ATTACHMENT D

Summary of Data Management Principles

The Alpha Magnetic Spectrometer on the International Space Station

14 November 2017

Brief description of the Alpha Magnetic Spectrometer (AMS):

AMS is a state-of-the-art precision particle physics detector designed to operate as an external module on the International Space Station (ISS). It uses the unique environment of space to study the universe and its origin by searching for antimatter, dark matter and new phenomena while performing precision measurements of cosmic rays composition and flux.

The key elements of AMS are the permanent magnet, the precision 9 layers of silicon tracker, the four planes of time of flight hodoscope, the array of anticoincidence counters, the Xe/CO2 transition radiation detector, the ring image Cerenkov detector and a 17 radiation length thick 3D electromagnetic calorimeter. AMS also contains more than 650 microcomputers with about 300,000 readout channels.

AMS was launched to the ISS on board STS-134 (space shuttle *Endeavour*) on May 16, 2011 and started to take data May 19, 2011. It is expected to continue through the lifetime of the ISS currently scheduled through 2024 and possibly to 2028. For 6 years of operations, more than 100 billions of cosmic ray events have been recorded and stored.

The main AMS Payload Operation and Control Center (POCC) and Science Operation Center (SOC) are situated at CERN, Geneva, Switzerland.

DOE's roles in the experiment:

DOE's Office of High Energy Physics under the DOE's Office of Science has the leading role in the AMS Experiment through an interagency agreement between NASA and DOE specifically for the AMS Experiment. DOE supports the AMS Experiment through two grants (DOE Grant Numbers DE-SC-0011848 and DE-SC-0011755 for Research and Operations respectively) with the Massachusetts Institute of Technology, Laboratory for Nuclear Science for a program in particle physics research. Under this grant, the AMS Principal Investigator (PI) Professor Samuel Ting has organized an international collaboration consisting of 45 institutes and universities from 15 countries.

Through the MIT grant and Principal Investigator, DOE is solely responsible for the science mission of AMS. To this end, DOE conducts regular, in depth scientific reviews of AMS by Blue Ribbon panels composed of leading U.S. physicists. These reviews have taken place in 1995, 1999, 2006, 2013 and 2016 and involve all the major international AMS partners. The entire

scope of detector operations and data analysis is reviewed at the same time. AMS is a unique DOE experiment as it is the only major, long duration, precision physics experiment conducted in space. It is therefore qualitatively different from traditional DOE experiments in accelerators.

In addition, DOE has provided support for the MIT scientists to design, construct and space qualify the electronics and data acquisition system as well as to develop the onboard and offline software. The DOE supported MIT scientists are responsible for the management and operations of AMS in space and lead the international data analysis effort.

Partnerships:

The AMS Experiment operates under the terms of the DOE interagency agreement with NASA. This document was signed originally in 1995 for a term of ten years. It was subsequently renewed and signed by the Associate Administrator for Space Operations Mission Directorate, NASA, Mr. William Gerstenmaier and by the former Director of the Office of Science, U.S. DOE, Dr. W.F. Brinkman, in January 2010.

Regarding data management, the Implementing Arrangement Part III, Section B. item 2 states that DOE (through the DOE-AMS Principal Investigator) is responsible for the management of "science data, analysis, distribution, and publication."

Organization – Agency/Lab level

In the current phase of the AMS Experiment, DOE, through the AMS-MIT team, is responsible for detector operations, data processing, setting the physics analysis topics and publications.

The DOE supported Laboratory for Nuclear Science (LNS) at M.I.T., under the leadership of Professor Bolek Wyslouch, provides administrative, infrastructure and support services for researchers in nuclear and particle physics. Since MIT is the lead U.S. university in AMS, LNS functions as the lead U.S. laboratory for DOE's role in AMS. Although there is a small contingent of MIT-AMS collaborators and students on campus at MIT, most of the AMS-MIT team are required to be in residence at CERN, Geneva where the AMS Payload Operations and Control Center (POCC) and the Science Operations Center (SOC) are located and where operations are continuous 24 hours a day, 7 days per week. LNS offices at MIT provide personnel, fiscal, property, and general services support to the MIT-AMS team. LNS provides the U.S. homebase which facilitates the MIT-AMS activities at CERN. In particular, the LNS Fiscal Office interfaces with the MIT Office of Sponsored Programs in the official processing of the MIT-AMS teams budget requests to DOE following MIT and DOE regulations and administers the follow-on DOE grants. Financial support for these services is provided through an administrative allocation applied to most activities administered by the Laboratory (salaries, materials and services, etc). Rates are reviewed annually and are added, as appropriate, in our DOE budget proposals.

Organization – Experiment level

The AMS organization chart is shown in Table I below.

