Ad-Hoc Task Force on Big Data
of the
NASA Advisory Council Science Committee

Meeting Minutes

Jet Propulsion Laboratory
Pasadena, CA
November 1-3, 2017

Charles P. Holmes

Charles P. Holmes, Chair

Gerald S. Smith, Executive Secretary

Report prepared by Joan M. Zimmermann
Ingenicomm, Inc.

Table of Contents

November 1, 2017

Opening remarks, Introduction

Mr. Gerald Smith, Executive Secretary of the NASA Advisory Council (NAC) Ad-Hoc Task Force on Big Data (BDTF), called the final meeting of the BDTF to order, detailed Federal Advisory Committee Act (FACA) rules on committee proceedings, and introduced Dr. Charles Holmes, Chair of the BDTF. Members introduced themselves around the table.

Dr. Holmes reported having briefed an interim TF report to the NAC Science Committee (SC) in July and submitted four recommendations, which entailed two sessions of meeting discussion. The SC has tabled any actions on these recommendations until the individual discipline science advisory committees can finish reviewing them. Dr. Holmes said he planned to attend the next SC meeting at the end of November to hear the results of any further deliberations, and to deliver the final report of the BDTF.

The goals of the current meeting are to review Jet Propulsion Laboratory (JPL) data science activities, complete the drafts of the focus-topic white papers, hear an update from the Caltech's Infrared Processing and Analysis Center (IPAC), and finish the TF's assessment of the data centers. The draft findings and recommendations to be introduced at this meeting include: creation of a new division in the Science Mission Directorate (SMD) on data science and computing; enhancing or bolstering Data Science and Computing (DS&C) representation on the SMD advisory committees; specific recommendations arising from focus topics; and an assessment of the SMD data programs. Dr. Holmes felt there might be a necessity for a half-day teleconference to clarify some final BDTF products, after the final meeting is held. Dr. Holmes noted that the Space Studies Board (SSB) is meeting this week as well, and will feature an open session on Big Data; JPL's Dan Crichton will attend the session and will provide a summary to the BDTF.

Member Reports

Members reported on their activities since the last meeting. Dr. Ray Walker noted that he and his colleagues had done a very thorough evaluation of every data system of the 12 missions undergoing the most recent Heliophysics Senior Review, including all data management plans from the proposals. Dr. Walker felt strongly that the BDTF should formalize an additional recommendation that NASA create a separate Senior Review activity to evaluate data management plans for the missions.

Dr. Reta Beebe said she'd been working on the Planetary Data System's new PDS4 version, which is based on XML language and can do a tremendous job accessing metadata. During the Cassini mission, it became evident that PDS3 is inferior for this purpose. She had also been working with JPL to gather specifications and create guidebooks on how to use mission data, including reduction algorithms so that future users can properly utilize these data. Her group had also developed Web pages on how to access Cassini data, and there is now a concentrated effort under way to revamp these pages and accommodate new PDS4 advantages for mining metadata. She was planning to interact further with Cassini team members on dissemination of PDS4 particulars, and felt this effort was going very well.

Dr. Holmes noted that he had arrived a day early and visited with IPAC and briefed JPL's Director and Deputy Director about the BDTF. Dr. Holmes was pleased that the JPL leadership expressed their enthusiasm for BDTF's activities and its accomplishments.

Dr. Neal Hurlburt described having attended a recent Astrophysical Data Conference, where he heard good discussions on server-side analytics, machine learning, data provenance and data calibration, PDS examples and evolution of bits, Jupyter notebooks, etc., indicating good agreement with BDTF's conclusions.

Dr. James Kinter noted that since the last meeting, he had reached out to people in the modeling workflow community, and had received a good response from Lars Bildsten of the Kavli Institute for Theoretical Physics, who provided updates on a trend in the Astrophysics modeling community, wherein numerous users were adapting a new tool called MESA (Modules for Experiments in Stellar Astrophysics). Bildsten felt the biggest hurdle was finding the right people to do the work (salary issues), and that it was important to have resources for compute and data storage to bring large data sets together. Dr. Kinter told a personal anecdote by way of example; he recently obtained a National Science Foundation (NSF) grant to acquire a large computer, with 6 PB of spinning disk, representing a substantial resource being shared across the university. He was still trying to get data onto the system, noting it is taking many more months than he had originally anticipated to get data onto the computer. Dr. Kinter affirmed that other groups are having similar problems transmitting large data sets, notably the National Center for Atmospheric Research (NCAR), which recently stood up a big D\data processing facility at the Wyoming center to host climate change simulations. They're trying to assemble all the data in one place so it can be analyzed, and latency has proven to be a big issue for them too.

Dr. Kinter also attended a conference in France on computing in the atmospheric sciences (which included major vendors— Cray, etc.), where several presentations emphasized the need for data storage and memory to be associated with manipulations of big data sets. He noted lastly that the National Science Foundation (NSF) is in the middle of a solicitation for a new flagship supercomputing capability.

JPL Overview

JPL Data Science Programs

Mr. Dan Crichton briefed the BDTF on JPL's efforts to create a more focused data science program, using data across many different areas, including engineering. He introduced Dr. Richard Doyle, Program Manager for Information and Data Science and for High Performance Spaceflight Computing. Mr. Crichton noted that the briefing would focus on how to expand the present data archives to make them more usable. Industry has been investing heavily in making data more useful, and JPL is leveraging ongoing efforts in machine learning. NASA archives must also prepare for the large increases in Earth Science data volumes that will result from future assets.

As a Federally Funded Research and Development Corporation (FFRDC), JPL can take advantage of technology development outside of NASA (NSF, e.g.) and infuse it back in to the Agency. JPL is able to take advantage of a lot of work in biology, defense and intelligence, and medicine, and is working on methodology transfer among the disciplines. A major question is how the Agency can evolve and create mechanisms to improve use and analysis of big data in the archives. Systems must be optimized for capturing data; the ultimate vision for big data is to connect the entire environment together, from capture to extraction and use. JPL is focusing on doing data science much more rapidly, building databases, making them searchable, building up data analytics, and cleaning up metadata.

Dr. Doyle described the guiding principles for data science at JPL, based on recurring challenges: the data lifecycle. There needs to be an awareness of the full end-to-end system flight system, to ground system, to archive. From a big data perspective, one needs to ask questions well before the data arrive at the archive. There is also the data ecosystem: no agency or institution owns all the relevant data. The data ecosystem is the reality check that the data is highly distributed, and therefore all solutions must be to get all the data to play together. Finally, there must be cross-cutting solutions that work in all the disciplines.

Within the data lifecycle, science event detection and response is critical. It is important to send alerts in the data package as quickly as possible, with appropriate support. Examples of such events are detections of dust devil on Mars, or wildfires on Earth. The strategy must include how to decide on which data should stay in the buffer. Users want to be able to flip between exploratory mode (such as visualization techniques), and analytic mode.

Mr. Crichton observed that archives are moving to collaborative environments. Internationalization is also a component, helping to build powerful models in all the disciplines. These methodologies can be used to start getting insights from data. This also requires the development of data-sharing infrastructure, "Big Data" infrastructure, intelligent data algorithms with common data elements and models, and evolution toward data analytics.

Dr. Doyle pointed out that the data science discipline is beginning to see more commercial providers, such as those for drones and self-driving cars, so NASA needs to track progress in the commercial sector and decide how it integrates into Agency strategy. The data science growth strategy at JPL includes a Data Science Working Group that is considering a set of use cases. The working group has brought in subject matter experts (SMEs) from engineering, formulation, human resources, and other business offices at NASA. JPL has launched pilot projects across those areas, and is beginning to look at the data challenges—two major ones that popped up are search analytics and machine learning. The group is starting to look at projects and trade analyses. JPL might want to pursue NASA-specific capabilities in machine learning for imagery, as one example. Partnering with universities, international partners, and commercial and open source providers (e.g., Amazon web services) will also be key.

Mr. Crichton said one JPL strategy was to apply data technologies across the ground environment to do better data discovery, and to better leverage PDS4 in particular. Dr. Holmes asked if there had been any success in bringing these new methodologies back to JPL. Mr. Crichton indicated that, yes, it just takes a little money to provide to the emerging participants who are interacting with data in new ways, driven by science. This can be seen in the Large Synoptic Survey Telescope (LSST) project, in how transient detection is being done, using algorithms to make classifications in real time.

The working group has two major recommendations:

- Use the mission science data life cycle as a way to organize a vision for data and computing, by partnering across SMD and with other agencies to explore opportunities for methodology transfer.
- Evolve to support use and data analytics for the community; this will drive the development of broad ecosystems and increase the use of data-driven approaches.

Dr. Kinter asked about the real utility of the commercial cloud (Cloud), as it seems to be everything, or nothing. Mr. Crichton felt there was an important role for the Cloud in some areas, and that NASA should use it where it makes sense. Dr. Doyle added that the Cloud needs to be part of the strategy, but it can't be used to push analytics to platforms. Its use should be considered, however, in setting up strategy properly.

Caltech Center for Data-Driven Discovery

Dr. George Djorgovski presented activities at the Caltech Center for Data-Driven Discovery. The Center supports the entire campus, and assists faculty in the formulation and execution of data-intensive projects. The Center was founded in the 1990s, with the advent of digital sky surveys, and the creation of a virtual observatory framework. It is a successful example of collaborative action between NASA and NSF. Today, every bit of data on Astrophysics is now available to anyone with an Internet connection. It is also recognized that this is a universal methodology; the question is how to more efficiently share expertise. Dr. Djorgovski noted that EarthCube is receiving significant support from Caltech through its contribution to software architecture for the Earth sciences.

