1. ***Develop a group of lead users / beta testers, and work with them to determine the accuracy and usefulness of indicators, and to identify the questions that users will want to have answered by the indicators.***

This task is listed first because it will develop and maintain a large group of lead users, experts, and beta testers that will be consulted throughout the project to establish the accuracy and usefulness of the results. Early work will address accuracy and identify features, indicators and use cases of high value to potential clients, and that will thus have economic impact. These interactions are needed to ensure that user needs and desires are well understood, that the project results will meet user needs, and that the indicator development leads to commercial value.

*Milestones:*

* Months 1-5 – Successfully recruit at least six scientific experts (from the list of internal and external affiliates of the Stanford Meta-Research Innovation Center (METRICS) and other researchers interested in meta-research) to provide feedback on the accuracy of coding of virtuous features from sample articles. These experts are in addition to the institutional lead users from the previous milestone.
* Months 6-12 – Identify a random sample of 100 coded articles and send 10 articles and their codings of virtuous features to each scientific expert for review. 70% inter-rater reliability will indicate sufficient accuracy to proceed with additional coding. Experts will also be asked if they can suggest candidate terms and logics that could be used to code features algorithmically.
* Months 14-17 – Identify a random sample of 100 articles that have been coded both manually (see task 5) and by text mining (see task 6) and send 10 articles and their codings of virtuous features to each scientific expert for review. Experts will assess both sets of codings. Results will be used to modify the algorithmic virtuous feature identification routines.
* Months 18-24 – Identify a random sample of 150 articles that have been coded both manually and algorithmically using the updated algorithms and send 15 articles and their codings of virtuous features to each scientific expert for review. Experts will assess both sets of codings. Results will be used to modify the algorithmic virtuous feature identification routines.

*Deliverables:*

A detailed accounting of the results from the study associated with each time period listed above will be delivered to NIDA.

1. ***N/A***
2. ***N/A***
3. ***N/A***
4. ***Code and validate virtuous features from a sample of scientific articles.***

The Subcontractor currently has a repository of 650 papers from PubMed published in 2000-2017 that has been assessed and coded for virtuous features. In this task, the Contractor will extend this assessment to an additional 500 papers published in 2015-2018 to get a more comprehensive insight of the most current status of these measures of the virtue of the research using the same processes detailed in our Phase I report.

*Milestones:*

* Months 1-6 – Manually code virtuous features (e.g., data and protocol availability, COI, novelty, reproducibility, potential for translation) for a random sample of 500 biomedical papers from PubMed published in 2015-2018. Coding will include the presence or absence of a feature, along with the location within the paper (or text from the paper) for features that are positively identified.
* Months 7-8 – Using the full set of 1150 papers (500 new + 650 repository), compare the efficacy of feature identification using full text and metadata, to determine those features that are easier and most straightforward and unambiguous/accurate to extract manually, and identify potential features (terms, logics, etc.) that could be used to algorithmically extract virtuous features from text.
* Months 6-10 – Conduct study with outside experts – see related milestone for task 1.

1. *Deliverables:*

Month 7 – A detailed accounting of the coded data from 500 papers will be delivered to NIDA.

Month 11 – An accounting of the evaluations of codings (task 1) by outside experts will be provided to NIDA.

1. ***Create an observatory of biomedical research that combines optimal field partitioning based on open sources, measures of virtue of research on expanded samples of PubMed articles and additional indicators of quality of methods, translational potential, and novel linkages. Feedback from end users and scientific experts from Task 1 will inform this effort to ensure the work is focused on indicators with both scientific and economic value.***

The output of task 5 can be expanded and scaled up to include an additional 3,000 PubMed articles published in 2015-2018. The exact number will depend on the efficiency of any algorithmic/automated tools that might have been achieved. The Contractor will focus on indicators that are easier and most straightforward and unambiguous/accurate to extract manually, including availability of raw data, full protocols, funding, and conflicts of interest disclosures. Random sampling will be used with in-depth data extraction and much larger samples of articles where information on virtue features may be collected with text-mining from abstracts and full-text articles.

The Contractor will explore different text mining approaches with different text term searches to identify those that may come closer to results obtained from in-depth manual extraction of these indicators. Text mining may be particularly useful to use for indicators that are difficult to extract manually, e.g. whether a study is a replication and whether it has been used in a systematic review/meta-analysis.

*Milestones:*

* Months 11-15 – Using data from task 5, explore multiple text mining approaches with different term or logic searches to identify approaches that may come closer to results obtained from in-depth manual extraction of these indicators. Precision and recall of the various approaches will be estimated by running the text mining approaches on the full set of 1150 coded papers and comparing text mining results to the manual result. The best approach for each indicator will be identified. The Contractor assumes that precision will be more important than recall for our users. Approaches with at least 50% precision and 25% recall may be candidates for further algorithmic development.
* Months 13-16 – Apply the text mining approaches to a new set of at least 500 articles using those approaches that meet the precision/recall criteria. A randomly sampled set of 1,000 machine identified features will be manually checked with the articles to establish the predictive power of these approaches.
* Months 14-17 – Conduct study with outside experts – see related milestone for task 1.
* Months 17-20 – Text mining approaches will be further refined using results from the first-year studies and will be applied to an additional 1,000 articles. A randomly sampled set of 2,000 machine identified features will be manually checked with the articles to establish the predictive power of these approaches. Approaches with at least 65% precision and 40% recall are sought.
* Months 19-22 – Conduct focused study with outside experts – see related milestone for task 1.
* Month 16-24 – Compile all previous verified results into a data observatory and make it publicly available. We will seek feedback about this observatory and about its detailed observations using surveys and communication with major funding agencies of biomedical research and with several leading institutions. This will include the previously identified experts and many others.
* Months 16-24 – Explore additional potential indicators based on a) misuse of p-values, and b) potential for translation, and c) unique combinations of virtuous features. This work will be exploratory and will inform future developments.

*Deliverables:*

Month 16 – A detailed accounting of the text mining approaches and how they each compare with manually coded data, along with related statistical analysis, will be provided to NIDA.

Month 22 – A detailed accounting of the refined text mining approaches and how they each compare with manually coded data, along with related statistical analysis, will be provided to NIDA.

Month 24 – The observatory will be completed and made publicly available.

Month 24 – An accounting of the exploratory work on new potential indicators will be provided to NIDA.