The Principal Investigator of AMS is Professor Samuel Ting of MIT and Deputy Principal Investigators include Dr. Javier Berdugo of CIEMAT, (Madrid, Spain), Professor Bruna Bertucci of Perugia (INFN and University of Perugia, Italy), Dr. Michael Capell, Senior Research Scientist, MIT, Dr. Andrei Kounine, Senior Research Scientist, MIT, Professor Shih-Chang Lee, (Academician, Academia Sinica, Taipei, Taiwan) and Professor Stefan Schael, (RWTH, I Institute of Physics, Aachen, Germany). These individuals are responsible for physics publications, operations of the experiment, and interfacing with DOE, NASA and CERN. All Deputy Principal Investigators have been involved in the experiment since its conception.

Professor Roberto Battiston, former Deputy Principal Investigator for many years is now President of the Italian Space Agency and acts as Senior Advisor to the AMS PI.

The AMS Principal Investigator also chairs the AMS Executive Committee that is made up of representatives from major collaborating countries.

AMS also has a Finance Committee on Common Fund chaired by the CERN Director of Administration and General Infrastructure, co-Chaired by a senior member of the AMS Collaboration (currently Professor Andrea Contin of the University of Bologna) and including representatives from each collaborating country. The purpose of this Committee is to determine the yearly common fund budget shared by the entire Collaboration for common items and services. The amount of each institute's contribution is calculated based on the number of authors from that institute listed on AMS publications.

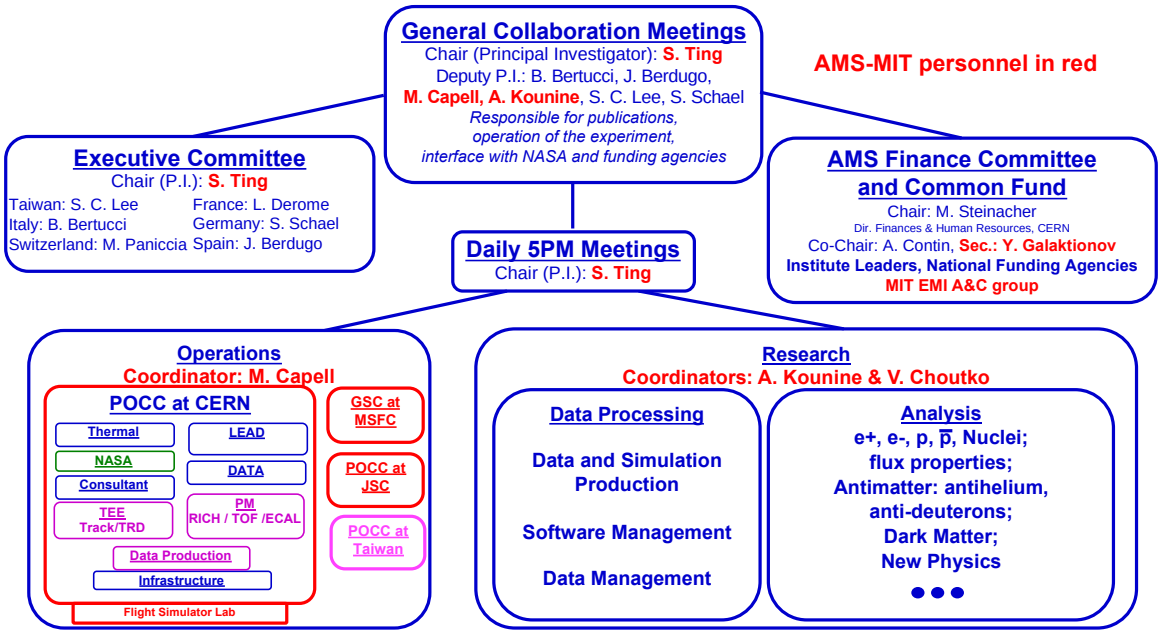

Table I
AMS Organization Chart

All members of the AMS Collaboration at CERN (and at MIT via video link) attend the daily 5PM meetings in the office of the Principal Investigator in the AMS POCC at CERN. Daily issues relating to AMS flight operations, detector status, data collection and analysis are reviewed. Once per week, the NASA AMS Project Management team joins the meeting via video connection from Johnson Space Center.

AMS operations and analysis are interrelated. Therefore, of particular note in the Organization Chart are the roles of the following MIT associated personnel in data management:

Dr. Michael Capell, responsible for AMS POCC operations.
Dr. Vitaly Choutko, responsible for AMS SOC operations.
Dr. Andrei Kounine, responsible for coordinating data analysis groups
Mr. Andrei Pashnin and Mr. Alexander Eline (AMS Data and user support group) are responsible for the AMS data transfer, replication and archiving.
V. Choutko, P. Zuccon et al, (AMS Software management group) are responsible for the AMS software maintenance and evolution.
Other MIT associated personnel with major responsibilities in data management and analysis are Dr. Weiwei Xu, Dr. Zhili Weng, Dr. Melanie Heil, Dr. Qi Yan and MIT graduate students.