The exploration of parameter space is the central problem of data science: clustering, classification, correlation and outlier searches. Challenges include algorithms and data incompleteness. Discovery takes place in high-dimensional parameter spaces: in some corner of the data space, there is something more than noise. Tools that pluck meaningful signals from noise will be important for disciplines like precision medicine. Tools that deal with the problem of data heterogeneity, in one case, have led to the discovery of supermassive black hole binaries.

LSST expects to see 10 million transient events per night. Follow-up will require a way to automatically classify transients effectively, on the basis of incomplete data. Bayesian networks and other techniques will be needed here. Optimizing feature selections, such as distinguishing the single star Lyra from a binary star system, involves machine-assisted discovery using Eureqa, which employs a symbolic regression method to classify binary stars. Automating the optimal follow-up method is also important. There are only small numbers of hours allotted for observatory time, and observers must learn how to efficiently use that time.

Data visualization is another potential tool: how to visualize complexity, and how to see beyond three dimensions. Caltech and JPL are using virtual reality techniques to allow users to interact within the data space; this effort has resulted in a start-up company that sprang out of the joint data center.

Caltech is working on methodology transfer with the medical community, which has resulted in an early detection research network (EDRN) software architecture for identifying biomarkers for cancer. Center work in real-time classification and response is being leveraged in seismology, as in a project that uses cell phones as a sensor network to detect seismic waves, and send out alerts. Sky survey tools are now being used in neurobiology, the result of which has produced the best clinical test for autism ever developed. Quantifying uncertainty, whether the data come from measurements or from model/simulation outputs, is starting to show progress, as well.

Caltech is also looking to train the next generation. To that end, Caltech held its first virtual summer school, with 25,000 students worldwide taking advantage of the course. This was a surprisingly large response. Dr. Holmes asked if Caltech had been interacting with an Ames Research Center effort on finding near-Earth Objects (NEOs). Dr. Djorgovski indicated that it had not, but that the Center would be open to the idea.

Dr. Kinter commended Dr. Djorgovski on a spectacular talk, and asked if there were a way to obtain more basic information on Caltech's efforts. Dr. Djorgovski took an action to provide BDTF with some summary papers, and an overview paper. Asked if there were a way to use machine learning to allow feedback to be passed from algorithm to algorithm in immersive virtual environments, Dr. Djorgovski said that if there's a human in the loop, yes, one can transfer information between algorithms in a modeling environment. He added that virtual reality and manipulation of three-dimensional data seem to trigger human pattern recognition pathways that a flat screen can't elicit; he felt this would become a very important way to process data. Dr. Holmes mentioned that BDTF member Dr. Eric Feigelson was very active in this community, and hoped to get him more involved in Caltech's efforts.

JPL Machine Learning/Data Science Methods and Applications

Earth Science

Dr. Lukas Mandrake presented a briefing on machine learning at JPL, which was inclusive of the data science life cycle, from formulation to archive. One can use machine learning to predict what science will result from a particular data set, which can help in calibrating instruments and assessing the quality of the data, ultimately using the data to help find what you're looking for, compressing it down into knowledge. Machine learning comes down to an autonomous loop; the human has control of this loop, helping decide whether the machine is seeing "truth." Machine learning is a way of explaining a phenomenon to a computer without using code: essentially giving a computer every case of data you want it to look at; this produces nonhuman written code. These are algorithms that inductively self-assemble from examples. Machine learning allows you to start asking: where should I start looking? What is likely to happen next? Show me the most interesting first. What inputs are most informative?

Dr. Mandrake presented an example of machine learning (HELM) that led to holographic life detection, which involves the use of digital holographic microscopes. These microscopes take an entire volume of liquid and turn it into a raw hologram. Life detection can be done on an instrument in real time. The tool can decide if the liquid contains life. In the future, holographic life detection will be able to identify species of bacteria. The HELM machine system looks for extremophiles on Earth, contamination in water sources, septicemia in blood, and analyzes ocean water for small single-celled organisms. (NASA's Astrobiology program did not provide support for this effort).

AVIRIS uses hyperspectral imagery from air-borne assets and was able to train the images to look for methane, while automatically correcting for atmospheric parameters. AVIRIS was flown over the sparsely populated Four Corners area in the US Southwest, and found a significant methane leak in a pipe that ran many feet underground. This discovery saved the pipeline owner millions of dollars, while helping the climate. AVIRIS has found multiple leaks.

Advances in quality/uncertainty estimation have benefitted the Orbiting Carbon Observatory (OCO-2), improving the accuracy of parts-per-million estimations of carbon dioxide concentrations. The Current OCO-2 Quality Estimation Product received a runner-up award for NASA Software of the Year. The product resulted in a spectrum of trusted data that identifies artifacts and provides sanity checks.

Orbital spectral analysis and super-pixel segmentation has been used for improving mineralogy maps, identifying crop types, recognizing diseased citrus crops, and estimating hurricane damage. The technique reduces the amount of bandwidth needed for data transfer, and can feed data to other assets. It is also used in spectral minority targets, and has proven useful for detecting sulfur from biosignature analog sites for Europa, in detection and tracking from orbit.

JPL has suggestions on how to supercharge data science at NASA: provide minor funding for seedling concepts, which then lead to demonstration and validation of a system. A mission then adopts the validated system, which is subsequently used on multiple missions and becomes heritage. If a mission Announcement of Opportunity (AO) says it is interested in data learning applications, the practice can dispel the "NASA hates it" myth. A small pot of money can be set aside for validation of data science and machine learning (DS/ML), and for supporting shared repositories for past DS/ML datasets. Dr. Holmes said the presentation melded very well with BDTF recommendations, as well as the Hurlburt/Feigelson BDTF white paper on data science methodologies. Dr. Beebe agreed that the AO should specify expertise, as the review committee needs to be formulated based on the requirements of the AO; this gets the right SMEs on the review panel.

Astronomy

Dr. Umaa Rebbapragada presented a briefing on advances that have resulted from JPL ML collaborations with the Palomar Transient Facility (PTF), Very Long Baseline Array (VLBA), Variables and Slow Transients survey, and MIT/Lincoln Laboratories work on the Space Surveillance Telescope. Machine learning has been hugely influential in advancing astronomical science. She noted that Big Data challenges will be coming up with the advent of LSST, the James Web Space Telescope (JWST), the Wide Field Infrared Survey Telescope (WFIRST), and the Transiting Exoplanet Survey Satellite (TESS). Data volumes will range from terabytes (TB)/night to TB/second. Huge catalogs are being collected from a number of all-sky surveys. Millions of detections per night were the norm in 2015 for PTF. The science team is a constraining factor in the era of increasing data volume. Manpower to analyze observations is limited, therefore it will be important to be able to automatically prioritize detections. For PTF, a machine learning algorithm used classifiers to filter false detections, asking the question: are these candidates optical or pipeline artifacts? These data-intensive sky surveys would be unusable without machine learning techniques.

Transient science requires real-time algorithms and real-time filtering to find the target early, filter out all irrelevant observations, and trigger rapid follow-up of scientifically rich targets. Typically, a machine learning (ML) algorithm will use a set decision threshold (e.g. 99%). Real-time filtering was used to find fast transients (pulsars) in the V-FASTR/VLBA surveys, and was considered to be very successful in automating and prioritizing observations. In other cases, real-time classification is used as a means of refinement as observations are being collected; science users can then contribute filters to downselect the types of objects they wish to examine.

In catalog science, or data mining of large catalogs, machine learning applications are commonly in stellar classification, star/galaxy separation, planetary transits, NEOs, and estimating cosmological parameters. Currently JPL is mining astronomical archives for weak lensing/strong lensing imagery (no spectroscopic images, however).

New and emerging techniques for astronomy include crowdsourcing and machine learning, such as those used to involve the efforts of amateur astronomers on the exoplanet identification task on NASA's citizen science website, Zooniverse. JPL has also used machine learning to help respond to instrument and survey changes, illustrated by how a pipeline upgrade changed data characteristics at PTF. Machine learning responded to unreported software changes in a pipeline, in this latter instance. Domain adaptation is another technique that computes mapping between source and target data sources that share common science goals.

Deep learning is starting to be used in astronomy; it has had a lot of success in the image domain by creating standardized ways of transforming data sets that can help capture contrast changes; this works natively in the raw format. It has also been found to help reduce error on image data sets, beyond human ability in some cases.

The challenge to machine learning is mainly that space telescope managers don't know that these techniques exist. Generally, machine learning proponents have to fight for funding in ROSES. The other challenge is cultural; there's an entrenched attitude that "it's just software," or "my post-doc can do it." As a result, data architectures are not as sophisticated as they could be.

Planetary Science

Dr. Kiri Wagstaff presented machine learning applications in planetary science (PS), centered primarily on content-based image classification, onboard data analysis and discovery, onboard data acquisition, and information extraction.

The NASA Planetary Data System (PDS) already houses 23 million images. The goal of machine learning in PS is to enable search by content, without pointing, through a combination of statistical analysis and deep learning classification. This enables a user to search by "crater," "dune," etc., and enables searches for particular features. This application enabled the discovery of evolving wheel damage on the Mars Curiosity rover.

