**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**

**Abstract**

**Background:**

Currently both device product codes and global medical device nomenclature (GMDN) do not provide the terms for breast implants surface type. It hinders significantly the data analysis when epidemiologists want to find out the association between the breast implant types and BIA-ALCL, which has been shown by literatures. A Breast Implant Ontology can categorize all the breast implant products and their features/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 will also be used as a backend dictionary for a *text mining tool* to help explore the patterns/trends from the MDR dataset and the social media data.

**Objective:**

The objective of our project was to create an extensive ontology for the management and comprehension of different breast implants and their product features, including but not exclusive 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 AccessGUDID database using the search query:

productCodes.fdaProductCode.productCode:(FWM) OR productCodes.fdaProductCode.productCode:(FTR)

Company catalogs were 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 will also be included.

Python libraries — including NumPy and Pandas – will be 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 AccessGUDID downloads. Protégé, an open source ontology application, will then be 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 onto the ontology.

**Result and Discussion**

A pattern design of the Breast Implant Ontology was developed prior to ontology development. 1739 breast implant data were downloaded from GUDID database on the date of June 28th, 2018. Additional information were added into the GUDID dataset, including a unique device name, device dimensions (width, height, projection, etc) from company’s catalogs and other product 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 device primary identification, version model number, catalog number, manufacturer’s device description, device ID, GMDN name, and GMDN definition are also included as annotations for each individual device.