ETI Thrust Area 1: Data Access and Collaboration

Jordan Stomps

February 10, 2020

0.1 Introduction

The purpose of this manual is to provide members, particularly those from university institutions, of the Consortium for Enabling Technologies and Innovation (ETI) details for collaboration with national labs. Data that has been gathered with potential interest to ETI collaborators is introduced, and instructions for accessing that data are provided. Hopefully this will serve as a catalyst for university members, like graduate students and their PIs, to partner with ventures at national labs that can benefit from academic research.

0.2 Multi-Informatics for Nuclear Operations Scenarios (MINOS)

This venture under NNSA Nonproliferation Research and Development (NA-22) is focused at Oak Ridge National Laboratory (ORNL). It is dedicated towards developing methods for identifying nuclear proliferation. To facilitate this study, MINOS acts as a compendium of datasets collected in various forms and at various locations. This section does not serve as an overview of the MINOS venture and its research efforts. Instead, this section will highlight examples of data that is collected as part of MINOS. For those interested in this data (perhaps for use in modelling and simulations), details will be presented on how to request access to this data and points of contact for collaboration with MINOS.

Examples of Available MINOS Data

An ETI Thrust Area 1 collaborators may be interested in MINOS data as a means for developing models and simulations that can be used in nonproliferation research. As it pertains to ETIs mission, MINOS provides real-world data, with ground-truth for some datasets, that can confirm or inform computational models. A brief, non-inclusive list of potential datasets are listed below:

- Seismo-acoustic
- Electro-magnetic

- Biota

- UHI Infrasound

- HFIR Measurements

- REDC Measurements

0.2.1 Requesting Data Access

Data is stored and managed by Lawrence Berkeley Laboratory (LBL). To begin the process, start by emailing Dr. Jared Johnson who is the ETI Point of Contact for MINOS. Please state that your affiliation (including your participation in ETI), your interest in collaboration with MINOS, and your intended use for the data. Upon approval, you will be asked to acknowledge and accept the MINOS Data Use Agreement. You will then be forwarded to register an account on the MINOS database and Dr. Brian Quiter of LBL will grant you access. Once this is complete you will be able to access, browse, download, and use MINOS data.

0.2.2 Browsing MINOS Data

While the MINOS website has many useful tools, this section serves to highlight a few. A great place to start for more detailed instructions is at the MINOS help page (found by going to the `Help` tab on the main page).

All datasets can be seen as a list by going to the `Datasets` tab on the webpage. This page allows keyword searches for specific datasets and has various tags to sort data based on its relevance. Each dataset has an affiliated organization, so selecting that on the left-hand side will filter only that data. Clicking on an entry will open a list of files associated with that dataset. There should be a JSON that contains summary information about what is contained on this page. Data is commonly tarballed into a large database format and a more reasonably sized HDF5 format. Clicking on any of the files brings the user to a download page for manually downloading files. Some data (for example biota) may be in another format like XLSX rather than those explained above.

Alternatively, data can be explored using the `Data Workspace` tab on the home page. Clicking `Workspace` allows the user to select any number of different data sets for download (remember to click `Save` after choosing

datasets). Once files are in the workspace, they can be selected and then reviewed by pressing `Load Data Collections`. The bottom portion of the page acts as a staging ground before downloading files. From here, there are three options for the user.

Download all files from data collections

Simply press `Submit Query` and all files checked in the workspace will be downloaded. The default file format downloaded will be the HDF5 datasets.

Download individual HDF5 datasets

This tab allows the user to individually select HDF5 files from the selected workspace datasets for download. A valuable feature here is the ability to add a time-selection visualization. Simply select the time filter for the datasets and a histogram will be generated with the frequency of entries at different time bins. This may be useful in understanding when data was being collected at maximum capacity. Subsets of the data can be selected by dragging an interval over the histogram. This will highlight the selection of times to be prepared for download.

Download individual non-HDF5 files

If there are any non-HDF5 files available for download in the selected workspace datasets, they can be specifically downloaded here.

0.2.3 Python Download API

If the user knows what data is needed, it can be downloaded without using the website. A Python API has been created that allows a user to download directly from code. To start, select
`Tools -> REST API Key Manager -> Generate New API Key`. Make sure to copy this key down and take note of its expiration date (new keys can always be generated). This section will not detail the rest of the instructions for using the API. Instead, these instructions are provided in the MINOS help documentation linked above.

This Python API can be executed from any desktop computer, but if it is more desirable it can also be executed from a JupyterHub portal found on the webpage at `Tools -> JupyterHub Portal`. This is run on a cluster at LBL

and allows a user to create Jupyter notebooks for data analysis. When the Python API is used here, data will be downloaded to the cluster, which may be more advantageous than downloading directly to a desktop computer.