Collaboration:

The AMS Collaboration is made up of 600 scientists and engineers from 45 institutions and 15 countries. Each country assigns a spokesperson representing the university teams from their home country according to their own internal procedures.

The majority of the institutes collaborating on AMS have worked with the DOE-MIT team on particle physics experiments for decades. The organization of AMS follows the model used in earlier DOE-MIT experiments (L3 at LEP/CERN, MARK-J at PETRA/DESY, etc.) For AMS, the NASA AMS Program Management Office at Johnson Space Center is an integral part of the AMS collaboration providing essential advice and support. The Principal Investigator of AMS convenes a general meeting of the entire collaboration every other month at CERN. The status of the analyses of various physics topics are presented in detail so that every member of the collaboration is fully informed.

In connection with AMS data management, the collaboration has five major remote computing centers in addition to CERN (France, Germany, Taiwan, China, and Italy) and their roles are detailed below.

Data policy management:

In accordance with the Implementing Arrangement between DOE and NASA Regarding AMS, DOE is responsible for the science mission of AMS. To this end, the AMS Principal Investigator, in consultation with the DOE Office of High Energy Physics and the AMS collaboration, is responsible for AMS data policy. As stated in Section VIII, item 2 of the DOE NASA

Implementing Arrangement: "In general, results of experiments will be made available to appropriate journals or other established channels as soon as practicable, consistent with good scientific practice."

Members of the AMS collaboration have unrestricted access to the data. Before AMS results are published, the findings must be independently validated by multiple (3 to 7) AMS research groups and meet the required level of precision. The data is included within the AMS publications and is therefore made available to the public. Indeed, a large volume of physics articles have been produced following the publication of original physics results from the AMS Collaboration using the AMS data.

The AMS Data Managment team led by V. Choutko of M.I.T. (see Table I) is responsible for the technical support for the AMS data management, including data transfer, simulation and reconstruction, data archiving and replicating in AMS regional computer centers.

Data Description & Processing:

Description:

Organization of original data collected by the detector.
Data collected on board the ISS by AMS consists of raw events, where each event is the collection of the information read out from the AMS detectors electronics after the latter are triggered by particle(s) crossing the AMS sensitive volumes. Raw events are logically grouped in sequences called runs, each run having a unique 32-bit identifier. Events inside one run are numbered in sequence. One run typically includes all events taken during a quarter of ISS Low Earth Orbit movements, i.e. ~23 minutes, and contains on average 700,000 events. On top of this, additional calibration information, like AMS electronics parameters, as well as various temperatures, pressures and other health and status data is recorded regularly.

Framed data.
Recorded data are packed into fixed length packets, called frames, and transmitted via the TDRS satellites to the ground at White Sands, NM, and then relayed to NASA Marshall Space Flight Center (MSFC), AL. At MSFC the framed data are archived (under NASA responsibility, for the lifetime of the ISS program) and simultaneously transferred to AMS relay computers at MSFC using the User Datagram Protocol (UDP). From the relay computers, the data is copied into the POCC at CERN and made available via the Network File Service (NFS) to the AMS Science Operation Center (SOC), also at CERN.

Raw data.
Data arriving at the AMS SOC are de-framed and checked for consistency. Then events are indexed and reassembled into the AMS raw files, one file per run. The resulting files are moved to permanent AMS disk storage, backed up through the CERN CASTOR system within the CERN Computing Center and registered in the AMS Oracle database. Database registration provides the following meta-data for every raw file: run time and path of the raw file, the sequence numbers and timestamps of the first and last events of the run, the status and the total number of events in the run, the size and Cyclic Redundancy Code (CRC)

checksum of the raw file.

Reconstructed data.
The format of the reconstructed AMS data is based on the CERN ROOT package. Reconstructed events are grouped together in an event summary file, usually one file per run. For every run the relevant time-based information, such as average geomagnetic cutoff, detector live time, ISS position parameters, temperatures, voltages of AMS electronics, etc. is also written to the event summary file. The timing precision of such data is about 100 milliseconds. In addition, the AMS conditional database, which contains the information of the time dependent properties of the AMS detectors and electronics, like pedestals, gains and detector alignments, is available through a collection of plain text files. The portions of the database which are relevant for a given run are also replicated in the event summary file.

