<table>
<thead>
<tr>
<th>Project Title</th>
<th>“TL;DR it”: Automating Article Synopses for Search Engine Optimization and Citizen Science.</th>
</tr>
</thead>
</table>
| Description   | Open Access initiatives have worked successfully to create open forms of scholarship, but the story of scholarship and the accessibility of research concepts/ideas in that scholarship remains out of reach for some audiences. To address this scholarly outreach gap, we are looking to build readable, accessible scholarship for citizen science and awareness. The first part of this work was completed with funds from an internal grant. It involved surveying and learning about summarization practices of academics and discerning how we might create a “snapshot” of a scholarly article - a concise and readable summarization of the thesis, methods, findings, major figures of an article - for a broad, reading audience. Fellows will have the opportunity to build on these findings and follow research into new deliverables that could include:  
- Applying generative and abstractive machine learning techniques to change reading levels of articles or other data sources.  
- Using natural language processing to identify and qualify jargon in articles or other data sources.  
- Designing the markup and structured data to encode the snapshot article for machine-readability and harvesting.  
- Establishing analytical metrics for assessing the reach and impact of the snapshot articles in search engines. |
| Problems/ Research Questions | Fellows will be encouraged to develop complementary research questions and solutions. Initial research questions include:  
- Can we use generative computing models to translate academic language into more accessible reading levels in these snapshot articles?  
- How can these snapshot articles be implemented or optimized for discovery in search engine result pages?  
- What are the metrics and analytics needed to assess the reach and impact of these articles? |
### Techniques
- Web scraping and text mining
- Natural language processing
- Generative computing techniques
- Semantic Web markup and optimization

### Tools/Languages used
The project team has expertise in Python, Javascript, Semantic Web optimization, and Excel. Some capacity exists to support fellows interested in network analysis and NoSQL/graph databases.

### Data
**Description:** Metadata and source articles that have been pulled from Open Access research. A "Big Data" source here could be a corpus of COVID-19 research in PubMed or any other data source that needs analysis/summarization.
**Data Type:** HTML, PDF, OCR Text
**Data Size:** MBs

### Outcome
Published papers, presentations, datasets, software code.

### Milestone Timeline
Six months per selected project

### References
GitHub repository for initial research:

https://github.com/msulibrary/msu-article-summarization

Fellow will have access to the initial research survey for text summarization practices (n=59 MSU Faculty)