Naming rules reduced 1,200+ unique ingredient names down to 850 (a 30% reduction).
- An ingredient standardization guide and a lookup table were produced.
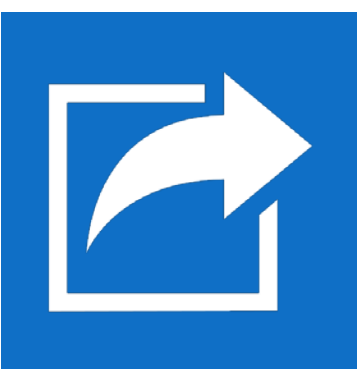- Established Ontology Governance Charter under CTP Data Standard Workgroup.

DISCUSSION

The Smokeless Tobacco Product Ontology helps both humans and machines. The Ontology demonstrates knowledge organization and immediate visual insights into concept relationships, which in this case led to the discovery that a particular ingredient is only used by one manufacturer for one brand of smokeless tobacco. Machine readability enables software to use the information in the ontology without the need for recoding. Inference engines can be employed to check logical consistency of relationships.

We will extend this Ontology to product consumption, potential exposure to harmful and potentially harmful constituents, biomarkers of exposure, and to individual health.

These efforts have already accelerated the development of the Cigarette Product Ingredient Ontology.

Ingredients Submitted to CTP → Sorted by Substance Type → Standardized Substance Name

An example of the power of ingredient standardization

The Smokeless Tobacco Product Ontology taxonomy shown as an indented list

---

This ontology portrays knowledge of smokeless tobacco products without the need for complex queries.
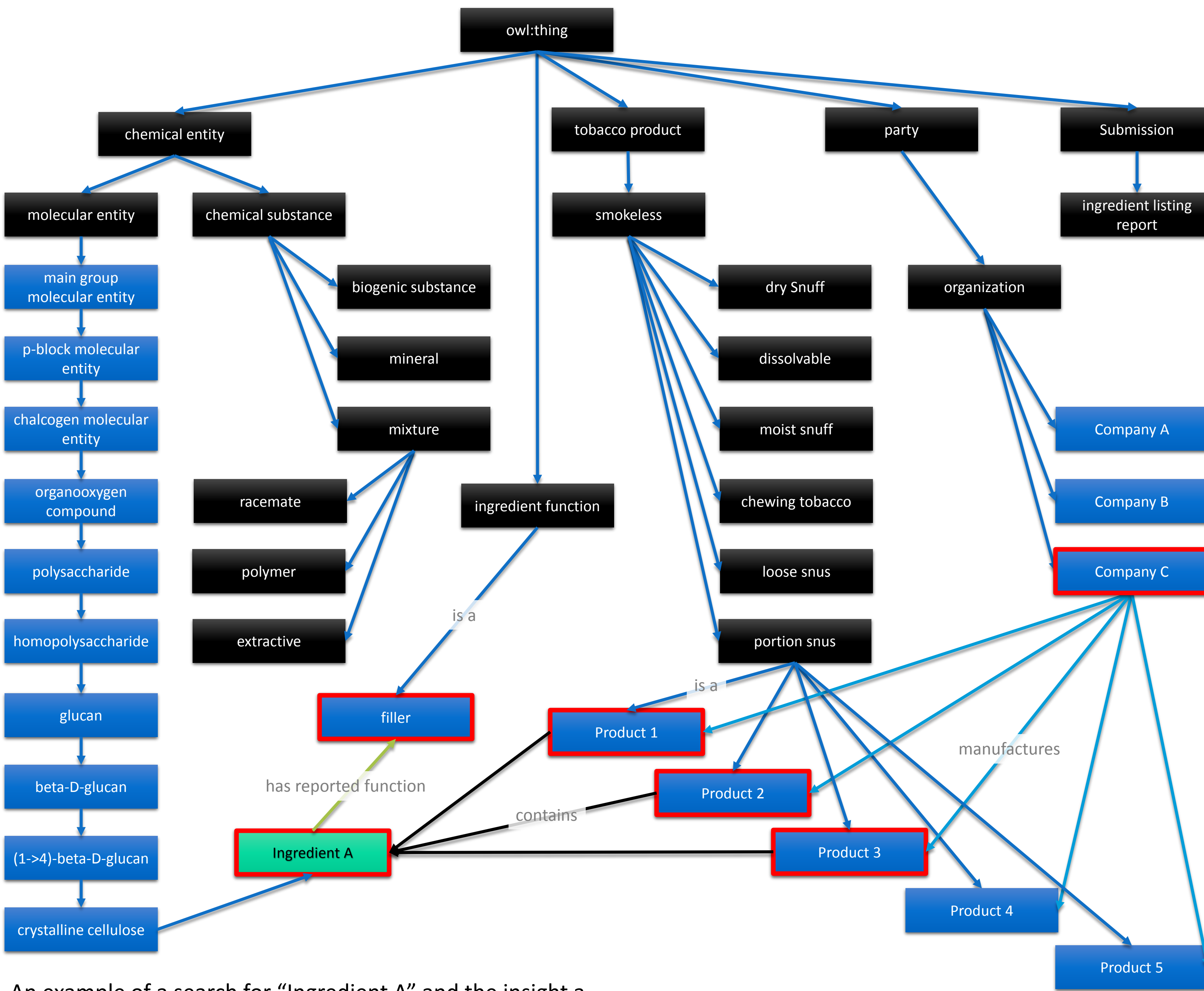
An example of a search for "Ingredient A" and the insight a visual graph of the ontology can bring