Examples of onboard science included a program to facilitate dust devil detection on Mars by preserving key frames, which consumes no data volume when no detections are under way. Onboard science is used also in thermal anomaly detection, polar cap tracking on Earth, and will be used in future applications for the Europa Clipper mission. For Curiosity, AEGIS, or the Autonomous Exploration for Gathering Increased Science, has been in use since May 2016, and is now considered standard operating procedure for the rover. AEGIS enables the rover to find the most interesting rock and collect spectra, which are then sent to Earth. Automatic pointing refinement is being developed for the Mars 2020 rover.

Machine learning is being used to extract information from planetary science observations. Machine learning have been used to read published papers to find compositional relationships for Mars targets, in one case: Which rocks on Mars contain hematite, magnetite? Automation takes the task from 30 minutes/document (human) to 5 seconds/document. JPL is in the process of integrating this technique with the Mars Science Laboratory (MSL) Analyst's Notebook. In order to increase support for machine learning and planetary science, Dr. Wagstaff recommended more deliberate support for machine learning plus planetary collaboration, particularly for downstream spacecraft infusion.

Q&A with Machine Learning Presenters

Dr. Walker resonated with the “it’s just software” statement, and pondered how to show the need for this expertise. He cited his experience with a conference in which a supposed practitioner could not simply explain machine learning (ML), indicating a cavalier attitude toward understanding. Dr. Rebbapragada commented that in ML, the training is not about the science, it’s about increasing a metric. In the long term, NASA needs more data scientists. In any proposal that has ML, the proposer should be able to write why it is needed. Dr. Walker agreed that each ML-related proposal needs those who are knowledgeable about ML to be in on the process. Other than having an expert in the loop, it isn’t clear how to solve the problem. Dr. Hurlburt noted a past rule that required an author to send a paper to a PI before it was published. Dr. Guhathakurta commented that scientists can’t look at this problem alone; it is an interdisciplinary issue, and bringing domain experts together will create a more capable generation.

Dr. Chris Mentzel noted that industry is using ML, and asked how NASA used tools that are not built in-house. Drs. Mandrake and Wagstaff explained that they leverage standards, and use a two-way open source practice, which benefits everyone. Dr. Rebbapragada said there is no one company that is dominating in ML, and that JPL has many PhD students in ML, so they are able to maintain ties to the discipline through conferences and literature. She felt there would be a convergence as these techniques are pushed into basic science. Dr. Mandrake said the ingredients of the “magic” is statistics, which is very robust and old, but it is still necessary to understand how it’s done. Dr. Wagstaff noted that one particular ML technique retrains data nightly in one case; the data itself changes over time, and so the system is getting feedback all the time. Dr. Kinter asked how overfitting was avoided. Dr. Wagstaff replied that methodology and cross-validation were applied. Dr. Rebbapragada said that in astronomy, some domain knowledge is needed to avoid overfitting. Avoiding overly biased samples is another part of it. Dr. Mandrake agreed that data must be thoroughly interrogated.

Public comment period

Dr. Graham Macintosh, citing his experience working with IBM’s AI and Deep Learning team, commented that even a “black-box” system that is hard to explicate can still function as an excellent hypothesis generator. For example, a system that identifies dark slope streaks on Mars can be queried, or “back-propagated,” at specific “neurons” to yield one of the features the system is looking for when it searches for streaks. In one case, the feature of note was a specific kind of rock, a result that could not otherwise have been anticipated. Serendipity can generate new lines of investigation. Similar AI systems that look for fraud and money laundering patterns have also yielded some surprising answers. Dr. Wagstaff added that this ability to provide explanation, in the discovery case, is valuable even when we don’t know what we’re looking for. You have to be able to say: why is it interesting (human-digestible labels for novelty). JPL’s machine learning groups are working hard on this aspect.

Big Data Analytics and Visualization

Sea Level Rise

Mr. Thomas Huang presented a briefing on NASA's Sea Level Portal website (sealevel.nasa.gov). JPL has been analyzing how users are using the data and how the user finds articles, and is identifying popular media outlines, and new and returning users. The Portal gets over 300,000 monthly page views, and is cited regularly by top technology news outlets. The Portal also provides tools to analyze sea level on the fly: tools include visualization, time series, scatter plot, maps of hydrological basins, etc.

The impact of big data is mainly in driving needs to scale computational and data infrastructures, and to support new methods for deriving science inferences. For NASA centers, downloading to local machines is becoming inefficient. Today, one must go beyond traditional data analysis. To address this, NASA created the NEXUS: Scalable Data Analytic Solution. NEXUS breaks files into small chunks, which are then infused into fast lookup search engines. NEXUS performance was tested with sample using Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua Daily data: the current Giovanni system took 20 minutes to process the data, while NEXUS took 2 seconds. Mr. Huang showed a sample analysis of an ocean anomaly, "The Blob," wherein the system can replay the anomaly and visualize it with other measurements, document and publish the anomaly, compute daily differences, and show "winners and losers" associated with the anomaly (e.g., whales).

JPL is developing information discovery solutions to provide the capability to match up *in situ* with satellites. NASA's Advanced Information Systems Technology (AIST) Oceanworks is establishing an Integrated Data Analytic Center, which focuses on technology integration, advancement, maturity, and deployment automation. The Science Data Analytic Platform is an implementation of Oceanworks, and has been donated to an Apache incubator; NASA is currently engaging the community in working with the Apache incubator. In summary, NASA needs to think beyond the archive, connect information to enable discovery, and push it as a community-driven solution. Data centers need to be in the business of enabling science. Mr. Huang hoped to see more investment in data and computational sciences, which requires a concerted effort between the science teams and the analytics teams.

Planetary Science

Ms. Emily Law and Mr. Shan Malhotra presented aspects of planetary science analytics at JPL. Ms. Law began by citing how big data can be leveraged to improve user experience and outcome. Interactive visualization and analytics are critical to this effort. The path forward is to increase the focus on data usability, invest and research in these areas, improve and scale up data usability to meet expectations, and partner with industry, as NASA cannot do this on its own. As examples, JPL has developed the Solar System Treks Projects, sponsored by both SMD and the Human Exploration and Operations Mission Directorate (HEOMD), but also while working closely with NASA's Solar System Exploration Research Virtual Institute (SSERVI). The Treks project is a set of web-based interactive portals for mission planning, scientific research, and public outreach, and it includes visualization and analytic tools, and data products from many past and current missions. Data-access Application Programming Interfaces (APIs) include virtual reality goggles and planetarium programs. The Treks can be used to explore Mars, Vesta, and other bodies of interest inside the Solar System. Treks in work include Titan, Icy Moons, Phobos, Comet CG, and Ceres. Treks face Big Data challenges with ever-increasing data volume (velocity), the usability of large volumes, and the variety of data sets. The Treks address big data challenges by adhering to general principles that govern architecture, data and systems. The Treks approach is applicable to other domains (e.g., Water Trek, Earth Trek).

Treks feature the ability to browse data products, and search and download Analysis tools include lighting, measurement, and sun and angle parameters. Visualization tools are also provided, such as 3D flyover and overlay; collaboration; a 3D print application; and the ability to compare data from past and current missions and various instruments. Users include missions, lunar scientists, teachers and students, and the general public. Mr. Malhotra displayed a video of a Solar System Trek, demonstrating how a high school student could generate lunar light maps based on Lunar Reconnaissance Orbiter (LRO) data, with the same fidelity a scientist would need. He demonstrated numerous tools such as crater detection, and "lensing," all of which use data from the Apollo through Clementine through the LRO mission: all the data is co-registered.

<u>Hydrology</u>
Dr. Jay Famiglietti presented a use-case driven application, which provided graphic illustration of regions in California during a drought as they dried, from June 2002 to June 2014 (using data from the NASA Grace mission time period, combined with USGS data). The application provides a lot of information on ground water, and can also show trends in freshwater availability in the US, as well as global aquifer depletion rates. JPL is now integrating the "water satellites" (TRMM, SMAP, GRACE-FO) to get a more holistic picture of the water cycle on Earth. The Western States Water Mission integrates key satellite data and links to models of agriculture, food production, climate, and ecology. The hope is to scale up to a Global Water Mission as a water "moon shot," which will enable users to select hydrologic regions of interest all over the world, and put tools into the hands of the community, and into accessible usable format for decision makers. Mr. Molhatra displayed video of Water Trek soil-moisture data from model outputs, at 1.7km pixel resolution. The video showed features such as transpiration at given loci, and changes in water table depth. JPL is in the process of building a device wherein a user can go out with a cell phone and query a site against the data, and correlate it with data coming from the Water Trek data set. Water Trek can pull up the flow rate on any river segment in cubic meters per second, and can also plot this data over time to produce a hydrograph. Synthetic Aperture Radar (SAR) data can be processed for subsidence data, which is also correlatable with well locations, and changes in *z* height (obtained from GPS sensors). Mr. Crichton noted that JPL had funded this effort, a good example of how to approach data science from a use case.

