Data-Driven Descriptor Models for Guiding Metal Nanoparticle Synthesis

Prof. X (PI), Prof. Y (co-PI), and Prof. Z (co-PI)

The University of Alabama

Data Management Plan

- **Types of data, samples, physical collections, software, curriculum materials; including the standards to be used for data and metadata format and content**

The computer simulation work performed by PI Prof. X described in this proposal will generate a number of files (KMC coordinate files, trajectory files, state-space model database, etc.) that then will be analyzed using using in-house computer codes (typically in Fortran). This analysis will result in additional data files that can then be imported into Microsoft® Office Excel® for further use (e.g., preparation of tables and figures). Snapshots from the simulations are typically stored in conventional graphic files (e.g., .tiff, .jpeg). Movies of the simulation trajectories can be produced and stored in conventional movie files (e.g., .avi, .mpeg, .mov formats). Prof. Y will be generating Objective-C codes for the iOS application development. This will also involve creating a user interface for the data acquisition process and creating an iOS version of the self-learning KMC simulator.

From the experimental side, data acquired in Prof. Y's laboratory will be stored in a variety of formats depending on the instruments used for data acquisition. This will primarily be data about the experimental parameters and characterization data (e.g., TEM images). This characterization data and TEM images will be stored as the raw images at maximum resolution, as well as in compressed forms and PDF files for later use. The experimental parameter data will be initially stored in MS Excel® spreadsheets until an iOS data acquisition process is developed.

- **Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements; policies and provisions for re-use, re-distribution, and the production of derivatives**

Preliminary analyses and summaries of these results by students and professors will be presented at internal (bi-weekly) group meetings. Reproducible results that provide new insight will then be promptly disseminated through presentations at professional conferences, as well as via peer-reviewed publications (including supplementary information). All relevant data and the most important metadata (e.g., main details of the input files used in the simulations, general description of apparatuses and raw materials, etc.) will be included in the body of our papers and their supplementary information. This has been common practice in all the papers historically published by the groups of Profs. X, Y, and Z. Contact information for the corresponding author will always be included in these publications. Any additional information or data will be available upon request, unless the information requested cannot be released (e.g., patentable inventions). If that is the case, the PIs will communicate these restrictions to the requesting party.

If there is intellectual property deemed patentable, this will be disclosed (via a University invention disclosure procedure), as this disclosure is required by UA. If a patent is pursued, our data release to the general public will be delayed until a provisional patent application has been filed.

New software is expected to be produced from the proposed self-learning KMC development. The software (source codes) will be available upon request under the GNU Lesser General Public License, and the binary iOS versions of the software will be made available via the Apple iTunes Store. We expect new self-learning KMC simulations to be of great value to other research groups. Thus, the new models will be fully described in our peer-reviewed publications.

Regarding re-use, re-distribution and production of derivatives, the websites of Profs. X, Y, and Z contain descriptions of their general research activities, as well as complete references to their papers. All the information posted on these websites is publicly available without restrictions.

- **Plans for archiving data, samples, and other research products, and for preservation of access to them.**

All of the analyzed data and necessary metadata (e.g., input files used in molecular simulations, descriptions of raw materials used in experiments) resulting from this research will be stored electronically in the local computing facilities of the individual research groups. In addition, summaries of analysis of data and necessary metadata will be compiled in individual student laboratory notebooks. The individual researchers regularly backup their data into standard storage devices (e.g., external hard drives, DVDs).

As a collaborative group, the simulation data and experimental data will be housed in a network-attached storage (NAS) data repository, which can hold approximately 2 TB of data. It has power redundancy and data redundancy (RAID level 5), and it allows data to be imported/exported readily via a web-browser interface. This will serve as our "group" data archive. If we receive external requests for our research data, we can easily grant NAS "access permissions", allowing an individual to browse the data and download any of our checkpoint files or trajectory files (from our simulations), as needed.

All data and metadata will be retained by Profs. X, Y, and Z for a minimum of three years after the conclusion of this award. If the PI or one of the co-PIs leaves during the award (or post-award) period, the remaining researchers accept the responsibility of acquiring and maintaining the necessary data, in order to ensure continuity.