Building a Comprehensive Breast Implant Ontology Leveraging GUDID and Unstructured Data Sources

Mark Jung[1,2], Michael Wu[1,3], Asiyah Yu Lin[1], Hongying (Helen) Jiang[1], Weiguang Wang[1], Zhou Feng[1], Dongyi (Tony) Du[1], Nilsa Loyo-Berrios[1]

[1]FDA/CDRH/Office of Surveillance and Biometrics, Silver Spring, MD 20993
[2]Montgomery Blair High School, [3]Wheaton High School

Background

- Currently, breast implant data is spread out across FDA PMAs, Global Medical Device Nomenclature, and Global Unique Device Identification Database (GUDID)
- This significantly hinders data analysis, such as in the BIA-ALCL project, where breast implant surface type may have a possible association with breast-implant associated anaplastic large cell lymphoma (BIA-ALCL)
- A domain ontology is a formal representation of categories, properties, and relations between entities in a single domain. An ontology can categorize and automatic classify terms.

Objective

- Create a complete ontology to classify breast implants based on their features.
- Use the ontology to support other BIA-ALCL related data analysis tasks, such as a backend dictionary to for a text mining tool to mine the breast implant related MDR data.
- Eventually merge with other related domain ontologies (e.g. OAE)
- Help improve accuracy and scope of GUDID

Methods

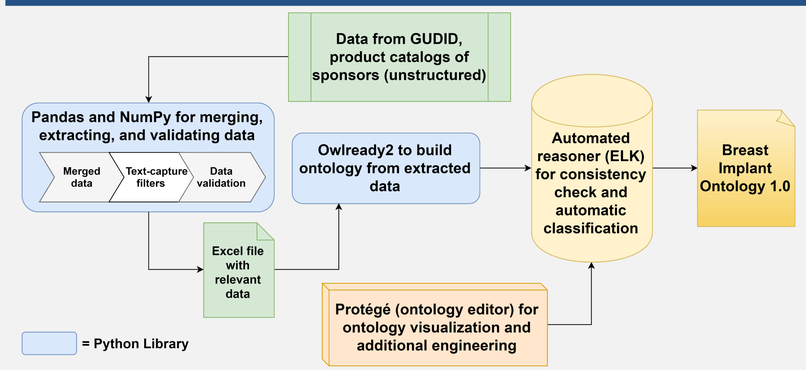

Figure 1. Ontology development workflow
Data was first downloaded, extracted, filtered, and validated before running the ELK reasoner to achieve the first version of the ontology.

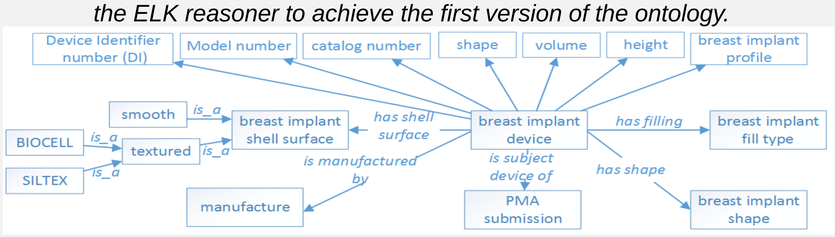

Figure 2. Initial ontology design pattern
Characteristics such as filling, surface, projection, shape, manufacturer, and PMA submission number were considered in the planning stages of the ontology development