Q&A Session

Dr. Mentzel asked about the makeup of the teams: who are they and what are their roles? Dr. May said that sets of scientists were typically matched with team capabilities. Mr. Huang noted that the Sea Level Portal was funded through the SMD Research and Analysis (R&A) program; in this instance, the JPL team interacted with scientists to assess whether the right tool was being built, which often entailed travel to relevant sites (such as Greenland). All work was done closely with the principal investigators, and the many ocean scientists at JPL. Dr. May said the procedure was basically to get the requirements from the scientists, and get feedback (on the tool or app) through SSERVI. Dr. Famiglietti noted his project had had weekly meetings with all the players, in a rigorous flight project framework. Dr. Kinter asked of Mr. Huang: in terms of chunking information, how specialized was the data restructuring in the NEXUS improvement over Giovanni? Mr. Huang noted two ways to speed up data processing: hardware, or a change in algorithm. With NEXUS, JPL did a little of both, and used partitioning of data. Software is open source; there is no secret sauce. It was more a matter of ingestion adaptation. Dr. Walker said he was very impressed with the flyover and overlay imagery in the Treks. He urged the team to document how these apps are being done and asked if they had thought about the issues of archiving the software. Mr. Huang said he had considered it: with the virtual machine (VM), one can take a snapshot of the entire machine. He was promoting archiving the VM, because compilers and operating systems will go away. Mr. Molhatra said that all flight-critical ground data was being moved to VMs for that reason. Data-level security also must be performed on some data, to have the ability to keep some data private to JPL. Dr. Kinter asked Dr. Famiglietti how the Western water effort was different from the national effort in Tuscaloosa. Dr. Famiglietti said there were components of the model that are the same. One major distinction is the incorporation of ground water and water management data.

Dr. Holmes commented that the JPL work sounds like a continuation of classic model: get smart people together to try something out. He asked that as BDTF is trying to push more science utility to the user, was there a good path forward to do this? Mr. Huang said that when he talks to Amazon Web Services (AWS), he has asked if he could throw an algorithm into the mix: it's inexpensive, it can be done, and can propel a move to an integrated data center concept. Dr. Holmes remarked that "bringing the code to the data" might need a universal charter. Mr. Crichton felt that one thing that is happening is that innovators are building data platforms and presenting them in different ways. The power is in bringing the computation and data together. He agreed that there needed to be a shift forward to bring the computation to the data. Mr. Molhatra observed that there are also OGC standards for data access, which gets us one step closer to enabling pushing algorithms to the platform. Mr. Huang thought the science community was beginning to see the value of not having to download enormous amounts of data to manipulate. Dr. Holmes asked the team to individually write down a few thoughts that he could incorporate into a TF white paper, with respect to moving these concepts forward.

<u>Discussion of Study Reports</u>
Members discussed progress on individual studies.

Dr. Holmes commented on the server-side analytics paper, noting that it seems to be a common notion at NASA that this is where it needs to go, but it remains unclear how to move it forward. How to get into large complex data systems is the issue. At Goddard Space Flight Center (GSFC), there is focused work on planetology. There is also some older work in Earth Science being used at NOAA. There are examples from the Solar Dynamic Observatory (SDO), and now some examples from astronomy, and digital sky surveys. The essence of a recommendation would be that SMD should do more to push these concepts through the ROSES competitions, conferences, and workshops, impelled by top-down guidance from SMD. Peer groups can be consulted to provide priorities.

Dr. Beebe, referencing how to show the diversity available in SMD, stressed that one solution does not fit all. The real deficiencies are in metadata; thus, NASA needs a systematic way to check that metadata is being done properly. The topic paper recommends that NASA reserve money to do a detailed data review at the ends of missions and produce guidebooks for data systems and instruments. This exercise must be done at the point just before the first Senior Review. Dr. Mentzel commented that it's more than integration—the value of the data after the missions are done is well worth a $30K investment; this strategic move can give the mission a whole new life.

Dr. Hurlburt reviewed the methodology topic paper, describing how advances over the decades have not been systematically incorporated at NASA. Data science must be embedded more rigorously into NASA missions. Each of the subdisciplines has slightly different needs. Opening the data in Heliophysics is one solution for enabling broad science. Deep learning algorithms and NASA data must also be combined. The paper also addresses the skill deficit in the community; NASA must make discipline scientists more cognizant of data science methods, through statistical summer schools, and more engagement with the NSF Hubs and Spokes program. The main recommendation is that NASA needs some procedures to embed this awareness in the community; someone in ROSES to oversee this effort. Dr. Beebe commented on the importance of workshops: JPL has had them for years, and they have proven effective for getting younger researchers to network and prepare for their futures.

Dr. Kinter discussed the paper on methods of modeling workflows, which haven't changed since the 1950s or 60s. He was trying to characterize specific problems in modeling, and make some tactical recommendations, such as for investments in hardware and software to support modeling workflows. Social issues include the salaries obtainable in the private sector (versus government/academia), and sustainability. There is also a cultural divide between scientists and IT developers; it will be necessary to convince people through demonstrations. As an example, it takes 6-7 years to get a new Earth system model out; ideally, we want to collapse that by an order of magnitude. Dr. Walker commented that older, experienced modelers are not generally happy with new notions, and will be hard to convince.

November 2, 2017

Mr. Gerald Smith opened the meeting and made administrative announcements. He then presented a number of HQ-commissioned mission art prints to each of the BDTF members to thank them for their service.

Carryover Items

Dr. Holmes reviewed the agenda for the day.

Assessment of Data Archives

Dr. Hurlburt presented the most recent draft of BDTF's assessment of NASA's data archives, and culled through the three major questions: how are the archives planning for the future; what do they want to stop doing; what steps are they taking to make their data interoperable with allied data set from other sites inside and outside of NASA? After having assessed a wide range of data, it appears that most archives rely on the Senior Review to help shape their visions of the future, and all archives seem to have some form of community feedback. Dr. Holmes asked if Headquarters regularly reviewed the Earth Science data community. Dr. Lee said there had been no formal review for some time, although the Earth Observing System Data and Information System (EOSDIS) does meet twice a year to receive community input; however, the meeting is not a programmatic assessment. Dr. Holmes felt this uncertainty about the Earth Science process could be added as a footnote to the paper.

Dr. Hurlburt said that another major point is that everything is changing rapidly. Old data is always an issue; NASA should get out of the business of storing data. Most archives are trying to move to interoperability, and virtual observatory models. Different divisions have different requirements. The VO model works reasonably well, although a big problem remains in finding a better way to get at data, and having a more uniform approach. Data must be reliable. A potential recommendation to NASA would be for the Agency to encourage the use of uniform, human and computationally readable data, and metadata description standards and protocols. Dr. Holmes felt BDTF might want to consider a separate recommendation that can be tied to recommendations from other papers, without duplicating them. Dr. Feigelson commented from an Astrophysics perspective, asking whether the archives would be functioning well in a decade, and would they open to moving to the Cloud? Should all the archives move to the Cloud? Dr. Hurlburt said the issue is complicated, but felt it had been addressed adequately in the assessment. Dr. Walker noted that the Cloud is a moving target; its use will require constant filtering back of the progress to the community so that they can make good decisions. Dr. Feigelson thought that the NASA-wide people seemed to pushing the Cloud, and asked whether BDTF wanted to comment on this trend. He felt the question was more about how much communication there is between JPL and Headquarters in terms of what is being done in data science. Dr. Doyle thought the challenge is looking at cross-cutting efforts in the NASA organizational structure.

Frontier Development Labs

Dr. Lika Guhathakurta presented a briefing on the activities of the NASA Frontier Development Lab (FDL), an applied artificial intelligence research accelerator and public/private partnership between NASA Ames Research Center (ARC) and the SETI Institute. Dr. Guhathakurta felt that FDL's utilization of exceedingly valuable NASA data was well worth the investment, given NASA's multibillion investments in missions.

FDL came about as a unique idea in 2016, originating with NASA's Jim Adams, to apply Artificial Intelligence (AI)-type tools to planetary defense; a partnership with ARC followed. To develop these applications, students attend an intense eight-week long course and tackle subjects important to both NASA and to humanity's future. FDL includes high-level partnerships with Intel, NVidia, IBM, and Lockheed Martin, among others. SETI functions to enable the public/private partnership. The FDL team is interdisciplinary, comprised equally of data scientists and space scientists. The latest effort tackled six topics, including planetary defense, long-period comets, solar storm prediction and solar terrestrial interaction, lunar volatiles, and AI applications in the space sciences. Some of the "magic" seen in AI solutions arises from the adjacent nature of the problem domains, which allows overlap of expertise and talent. FDL also benefits from the vast GPU compute power provided by the private sector. IBM's Executive Project Manager is engaged with FDL in providing compute resources for the team. FDL benefits also from the numerous experts who are at hand in nearby Silicon Valley.

Results from some of these exercises include improvements in solar storm prediction. In this case, the FDL team developed a neural network, called FlareNet, that connected solar ultraviolet images taken by the SDO, with the forecasts of maximum x-ray emissions, and developed a technique that has the potential to improve both the reliability and accuracy of solar flare predictions. The result illustrates that the application of AI tools to the "black box" of magnetohydrodynamics can lead to real progress. Solar flares present a significant threat to communications, Earth-orbiting satellites, and the terrestrial power grid. Because these flares travel at the speed of light, they can cause damage with little to no warning. Applying AI to the prediction of solar flares, particularly X-class solar flares, can reduce risk by providing some predictive capability to flare classification.

Deep learning (DL) has revolutionized image classification in general, and has dramatically improved the accuracy of image identification. DL has superseded the human approach. To achieve breakthroughs with DL, one must first prepare datasets and build the scientific process through software; this sometimes provides new physical insights. In the FlareNet study, DL was used to connect SDO images with flare strength, and several convolutional layers allowed the neural network to recognize features of increased complexity.

FlareNet's first goal was to see if the network could connect solar images with flare x-ray amplitude (training data). Only pre-2015 flares were used for training. At present, FlareNet's neural network seems able to generalize for weak flares (C-class), but it does not perform as well in predicting stronger flares. The current incarnation of FlareNet underperforms because it needs a more robust (larger *n*) data set. Dr. Kinter asked for a definition of the quality metric. Dr. Macintosh answered that the metric was the aggregate Euclidian distance between predicted and observed. Dr. Guhathakurta pointed out although FlareNet is not a perfect tool, these results were produced after a mere 8 weeks of student effort.

