6 Month Evaluation: Instructions for Performers on Timing and Work Products

During the course of the evaluation, you will be given a set of scenarios and asked to produce a variety of work products/outputs with (approximate) timing information on how long it took to generate the specific product.

We recognize that generating these products will be iterative and collaborative – and that documenting what you did and producing the requested work products will slow things down and distort the timing. Don't worry about this and just do the best you can. Keep in mind that the purpose of the 6-month evaluation is to identify where the tools are working, and where there are gaps.

Below is a list of types of work products being asked for, and what kind of output we would like. The timing log to help everyone track the time spent working on specific tasks/work products, is located here:
https://docs.google.com/spreadsheets/d/1sfNWqdgdSIShRSndwinSVU_5YXsIhYuE/edit?usp=sharing&ouid=109251687910426835850&rtpof=true&sd=true
The work products for the evaluation should be deposited within the DARPA ASKEM GitHub repository (https://github.com/DARPA-ASKEM/program-milestones/tree/main/6-month-milestone/evaluation).

- **Timing Log**: The log is a running record of time spent on tasks associated with a work product. Given the iterative nature of development, it may be easiest to keep a running log of your activities, and then note the point at which a particular product is completed; this would allow timing to be computed for a given work product even if you jump around or do some iterative refinements. Where appropriate, record separately the execution time for a program that takes 15 min or longer, to distinguish this time from the time you spend getting things set up to run.
- **(TA1) Search and Discovery for models, publications, parameter values, data**: The task is to find relevant models, publications, parameter values, or datasets; the output product will briefly document the search strategy, include links to important references (e.g. papers or datasets, note the ones you actually ended up using for the rest of the scenario), and briefly describe how the model or dataset(s) addresses the specific requirements from the scenario.
- **(TA1) Model Extraction**: The task is to extract a description of the model and provide it in an executable form that is grounded to the DKG; the output work product will include:
  - Pointer to executable model representation (e.g. GroMet Function Network or Bilayer)
  - Model representation must have some output that is human readable/inspectable, in terms of equations, variables, parameters, with definitions, values, annotations, and any other metadata

- o Provide default values for parameters, initial values for variables, whatever else is needed to initiate/run the model
- o Grounding to DKG concepts and linkages/alignment between data, publications, code, and GroMet components should be inspectable and understandable by a human
- **(TA1) Model Execution/Unit Testing**: The task is to execute the model in its most basic form, for purposes of debugging and unit testing. The work product includes a pointer to GroMEt representation (including initial values and parameter settings) with execution outputs together with a brief commentary on the extent to which the model "passes" the unit test.
- **(TA2) Model Extension/Transformation**: The task is to alter or extend an existing model. The output work product should include a pointer to the revised model, along with a summary of how the revised model meets the requirements of the scenario.
- **(TA2) Model Comparison**: The task is to compare the structure of two or more models and provide a description of the similarities and differences. The format of the output product is flexible, and could include descriptions of new compartments, differences in parameters or transition pathways, diagrams, etc.
- **(TA2) Model Space Exploration**: Explore a space of closely related models (structurally speaking). For each model, calibrate parameters as specified by scenario, and do model selection based on how well the fitted model output compares with data. The final output of this type of task is the final model selection with calibrated parameters, a summary of the model space explored (e.g. what kinds of models were considered), and a brief explanation of how you came to your final selection.
- **(TA3) Simulation Workflow**: The task is to do simulation tasks with the model under different conditions, as defined in the scenario. These could include basic forward simulation (forecasting, backcasting, etc.), calibrating model parameters with data, implementing interventions, comparing different simulation results, doing a sensitivity analysis, optimization, creating ensembles, etc. The output work product should include a representation of the simulation workflow (e.g. string diagram), the simulation outputs, and a brief description of how the simulations are set up to address the specific question in the scenario.
- **(TA3) Answers to Scenario Questions**: The task is to address an explicit question in the scenario. The output should include the answer, simulation outputs and explanations to support the answer provided.
- **(TA1-3) Summary of Scenario Experience**: When you have completed your answers to a Scenario, we ask that you provide a short (1-2 paragraph) summary of what went well, what was challenging, where you hit roadblocks, etc.