Results

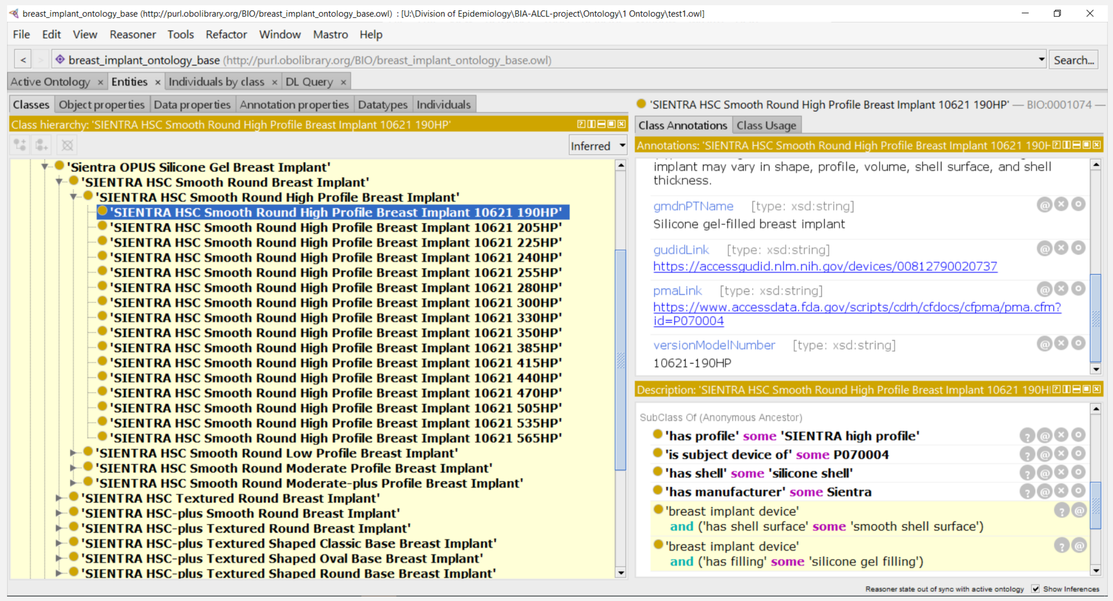

Figure 3. Class Tab in Entities View in Protégé 5
Individual devices (lowest level classes) are subclasses of specific relationships (e.g. 'has profile' some 'SIENTRA high profile') and are also annotated with additional information, e.g. GUDID link and version model number.

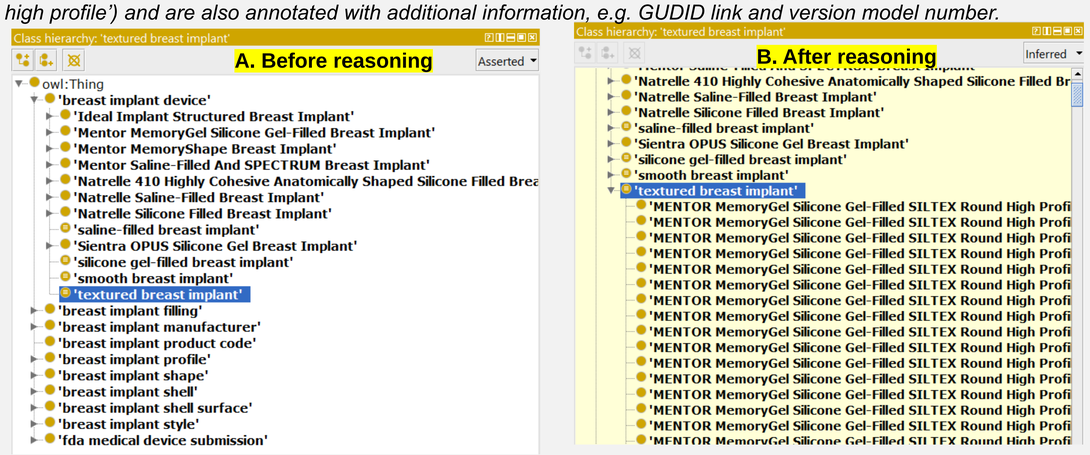

Figure 4. Before (A) and after (B) running the ELK reasoner on the Breast Implant Ontology
Defined classes, e.g. 'saline-filled device', 'silicone gel-filled device', 'smooth device', and 'textured device', need not be manually populated; instead, supplying the relevant axioms to an automated reasoner allows it to infer subclasses.

- An Initial ontology design pattern was created to lay out the attributes of breast implants and how they will be related to each other (Figure 2).
- 1738 breast implant products (individual devices) from GUDID are included in the Breast Implant Ontology (BIO).
- The hierarchy of BIO is as below:
  — breast implant device
    — breast implants with PMA number (8)
      — company's brand
        — style
          — individual breast implant
- In addition to class-level features, annotations such as product code, device catalog number, version model number, GUDID link, PMA link, GMDN name, GMDN definition, and device publish dates were included in the ontology (Figure 3).
- Artificially defined classes were created to facilitate the automatic classification:
  - Saline-filled breast implant
  - silicone gel-filled breast implant
  - smooth breast implant
  - textured breast implant
- For example, by defining "textured breast implant" with axioms "is a breast implant device" and "has a textured shell surface", an automated reasoner such as ELK can automatically populate "textured breast implant" with all textured devices.
- After reasoning, ELK classified 906 textured breast implants and 836 smooth breast implants from GUDID (Figure 4).

Conclusion

The current Breast Implant Ontology (BIO) is in its preliminary stage. It extends current GUDID breast implant information by adding more features (*e.g.* surface type). To make BIO useful for the text mining on MDR datasets, off-market historical breast implants need to be added. Future work also includes adding device dimension measurements (*e.g.* width/diameter, height, fill volume, projection, etc.) into the ontology. BIO will also be released as an open source to other stakeholders, e.g. researchers, industries, and organizations.

Acknowledgements

We thank the valuable support from OSB and DSD teams, particularly Danica Marinac-Dabic, Behnaz Minaei, Terrie Reed, Binita S Ashar, Karen Nast, Sung Yoon, Trisha Eustaquio, and Tajanay Ki.