FDL also built STING, which discovered the imprint of the magnetospheric ring current (Kp Index) in precursors of geomagnetic storms. In this case, machine learning methods were able to extract important physical parameters without *a priori* knowledge of the system. Similar techniques have also improved radar determinations of NEOs.
Dr. Feigelson commented that SMD appeared to be sparsely involved in FDL, and should contribute more. Dr. Guhathakurta agreed with this assessment.

In closing, Dr. Guhathakurta noted that application-oriented AI can provide important benefits to NASA. Moreover, there is a strong incentive for the private sector to participate in AI advances, providing a clear risk- and cost-reduction benefit to the human exploration space community.

Dr. Lee commented that AI has been showing great promise, while at the same time, the effort represents just the beginning of the discipline. He said he had just returned from a review panel of many proposals using AI techniques, where he heard the most common criticism to be an objection to the error metric not being quantified; NASA must address this concern in its existing programs. Dr. Walker agreed that AI and machine learning systems need better confidence levels.

Dr. Kinter said he frequently hears about AI's potential to yield new insight, but cautioned against these techniques being used to "rediscover the obvious." Are there standard tools that are first applied that can identify and remove known relationships? Dr. Guhathakurta agreed that this needs to happen, and it's what the scientists are bringing into the discussion. AI is not a solution, but an approach. Dr. Doyle felt that BDTF might want to comment on the explanability aspects of AI. Dr. Mentzel commented that he was not sure NASA was the right Agency to invest in explanability, and that academia would be more appropriate. BDTF could, however, approach this concern in its methodology white paper. Dr. Feigelson felt that BDTF could recommend further education in this area; teaching it from the ground up, linking the most elementary components of machine learning to the most advanced.

<u>Big Data at IPAC</u>
Dr. David Imel presented an update on IPAC in the context of Big Data. IPAC, a Caltech Astrophysics Science Center, focuses on cosmology, exoplanets, asteroids and the solar system, and infrared submillimeter Astrophysics. The Center puts forth a large public outreach effort, and boasts a vibrant research environment. For 32 years, IPAC has supported science operations for NASA missions, including the Wide-field Infrared Survey Explorer (WISE) and NEOWISE. IPAC is also the US data center for NASA/European Space Agency (ESA) partnerships, and provides science operations and

archives for ground-based observatories for missions such as the Two Micron All-Sky Survey (2MASS) and PTF/ZTF (PTF is transitioning to the Zwicky Transient Facility).

Since the last meeting of BDTF, Spitzer discovered the TRAPPIST exoplanet system, a discovery that was the subject of a Google doodle and which made the front page of the New York Times. There was also a discovery of the first gravitational radiation signal that was coincident with an electromagnetic signal, and the recognition of debris disks as good future targets for JWST. IPAC's NASA Extragalactic Database, or NED, is considered the "Google for Galaxies." NED is building the census of the Universe, and is approaching a billion sources. NED is also a data discovery engine, containing original sources for objects of interest. The NASA/IPAC Infrared Science Archive (IRSA) is used for original Astrophysics research and can cross-correlate data from many missions. Catalogs, images and spectra from many missions are regularly used to enable new science discoveries. IPAC holds NASA's official exoplanet database, data related to ground transit surveys, and also supports the Exoplanet Follow-up Observing Program (ExoFOP), a way for the science community to contribute follow-up data on exoplanet discoveries. The archive also supports analysis (interactive table searches, e.g.), and will support future exoplanet missions such as TESS and JWST.

Big Data hardware at IPAC physically comprises 3 rooms and 76 racks, and will be increasing to 7000 cores. IPAC will grow from 12PB of disk to 30PB in the next few years. Dataset volumes at IPAC have increased by a factor of 100 in the last decade. Data complexity is increasing, and query rates and holdings growth are also accelerating. IPAC must focus on reliable operations and guarantee information technology (IT) security to its users. IPAC's highest priority for its resources is serving NASA missions.

Ten lessons being learned at IPAC:
Large holdings: Strategic organization can be much more important than technological advances when dealing with large data holdings. Critical high-demand data should be put in faster storage, for example. Tables should be organized by common-use cases; IPAC is looking at tessellation schemes for organization. Some tables are growing fast, as represented by a case study in ZTF light curves. ZTF is expected to generate a table with more than a trillion entries; here IPAC is using a hybrid approach that would be a cost-effective solution, allowing rapid response to queries, but limiting available queries. Steve Groom pointed out in this context that while this approach is a response to specific project requirements, it may present problems for future data analysis.

Constant data ingest: Invest early in scalable design; follow details end-to-end.

Data complexity and variability: Interface and metadata standards are needed to search across archives. IPAC has adopted Virtual Observatory (VO) protocols, and is working with other archives to adopt the Common Archive Observation Model. This model accommodates co-registration of moving objects, enables cross-identification of objects between datasets (wavelength, resolution), and allows evolution of knowledge of the relationships between objects. IPAC is currently ingesting the 2MASS catalog, (470 million sources); the next catalog will have 750 million sources. ExoFOP images, spectra, notes, and catalogs, all need to be ingested and correlated, and then made accessible.

Machine Learning: Machine Learning is helping to solve IPAC Big Data challenges. Research applications include self-organizing maps, *t*-distributed stochastic neighbor embedding for galaxy colors, and classification of periodic variable stars. ML is also used for transient detection in PTF/ZTF and will be used for the upcoming NEOCam mission (for tracklet identification). For literature extraction, NED has evaluated several ML packages, but with limited success.

Data Visualization: Good interactive graphics provide intuition about the data. For looking at time domain data one can use folded-viewing, and periodograms for moving objects. For massive data sets, data visualization can go from symbol representation to continuous quantities. There is also the use of data cubes, exploring 3-D and N-D representations, and maybe virtual reality, eventually. A case study of intuition through visualization at IPAC is in parallax views for ultra-cool dwarf stars.

Data discovery: Ease of access is more important than a single point of access; the use of data in the NASA archives can double the science of the original mission. IPAC is part of the NASA Astronomical Virtual Observatories (NAVO) collaboration of NASA archives, providing uniform access to data via VO protocols.

Virtualization: IPAC is using lessons learned from processing of Herschel data. It is important to make analysis available near the data to provide a safe environment. IPAC Herschel provides virtual machines to run memory-intensive and complicated analysis software.

Commercial Cloud: IPAC is finding the Cloud useful for ephemeral computer needs (debugging, sandboxes, surging computing needs, e.g.). However, the Cloud does not necessarily reduce system administration costs; data sets still have to be managed, and connections between servers must still be maintained. The IPAC center is more cost-effective for systems in long-term and mostly full-time use. In a case study on surge computing, IPAC found that the Cloud can provide an agile approach to compute-intensive tasks (e.g., computing light curves).

Interoperability: Increasing interoperability can allow a divide-and-conquer approach to Big Data. In a case study, NED had been using an Aladdin Java applet, not supported by modern browsers, so IPAC adopted an open-source Firefly image viewer service, which brings not only image viewing, but overlays and plotting.

Opportunities and Challenges: IPAC hosts archives that effectively function as observatories themselves, as illustrated by the recent discovery of a new class of galaxies using only data in the archive. IPAC is also leading a study in joint processing from large surveys, with the goals of improved science measurements, cross-system systematic checks, and suppression of spurious objects. IPAC is also working with the high-speed network, the Pacific Research Platform, and is just beginning discussions with the Open Science Grid. As to challenges, data analytics can inform strategic investment for IPAC (using information derived from Google/Apache hits). Bringing analysis to the data is both a challenge and an opportunity. An IPAC team is currently integrating the three LSST Science Platform Aspects to help workflows move across data domains.

Discussion

Dr. Beebe asked for more specific information about a Canadian facility referenced in Dr. Imel's talk (the CADC; Canadian Astrophysics Data Center). Dr. Imel took an action to provide a reference for her. Dr. Walker asked if IPAC had access to astronomy VO protocols. Dr. Imel said that yes, IPAC was adopting these. Dr. Groom added that NAVO is part of a collaboration to implement VO protocols in a consistent way, and is benefiting from additional participation from the Chandra data center, and a Johns Hopkins University VO effort. Dr. Holmes asked about progress in metadata, and in lining up metadata models. Dr. Imel replied that in general, IPAC was converging on standards. The CADC, for one, can now give projects a blueprint (CAOM metadata standard) for ingesting their metadata. Dr. Doyle thought that IPAC has a great handle on standards, and was in favor of an idea that would apply some tens of millions of dollars straight to the archives to push discovery as far as it will go. Dr. Holmes thought that funneling the money to principal investigators might reduce "sandboxing," and help research to stay relevant. Dr. Imel commented that in bringing analysis to the data, one would ideally want to have discovery but also structure; and also, would want to maintain semantic information capture. Dr. Holmes asked if IPAC were using machine learning for detecting transients. Dr. Imel noted that Umaa Rebbapragada was a collaborator on the published machine learning transient methods which will be applied to NEOCAM (now in pre-phase A).

Public Comment

No comments were noted.

BDTF Studies

Dr. Kinter reviewed a topic paper on modeling workflows, which was essentially taking the argument that aging workflows are hindering improvements in modeling and simulation in the Earth and space sciences. The paper provides some examples in solar physics, data volumes and variety, and problem statement details, and includes a potential finding on aging workflow, as well as the mismatch between scientist training in data science and the need to employ data science in missions. Three recommendations flow from these findings: NASA should make the necessary investments to accelerate modeling workflows; NASA should target immediate efforts in modularization; and NASA should make investments in virtual environments and other techniques that enhance modeling workflow.

