**Build a Comprehensive Breast Implant Ontology Leveraging GUDID and Unstructured Data Sources.**

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

**Division of Epidemiology, OSB/CDRH/FDA**

**Abstract**

**Background:**

Individuals with breast implants have a risk of developing breast implant-associated anaplastic large cell lymphoma (BIA-ALCL), which based on published literature, appear to be associated with breast implants that have textured surface. The FDA product codes and the Global Medical Device Nomenclature (GMDN) do not capture breast implants surface type. This significantly limits our ability to assess the potential association between the breast implant surface types and BIA-ALCL. A Breast Implant Ontology can categorize all the breast implant products and their attributes under an ontological structure, which can be used by a *semantic reasoner tool* to automatically classify breast implants features (e.g. smooth or textured surface types). This ontology could also be used as a backend dictionary for a *text mining tool* to help explore the patterns/trends from BIA-ALCL cases reported through the Medical Device Reporting System or other data sources.

**Objective:**

The objective of this project is to create an extensive ontology to categorize the different breast implants and their product features, including but not limited to: manufacturer, brand, filling, and so on. The ontology will help FDA determine possible relationships between BI structures and patient symptoms that correspond with BIA-ALCL.

**Method:**

Breast implant data on implant characteristics were gathered primarily from the GUDID database using the FDA product codes FWM and FTR. Current catalogs from four breast implant sponsors were also used for cross-verification of data as well as providing additional information, such as device dimensions, diameters, heights, projection, etc. Free text such as literature review and research articles were also included.

Python libraries — including NumPy and Pandas – were used for piping data through text-capture filters into the Pandas DataFrame data structure and finally into an Excel file. This will be automated as much as possible to facilitate future GUDID downloads. Protégé, an open source ontology application, was used to help categorize each implant based on their properties to create a logical structure. To increase efficiency, the Python Owlready2 library was used to load all content from the excel sheets into the ontology.

**Results and Discussions**

A pattern design of the Breast Implant Ontology was developed prior to ontology development. A total of 1,739 breast implant data, consisting of 33 unique fields including – but not limited to –device identification (DI), record status, publish date, brand name, model number, company name, and device description were downloaded from GUDID database on June 28th, 2018. Additional information was also retrieved, including a unique device name, device dimensions (width, height, projection, etc) from sponsor’s catalogs and other PMA approval orders. Overarching classification categories that were ultimately included are device manufacturer, brand, style, filling, profile/dimensions, size, shape, shell, shell surface, and product code. Data such as DI, model number, catalog number, manufacturer’s device description, GMDN name, and GMDN definition are also included as annotations for each individual device.

The ontology is searchable with SPARQL queries or Protégé’s DL Query tab, allowing users to find individual data as well as their relationships. This may simplify text mining in the context of discovering possible correlations between certain types of breast implants and BIA-ALCL. The ontology is also available as open source to aid other researchers and organizations.