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

**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, appears to be associated with breast implants that have a 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.

**Methods:** A theoretical 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 was 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 eventually 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.

**Results:** 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.

**Keywords**

Breast implants, Ontology, Automatic classification, ALCL, OWL

**Background**

Breast implants are very common medical devices that are implanted under breast tissue and are used for breast augmentation or reconstruction. There are currently 1738 different types of breast implants available on the market: all varying in size, thickness, texture, shape, and fill-types. However, the details for every specific breast implant is scattered across FDA PMAs, the GDMN, and the GUDID. This disorganization often leads to redundant entries and significantly hinders data analysis, slowing down future research on breast implants. This is best shown in the BIA-ALCL project. BIA-ALCL is a type of non-Hodgkin’s lymphoma (immune system cancer) and is often found near the scar tissue and fluid of the implant. Current research suspects that breast implant surface textures may have a correlation with BIA-ALCL; but, future studies cannot be efficiently conducted without a common database to help provide detailed information on each breast implant. A domain ontology is a formal representation of categories, properties, and relations between entities in a single domain. It is an effective method to categorize and automatically classify terms in a logical structure, showing the relationships between each property. For these reasons, the use of ontologies may help to organize breast implants and their properties that allows for efficient analyzation.

**Methods**

The ontology design is as follows: A class named **breast implant device** would contain each individual breast implant, organizing them first by FDA premarket application (PMA) name and then by brand specific sublevels. Breast implant properties (if classifiable) would be sibling classes to **breast implant device**. These would include breast implant filling, manufacturer, product code, profile, shape, shell, shell surface, style, PMA, and PMA supplement. Corresponding OWL Object Property names would follow OWL convention (e.g. “has filling,” “has manufacturer”) except for PMAs and PMA supplements, which are related to breast implant devices using “is subject device of” and “mentions”, respectively. Non-classifiable breast implant data would be included in BIO either as annotations or data properties. For breast implant device annotations, these would include catalog number, version model number, device description, device ID, device publish date, GMDN name and definition, PMA link, and GUDID link. Several PMA annotations would also be added, including advisory committee, applicant, applicant address, approval order statement, date received, decision code, docket number, federal regulation notice date, generic name, review grant status, supplement reason, supplement type, and trade name.

The Global Unique Device Identification Database (GUDID) was used to provide base information for the ontology. GUDID is an FDA managed database that stores all medical device information along with a unique identifier. Initial data were downloaded as delimited text files from the FDA public online portal AccessGUDID using the search query, *productCodes.fdaProductCode.productCode:(FWM) OR productCodes.fdaProductCode.productCode:(FTR)*. This resulted in 1745 rows of data. Later it was discovered that 7 devices listed as saline-filled breast implants were in fact saline pumps or injection domes – these devices were removed from consideration, giving a total of 1738 FDA approved breast implants. Relevant columns from GUDID that were used to build the include the following: PrimaryDI, devicePublishDate, brandName, versionModelNumber, catalogNumber, companyName, deviceDescription, sizeText, productCode, productCodeName, deviceId, gmdnPTName, gmdnPTDefinition. While some columns went directly into BIO, others were used to infer information. Most of these data were explicit – although their location varied across sponsors – and some were implicit. Mentor MemoryShape implant heights (text) and profiles were explicitly stated in GUDID’s device description column. Texture (SILTEX) and shape (teardrop) are implied and as such, not in GUDID. On the other hand, Mentor saline-filled/SPECTRUM implants provide texture, shape, and profile in the brand name column. Allergan (Natrelle) breast implants have their information listed in the device description column. However, because property naming was slightly inconsistent and a pattern was found regarding the front half of Allergan catalog numbers, a dictionary mapping that half to corresponding breast implant properties was created. A Natrelle catalog was used for this task. The same was done for Sientra breast implants. Ideal Implant were much easier, having only 14 devices total, all of which have a smooth surface, saline filling, and round shape.

Sponsor catalogs were used to provide verification for GUDID catalog numbers as well as to provide additional data. GUDID does not provide individual breast implant device dimensions, such as width, height, diameter, or projection. Similarly to finding breast implant class properties, catalog numbers were used in linking GUDID with device dimensions, which were added as data properties.

FDA PMA data were added to BIO to include pertinent info for users doing research or reviews. All PMA data were first downloaded into a single file. Relevant PMAs related to breast implants were then filtered locally.

All data were compiled into a single Excel file with four spreadsheets, a classification and dimensions sheet, an annotations sheet, a unique classifications and dimensions sheet, and a PMA data sheet.

**Results**

This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

**Discussion**

For research articles this section should discuss the implications of the findings in context of existing research and highlight limitations of the study. For study protocols and methodology manuscripts this section should include a discussion of any practical or operational issues involved in performing the study and any issues not covered in other sections.

**Conclusions**

This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study to the field.

**List of abbreviations**

GUDID — Global Unique Device Identification Database

GDMN — Global Medical Device Nomenclature

BIA-ALCL — Breast Implant-Associated Large Cell Lymphoma

PMA — Pre-Market Approval

OWL — Web Ontology Language

SPAQRL — Simple Protocol and RDF Query Language

**Citations**