Dr. Lee commented that the paper seems to suggest that NASA invest in improving workflows in a flat budget environment, and asked which tradeoffs NASA should make to do accommodate the investment. Dr. Kinter noted that researchers are getting less

than 5% of peak performance in current modeling efforts; with just need a factor of 2 or 3 improvement, one could obtain 10 or 15% of peak.
Dr. Holmes suggested pulling the findings into the Executive Summary to grab attention, streamlining the language, adding some quantitative statements, and highlighting a systems approach to improving workflows. Dr. Feigelson felt there were a few areas where dramatic, order-of-magnitude (OOM) improvements have been made in the computing field: one is compressive sensing, which has enabled essentially a 100-fold faster image processing improvement. The improvement stems not from a coding change but from an advance in mathematics. Another is a method called "designer computer experiments" that uses Latin squares (Latin hypercube sampling), resulting in OOM improvements in some cases. Dr. Feigelson suggested that NASA peer reviews include a reviewer who is an expert in such experiments, when evaluating proposals. Disseminating knowledge of these new techniques is also an educational issue; NASA needs to get these ideas into the mainstream, and researchers themselves need an advanced education in methodology to make progress.

Dr. Beebe reviewed the topic paper on data discovery, the idea of which was to address the Directors of the SMD discipline divisions and recommend a separate review for data management plans associated with missions. Any proposal would have to be readily absorbed into the current budget (i.e. an extra Senior Review to produce user's guides towards the end of a mission, at the cost of about $30k per instrument, to assure the impact of the mission would extend well beyond its lifetime). Dr. Holmes advised getting the paper's recommendations succinctly into the Executive Summary.

Asked if there were a documented policy in Astrophysics for data management, Dr. Feigelson felt that the discipline had had a very good track record on data management for decades, but could not point to a specific policy direction. Dr. Guhathakurta thought such information should be extractable out of any AO. Dr. Feigelson added that an Astrophysics AO generally states that a mission must have a data archiving plan, as well as data in a standard format. The issue-quality of Level 3 data varies somewhat, and sometimes documentation is poor. In the Heliophysics Division (HPD), data management plans were seen to be encoded as data policy. Dr. Doyle felt that APD seemed to be best at self-policing. The Planetary Data System (PDS) is moving along more on metadata. Dr. Feigelson noted that the Astrophysics field has interoperability conferences every 6 months, and that data standardization tends to be an international effort.

<u>Commercial Partnering in Cloud Computing</u>
Mr. Jim Rinaldi, Chief Information Officer (CIO) for JPL, presented a briefing on JPL efforts in Cloud Computing (CC), which has helped to support digital transformations, working with others, out-of-the-box thinking, on-demand scalable capacity, new technology options, new business models, rapid low-risk experimentation, and thinking of "everything as a service." The Cloud is affecting not only what JPL does but how it works. The Lab started experimenting with Cloud in 2007-8, starting with 10 providers. Today, flagship missions are leveraging the Cloud, helping to save NASA resources.

There are challenges to employing the Cloud. Some are related to human nature, which tends toward protecting legacy and resisting adaptation to new architectures. JPL chose to solve these challenges with early visioning and partnering on game-changing technologies. The Amazon Cloud accounts for missions and IT have increased to about 100 since 2010. For example, image processing from the Mars Science Laboratory (MSL) is done on Cloud accounts. For the future Earth Science missions Surface Water and Ocean Topography (SWOT) and the NASA/ISRO Synthetic Aperture Radar (NISAR) mission, JPL has no computer big enough to handle the expected volume of data. JPL is also preparing for the next workforce by hosting students who are proficient at handling Cloud technology.

JPL is testing Alexa as a means of increasing interest in space, and for using lessons learned: NASA Mars, NASA's first Alexa app, answers questions such as "How cold is Mars?" Amazon Web Services (AWS) and JPL partnered in the development of GovCloud services, Glacier (storage platform) and Eon, and are also working together to evolve Alexa in the enterprise, such as Mars 2020 in AWS, Europa mission in AWS, and in evolving serverless computing. JPL must also adhere to transparent cost models to keep things affordable. There can be some economy of scale in cloud computing, when adding in energy and facility costs. There is cost in unused capacity.

JPL is considering potential projects such as Augmented Intelligence, moving Data Centers, and edge computing. The next big thing, the Internet of Things (IoT), uses a lot of sensor data, so JPL is working with Microsoft and Redhat in trying to understand hacking attempts, controlling robots with voice commands, and using Alexa as a virtual helpdesk and phone. In this area, strong partnerships with vendors are as critical as the services they provide.

In response to several questions, Mr. Rinaldi said he was involved with Mike Little's efforts at Earth Science's AIST, and the Pacific Research Platform. He was also meeting regularly with the CIOs from the Federal National Laboratories as a good opportunity to network, and to plug into the Department of Energy's (DOE) ESNet, and to test GPU computing. Dr. Mentzel asked Mr. Rinaldi how he saw the needs of the workforce changing. Mr. Rinaldi felt that the workforce would require lots of training, and agreed with the idea of bringing in experts to get people up to speed. He felt the Cloud and evolving technology would influence more open software development at JPL, which in turn can bring significant improvement to turnaround time and implementation. JPL is already generating new titles, like "site engineers." JPL also regularly talks to vendors to obtain Lessons Learned; e.g., how Netflix uses the Cloud for their services. Dr. Feigelson commented that some Astrophysics users felt that Cloud was too expensive for storage. Mr. Rinaldi suggested bringing the partner in to see why storage was expensive, such as egress costs. Architectural approaches can solve these problems. At other times, the Cloud is just not ready for some services. Mr. Rinaldi did not think everything was going to go on the Cloud. He also asked: is AWS too big to fail? It's important to ensure integrity and accessibility of data. The next challenge will be: Where's my data? This problem is yet to be solved, but JPL is beginning to understand requirements. Amazon creates 1000 new services per year; it will be important to avoid getting stale.

Compliance is another issue to watch. Partnership is also a two-way street in terms of opportunities; by taking on a recent JPL project for using a smartphone to log into computers, the provider managed to leverage the whole government as a market.

Science Data System (SDS) Considerations for SWOT and NISAR

Mr. Hook Hua, Science Data System Architect for the future SWOT and NISAR missions, presented a briefing on the anticipated requirements of two very data-intensive, synthetic aperture radar missions. NISAR especially will be pushing the boundaries on data needs. The estimated daily volume from NISAR is expected to be nearly 100 TB per day, two orders of magnitude more than Soil Moisture Active Passive (SMAP) in 2015. This jump in data volume has implications for data storage, data movement, costs, and agility (to respond to changing requirements). The complexity of workflows for SWOT and NISAR range from medium to high. Science users will likely need to adopt new strategies to interact with this high-volume data, which is acting as a forcing function to change. The missions will have to both continuously process the data into actionable products, as well as keep the data stream moving. This will require concurrent processing pipelines, where Cloud will be a part of the solution. The premise is to route science data products into AWS S3 object storage; it currently looks like the Cloud will be able to meet needs. JPL is also looking at GPU-accelerated data product generation.

Forward processing in AWS Cloud will use Virtual Private Clouds to help organizations retain their jurisdiction and IT security requirements. Another novel approach being considered is "containerizing" product generation executives (PGEs) into self-contained Docker Containers to reproduce all existing data products. "Use what you store and store what you use."

Relevant cost components such as compute, storage, data movement, etc., are being rolled up into total cost assessments. There are also ways to optimize costs of storage by tiering data into "hot" storage for frequently accessed data, versus "cold" storage for less dynamic data. Negotiating rates with AWS is also a cost factor; the size of the data system compute nodes can automatically grow and shrink on demand, and can be easily scaled up. There is also the AWS Spot Market, which is likened to Priceline for compute nodes. Users can bid for a price they're willing to tolerate, but this will entail some tolerance of volatility (such as getting bumped off the market when a higher bidder intervenes). This activity forces the data system to get more resilient, and JPL is investigating a number of ways to mitigate the impacts of the Spot Market.

JPL is currently validating Cloud adaptation, first by collocation of SDS and DAAC in the Cloud. The idea is to move the analysis instead of the big data, and rely on process migration instead of data migration. The concept visualizes a "lake" of data containing all of NASA's missions, with services built around it. Dr. Lee noted he had made a presentation in 2014 to all the space agencies on the concept of collocation, but the idea was not widely accepted. He also pointed out that NOAA is now putting data into the Cloud, where NASA can get in and process it (using MODIS algorithms with current Geostationary Operational Environmental Satellite Program/GOES data). Mr. Hua said much of this work had funded by NASA's Little and Murphy. He described another example of using Sentinel data in the Cloud for emergency response, which enabled NASA to get critical products to the Federal Emergency Management Agency (FEMA) within hours. While the process is not technically operational, it is indeed being used.

