

## Attachment 1 – DAGSI Research Topic Template

1. **Research Title:** Semantic Labeling of Data Manifold Regions for Human Understanding
2. **Individual Sponsor:** List the AFRL research topic sponsor's contact information

Example:

Dr. Paul R. Havig II, AFRL 711HPW/RHCV  
711 HPW/RHCV  
2255 H ST.  
WPAFB, OH 45433  
[Paul.Havig@us.af.mil](mailto:Paul.Havig@us.af.mil)

3. **Academic Area/Field and Education Level**

Computer Science/Mathematics (BA/BS, MS or PhD level)

4. **Objectives:** The objective of this research is to investigate a novel means of assigning semantic labels using algorithms, to the output of topological data analysis (TDA), that take broader data domain knowledge into account. Assigning such labels would make the output of the TDA (which is very mathematically rigorous) more consumable and understandable by users who may not have the mathematical background of TDA.
5. **Description:** TDA is a subfield of data mining that uses tools from the mathematical field of Topology to approximate, visualize, and exploit the manifold of a dataset to find insights. The *mapper* algorithm, in particular, is proving to be an effective applied TDA algorithm for recovering the k-simplex of the manifold of a data. Mapper has utility in the analysis of the k-simplex, which can be seen as a graph where vertices represent data clusters and edges correspond to an overlap between up to k- clusters. Dense regions of the simplex thus represent regions of the data manifold that is dense with data having some kind of commonality, be it the presence of an attribute, a common distribution of continuous feature values, or a strong correlation between the values of subsets of features. The immediate assignment of a semantic label that characterizes the commonality of such data would augment existing k-simplex visualization methods with labels that describe, in a human understandable way, what different "regions" of the data manifold qualitatively represent. This would thus provide an immediate "map" of a dataset for a human analyst, which could be quickly consumed and lead to fast insights as to the broad properties of a dataset and the fast interpretation of a datum exhibiting particular feature values.
6. **Research Classification/Restrictions:** None.
7. **Eligible Research Institutions:** Wright State University, Ohio State University