Monte Carlo simulated data.
Monte Carlo simulated events are produced using a dedicated program developed by the collaboration based on the GEANT-4 package. This program simulates electromagnetic and hadronic interactions of particles in the material of the detector and generates detector responses. The digitization of the signals is simulated precisely according to the measured characteristics of the electronics. The response of the trigger logic is fully simulated and digitized. The simulated events then undergo the same reconstruction procedure as used for the data. On top of the usual output data file with the reconstructed events an additional file per simulation job with format equivalent to AMS raw data is produced. These files can be used in subsequent reruns of the simulated data with different versions of AMS offline reconstruction software as needed.

Processing:

First pass data production.
Freshly arrived data are put into first pass production, which runs in a fully automated manner and produces the data summary files for a quick detector performance evaluation. Usually the reconstructed data are available for the analysis within two hours after the raw data are available in the SOC. The data summary files from the first pass production are used subsequently to produce detector calibrations, which are required for second pass production.

Second pass data production.
Second pass production uses all the available calibrations, alignments and ancillary data from the ISS as well as monitoring values (temperatures, pressures, and voltages) to produce datasets ready for physics analysis. Second pass production typically runs every 6 months. Second pass production uses the computing power of both CERN and the remote computing centers (JUROPA at Juelich, Germany; Academia Sinica at Taipei, CNAF at Bologna, Italy, and IN2P3 at Lyon, France). All data produced in remote computing centers is returned to CERN for validation, archiving and distribution to the collaboration.

The full set of AMS data is replicated in the AMS remote centers. This includes the raw data collected from the ISS and the Monte Carlo simulation output and the subsequent output from the second pass production.

Data reproduction.
Consolidation releases of the disparate parts of the evolving software require that, on a 2 to 3-year time scale, the full data set (from the ISS and the simulation) is rerun through second pass production.

Monte Carlo production.
Monte Carlo production is done at both CERN and the AMS remote computing centers, including computer center in Germany (Juelich), two computing centers in the U.S. (Berkeley, Argonne) and computer center in China (Nanjing). All of the simulated data is transferred to CERN, validated, processed and archived. Then it is made available to the collaboration and all of it is replicated in AMS remote centers.

Data Products and Releases:

Raw data is updated constantly, after the validation is done. Data is accessible to the collaboration, including remote access, through CERN EOS file system and archived by the CERN CASTOR tape system.

Reconstructed data from the first pass production is updated constantly, after the validation is done, and typically the reconstructed data is available on CERN EOS file system within two hours after the arrival of the raw data. It is also archived on the CERN CASTOR file system.

Reconstructed ISS data from the second pass production is updated every 6 months; simulated data is updated as available. These data are published on the CERN EOS file system to the collaboration after the validation and integrity checking is done, and archiving on CASTOR is done immediately after the validation.

No confidential or personal information is stored, archived or replicated.

Plan for Serving Data to the Collaboration and Community:

All data is accessible to the entire collaboration, at CERN and in the major AMS remote computing centers in France, Italy, China, Germany, Spain and Taiwan. This is planned through the lifetime of the ISS, the year 2024 and beyond. Such data sharing enables several independent analyses of many physics topics by different research groups to be conducted simultaneously. According to the AMS scientific policy, the research findings are published only when the analysis results are compatible between all (2 to 7) research groups. The results of different group analyses are stored be means of CERN INDICO system, available online for any collaboration member.

AMS publications include the data analyzed and are therefore made available to the public. In addition, all data points shown in the publications are available in a machine-readable form in a refereed journal or as supplemental material, accessible online and referenced in the publications.

The AMS-NASA radiation group collaboration is another example of serving AMS data beyond the high-energy physics community. The NASA radiation group will use the AMS-02 monthly charged cosmic ray fluxes to improve the Galactic Cosmic Ray (GCR) models employed to predict the radiation dose absorbed by astronauts for both ISS operations and long duration missions.

Finally, according to the NASA-DOE agreement, data will be released to the public one year after the termination of the experiment.

Plan for Archiving Data:

Original framed data is backed up at NASA MSFC and will be preserved up to year 2024 and beyond.
All the data is archived within the CERN CASTOR storage facility, the raw data with two tape replicas.
Raw data and second pass reconstructed data is equally replicated at AMS remote computing centers.
Monte Carlo simulated data is backed up in the original production center for a certain period, according to AMS scientific needs and also replicated at AMS remote computing centers.

The AMS Data Management team is in close (daily) contact with the CERN Computing Center and the AMS remote computing centers. As technologies for archiving data evolve, the AMS team will work with each of the computing center teams to ensure that all the AMS remains appropriately archived.

Plan for Making Data Used in Publications Available:

All data points shown in the publications are and will be available in a machine-readable form in a refereed journal supplemental material, accessible online and referenced in the publications. In addition, all data points are available in IN2P3 and ASI Cosmic Ray Databases. We are also planning to make available online the underlying digital research data used to generate the displayed data via the AMS web page .

Responsiveness to SC Statement on Digital Data Management

This data management plan fully follows the SC Statement on Digital Data Management.