Why not just generate products on demand, instead of doing all this collocation? It is a trade-off between storage and compute. Storage is getting cheaper and cheaper to process, to the point where it's better to process the data once and store it indefinitely. On-the-fly processing runs into re-validation issues. Processing the data once makes everything permanent and versioned so there is no ambiguity in reprocessing. In order to address the needs of most users, Mr. Hua said he and his team were actively working with the science teams and collecting use cases, and trying to ease them into this new paradigm. Dr. Kinter asked if JPL was looking at multiple storage paradigms, such as granularity, chunking, and transposing. Mr. Hua said that storage paradigms would more of a Level 3 product. Dr. Mentzel suggested that one approach could be to combine data with existing services such as Google Earth's search engines. Is JPL thinking about vendor lock-in? Mr. Hua said that while JPL is trying to be vendor-agnostic, AWS is currently the best choice, although Google and others are catching up. JPL is also very aware of data lock-in; one can't just pack up 100 PB and go. Keeping separate copies outside the Cloud is one strategy. Dr. Doyle pointed out that one concern would be that the vendor has analyzed your cost model better than you have. The other is that there are different usage models. The way to win the cost game is predicting usage patterns further into the future.

Dr. Holmes noted that a new paragraph for the server side analytics paper is warranted by this last presentation. Dr. Walker said the use of the Cloud in this manner struck him as risky, but as potentially high-reward. Mr. Hua said he had listened to a lot of chatter in the SAR community, and was really trying to create products that will make 80% of the users happy. He noted that GPS users have gone through the same evolution in coming to accept Level 2 products, instead of demanding Level 0 data to process themselves.

<u>BDTF Studies</u>
Dr. Feigelson reviewed the TF paper on data science methodologies, the gist of which was to encourage professional development for SMD scientists to learn informatics and statistics, bring in expertise, and bring modern software engineering into NASA's internal and funded software enterprise, and to include information scientists as staff or consultants. Specifications and performance reviews for satellite mission software should include high standards for computer algorithms and statistical methods, and include regular evaluation by cross-disciplinary experts. Dr. Lee commented on the software engineering issue raised in the paper. He said NASA has tried to ask for software engineering in calls in the past, and the response had been underwhelming. He hoped for more elaboration on what software engineering means. Dr. Mentzel thought it was a matter of bringing software engineering expertise to graduate students (software carpentry is the term). Dr. Doyle supported an open source paradigm, although it has its own challenges, based on whether you are you a provider (liability,

licensing issues) or consumer. He noted that he had found that freshouts tend to use open source as a proxy for publishing.

Discussion
The BDTF briefly reviewed recommendations and adjourned for the day.

November 3, 2017
Discussion
Dr. Holmes said he was writing up a finding on Dr. Guhathakurta's presentation on FDL. Dr. Walker said he'd been thinking about Mr. Hua's talk on AWS, and wondered how JPL intended to safeguard data against breaches. Dr. Holmes felt that Mr. Rinaldi and other NASA CIOs have an extensive list of security questions, and felt they were asking the right questions. He wasn't sure BDTF had the expertise to address the issue. Dr. Smith agreed the NASA was on top of the situation. Mr. Crichton reported that the Big Data session at SSB had addressed the issue, and had presented thoughts about architecting security, as well as retention security. There was a serious issue with NOAA, which had reported that fake weather maps were being passed around during the recent hurricanes. Dr. Beebe thought this was the beginning of unfunded mandates to participate on all the data science review panels on ROSES; it's serious business and it's a workload. Dr. Feigelson noted that the National Institutes of Health (NIH) have had a policy that gives a fraction of funds to mathematicians/statisticians to assist biomedical researchers, which been going on for decades. NASA did once have a requirement for each mission to have a 1% tax for Education and Public Outreach. Could a similar "tax" could be instituted for infusing data science expertise, to build in the funds to bring in cross-disciplinary experts. It's not impossible, it's been done. Dr. Holmes felt these concerns had been covered in the overall recommendations. BDTF must additionally point out that these changes will require leadership from the top.

Mr. Crichton reported on a discussion session at SSB on Big Data, which had centered on topics of scale, access of archival data, and increased data visualization. Ed Kearns, Chief Data Officer at NOAA, reported on the results of having moved NOAA observational data into the Cloud (roughly 30 PB), which led to an increase in traffic to the data, and an increase in use. NOAA is now getting questions on how to download data from the Cloud. There still needs to be a shift in thinking about how to do analysis. There was a talk on sea level rise projects and what's going on in plugging in iPython and Jupyter notebooks. A ML talk from Stanford focused on the use of deep learning and neural networks in detecting solar flares. Sara Gilbertson gave a talk about an SSB session that had overlaps with BDTF, on long term synoptic data records, Heliophysics and deep learning. The summary message is that SSB wants to spend more time on Big Data, via workshops. Dr. Holmes reported having had many conversations with Dr. David Spergel, a former Chair of SSB, who is also developing data approaches to Astrophysics. He is onboard with the BDTF message, and has been pushing SSB to start thinking about Big Data.

Finalize Studies, Findings and Recommendations

Dr. Holmes displayed the draft paper on server-side analytics for review. Mr. Crichton commented on modeling scalability and how to make the decision on when to go server-side. He thought the decision could be described quantitatively and mentioned an upcoming Caltech workshop on the subject, scheduled for March 2018. Dr. Beebe brought up the fact that Congress is mandating availability of data, and the question of how NASA proposed to manage this. Dr. Holmes felt that question was beyond our BDTF's scope. Dr. Kinter thought that one big issue is that server-side analytics (SSA) means different things to different people. Some need the next level of generic statistics. Others don't know what calculation they want to do, but they want to use packages like R or linear algebra routines; these are wildly different needs. It would probably be worth including this distinction in the white paper; Dr. Kinter agreed to draft some language to this effect.

Dr. Holmes returned to Dr. Beebe's point on how the Congressional mandate on data would affect how final research products will be made available to the public: these are 100s of PB coming down the road. How will NASA do this? Dr. Lee said that NASA does not have a strategy, and is not archiving individual PI data yet. Some programs have data management plans, but these also are not uniformly applied. Dr. Holmes asked each of the teams to do some more editing on the white papers, to be finalized over the next few weeks. He suggested no individual attribution, to which the BDTF agreed.

Dr. Lee, Program Officer for High Performance Computing at ARC, gave a brief summary of the results of an Ames survey that had been sent out to the broader research community. In 2013-14, the Ames Supercomputing division surveyed a select group of users to derive requirements for meeting Big Data challenges. The interviews were done face-to-face or by phone, and the results were published in a NASA technical report, which is available on the NASA website, and is easily downloadable. The results indicated several big challenges: data discovery (where is it, how do we process), data management (transferring data is a big problem, highlighting the need for investment to develop tools, models, and algorithms), workflows (increasing complexity of processing pipeline, which will require validation and review of workflows), infrastructure (for SSA, e.g.), dissemination. This year, NASA hired a new data scientist at the Ames Supercomputing Division, and is doing another survey to understand more detailed requirements to support NASA missions. The results will follow on from the previous study, and expand the audience to NASA-wide, and getting into more detail. ARC is developing benchmark cases to help design systems and guide acquisition. Websites, and communications with NASA's internal Big Data Working Group are being used to reach out to HEOMD, OCIO, etc., and not just the scientists. Dr. Holmes suggested broadcasting the new survey on NSPIRES and via the next American Geophysical Union (AGU) conference.

Public Comment

No comments were noted.

Final Discussion/Open Action Items

Dr. Hurlburt reviewed findings and recommendations on NASA Archives, which were seen to be working well and providing ample fuel for producing scientific papers. Dr. Holmes asked that metrics on the increasing numbers of papers be added. The essential recommendation to SMD is to view its set of data archives at an equivalent rank to its flight missions. BDTF reviewed the recommendation and approved the language.

Dr. Beebe reviewed a recommendation on data accessibility and archives: i.e. near the end of the prime mission, NASA should review the data entering the archives, and the quality of the calibration. In addition, the team should prepare or update a user's guide for each instrument, detailing its use. BDTF approved the recommendation.

Dr. Kinter reviewed a finding on aging workflows leading to limited performance, and the inadequate preparation of scientists involved in modeling, and presented a recommendation on making investments in appropriate training, workshops, and special collections to drive a cultural change. The recommendation included immediate efforts and longer-range investments. Dr. Holmes asked that the recommendation be broadened to include the other federal agencies- DOE, NOAA, NSF, etc. BDTF concurred with the final language.

Dr. Holmes finalized a recommendation to NASA to move to SSA architectures. BDTF concurred with the final language.

Dr. Holmes proposed a new finding on how NASA should lead the way in promoting student workshops and education, to help attack real world problems, train new cadres of data scientists, and help young researchers network. The finding would be couched as an Attaboy with an implied recommendation, followed by instructive examples. Dr. Walker added that it was important to note that these students are also being taught new, emerging techniques. BDTF reached consensus on this finding, with the final wording to be refined.

Dr. Hurlburt reviewed findings and recommendations on methodology, noting that strides in methodology are often not incorporated into NASA data analysis programs. BDTF recommends that NASA make the necessary changes in training and education to implement critical capabilities that new and evolving data science algorithms can provide. BDTF concurred with the language, leaving specific examples to be more detailed in the topic white paper.

A recommendation on establishing a Data Science and Computing Division in SMD was finalized and accepted.

Dr. Holmes requested that Dr. Hurlburt's assessment of NASA data archives become a fifth white paper.

A recommendation to create more DS&C advisory positions was given to be passed on informally. Dr. Holmes gave an action to BDTF members to evaluate the rosters of the advisory committees and see who qualifies as an SME. There should be at least one seat at the table for each of the discipline committees, as well as the Science Committee. Dr. Holmes felt the implicit recommendation is that there should be a new FACA committee devoted to DS&C, and said he would pass this on privately.