---

APPLICATIONS

- Moving from term-based to concept-based document retrieval
- An ingredient data service for CTP's new Integrated Review System
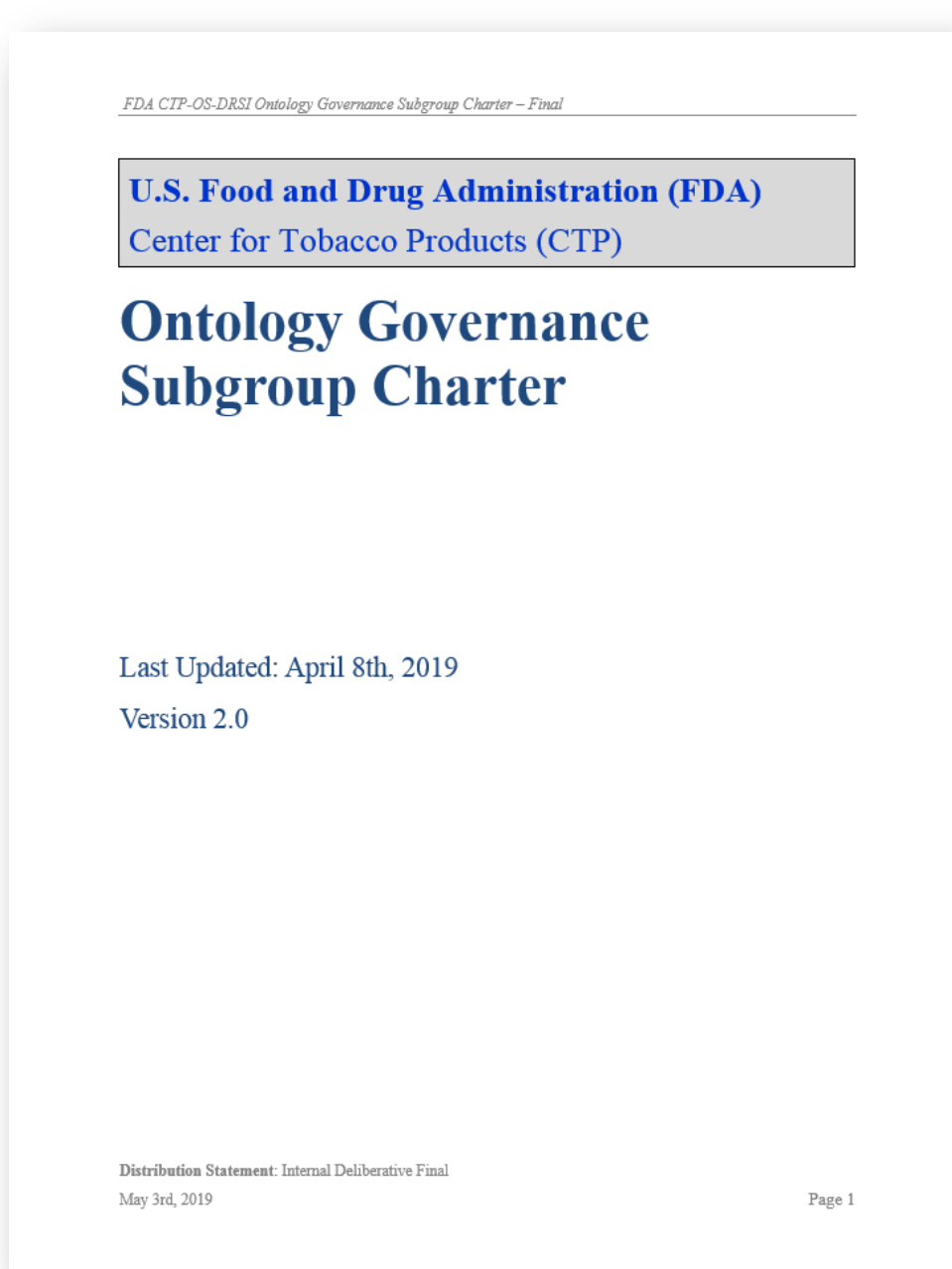- Revising the structured information repository
- Informing migration of legacy data from the old to new repository
- Making ingredient collection uniform across submission types

ADDITIONAL WORK PRODUCTS

U.S. Food and Drug Administration (FDA)
Center for Tobacco Products (CTP)

Ontology Governance Subgroup Charter

Last Updated: April 8th, 2019
Version 2.0

U.S. Food and Drug Administration (FDA)
Scientific Computing Support
Center for Tobacco Products (CTP)

Guide for Standardizing Ingredient Data Draft

Task Order 6.0
Task Area 4.1 – Building Knowledge
Contract # HHSF223018277

Last Updated: June 24th, 2019
Version 4.2

Ingredient Lookup Table