**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:**

Breast implants (BI) have become a ubiquitous treatment in modern society — a common practice for cosmetic augmentation or medical reconstruction — but recently have experienced withdrawals due to a serious complication known as Breast Implant-Associated Anaplastic Large Cell Lymphoma (BIA-ALCL). BIA-ALCL is a rare T-cell lymphatic cancer, which causes enlarged lymph nodes near the surface of the skin in the breast tissue due to implant insertion. It’s also categorized as a non-Hodgkin’s lymphoma by the World Health Organization, and groups like the Plastic Surgery Foundation (PSF) and National Comprehensive Cancer Network (NCCN) have published information to provide diagnoses and preventative measures. However, there is still limited understanding of what causes this disease to occur.

**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:**

Data on implant characteristics will be gathered primarily from the AccessGUDID database using the search query:

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

Company catalogs will be useful for cross-verification of data as well as providing additional information, such as device dimensions. Free text such as literature review and research articles will also be included.

The AccessGUDID information will be transformed in three main steps:

1. Merging the provided text to a Microsoft Excel document consisting of their respective sheets.
2. Extracting the relevant columns from the Excel file.
3. Extracting relevant data (e.g. device style, shape, filling) from the columns.

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 will be used to help directly load all content from the excel sheets onto the ontology.

**Result and Discussion**

Familiarity with ontology development strategies as well as the Protégé editing tool was established before development on the breast implant knowledge graph began. During development, class structure and relationships played a fundamental role in discussions as standards within companies were found to be varying and/or inconsistent with government data. The role of the team itself was also considered when assigning a standardized value across devices, such as a unique device name, and when assigning device dimensions (width, height, projection, etc) which were not provided by the AccessGUDID database.

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.