Final thoughts around the table:
Dr. Feigelson thought a new NASA DS&C officer should convene a Task Force on the subject every 3-5 years. ROSES should also target the informatics topic. He felt that the new division was the most expensive recommendation, while the others were cheap and doable.
Dr. Walker commented that while it is widely recognized that data science is important, the next battle is to convince the universities that this is so.
Dr. Beebe felt BDTF had produced good, actionable succinct recommendations. She was doubtful NASA could make the recommended appointments in the near future, but hoped the Agency could appoint a current full-time employee (FTE) to take on the task.
Dr. Holmes felt the recommendations were obtainable, overall.
Dr. Mentzel said it had been an honor and privilege to work with BDTF, which had produced some solid recommendations. He hoped NASA would take heed.
Dr. Kinter commented that as he was part of a group that's been in the trenches on these issues, all of these issues are hard rows to hoe. If any of the recommendations are accepted, they will benefit the entire community.
Dr. Hurlburt thought BDTF issues were already resonating in the community, and felt these recommendations would find acceptance.

Dr. Holmes reminded BDTF that communication is a two-way street, and left it to the members to broadcast the word. He asked for a formal email from Mr. Smith to promulgate forward to the community, to help to promote and expand the readership for BDTF conclusions.

Dr. Holmes concluded the meeting with a recitation of the Scope of Task Force statement:
"The scope of the Task Force includes all NASA Big Data programs, projects, missions, and activities. The Task Force will focus on such topics as exploring the existing and planned evolution of NASA's science data cyber-infrastructure that supports broad access to data repositories for NASA Science Mission Directorate missions; best practices within NASA, other Federal agencies, private industry and research institutions; and Federal initiatives related to big data and data access."**

Dr. Holmes praised the efforts of BDTF, noting with great pride that all of its members had gone above and beyond in carrying out the Task Force charter. He adjourned the meeting a little before 1 pm.

***Abstracted from the Terms of Reference, Ad Hoc Task Force on Big Data, signed by the NASA Administrator on Jan. 8, 2015.*

Appendix A
Attendees

Ad Hoc Big Data Task Force Members
Charles P. Holmes, ***Chair***, Big Data Task Force
Reta Beebe, New Mexico State University
Eric Feigelson, Pennsylvania State University
Neal Hurlburt, Lockheed Martin
James Kinter, George Mason University
Chris Mentzel, The Moore Foundation
Raymond Walker, University of California at Los Angeles
Gerald Smith, ***Executive Secretary***, NASA HQ

NASA Attendees
Dan Crichton, NASA JPL
Richard Doyle, NASA JPL
Jay Famiglietti, NASA JPL
Hook Hua, NASA JPL
Thomas Huang, NASA JPL
Emily Law, NASA JPL
Tsengdar Lee, NASA ARC
Lukas Mandrake, NASA JPL
Lewis John McGibbney, NASA JPL
Shan Molhatra, NASA JPL
Lika Guhathakurta, NASA ARC
Ryan McGranaghan, NASA JPL/UCAR
Umma Rebbapragada, NASA JPL
Nga Quach, NASA JPL
Kiri Wagstaff, NASA JPL

Non-NASA Attendees
David A. Imel, Caltech
Bill Diamond, SETI
Steve Groom, Caltech
Graham Mackintosh, STC
Amy Reis, Ingenicomm
Joan Zimmermann, Ingenicomm

Webex Attendees
Sara Gilbertson, NASA
Alfreda Hall, NASA SMD
Nick Perlongo, Aerospace Corp.
David Sabol, JHU/APL

Nick Serra, Space Corp.
John Sprague, NASAOCIO
Chris Shenton
Rick Wilson, NASA JPL
Jay Wyatt, NASA JPL

Appendix B
Membership

Charles P. Holmes, Chair
Retired
Formerly at NASA HQ
Science Mission Directorate

Gerald Smith, Executive Secretary
Science Mission Directorate
NASA Headquarters

Reta F. Beebe
Professor, Department of Astronomy
New Mexico State University

Eric D. Feigelson
Department of Astronomy and Physics
Pennsylvania State University

Neal E. Hurlburt
Research Science Manager,
Solar and Astrophysics Laboratory
Lockheed Martin Space Systems Company

James L. Kinter
Director, Center for Ocean-Land
Atmosphere Studies
George Mason University

Chris Mentzel
Program Director, Data-Driven Discovery
Gordon and Betty Moore Foundation

Clayton P. Tino
Software Architect, Virtustream Atlanta
Virtustream Incorporated

Raymond J. Walker
Professor, Institute of Geophysics and
Planetary Physics
University of California, Los Angeles

Appendix C
Presentations

1. JPL Data Science Programs; *Dan Crichton, Richard Doyle*
2. Caltech Center for Data-Driven Discovery; *George Djorgovski*
3. Machine Learning Applications in Earth Science; *Lukas Mandrake*
4. Machine Learning Applications in Astronomy; *Umaa Rebbapragada*
5. Machine Learning Applications in Planetary Science; *Kiri Wagstaff*
6. Big Data Analytics for Sea Level Rise; *Thomas Huang*
7. Big Data Visualization for Planetary Science; *Emily Law, Shan Maholtra*
8. Big Data Analytics for Hydrology; *Jay Famiglietti*
9. Frontier Development Labs; *Madhulika Guhathakurta*
10. Big Data at IPAC; *David Imel, Steve Groom*
11. Commercial Partnering in Cloud Computing; *Jim Rinaldi*
12. SDS Plans for SWOT and NISAR; *Hook Hua*

Appendix D
Agenda
Ad Hoc Big Data Task Force
of the
NASA Advisory Council Science Committee

November 1-3, 2017
Jet Propulsion Laboratory (JPL), Theodore von Kármán Auditorium,
4800 Oak Grove Drive, Pasadena, CA 91011

Agenda
(Pacific Daylight Time)

Wednesday, November 1st

8:30 – 9:00	Opening Remarks/Introduction	Dr. Charles Holmes Mr. Gerald Smith
9:00 – 9:30	Member Reports	Membership
	JPL OVERVIEW	
09:30 – 09:50	JPL Data Science Programs	Mr. Dan Crichton, Dr. Richard Doyle
09:50 – 10:10	Caltech Center for Data-Driven Discovery	Prof. George Djorgovski, Caltech
10:10 – 10:30	***BREAK***	
	MACHINE LEARNING/DATA SCIENCE METHODS AND APPLICATIONS	
10:30 – 10:50	Machine Learning Applications in Earth Science	Dr. Lukas Mandrake
10:50 – 11:10	Machine Learning Applications in Astronomy	Dr. Umaa Rebbapragada
11:10 – 11:30	Machine Learning Applications in Planetary Science	Dr. Kiri Wagstaff
11:30 – 11:50	Machine Learning Wrap Up; Q&A	
11:50 – 12:00	Public Comment	
12:00 – 13:00	***Lunch***	
13:00 – 14:30	Non-FACA JPL Tour	Mr. Dan Crichton
	BIG DATA ANALYTICS AND VISUALIZATION	
14:30 – 14:50	Big Data Analytics for Sea Level Rise	Mr. Thomas Huang
14:50 – 15:10	Big Data Visualization for Planetary Science	Ms. Emily Law Mr. Shan Malholtra
15:10 – 15:30	Big Data Analytics for Hydrology	Dr. Jay Famiglietti Mr. Shan Malholtra
15:30 – 16:00	Big Data Analytics Wrap Up; Q&A	
16:00 – 17:00	First discussion of study reports – summarize updates since the	Membership

	June tele-meeting.	
17:00	***ADJOURN FOR DAY 1***	

Thursday, November 2nd

8:30 – 9:00	Assemble/Day 1 Carryover Items	Dr. Charles Holmes Mr. Gerald Smith
9:00 – 9:30	Assessment of Data Archives	Dr. Neal Hurlburt
9:30 – 10:15	Frontier Development Labs	Dr. Madhulika Guhathakurta
10:15 – 10:30	***BREAK***	
10:30 – 11:50	Big Data at IPAC (Infrared Processing & Analysis Center)	David Imel Steve Groom
11:50 – 12:00	Public Comment	
12:00 – 13:00	***LUNCH (in place)*** **12:15-12:55 Results from Cassini**	 Prof. Dave Stevenson, Caltech
13:00 – 14:30	BDTF Studies – Finish	Membership
	SCIENCE DATA PROCESSING AND INFRASTRUCTURE	
14:30 – 14:50	Commercial Partnering in Cloud Computing	Mr. Jim Rinaldi, JPL CIO
14:50 – 15:10	SDS Plans for SWOT and NISAR	Mr. Hook Hua
15:10 – 15:30	***BREAK***	
15:30 – 17:00	BDTF Studies – Finish	Membership
17:00	***ADJOURN FOR DAY 2***	

Friday, November 3rd

8:30 – 9:00	Assemble/Day 2 Carryover Items	Dr. Charles Holmes Mr. Gerald Smith
9:00 – 11:00	Finalize Studies, Findings and Recommendations	Membership
11:00 – 11:05	Public Comment	
11:05 – 12:00	Final Discussion/Open Action Items	Membership
12:00	***ADJOURN FOR DAY 3***	

Dial-In and WebEx Information

For entire meeting November 1-3, 2017

Dial-In (audio): Dial the USA toll-free conference call number 888-324-9653 or toll number 1-312-470-7237 and then enter the numeric participant passcode 3883300. You must use a touch-tone phone to participate in this meeting.

WebEx (view presentations online): The web link is https://nasa.webex.com, the meeting number is 991 009 965, and the password is BDTFmtg#6 .

**** All times are Pacific Daylight Time ****